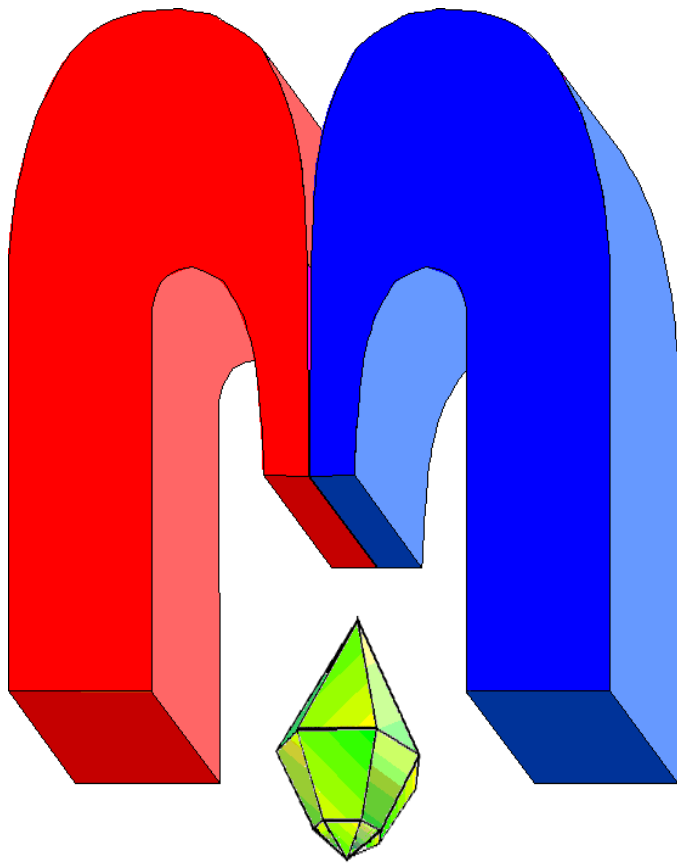


ISSN 2072-5981

doi: 10.26907/mrsej



***magnetic
Resonance
in Solids***

Electronic Journal

Volume 21

Special Issue 3

Paper No 19308

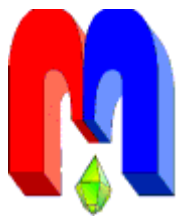
1-18 pages

2019

doi: 10.26907/mrsej-19308

<http://mrsej.kpfu.ru>

<http://mrsej.ksu.ru>



Established and published by Kazan University
Endorsed by International Society of Magnetic Resonance (ISMAR)
Registered by Russian Federation Committee on Press (#015140),
August 2, 1996
First Issue appeared on July 25, 1997

© Kazan Federal University (KFU)*

"Magnetic Resonance in Solids. Electronic Journal" (MRSej) is a peer-reviewed, all electronic journal, publishing articles which meet the highest standards of scientific quality in the field of basic research of a magnetic resonance in solids and related phenomena.

Indexed and abstracted by
Web of Science (ESCI, Clarivate Analytics, from 2015), Scopus (Elsevier, from 2012), RusIndexSC (eLibrary, from 2006), Google Scholar, DOAJ, ROAD, CyberLeninka (from 2006), SCImago Journal & Country Rank, etc.

Editor-in-Chief

Boris **Kochelaev** (KFU, Kazan)

Honorary Editors

Jean **Jeener** (Universite Libre de Bruxelles, Brussels)


Raymond **Orbach** (University of California, Riverside)

Executive Editor

Yurii **Proshin** (KFU, Kazan)
mrsej@kpfu.ru



This work is licensed under a [Creative Commons Attribution-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-sa/4.0/).

 This is an open access journal which means that all content is freely available without charge to the user or his/her institution. This is in accordance with the [BOAI definition of open access](https://www.boai.ru/).

Editors

Vadim **Atsarkin** (Institute of Radio Engineering and Electronics, Moscow)

Yurij **Bunkov** (CNRS, Grenoble)

Mikhail **Eremin** (KFU, Kazan)

David **Fushman** (University of Maryland, College Park)

Hugo **Keller** (University of Zürich, Zürich)

Yoshio **Kitaoka** (Osaka University, Osaka)

Boris **Malkin** (KFU, Kazan)

Alexander **Shengelaya** (Tbilisi State University, Tbilisi)

Jörg **Sichelschmidt** (Max Planck Institute for Chemical Physics of Solids, Dresden)

Haruhiko **Suzuki** (Kanazawa University, Kanazava)

Murat **Tagirov** (KFU, Kazan)

Dmitrii **Tayurskii** (KFU, Kazan)

Valentine **Zhikharev** (KNRTU, Kazan)

Technical Editors of Issue

Maxim **Avdeev** (KFU)

Alexander **Kutuzov** (KFU)

* In Kazan University the Electron Paramagnetic Resonance (EPR) was discovered by Zavoisky E.K. in 1944.

New approach in correlation analysis

R.R. Nigmatullin

Kazan National Research Technical University (KAI), Kazan 420111, Russia

E-mail: renigmat@gmail.com

(Received March 16, 2019; revised March 20, 2019;
accepted March 31, 2019; published April 19, 2019)

In this paper, a new method for restoration of the desired correlations is proposed. It allows to evaluate the “content” of the external factors ($l = 1, 2, \dots, L$) setting in the form of data arrays $y_m^{(l)}(x)$ ($m = 1, 2, \dots, M$) inside the given $Y_m(x)$ function that is supposed to be subjected by the influence of these factors. As contrasted to the conventional correlation analysis, the proposed method allows finding the “influence” functions $b_l(x)$ ($l = 1, 2, \dots, L$) and evaluating the “remnant” array $G_m(x)$ that is remained as a “quasi-independent” part from the influence of the factors $y_m^{(l)}(x)$. The general expression works as a specific “balance” and reproduces the well-known cases, when $b_l(x) = C_l$ (it is reduced to the linear least square method with $G_m(x) \cong 0$) and coincides with the remnant function $Y_m(x) \cong G_m(x)$, when the influence functions becomes negligible ($b_l(x) \cong 0$). The available data show that the proposed method allows to extract a small signal from the “pattern” background and it conserves its stability/robustness in the presence of a random fluctuations/noise. The method is rather flexible and allows to consider the cases of strong correlations, when the external factors act *successively*, forming the cause-and-effect chains. It can be generalized for expressions containing the bonds expressed in the form of memory functions. The proposed method adds new quantitative ties expressed in the form of the desired functions to the conventional correlation relationships expressed in the form of the correlation coefficients forming, in turn, the correlation matrices. New relationships allow to understand deeper the existing correlations and make them more informative, especially in detection of the desired deterministic and stable bonds/laws that can be hidden inside.

PACS: 02.50.-r, 02.60.-x, 02.70.Rr, 05.40.Ca.

Keywords: correlations, functional least-square method, extraction of a small signal, influence and remnant functions.

This original paper I would like to devote to my Teacher Prof. Dr. Boris I. Kochelavae

Preface

When I was the PhD student (1970-1973) I was left by my Teacher with the problem alone, because he was invited to spent his “sabbatical leave” in the USA again. This time was very difficult from one side and was instructive for me from another side.

Finished the Physical faculty of Kazan State University with red diploma and became the PhD student, I should prove to myself that I am able to solve a difficult problem being alone. During two years, I solved a problem of relaxation in liquid He^{3,4} mixtures and obtained the desired expressions. However, absolutely the same result was obtained by another physicist in the frame of the Zubarev’s NSO approach from Krasnoyarsk institute of Physics and I forced to change a subject for consideration of the processes of relaxation in paramagnets at low temperatures. During one year, I solved another problem and received the desired degree in 1974 year.

I never forget these “testimony” years and I would like to express my acknowledgements to my Teacher for this “lesson” that I received in that time. These years convinced me that I am able to work alone and solve the problems that I can meet in my life. Now I work in the Technical University (KNRTU-KAI) and participate in solution of engineering problems. I consider them as a new challenge that tests my ability and competence. I can consider myself as successful

scientist because my Hirsch index is 27 and number of citations exceeds 3000 (you can visit the site: <http://expertcorps.ru/science/whoiswho/ci86> for more details). Thanks to my Teacher when he left me alone, I had a possibility to test myself for the usefulness to the chosen scientific carrier.

A few words about the original paper that I want to suggest to an attentive reader. I wrote more than 250 papers and each of the published paper is considered as my “scientific child”. Some of them received a happy fortune; some of them were forgotten. I consider myself as an “armor” who proposes new methods for attacking the difficult problems that needs to be “defeated” with the help of new “arms”. However, now I work actively in three “hot” regions: (a) dielectric spectroscopy, (b) fractals and fractional calculus and (c) development of new statistical methods in extraction of deterministic information from the random signals. The original paper presented below belongs to my last but not the least section (c). It solves one interesting and general problem of the mathematical statistics and it will be useful for a wide circle of engineers and experimentalists, which deal with treatment of different data. I do hope also that this paper will attract an attention of the specialists that work in NMR/EPR applications as well, especially in detection of weak signals affected by a “noise”.

