ISSN 1995-0802, Lobachevskii Journal of Mathematics, 2024, Vol. 45, No. 8, pp. 3956–3971. © Pleiades Publishing, Ltd., 2024.

Quasi-reproducible Experiments: Universal Fitting Function for Quantitative Description of Complex Systems Data

R. R. Nigmatullin[1]*, A. A. Litvinov[1,2], and S. I. Osokin[2]*****

(Submitted by A. M. Elizarov)

[1] *Tupolev Kazan National Research Technical University, Kazan, 420111 Russia*

[2] *Kazan Federal University, Institute of Information Technologies and Intellectual Systems, Kazan, 420008 Russia*

Received May 22, 2024; revised June 28, 2024; accepted July 10, 2024

Abstract—This paper provides the foundations of an original theory of quasi-reproducible experiments (QRE) based on the testable hypothesis that there exists an essential correlation (memory) between successive measurements. Based on this hypothesis, which the authors define for brevity as the verified partial correlation principle (VPCP), it can be proved that there exists a universal fitting function (UFF) for quasi-periodic (QP) and quasi-reproducible (QR) measurements. In other words, there is some common platform or "bridge" on which, figuratively speaking, a true theory (claiming to describe data from first principles or verifiable models) and an experiment offering this theory for verification measured data, maximally "cleaned" from the influence of uncontrollable factors and apparatus/soft wire function, meet. The proposed theory has been tested on eddy covariance ecological data, specifically measuring only the concentration of CH_4, CO_2 and water vapors of H_2O in the local atmosphere where the corresponding detectors for measuring of the desired gases content are located. For these tested eddy covariance data associated with the presence of two gases CH_4, CO_2 and H_2O vapors in atmosphere there is no simple hypothesis containing a minimal number of the fitting parameters, and, therefore, the fitting function that follows from this theory can serve as the only and reliable quantitative description of this kind of data belonging to the tested complex system. Applications of this theory to practical applications, the place of this theory among other alternative approaches, (especially touching the professional interests of ecologists) and its further development are discussed at the end of this paper.

DOI: 10.1134/S1995080224604739

Keywords and phrases: *quasi-reproducible experiments, complex systems, verified partial correlation principle, universal fitting function, quasi-periodic measurements, quasi-reproducible measurements, memory effects, eddy covariance*

The list of the main abbreviations

They are given in the alphabetic order: AFR—amplitude-frequency response; (F)LLSM—(functional) linear least square method; GPD—the generalized Prony decomposition; IM—Intermediate Model; IE—"Ideal Experiment"; QRE(s)—Quasi-reproducible experiment(s); QPE(s)—Quasi-periodic experiment(s); VPCP—Verified partial correlation principle.

1. INTRODUCTION AND FORMULATION OF THE PROBLEM

Is it possible to construct a "universal" fitting function for a nonstationary QR experiment? Such a question will seem absurd and meaningless to any experienced researcher. Everybody knows how the traditional interaction between theory and experiment takes place. Theory proposes models, hypotheses based on some assumptions and postulates. The experiment, in turn, tests these hypotheses, trying

*E-mail: renigmat@gmail.com

**E-mail: litvinov85@gmail.com

***E-mail: s.osokin@it.kfu.ru

to eliminate to the maximum extent possible the influence of uncontrollable factors and distortions (interferences) introduced by the measuring equipment (it is usually defined as a hardware/apparatus function). What can be introduced fundamentally new to this traditional scheme? What if we imagine that it could find some verified or tested principle that virtually all measurements obey? If such a principle is found, then from its mathematical formulation it would be possible to derive some "universal" fitting function, which would make it possible to describe all measurements. The authors define this principle as the verified partial correlation principle (VPCP). Thus, based on the VPCP, it will be possible to obtain a general mathematical model (defined as an intermediate model—IM) to which all measurements satisfying this principle will have to obey. What is this principle that most measurements satisfy? If we unfold the VPCP, it is formulated as follows: successive measurements retain partial correlation (memory) among themselves and remain fully or partially correlated as a result of a series of successive measurements. Of course, here it is necessary to make clarifications about the nature and type of these measurements and to translate this imprecise verbal formulation into a strict mathematical language. Therefore, it is necessary to first conduct several concepts and then translate this principle into the language of mathematical formulae. By an "ideal" experiment (IE) the authors understand such an experiment when a sequence of measurements m ($m = 0, 1, 2, ..., M-1$) carried out for some average period T with respect to the controlling external variable x leads to the same value of the measured response function $F(x)$. In this sense, all measurements corresponding to an IE are completely or absolutely correlated. Mathematically, this statement looks as follows