1. Introduction and formulation of the problem

If any reader have a look at the wide set of books and papers meant for practitioners [1-10] related to correlation analysis then he has a right to pose a reasonable question: what kind of new elements the ambitious author of the paper is trying to introduce in order to convince the skeptically tuned expert that these innovative elements are really *new* and will be *useful* for practical applications? If a reader will print in his computer “correlation and regression analysis” then he can find about 25 million results (!) associated with this subject. This huge “information wall” can serve as a real obstacle in attempts to suggest some new and original elements that “go out” from the conventional Pearson correlation analysis and its existing generalizations.

Nevertheless, the analysis of the basic literature [5-15] related to the establishing of the desired relationships between two correlated random sets of data allows finding the desired *new* and important elements that can be helpful and useful in many applications. In order to be more specific let us formulate the problem in details, including some necessary definitions.

Let the set $\{x_j\}$ ($j = 1, 2, \dots, N$) coincides with the input variable and $Y_m(x_j)$ determines the rectangle matrix of the corresponding responses (outputs), where the parameter $m = (1, \dots, M \leq N, M > 1)$ determines the number of successive measurements related to the given input set.

In accordance with the general theory related to the quasi-reproducible experiment (GTQRE) [16-17] one can derive a general functional dependence for the set of $Y_m(x_j)$ and this dependence can be expressed in the form of the *generalized* fitting function related to the Prony function. In many practical applications, it is necessary to establish the correlation relationships between measurements that were obtained for the *same* set of input variables and corresponding to similar experimental conditions. We define these output variables as: $y_m^{(l)}(x_j)$, $l = 1, 2, \dots, L$, where parameter l determines the number of the external factors. Mathematically, the problem that we are going to solve can be formulated as: Is it possible to relate at least approximately the set of the functions $Y_m(x_j)$ with other functions $y_m^{(l)}(x_j)$ for removing their correlation dependence and separate its “remnant” or almost independent part?

We try to find the original solution of this problem that will have large applications related to various data.

2. Description of the proposed algorithm

In many cases, it is impossible to express the single-valued variable x_j through the chosen function $y_m^{(l)}(x_j)$ (except the cases of the simplest dependences), because the output $y_m^{(l)}(x_j)$ for any fixed m expressed in the form of the multi-valued function and the analytical transformation $x_j \rightarrow F(y_m^{(l)})$ becomes impossible. Usually, in this case one can use the statistics of the fractional moments [18, 19] and calculate the complete correlation factor that generalizes the conventional Pearson correlation coefficient over all space of the fractional moments. However, in many cases, it is necessary to know the “share” of correlations of one random variable with respect to another one in the form of an “influence” function and the problem in this formulation, as far as we know, is *not* solved. Mathematically, this correlation relationship (in the case when different external factors are considered independently from each other) can be expressed approximately in the form of the following linear combination:

$$Y_m(x_j) \cong \sum_{l=1}^L \langle b_l(x_j) \rangle y_m^{(l)}(x_j) + G_m(x_j), \quad (1)$$

where the sets $b_l(x_j)$, $G_m(x_j)$ determine the *unknown* functions that are needed to be evaluated. In expression (1) we assume that the “influence” functions are expressed in the averaged sense $b_{ml}(x) \rightarrow \langle b_l(x) \rangle$ and $G_m(x) \rightarrow \langle G(x) \rangle$ (for simplicity we omit the index j) that implies that they do *not* depend on index m . We suppose also that variables $Y_m(x_j)$, $y_m(x_j)$ can be close in the statistical sense and, therefore, they are correlated to each other in some extend. For their evaluation, one can use the functional linear least square (FLLS) method, proposed in papers [16-17], requiring that the functional dispersion between the functions figuring in (1) achieve a minimal value:

$$\sigma(x) = \min \left[\frac{1}{M} \sum_{m=1}^M \left(Y_m(x) - \sum_{s=1}^L \langle b_s(x) \rangle y_m^{(s)}(x) - \langle G(x) \rangle \right)^2 \right]. \quad (2)$$

Taking the functional derivatives with respect to unknown functions $\langle b_l(x) \rangle$, $\langle G(x) \rangle$ we obtain:

$$-\frac{\delta \sigma(x)}{\delta \langle b_l(x) \rangle} = \frac{1}{M} \sum_{m=1}^M \left[y_m^{(l)}(x) \left(Y_m(x) - \sum_{s=1}^L \langle b_s(x) \rangle y_m^{(s)}(x) - \langle G(x) \rangle \right) \right] = 0, \quad (3)$$

$$-\frac{\delta \sigma(x)}{\delta \langle G(x) \rangle} = \frac{1}{M} \sum_{m=1}^M \left[\left(Y_m(x) - \sum_{s=1}^L \langle b_s(x) \rangle y_m(x) - \langle G(x) \rangle \right) \right] = 0.$$

Here we apply the averaging procedure over all admissible measurements ($m = 1, 2, \dots, M$) supposing that the unknown functions $\langle b_l(x) \rangle$ do *not* depend on the current index m . Defining the pair correlation functions

$$Q_l(x) = \frac{1}{M} \sum_{m=1}^M Y_m(x) y_m^{(l)}(x), \quad q_{s,l}(x) = \frac{1}{M} \sum_{m=1}^M y_m^{(s)}(x) y_m^{(l)}(x), \quad (4)$$

$$\langle Y(x) \rangle = \frac{1}{M} \sum_{m=1}^M Y_m(x), \quad \langle y^{(l)}(x) \rangle = \frac{1}{M} \sum_{m=1}^M y_m^{(l)}(x), \quad s, l = 1, 2, \dots, L,$$

one can obtain the system of linear equations with respect to unknown functions $\langle b_s(x) \rangle$ figuring in (3)

$$\sum_{s=1}^L q_{s,l}(x) \langle b_s(x) \rangle + \langle G(x) \rangle \langle y^{(l)}(x) \rangle = Q_l(x),$$

$$\sum_{s=1}^L \langle b_s(x) \rangle \langle y^{(s)}(x) \rangle + \langle G(x) \rangle = \langle Y(x) \rangle.$$
(5)

From linear system (5) one can calculate the desired functions $\langle b_s(x) \rangle$ and $\langle G(x) \rangle$ and, finally, find the solution of equation (1). Here we want to notice that for calculation of the functions $\langle b_s(x) \rangle$ we need to know the value of $\langle G(x) \rangle$, however from (1) one can restore directly the set of desired functions $G_m(x)$ for any fixed m . One can simplify equation (5) also, if one can express the unknown function $\langle G(x) \rangle$ from the second line and insert this expression into the first line of (5). After simple algebraic manipulations, one can obtain the following expression:

$$\sum_{s=1}^L \tilde{q}_{s,l}(x) \langle b_s(x) \rangle = \tilde{Q}_l(x),$$

$$\tilde{q}_{s,l}(x) = q_{s,l}(x) - \langle y^{(s)}(x) \rangle \langle y^{(l)}(x) \rangle,$$

$$\tilde{Q}_l(x) = Q_l(x) - \langle Y(x) \rangle \langle y^{(l)}(x) \rangle,$$

$$\langle G(x) \rangle = \langle Y(x) \rangle - \sum_{s=1}^L \langle b_s(x) \rangle \langle y^{(s)}(x) \rangle.$$
(6)

These expressions solve the problem of calculation of the set of *true* quasi-independent functions $G_m(x)$ from (1) based on the calculated set of the “influence” functions $\langle b_s(x) \rangle$ from (6). It is instructive to write down some expressions for the simplest cases, having large practical importance.

1. *The influence of one factor* ($L = 1$).

For this case system (6) has the following solution:

$$\langle b_1(x) \rangle = \frac{\tilde{Q}_1(x)}{\tilde{q}_{11}(x)},$$

$$\langle G(x) \rangle = \langle Y(x) \rangle - \langle b_1(x) \rangle \langle y^{(1)}(x) \rangle,$$

$$\tilde{q}_{11}(x) = \frac{1}{M} \sum_{m=1}^M \left(\Delta y_m^{(1)}(x) \right)^2, \quad \Delta y_m^{(1)}(x) = y_m^{(1)}(x) - \langle y^{(1)}(x) \rangle,$$

$$\tilde{Q}_1(x) = \frac{1}{M} \sum_{m=1}^M \Delta Y_m(x) \Delta y_m^{(1)}, \quad \Delta Y_m(x) = Y_m(x) - \langle Y(x) \rangle.$$
(7)

It is useful to consider the limiting cases.

(a) Let us suppose that

$$\Delta Y_m(x) \cong \lambda \Delta y_m^{(1)}(x).$$
(8)

In this case, (after averaging (8) over all measurements $m = 1, 2, \dots, M$) we have

$$\langle \Delta Y(x) \rangle = \lambda \langle \Delta y^{(1)}(x) \rangle + \varepsilon(x),$$
(9)

and the calculation of the unknown parameter λ is reduced to the linear least square method (LLSM). The mean value in (9) is replaced by numbers that are calculated over the input variable x . If $\lambda = 1$ (it implies the complete coincidence of the function $Y_m(x)$ with $y_m^{(1)}(x)$) then $G_m(x) = 0$. For $\lambda \cong 1$, the influence of the function $G(x)$ becomes negligible and it is replaced in (9) by a fitting error function $\varepsilon(x)$.

(b) Let us suppose that the number of measurements is small $Y_m(x) \cong \langle Y(x) \rangle$ (it corresponds the case of an “ideal” experiment considered in [16-17]). In this case, the influence of correlation with the function $\langle y_1(x) \rangle$ becomes negligible ($\langle b_1(x) \rangle \cong 0$). In practice, this condition can be tested easily if we analyze the condition $\langle Y(x) \rangle \cong \langle G(x) \rangle$ with the help of the Pearson correlation coefficient. If this coefficient (PC) is located in the interval $0.9 < PC < 1$, then one concludes that the arrays $Y_m(x)$ and $y_m^{(1)}(x)$ do not correlated with each other or, in other words, they practically independent from each other. Therefore, expression (7) covers the cases known earlier.

2. The influence of two factors ($L = 2$).

For this case from system (6) one can receive the following expressions for $\langle b_{1,2}(x) \rangle$:

$$\langle b_1(x) \rangle = \frac{\tilde{Q}_1(x)\tilde{q}_{2,2}(x) - \tilde{Q}_2(x)\tilde{q}_{1,2}(x)}{\tilde{q}_{1,1}(x)\tilde{q}_{2,2}(x) - (\tilde{q}_{1,2}(x))^2}, \quad \langle b_2(x) \rangle = \frac{\tilde{Q}_2(x)\tilde{q}_{1,1}(x) - \tilde{Q}_1(x)\tilde{q}_{1,2}(x)}{\tilde{q}_{1,1}(x)\tilde{q}_{2,2}(x) - (\tilde{q}_{1,2}(x))^2}. \quad (10)$$

Here the matrix components are defined by the following expressions:

$$\tilde{q}_{s,l}(x) = \frac{1}{M} \sum_{m=1}^M \Delta y_m^{(s)}(x) \Delta y_m^{(l)}(x), \quad \tilde{Q}_l(x) = \frac{1}{M} \sum_{m=1}^M \Delta Y_m(x) \Delta y_m^{(l)}(x), \quad (11)$$

$$\Delta f_m(x) = f_m(x) - \langle f(x) \rangle, \quad s, l = 1, 2.$$

The “remnant” function (weakly correlated part of $Y_m(x)$) is given by expression:

$$G_m(x) \cong Y_m(x) - \sum_{l=1}^2 \langle b_l(x) \rangle y_m^{(l)}(x). \quad (12)$$

Expressions (10)-(12) are considered as the final expressions for the case of influence of two additive factors. In any case considered above, it is supposed that the number of similar measurements M exceeds the number of external factors L ($M > L$).

The previous calculations were associated with the case when all external factors act independently/additively from each other. This case is expressed mathematically by expression (1). Is it possible to consider the case when the factors are strongly correlated and act *successively*, i.e. when one dominant factor (cause) evokes the appearance of another factor (effect)? In this case, locating the cause-and-effect relations in the right order one can write the following relationship:

$$Y_m(x) = G_m(x) \prod_{s=1}^L \left(y_m^{(s)}(x) \right)^{\langle b_s(x) \rangle}. \quad (13)$$

Taking the natural logarithm from the both parts, we obtain

$$\ln(Y_m(x_j)) \cong \sum_{l=1}^L \langle b_l(x_j) \rangle \ln \left(y_m^{(l)}(x_j) \right) + \ln(G_m(x_j)). \quad (14)$$

Formally, equation (14) is similar to expression (1), however, in addition one can determine the values of the logarithmic functions, when the function $y_m^{(l)}(x_j)$ becomes negative and equals zero. Therefore, all calculations considered above can be applicable for this case too. The proposed method allows also combining these two cases and considering more complex problem. In order to show its generality we rewrite expression (13) in the following form

$$F_m^{(l)}(x) = G_m^{(l)}(x) \prod_{s=1}^l \left(y_m^{(s)}(x) \right)^{\langle w_s(x) \rangle}, \quad s = 1, 2, \dots, l. \quad (15)$$

Equation (15) allows finding the set of the unknown functions $G_m^{(l)}(x)$ and the weighting factors $\langle w_s(x) \rangle$ with the help of equation (14). After that, one can apply *again* equation (1) representing it in the form:

$$Y_m(x) \cong \sum_{l=1}^L \langle b_l(x) \rangle G_m^{(l)}(x) + R_m(x). \quad (16)$$

From this equation, it is possible to find the additive weighting factors $\langle b_l(x) \rangle$ and the remnant function $R_m(x)$ that is defined again as the weakly-correlated part of $Y_m(x)$.

In the case of cause-and-effect relations, a special attention should be paid for the *negative* values of functions figuring in expressions (13) and (15). The negative values of a power-law function $f(x)$ in the space of real numbers can be defined as

$$\text{Re}(V(x)) = \text{Re} \left(\left[(-f(x))^{\pm\alpha(x)} \right] \right) = |f(x)|^{\pm\alpha(x)} \cos(\alpha(x)\pi). \quad (17)$$

The region when $f(x) \equiv 0$ requires some regularization procedure in any partial case, when $\alpha(x)$ becomes negative or close to zero value ($\alpha(x) \cong 0$). In conclusion of the section 2, one can show also another possibility, where the proposed algorithm could be applied also. Let us suppose that the measured function $F_m(t)$ is formed by a linear combination of the correlated functions related to the initial one by means of the convolution operation, i.e:

$$F_m(t) = \sum_{l=1}^L \int_0^t \langle \mu_l(t - \tau) \rangle f_m^{(l)}(\tau) d\tau + \Phi_m(t). \quad (18)$$

In expression (18) the set of the functions $\langle \mu_l(t) \rangle$ determines the so-called “correlation memory” that arises between the correlation function $f_m^{(l)}(t)$ evoked by the factor l ($l = 1, 2, \dots, L$) and the initial function $F_m(t)$ that can be affected by this factor. These correlations are important in cases when they are expressed in the form of the Riemann-Liouville fractional integral [20]. Applying the Laplace transform to the both parts of (18), we obtain:

$$F_m(p) = \sum_{l=1}^L \langle \mu_l(p) \rangle f_m^{(l)}(p) + \Phi_m(p), \quad (19)$$

$$J(t) := J(p) = \int_0^{\infty} J(t) \exp(-pt) dt.$$

The first line in (19) is similar to expression (1) and, therefore, one can apply the algorithm developed above for this nontrivial case. From our point of view, expression (19) can be useful in analysis of different dielectric “mixtures”, when the total complex permittivity can be presented approximately as an additive combination of different complex permittivities associated with different dielectric materials.

In order to outline the limits of the proposed algorithm one can find the answer to the following question: how many minimal measurements (that is given by the value of M) are necessary for its application? The analysis shows that for calculation of the mean values given by expressions (4) only *two* repetitions of the *same* measurement ($M = 2$) are necessary. However, these two successive measurements cannot create the statistically significant sampling and, therefore, it is necessary to collect a “rich” sampling with number of measurements exceeding $M > 10$, as minimal. How to do it artificially if a researcher cannot realize the representative sampling? In this case, one can suggest an approach that is described below in section 3.

3. Creation of a representative sampling

Let us suppose that we definitely know only two measurements coinciding with the limits of a possible sampling $Ydn(x)$ and $Yup(x)$. These functions define the “down” and “up” limits of the desired sampling, respectively. Let us define the variable from the interval $[0, 1]$

$$v_m = \frac{m-1}{M-1}, \quad m = 1, \dots, M. \quad (20)$$

Then we define the function from the same interval $F(v_m) = 3(v_m)^2 - 2(v_m)^3$. This specific choice is stipulated by the fact that at $F(0) = 0$, $F(1) = 1$, $F(1/2) = 1/2$ and, therefore, the desired array can be obtained as:

$$Y_m(x_j) = Ydn(x_j) + F(v_m)(Yup(x_j) - Ydn(x_j)), \quad (21)$$

$$Ydn(x) \leq Y_m(x) \leq Yup(x), \quad j = 1, 2, \dots, N, \quad m = 1, \dots, M.$$

Let us suppose that *a priori* 3 functions are known: $Ydn(x)$, $Yup(x)$ and some intermediate function $f_1(x)$. We look for the function $F(v)$ in the form of the polynomial of the second order:

$$F(v) = av + bv^2. \quad (22)$$

For the finding of the unknown functions, $a(x)$ and $b(x)$, we take into account the following conditions: $F(1) = 1$, $F(1/2) = w_1(x) = (f_1(x) - Ydn(x)) / (Yup(x) - Ydn(x))$. After simple algebraic transformations we obtain again the structure (21) with the function $F(v_m, w_1(x))$ from (22), where the functions $a(x)$ and $b(x)$ are defined as

$$a(x) = 4w_1(x) - 1, \quad b(x) = 2 - 4w_1(x). \quad (23)$$

By complete analogy with this result, one can obtain the desired expressions for *two* known intermediate functions: $f_k(x)$ ($k = 1, 2$). We subordinate the function $F(v)$ to the following conditions: $F(1) = 1$, $F(1/3) = w_1(x)$ and $F(2/3) = w_2(x)$ with

$$w_k(x) = (f_k(x) - Ydn(x)) / (Yup(x) - Ydn(x)), \quad k = 1, 2$$

and consider the polynomial of the third order:

$$F(v) = av + bv^2 + cv^3. \quad (24)$$

After some algebraic manipulations, one can find the desired functions from the linear system of equations

$$a(x) + b(x) + c(x) = 1,$$

$$\frac{1}{3}a(x) + \frac{1}{9}b(x) + \frac{1}{27}c(x) = w_1(x), \quad (25)$$

$$\frac{2}{3}a(x) + \frac{4}{9}b(x) + \frac{8}{27}c(x) = w_2(x).$$

The system admits the solution:

$$a(x) = \frac{1}{2}(2 + 18w_1(x) - 9w_2(x)),$$

$$b(x) = -\frac{9}{2}(1 + 5w_1(x) - 4w_2(x)), \quad (26)$$

$$c(x) = \frac{9}{2}(1 + 3w_1(x) - 3w_2(x)).$$

Finally, we obtain the same mathematical expression (21), where the function $F(v)$ should be replaced by expression (24) with the functions from (26). More complex cases with intermediate functions $f_k(x)$ ($k = 1, 2, \dots$) can be considered and, therefore, the problem of generation of an artificial array in general case can be solved by a similar way.