$$F(x + mT) = F(x), \quad m = 0, 1, \ldots, M-1. \tag{1}$$

Here the controlled (manipulated) variable x can coincide with a time variable (t), frequency (ω), wavelength (λ), etc. Since the experiment performed on this set of variables is a single-factor experiment, the other controlled variables affecting the response function are assumed to be unchanged within some range of their values during the single-factor experiment. The solution to this functional equation (1) is a periodic segment of the corresponding Fourier series. For discrete data, the Fourier series segment, is usually written in the form

$$F(x) \cong \Pr(x) = A_0 + \sum_{k=1}^{K \gg 1} \left[Ac_k \cos\left(2\pi k \frac{x}{T}\right) n + As_k \sin\left(2\pi k \frac{x}{T}\right) \right]. \tag{2}$$

The parameter T defines some average measurement period with respect to the input variable x. It follows from this simple equation (1) that expression (2) can be used as a fitting function for the response function in the IM. In this idealized case, the intermediate model (IM) coincides with the solution of the Fourier series segment, and the coefficients of this expansion can act as fitting parameters corresponding to the IE. Actually, the fitting parameters can form the desired AFR. It is quite obvious that the IE requirement (1) is not realized in reality and analysis of various data shows that instead of equation (1) a more general functional equation should be written

$$F(x + LT) = \sum_{l=0}^{L-1} a_l F(x + lT) + b. \tag{3}$$

It can be interpreted as follows: starting from some measurement L ($L < M$), some measurements become partially correlated and the realized experiment cannot generate fundamentally new measurements. "New" measurements (starting from some L, $M > L$) become already dependent (correlated) on the previous/"past" ones realized earlier, and the desired set of correlation coefficients a_l ($l = 0, 1, ..., L-1$) can be found using the least squares method (LLSM). The solutions of this functional equation are expressed as a generalized Prony decomposition (GPD). We consider this a fundamental finding, as it shows the *inapplicability* of the formal Fourier decomposition for analyzing real data. The adequate IM for data where the set of constants a_l ($l = 0, 1, ..., L-1$) does not depend on the external variable x, is exactly the solution of equation (3), acting as an alternative fitting function. Before giving the general solution of equation (3) it is instructive to consider the partial case. Let $l = 0$, $L = 1$ in (3). Therefore, we have

$$F(x + T) = a_0 F(x) + b. \tag{4}$$

Solution of (4) can be written as

$$F(x)=\left(\begin{array}{cc}(a_0)^{x/T}\mathrm{Pr}(x)+b/(1-a_0), & a_0>0,\ a_0\neq 1\\ |a_0|^{x/T}\cos\left(\pi\frac{x}{T}\right)\mathrm{Pr}(x)+b/(1+a_0), & a_0<0,\ a_0\neq -1,\\ \mathrm{Pr}(x)+b\left(\frac{x}{T}\right), & a_0=1.\end{array}\right). \tag{5}$$

Here $\mathrm{Pr}(x)$ is a periodic function that is determined by expression (2). The most interesting conclusion follows for $a_0>1$. In this case, the exponential argument in (5) $\ln(a_0)>0$ and, finally, we obtain the increasing function. In means that the next measurement in (4) becomes important, while for $a_0<1$ the next measurement becomes decreasing. Even the appearance of a constant b at $a_0=1$ leads to increasing solution proportional to x/T. How to understand (3) from another point of view? Let to consider the following case

$$F(x)=f_1(x)+f_2(x), \tag{6}$$

where functions $f_1(x)$ and $f_2(x)$ in (6) are periodic with the same period T (for simplicity ones put $b=0$). Taking into account equation (4) one can write

$$\begin{gathered}E_TF(x)\equiv F(x+T)=\lambda_1f_1(x)+\lambda_2f_2(x),\\ (E_T)^2F(x)\equiv F(x+2T)=(\lambda_1)^2f_1(x)+(\lambda_2)^2f_2(x).\end{gathered} \tag{7}$$