Expression (21) generates a set of the functions from the interval $[Ydn(x), Yup(x)]$. If necessary, one can add the intermediate functions $f_k(x)$ ($k = 1, 2, \dots$) and use more complicated expressions (22) and (24) shown above. These expressions can be applied for generation of the desired data in cases, when the creation of the arrays becomes impossible. This situation is observed in economics, meteorology and biological/medical sciences.

Let us suppose that we fixed the random curve $Ydn(t_0)$ at the moment t_0 and to the moment t_M it accepted the form $Yup(t_M)$. Then, introducing the normalized temporal variable

$$v_m = \frac{t_m - t_0}{t_M - t_0}, \quad t_m = t_0 + \frac{m}{M} (t_M - t_0), \quad m = 0, 1, \dots, M, \quad (27)$$

one can create an artificial data array with the help of expression (21) and its generalizations given below. Actually, expression (27) can be considered as a transformation of one random function to another one and this transformation can be useful in many practical applications.

In conclusion of this section, we would like to show how to obtain the *quantitative/numerical* estimations for the influence functions $\langle b_l(x) \rangle$ and the remnant function $G_m(x)$. For this purpose, we take the average values from the functions $Y_m(x)$, $y_m^{(l)}(x)$ and $G_m(x)$

$$\langle Y(x) \rangle = \frac{1}{M} \sum_{m=1}^M Y_m(x), \quad \langle y^{(l)}(x) \rangle = \frac{1}{M} \sum_{m=1}^M y_m^{(l)}(x), \quad \langle G(x) \rangle = \frac{1}{M} \sum_{m=1}^M G_m(x), \quad (28)$$

and then we replace in expression (1) the functions $\langle b_l(x) \rangle \rightarrow b_l$ by the set of unknown constants and $\langle G(x) \rangle$ by the constant g . In the result of this transformation we obtain

$$\langle Y(x) \rangle \cong \sum_{l=1}^L b_l \langle y^{(l)}(x) \rangle + g. \quad (29)$$

The set of the desired constants b_l ($l = 1, 2, \dots, L$) and g can be evaluated from (29) with the help of the linear least square method. In particular, for $L = 1$ the constants b_1 and g coincide with the value of the slope and intercept, accordingly.

4. Examples

4.1. Verification of expressions (7) on mimic data

Let us consider the following problem. From some measurements we have a “pattern” background that is described mathematically by an array $y_m^{(1)}(x)$. This background is distorted by a “small” signal $Sgn(x)$ that was not appeared in the background earlier. Is it possible to “notice” the presence of the function $Sgn(x)$ inside the function $Y_m(x)$ and extract this small signal in the presence of a noise? We should stress here that this common problem is solved in many branches of natural/technical sciences:

1. Detection of a small additive $x(c)$ on the “pattern” background – electrochemistry, microbiology, chemical technology etc.
2. Detection of a small signal $Sgn(t)$ on the radio-interference background.
3. The general problem “friend-or-foe” identification in military science.

For demonstration purposes, we choose the following functions:

- (a) “Pattern” array will be identified with $y_m^{(1)}(x)$ and generated by expression (21) with the limiting functions: $Ydn(x) = \sin(0.4x - \pi/8)$, $Yup(x) = \cos(0.67x - \pi/8) + 3$; the array size: $M = 100$; number of discrete points: $j = 0, 1, \dots, N = 500$; the interval of the input variable: $x_{\min} = 10^{-3}$, $x_{\max} = 10$, $x_j = x_{\min} + (j/N)(x_{\max} - x_{\min})$.

- (b) The tested massive $Y_m(x)$ is generated again with the help of (21) and located in the interval

$$Ysdn(x) = Ydn(x) + Sgn_1(x), \quad Ysup(x) = Yup(x) + Sgn_2(x),$$

$$Sgn_{1,2}(x) = a/[(x - x_{1,2})^2 + \gamma^2], \quad a = 0.01, \quad x_{1,2} = 6, 5, \quad \gamma^2 = 0.05,$$

$$Ysup(x, w_1) = w_1 \times Ysup(x) + w_2 (Yrnm(x) + 3),$$

where, in turn, $w_1 + w_2 = 1$, $Yrnm(x)$ defines the generator of the uniform random numbers from the interval: $[-1, 1]$.

Having these functions, it is naturally to solve the problem of extraction of useful signals $Sgn_{1,2}(x)$ from the array functions $Y_m(x)$ at different values of the parameter w_1 ; it admits a “mixture” of a noise ($w_1 = 1$ (absence of the noise), $w_1 = 0.9$ (small influence of a noise)). The figures 1-4 demonstrate the results of this simple analysis. In Fig. 1, we show the location of two small signals in the absence of a noise $w_1 = 1$ on the corresponding background. In Fig. 2 we show the extraction of two weak signal with the help of expressions (7). If the background remains stable ($\langle b_1(x) \rangle$ is located near the unit value, $\langle G(x) \rangle \cong 0$) then the weak signal is detectable easily. The influence of the random fluctuations is shown in Figs. 3a, 3b. Even a mixture of a small noise (Fig. 3a) $w_2 = 0.1$ distorts the $Sgn_2(x)$. With increasing of the parameter $w_2 = 0.5$ the $Sgn_2(x)$ completely disappears, however the $Sgn_1(x)$ conserves its initial location (see Fig. 3b). Therefore, having a stable and “pure” background that serves as a specific detector one can notice the presence of a small “entity” comparable with the noise or it can be detected as a signal.

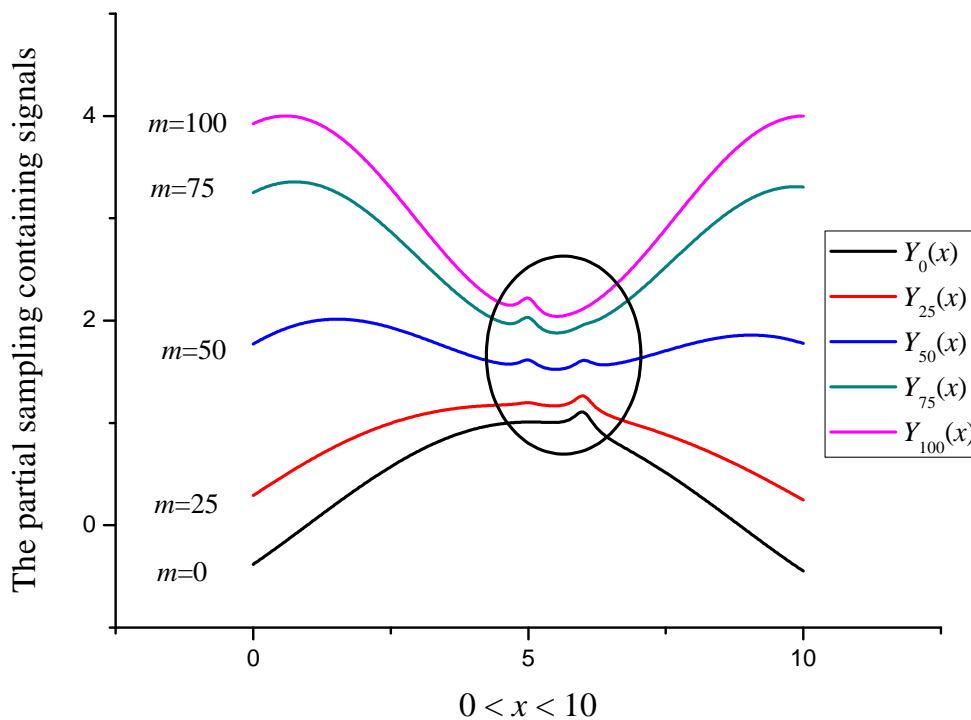


Figure 1. Here we demonstrate the location of two artificially created small signals. The array was created with the help of expression (21). $M = 100$.

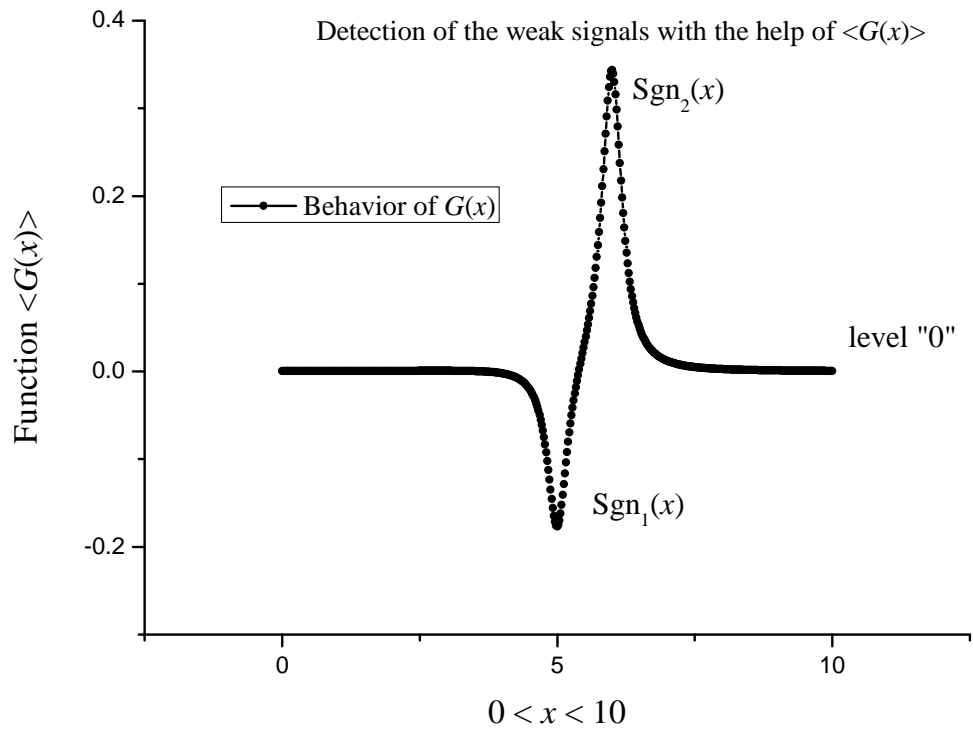


Figure 2. For this “ideal” background two small signals (where dependence $\langle G(x) \rangle \cong 0$ serves as a specific “ruler”) are detected easily with the help of expressions (7).

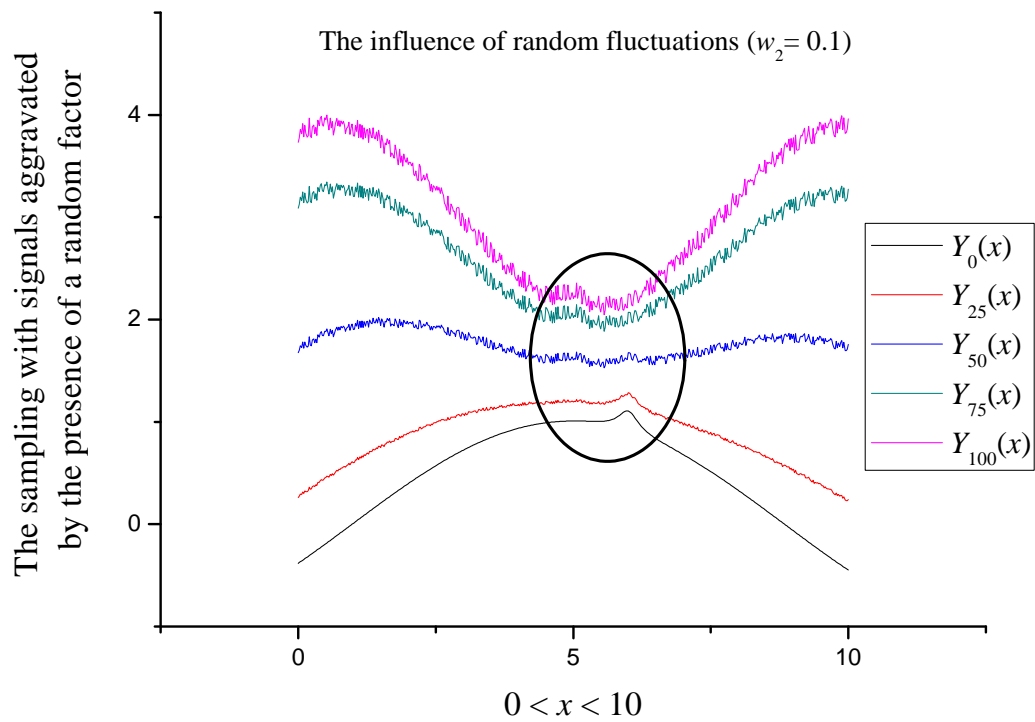


Figure 3a. The presence of a “noise” (even in a small amount $w_2 = 0.1$) in the chosen sampling creates some certain difficulties in detection of small signal.

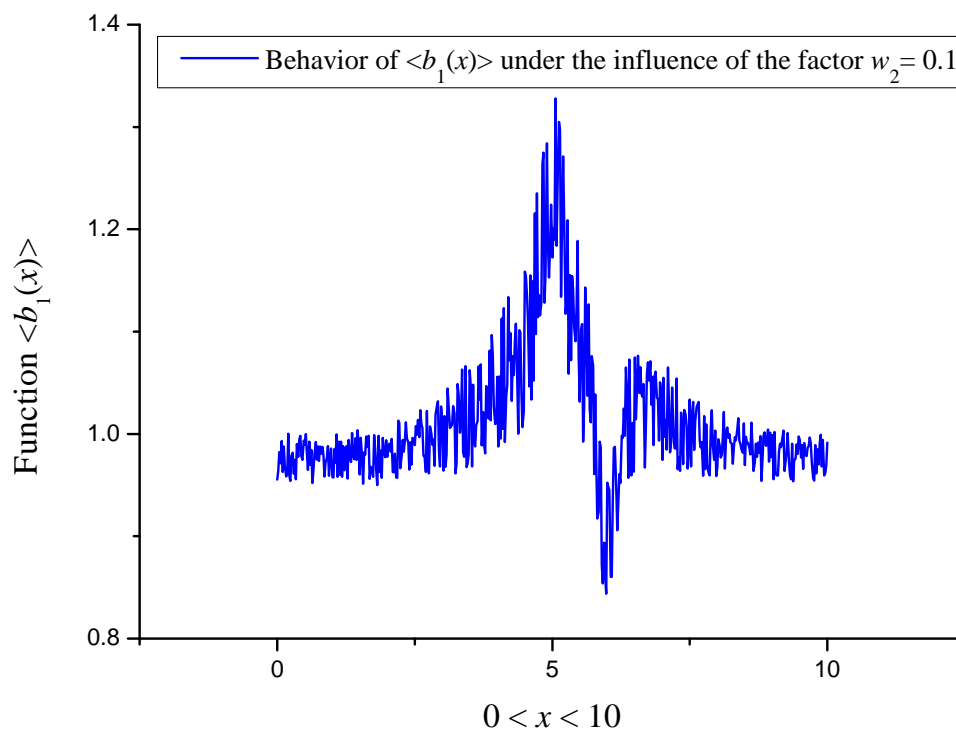


Figure 3b. The view of the distorted signals under the influence of a noise. The form of signals are distorted, while their locations are still conserved. The distorted influence function $\langle b_1(x) \rangle$ is shown in this figure.

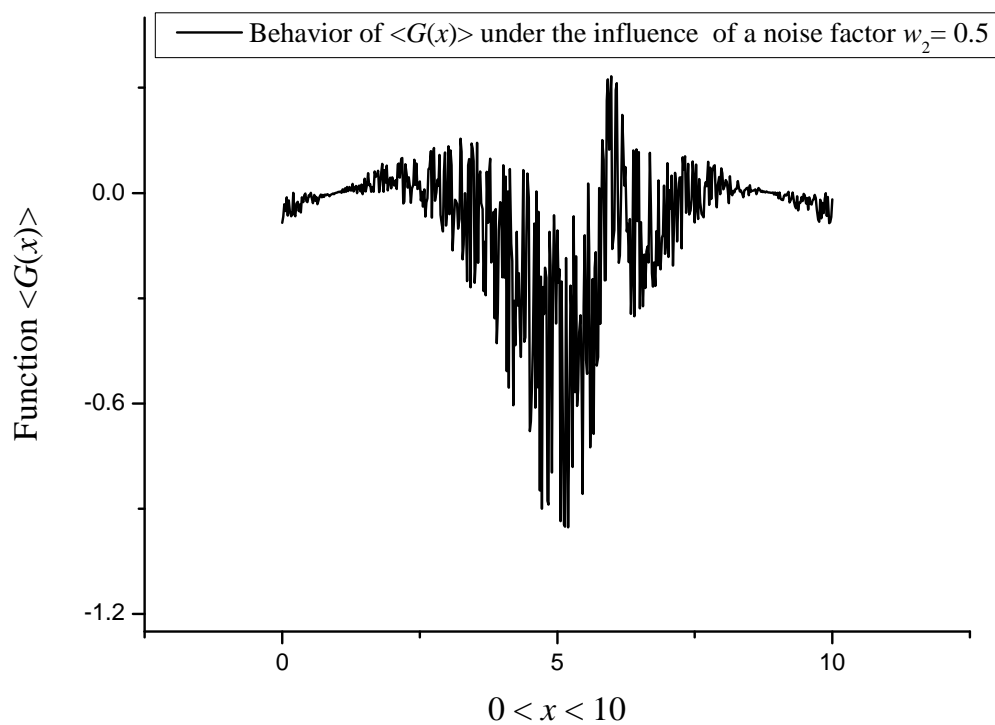


Figure 3c. The further increasing of random fluctuations ($w_2 = 0.5$) leads to the complete disappearance of the small signals and their locations.