Here we introduce the shifting operator E_T associated with period T. With the help of this operator one can rewrite the last line in (7) in very compact and clear form

$$(E_T-\lambda_1)(E_T-\lambda_2)F(x)=0\quad\text{or}\quad F(x+2T)=a_1F(x+T)+a_0F(x). \tag{8}$$

Therefore, equation (3) can be presented in the equivalent form with the help of application of the shifting operators E_T as

$$F(x+LT)=\prod_{l=0}^{L-1}(E_T-\lambda_l)F(x)+b. \tag{9}$$

Multiplying the terms in (9) it is easy to relate the set of the roots λ_l ($l=0,1,\ldots,L-1$) with coefficients a_l using Vyet's theorem, relating the coefficients of a polynomial to their roots. It makes sense to give the solution of (3) in full, devoted to the consideration of this nontrivial problem. The solution of equation (3) has the form

$$\begin{gathered}(A)\ \sum_{l=0}^{L-1}a_l\neq 1,\quad F(x)=\sum_{l=1}^{L}[\lambda_l]^{x/T}\mathrm{Pr}_l(x)+\frac{b}{1-\sum_{l=0}^{L-1}a_l},\\ (B)\ \sum_{l=0}^{L-1}a_l=1,\quad F(x)=\sum_{l=1}^{L}[\lambda_l]^{x/T}\mathrm{Pr}_l(x)+\frac{b}{L-\sum_{l=0}^{L-1}la_l}\frac{x}{T}.\end{gathered} \tag{10}$$

Here the functions $\mathrm{Pr}_l(x)$ are the set of periodic functions (2), the set of constants λ_l ($l=1,2,\ldots,L$) is found as roots of the polynomial from equation

$$P_L(\lambda)\equiv\lambda^L-\sum_{l=0}^{L-1}a_l\lambda^l=0. \tag{11}$$

Here we do not give solutions for complex-conjugate roots and the degenerate case (when the polynomial $P_L(\lambda)$ has congruent roots); they are given in the publications [4, 5] cited above. We only note that for the case of a negative root ($\lambda_s<0$) the periodic function $\mathrm{Pr}_s(x)$ becomes already anti-periodic, and the solution for this case can be given in the form of the expression

$$F(x)=\Big[|\kappa_s|\Big]^{x/T}\cos\left(\pi\frac{x}{T}\right)\mathrm{Pr}_s(x). \tag{12}$$

Thus, for the case of QP measurements satisfying equation (4), the GPD segment (not Fourier!) written in the form (4) acts as the IM. Further development of this idea and its verification on available data were

given in [5]. In particular, there it was shown how to go beyond the acceptable observation interval given by the variable x and to suppress the partial memory influence occurring between measurements in order to extract from the solutions of (4) only purely periodic functions that can correspond to IE and be presented to a theoretical model (hypothesis) for verification. Of course, hypothesis (3) has some limit of applicability, consisting in the fact that the set of parameters a_l $(l = 0, 1, \ldots, L-1)$ does not change during the experiment. In reality, in many experiments this set can change with the changing of the external variable x, and the number of measurements in many experiments cannot be large. Therefore, in order to go beyond equation (3), it is necessary to obtain analytical solutions for the case when $a_l \to a_l(x)$ and to find the minimal limit of the number of measurements when additional information expressed in the form of repeatability of measurements is important for finding the fitting function following from IM. Recently, these solutions have been found and they are the ones that allow making the proposed theory more flexible and adaptable to describe non-stationary experiments, when the influence of uncontrollable factors in the measurement process may be significant. Therefore, it makes sense to introduce the following distinctions in the terminology used. Under quasi-periodic (QP) measurements we understand such measurements, when the experimental conditions allow to keep the uncontrolled parameters strictly within the specified limits of permissible intervals and, therefore, they can be considered approximately stationary (a_l are stayed const). Under quasi-reproducible (QR) measurements we understand such measurements when the influence of uncontrollable factors becomes *significant* ($a_l(x)$ depends on the current variable x) and the experimental conditions are already non-stationary. In this case, it becomes necessary to find solutions of a more general functional equation