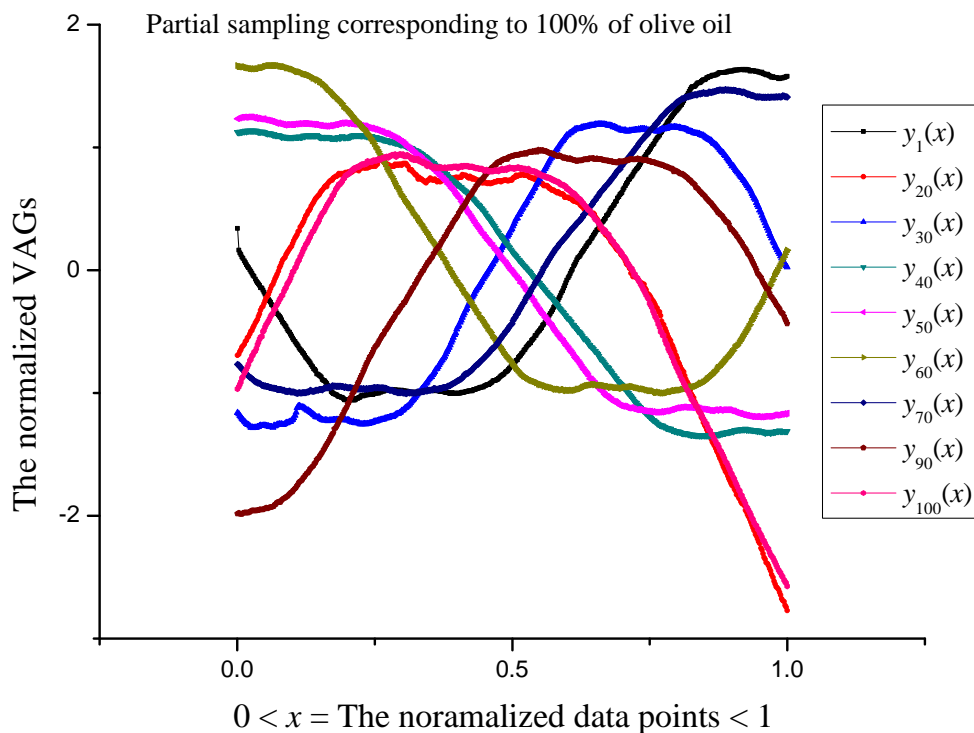


Figure 4. Here we demonstrate the selected sampling of the VAGs corresponding to 100% of olive oil.

4.2. Verification of expressions (7) on real data

As one can notice from the previous analysis, the creation of a stable “pattern”/ideal background plays a *key* role in any original research. If the reliable and sensitive background can be created then a researcher will obtain a sensitive instrument for detection of a “strange” signal/reality that can appear in the given background. I asked doctor of chemical science Artem V. Sidelnikov from Ufa Petroleum Institution to send me some electrochemical data for short analysis. He asked his colleagues to perform a typical electrochemical experiment: to choose a background solute and add some “ x ” substance for its detection. Actually, he and his colleagues sent me VAGs of two experiments:

Case 1. The background contained 100 measured voltammograms (VAGs) corresponding to 100% pure olive oil. Another experiment contained the mixture between 80% of olive oil and 20% of rape oil. Other parameters were not so important for analysis. Only one basic question is remained: is it possible to notice the difference/direct correlations between pure olive oil chosen as a background (in our notations it is defined as an array $y_m(x)$) and mixture $Y_m(x)$ when 20% of olive oil was replaced by 20% of rape/(rapeseed) oil?

Case 2. As a background, the selected carbon electrode located near the Faraday region measured the VAGs of a background (the conventional buffer solution was chosen). Then after 100 measurements, a small amount of water was added (about 1%) and the “modified” background with water was measured again.

Two experiments were performed in the *same* experimental conditions. It implies that any surrounding parameters as: temperature (T), humidity (H), pressure (P) keep their given values.

Is it possible to notice possible changes evoked by the presence of “ x ” substance inside the given background, if each experiment was repeated $M = 100$ times? This repetition is necessary for the creation of the representative array.

In this real experiment we associate the background array with the functions $y_m(x)$ ($L = 1$), where $y_m = J_m$ is the registered current, $x =$ (normalized data points from the interval $[0, 1]$) and arrays $Y_m(x)$ for cases 2 and 3 (as affected arrays with the presence of “ x ” substance). For this case, it is sufficient to use simple expressions (7) as in the previous mimic experiment explained above. For both experiments, we used the normalized VAGs, which are defined as

$$i_m(x) = \frac{J_m(x) - \langle J(x) \rangle}{\text{stdev}(J_m(x))},$$

$$\langle J(x) \rangle = \frac{1}{M} \sum_{m=1}^M J_m(x).$$
(30)

The figures 4-8 demonstrate the details that are explained in the corresponding captions. The most important figures, from our point of view are figures 6 and 8, where we show two different behaviors of the functions $\langle G(x) \rangle$ and their mean values $\langle Y(x) \rangle$ averaged over the arrays containing “ x ” substance. For the first experiment, we observe *weak* correlations between two sampling compared, while for the second experiment we observe the *strong* correlations between them. The proposed approach forces to reconsider the experimental results that are obtained in the result of this independent analysis, when a competitive theoretical hypothesis is used for the fitting of experimental data. We omit the figures demonstrating the corresponding arrays $G_m(x)$ because they are unimportant for this case.

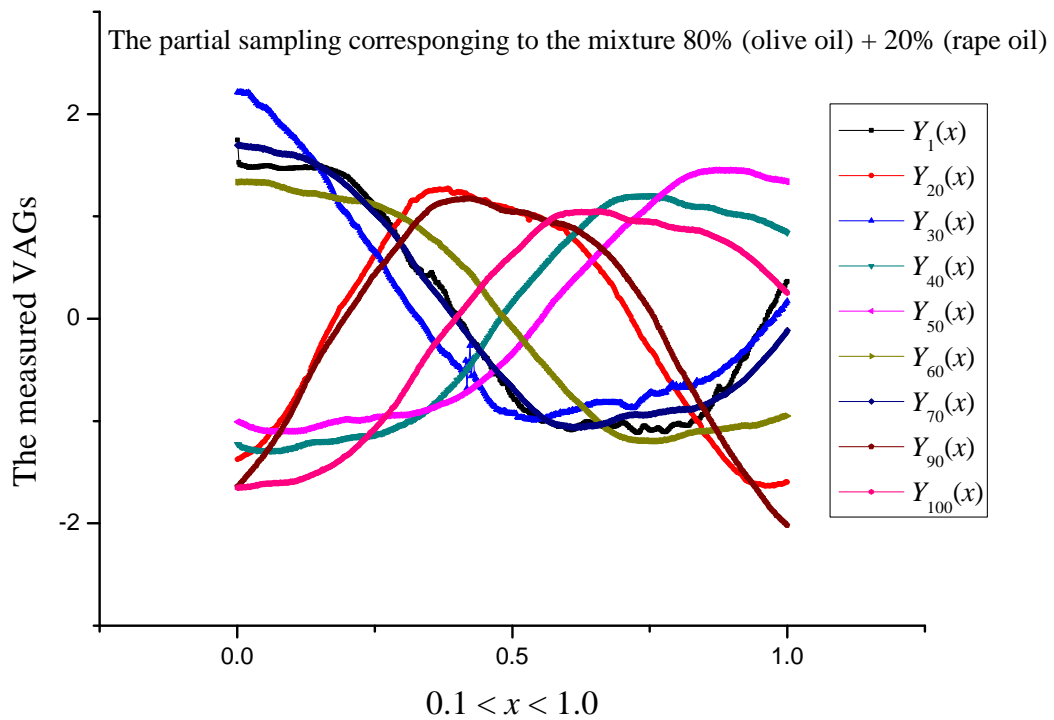


Figure 5. For comparison with the previous figure we show the VAGs corresponding to the mixture: 80% olive oil and 20% of the rapeseed oil. All VAGs are normalized with accordance with expression (30).

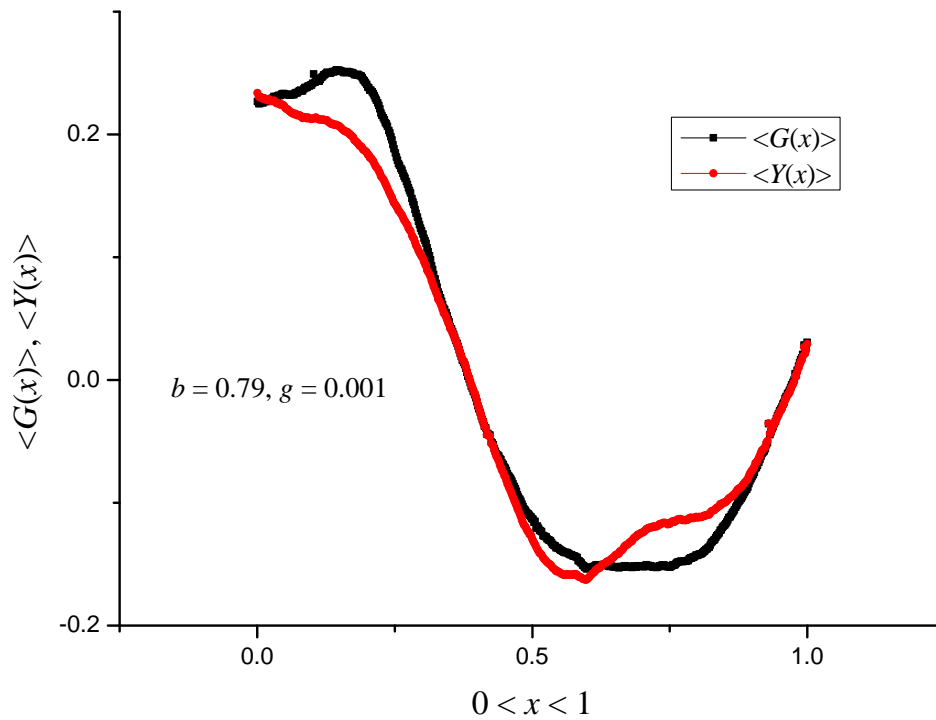


Figure 6a. As one can notice from the central figure the curve $\langle G(x) \rangle$ (in black color) close to the mean curve $\langle Y(x) \rangle$ (red color). It means that the correlation between the VAGs corresponding to 100% and the mixture is weak (these oil mixtures can be considered as the independent from each other).

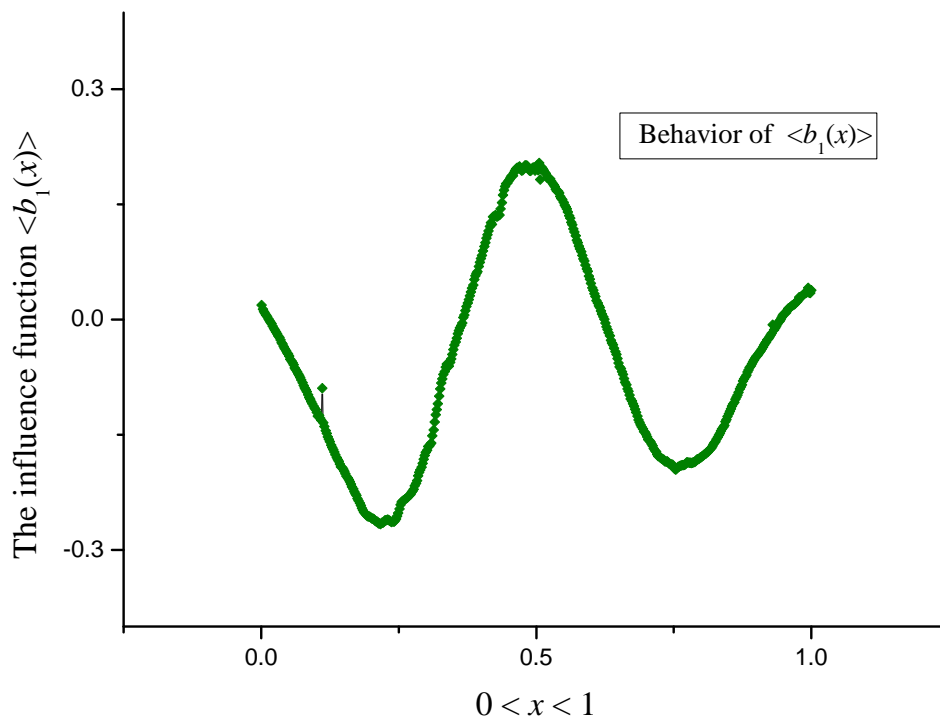


Figure 6b. It is confirmed also by a weak dependence of the influence function $\langle b_1(x) \rangle$ depicted in this figure.

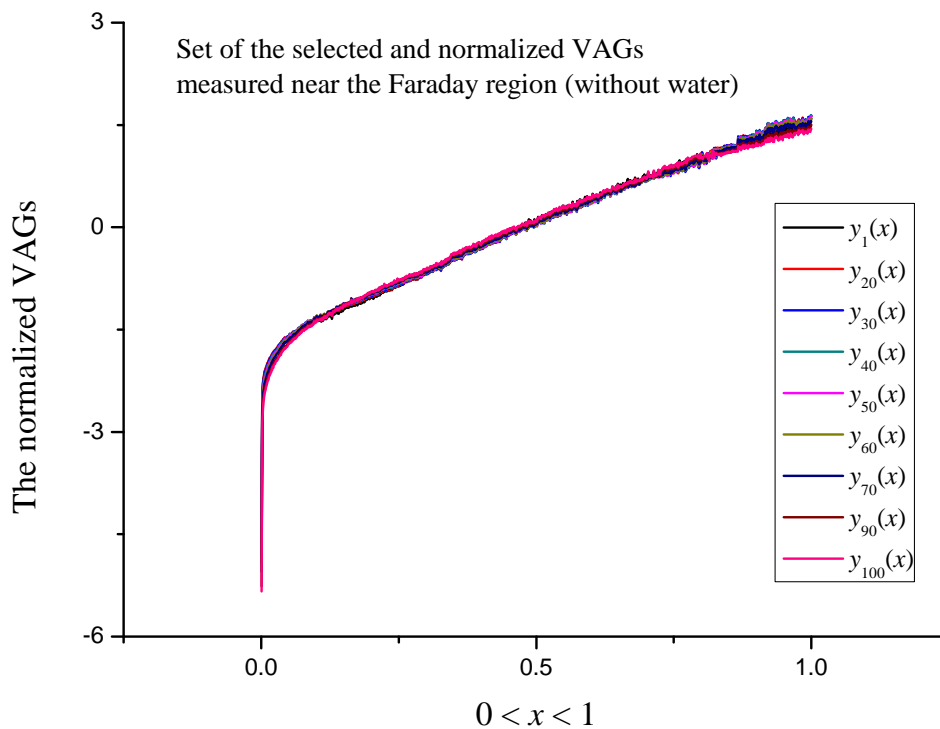


Figure 7a. Here we demonstrate the selected samplings corresponding to the measured VAGs in the given background solute for the Faraday region without water.

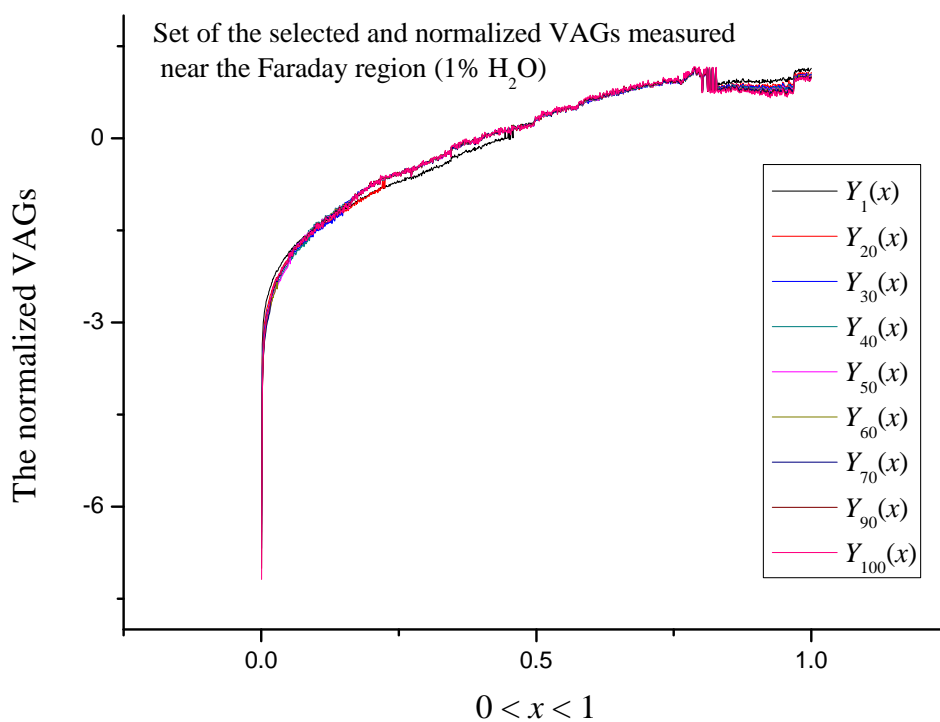


Figure 7b. VAGs distorted by an addition of a small amount of water (about 1%) shown in this figure. One can notice some small distortions by an inexperienced eye and concludes that the difference between these two arrays (compare it with the neighboring Fig. 7a) will be small.