$$F(x + LT) = \sum_{l=0}^{L-1} \langle a_l(x) \rangle F(x + lT). \tag{13}$$

It turns out that it is possible to find analytical solutions of equation (13) for a wide class of functions satisfying the periodicity condition. It is to outline the foundations of this more general theory, based on a series of consecutive measurements, applicable to the description of QR experiments, its possible generalizations and verification on available data, that this paper is mainly devoted. Of course, any theory will be incomplete if it has not been tested by experiment. Therefore, the authors give a non-trivial example based on the description of eddy covariance data in the frame of the proposed theory. Moreover, the proposed algorithms are general enough, they can be applied to analyze many similar QR experiments too.

2. BASIC THEORY OF THE QR EXPERIMENTS

2.1. Self-Consistent Solutions of Equation (13)

We can obtain solutions of the functional equation (13) provided that the "length" L, which characterizes the memory between measurements, is assumed to be known. So, let us assume that all consecutive measurements satisfy the equation

$$F_{L+m}(x) = \sum_{l=0}^{L-1} \langle a_l(x) \rangle F_{l+m}(x), \quad m = 0, 1, \ldots, M-1. \tag{14}$$

In order to find the unknown functions $\langle a_l(x) \rangle$ $(l = 0, 1, \ldots, L;\ L < M)$, one can generalize the LLSM and require that the functional dispersion accepts the minimal value

$$\sigma(x) = \frac{1}{M-L} \sum_{m=0}^{M-L-1} \left[F_{L+m}(x) - \sum_{l=0}^{L-1} \langle a_l(x) \rangle F_{l+m}(x) \right]^2 = \min. \tag{15}$$

In order to get the desired solution, it is necessary to take the mean value over the remaining measurements $(l = 0, 1, \ldots, M-L-1,\ L;\ L < M)$. Taking the functional derivatives with respect to unknown functions $\langle a_l(x) \rangle$, we obtain

$$-\frac{\delta\sigma(x)}{\delta\langle a_l(x) \rangle} = \frac{1}{M-L} \sum_{m=0}^{M-L-1} \left[F_{l+m}(x) \left(F_{L+m}(x) - \sum_{s=0}^{L-1} \langle a_s(x) \rangle F_{s+m}(x) \right) \right] = 0. \tag{16}$$

Here we apply also the averaging procedure over all set of admissible measurements supposing that the set of the functions $\langle a_l(x)\rangle$ ($l=0,1,\ldots,L$; $L<M$) does not depend on the index m. Introducing the definitions of the pair correlation functions

$$K_{L,l}=\frac{1}{M-L}\sum_{m=0}^{M-L-1}F_{L+m}(x)F_{l+m}(x),\quad K_{s,l}=\frac{1}{M-L}\sum_{m=0}^{M-L-1}F_{s+m}(x)F_{l+m}(x),\qquad(17)$$

$s,l=0,1,\ldots,L-1$, one can receive the system of linear equations for the calculation of unknown functions $\langle a_l(x)\rangle$

$$\sum_{s=0}^{L-1}K_{s,l}(x)\langle a_s(x)\rangle=K_{L,l}(x).\qquad(18)$$

It makes sense to define this procedure as the functional linear least squares method (FLLSM), which includes the ordinary LLSM as a partial case. Now let us return to equation (13). The solution of this equation is sought in the form

$$F_0(x)=\Big[\kappa(x)\Big]^{x/T}\mathrm{Pr}(x),\qquad F_m(x)=\Big[\kappa(x)\Big]^{m+x/T}\mathrm{Pr}(x).\qquad(19)$$