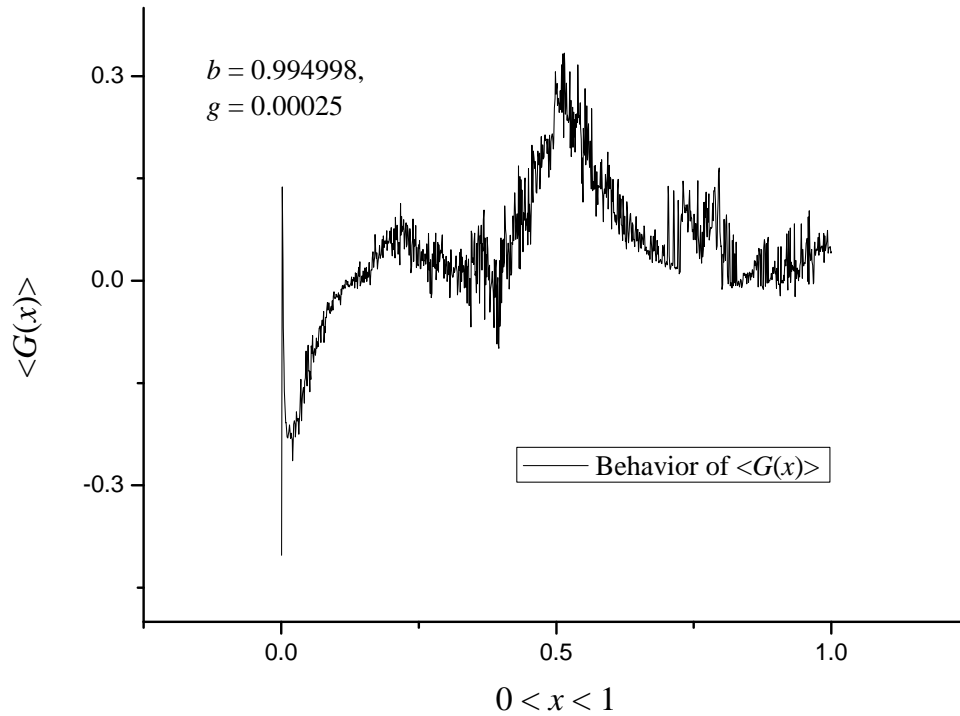


Figure 8a. This figure can be considered as the central one in the frame of the second experiment. Here we show the small signal corresponding to “water” influence. It is small; however, it can be detected.

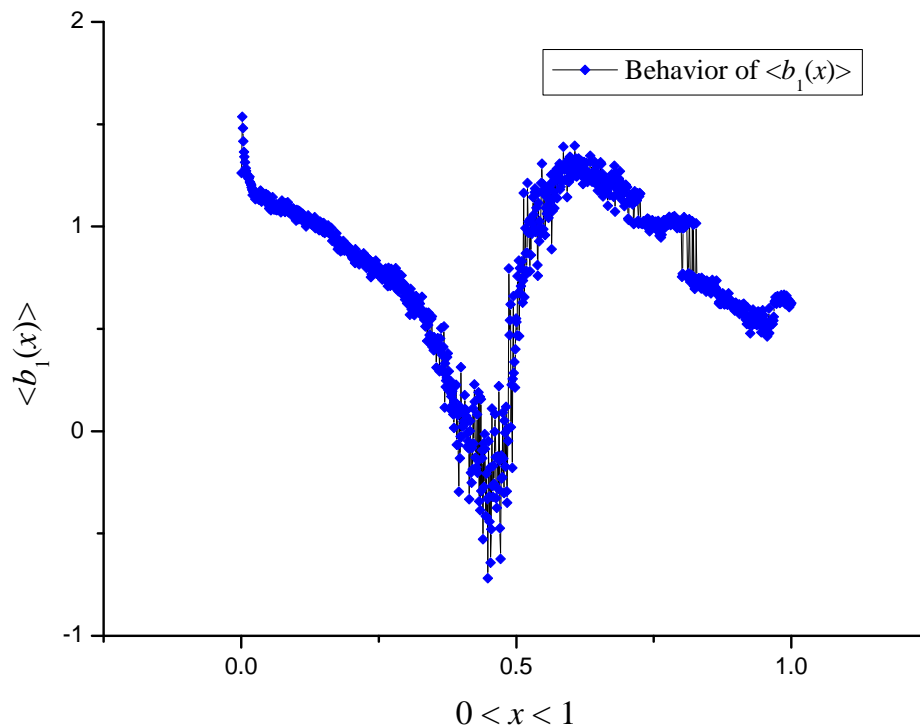


Figure 8b. The behavior of the influence function $\langle b_1(x) \rangle$ reflecting the strong correlations by the initial array $y_m^{(1)}(x)$ is shown in this figure. Comparison with functions $Y_m(x)$ and $\langle Y(x) \rangle$ depicted in Fig. 7b shows that they are different and, therefore, these two arrays are strongly correlated with other.

5. Analysis of the obtained results

In conclusion, we want to stress the advantages of the proposed approach:

1. We generalized the existing expressions related to analysis of correlations and, in the first time, we obtain the solution in explicit form that allows to separate *directly* the correlated factor (taken in the form of background array) from a “mixture” of other factors and to obtain the desired uncorrelated/remnant part $G_m(x)$ in the form of an array.
2. Expression (1) can be generalized and allows considering the factors having a temporal memory (18) and nonlinear factors that are related by the cause-and-effects relations (expressions (15) and (16)).
3. In the limiting cases the general expression (1) is reduced to the linear least square method (when $\langle b_l(x) \rangle$ can be replaced by the constant values) and, in another case (when we have approximately $Y_m(x) \cong \langle Y(x) \rangle$) it restores directly the uncorrelated part $G_m(x)$, when the degree of correlations becomes negligible $y_m^{(l)}(t) \cong 0$.
4. The examples based on mimic and real data show and confirm some important details that can be obtained in the result of the proposed approach. Special attention should be paid for the receiving of the stable/robust *background* serving as a specific detector for registration of any small signal that could appear in the result of an unexpected “ x ” perturbation. This registered perturbation can help in detection of the desired “signal” (amount of water in the second experiment) that separates this new “quality” from the random fluctuations containing in the “pattern” background.
5. New approach can be helpful in selection of the array that can be associated with the correlated cluster or not. Really, if we obtain the result, when $\langle b_l(x) \rangle \cong 1$, $\langle G(x) \rangle \cong 0$, then it implies that the array $Y_m(x)$ is strongly correlated with the compared arrays $y_m^{(l)}(x)$, while for the opposite case $\langle b_l(x) \rangle \cong 0$, $\langle G(x) \rangle \cong \langle Y(x) \rangle$ these arrays can be considered as almost independent from each other. This simple comparison adds more certain information to the conventional analysis based on the Pearson correlation analysis and the statistics of the fractional moments [16, 18, 19].

It is interesting to note also that the first experiment demonstrates some *unexpected* result: the VAGs obtained for the pure olive oil and for the mixture of two oils do *not* correlated with each other. It means that the oil mixture forms an independent chemical substance that differs from simple mechanical mixture of two oils. For the second experiment, we obtain also important result; the proposed approach allows to extract a “signal” corresponding to a small amount of water (Fig. 8) and proves that VAGs obtained for the both cases are strongly correlated.

Acknowledgments

I would like to express my gratitude to Dr. Artem V. Sidelnikov from Technology Department of Ufa State Petroleum Technological University (USPTU) for the electrochemical data that were sent to me for this analysis.

References

1. Haghghat M., Abdel-Mottaleb M., Alhalabi W. *IEEE Transactions on Information Forensics and Security* **11**, 1984 (2016)
2. Sen P.K. *Journal of the American Statistical Association* **81**(394), 560 (1986)
3. Schervish M.J. *Statistical Science* **4**, 396 (1987)
4. Anderson T.W. *An Introduction to Multivariate Statistical Analysis*, Wiley, New York (1958)

5. Mardia K.V., Kent J.T., Bibby J.M. *Multivariate Analysis*, Academic Press (1979)
6. Feinstein A.R. *Multivariable Analysis*, New Haven, CT: Yale University Press (1996)
7. Hair J.F.Jr. *Multivariate Data Analysis with Readings*, 4th ed., Prentice Hall (1995)
8. Johnson R.A., Wichern D.W. *Applied Multivariate Statistical Analysis*, 6th ed., Prentice Hall (2007)
9. Schafer J.L. *Analysis of Incomplete Multivariate Data*, CRC Press (1997); Saeed V. *Advanced Digital Signal Processing and Noise Reduction*, 2nd ed., John Wiley & Sons, Ltd (2001)
10. Haghighat M., Abdel-Mottaleb M., Alhalabi W. *IEEE Transactions on Information Forensics and Security* **11**(9), 1984 (2016)
11. Haghighat M., Abdel-Mottaleb M., Alhalabi W. *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* 1866 (2016)
12. Lee Y.W. Cheatham T.P.Jr., Wiesner J.B. *Proceedings of the IRE* **38**(10), 1165 (1950)
13. Buciu I., Gacsadi A. *IEEE, Proceedings ELMAR*, Zadar, 21 (2011)
14. Boashash B. *IEEE Transactions on Acoustics, Speech, and Signal Processing* **36**(9), 1518 (1988)
15. Boashash B. *Proceedings of the IEEE* **80**(4), 519 (1992)
16. Nigmatullin R.R., Maione G., Lino P., Saponaro F., Zhang. W. *Communications in Nonlinear Science and Numerical Simulation* **42**, 324 (2017)
17. Nigmatullin R.R. in *Complex Motions and Chaos in Nonlinear Systems, Nonlinear Systems and Complexity*, edited by V. Afraimovich et al., Vol. **15**, Ch. 1, p. 1, Springer (2016)
18. Nigmatullin R.R., Toboev V.A., Lino P., Maione G. *Chaos, Solitons & Fractals* **76**, 166 (2015)
19. Nigmatullin R.R., Ceglie C., Maione G., Striccoli D. *Nonlinear Dynamics* **80**(4), 1869 (2015)
20. Uchaikin V.V. *The Method of the Fractional Derivatives*, Chapter 5, “Artishok” Publ. House (2008) (*in Russian*)