The functions $\langle a_l(x)\rangle$, $\kappa(x\pm T)=\kappa(x)$, $\mathrm{Pr}(x\pm T)=\mathrm{Pr}(x)$ in accordance with the suppositions made above can be expressed approximately by the segment of the Fourier series in complete analogy with expression (2)

$$\Phi(x)=A_0+\sum_{k=1}^{K\gg1}\left[Ac_k\cos\left(2\pi k\frac{x}{T}\right)+As_k\sin\left(2\pi k\frac{x}{T}\right)\right].\qquad(20)$$

It is obvious that the decomposition coefficients Ac_k, As_k ($k=1,2,\ldots,K$) depends on the specific form of the decomposed function. Inserting the probe/inoculating solution (16) in equation (14), we derive equation for calculation of unknown functions $\kappa(x)$ figuring in (19)

$$\Big[\kappa(x)\Big]^{L}-\sum_{l=0}^{L-1}\langle a_l(x)\rangle\Big[\kappa(x)\Big]^{l}=0.\qquad(21)$$

If the functional roots $\kappa_q(x)$, $q=1,2,\ldots,L$ can be calculated from (21), then the general solution for the function $F_m(x)$ is written in the form

$$F_0(x)=\sum_{q=1}^{L}\Big[\kappa_q(x)\Big]^{x/T}\mathrm{Pr}_q(x),\quad F_m(x)=\sum_{q=1}^{L}\Big[\kappa_q(x)\Big]^{m+(x/T)}\mathrm{Pr}_q(x),\quad m=0,1,...,M-1.\quad(22)$$

The number of periodic functions $\mathrm{Pr}_q(x)$ should coincide with number of functions $\kappa_q(x)$, $q=1,2,\ldots,L$, entering in the last expression (22). It is this expression that can be considered as a general solution of the functional equation (8). This solution can be interpreted as follows: if successive measurements are partially correlated with each other ("remember" each other) and can vary during the average measurement period T, then the fitting function to describe these measurements is self-consistent and is determined by the whole set of random measurements taking part in this process. Obviously, this new result generalizes previous results [5] obtained for the case when the functions $\langle a_l(x)\rangle$ can be approximated by constants a_l. It would be desirable to obtain solutions of equation (14) for the case when the functions $\langle a_l(x)\rangle$ are not completely periodic or obtained *a priori* from other conditions. But, to the best of the authors knowledge, the mathematical theory of solutions of functional equations is practically undeveloped [6] compared, for example, with a full-fledged theory of solutions of differential or integral equations. Therefore, the proposed theory defines a new direction for mathematicians working in the field of functional analysis and personally aiming at applications of their results in physics, chemistry and engineering. For practical applications it makes sense to consider in more detail the case of short memory with ($L=2$), since the number of fitting parameters for this case is minimal. As one can see below the long memory case $L>2$ can be reduced also to the short memory

case also. It is the results for this case that will be needed to describe real measurements, which are given in the next section. For the case of short memory when $L = 2$ we obtain

$$F_{2+m}(x) = \langle a_1(x)\rangle F_{1+m} + \langle a_0(x)\rangle F_m, \quad m = 0, 1, \ldots, M-1. \tag{23}$$

Equation (18) for this case accepts the form

$$\begin{aligned} K_{00}(x)\langle a_0(x)\rangle + K_{10}(x)\langle a_1(x)\rangle &= K_{20}(x), \\ K_{10}(x)\langle a_0(x)\rangle + K_{11}(x)\langle a_1(x)\rangle &= K_{21}(x). \end{aligned} \tag{24}$$

Solution of equation (23) is written as

$$F_0(x) = \Big[\kappa_1(x)\Big]^{x/T}\mathrm{Pr}_1(x) + \Big[\kappa_2(x)\Big]^{x/T}\mathrm{Pr}_2(x), \quad \kappa_{1,2} = \frac{\langle a_1(x)\rangle}{2} \pm \sqrt{\left(\frac{\langle a_1(x)\rangle}{2}\right)^2 + \langle a_0(x)\rangle}. \tag{25}$$

If one of the roots in (25) becomes negative (for example, $\kappa_2(x) < 0$), then the general solution for this case (with extraction of the real part) cab be written as

$$F_0(x) = \Big[\kappa_1(x)\Big]^{x/T}\mathrm{Pr}_1(x) + \Big[|\kappa_2(x)|\Big]^{x/T}\cos\left(\pi\frac{x}{T}\right)\mathrm{Pr}_2(x). \tag{26}$$

If the order of measurements to assess the effect of nonstationary of the process as a whole is significant, then the proposed theory allows us to recover the whole nonstationary sequence according to the relations:

$$\begin{aligned} F_m(x) &= \Big[\kappa_1(x)\Big]^{m+(x/T)}\mathrm{Pr}_1(x) + \Big[\kappa_2(x)\Big]^{m+(x/T)}\cos\Big(\pi(\frac{x}{T}+m)\Big)\mathrm{Pr}_2(x), \\ \mathrm{Pr}_{1,2}(x) &= A_0 + \sum_{k=1}^{K_m}\left[Ac_{k,1,2}^{(m)}\cos\left(2\pi k\frac{x}{T}\right) + As_{k,1,2}^{(m)}\sin\left(2\pi k\frac{x}{T}\right)\right], \quad m = 0, 1, \ldots, M-1. \end{aligned} \tag{27}$$

Here the functions $\mathrm{Pr}_{1,2}(x \pm T) = \mathrm{Pr}_{1,2}(x)$ keep their periodicity over the mean period T, however, the decomposition coefficients $Ac_{k,1,2}^{(m)}$ and $As_{k,1,2}^{(m)}$ $(k = 1, 2, \ldots, K_m)$, figuring in (27), can be differed from the case $m = 0$ and reflect the influence of possible instability during the whole measurement process. If the true sequence of measurements is not essential and the results of measurements remain invariant with respect to permutations of all measurements with each other, one can group all measurements into three independent groups (a specific triad), and the case of long memory is reduced to the short memory case again, considered above. This simple idea allows us to significantly reduce the number of fitting parameters and to obtain again a fitting function with a minimum number of fitting parameters. This procedure related to the formation of the necessary triad is described in the next subsection.

2.2. Clusterization Procedure and Reduction To an "Ideal Experiment"

As it was emphasized earlier in [5], the estimation of the "true" value of L based on a common criterion is an *unsolved* problem. If the assumption regarding permutations of measurements with each other can be justified and seems quite reasonable, one can propose the following procedure for clustering and partitioning all measurements into three groups (creating a specific triad). For this purpose, we will consider the distribution of slopes (angle tangents) of each measurement with respect to their average measurement whose angle tangent is equal or close to one

$$Sl_m = slope(\langle y\rangle, y_m) \equiv \frac{(y_m \cdot \langle y\rangle)}{(\langle y\rangle \cdot \langle y\rangle)}, \quad \langle y\rangle = \frac{1}{M}\sum_{m=0}^{M-1} y_m, \quad (A \cdot B) = \sum_{j=1}^{N} A_j B_j. \tag{28}$$

The bracket in (28) defines the scalar product between two functions having $j = 1, 2, \ldots, N$ data measured points. We assume that the random measurements $y_m(x)$ for $m = 0, 1, \ldots, M-1$ approximate the functions $F_m(x)$ $(y_m(x) \cong F_m(x))$, appearing in equation (27). If we construct a random distribution slope function Sl_m depending on the measurement number m, then arrange all measurements in descending order $Sl_0 > Sl_1 > \cdots > Sl_{M-1}$. This distribution slope function can be divided into three groups. The top group of measurements "up" has slopes localized in the interval $(1 + \Delta, \max(Sl_m))$;

the middle group (defined as "mn") contains measurements with slopes in the interval $(1-\Delta, 1+\Delta)$; and finally, the bottom group (denoted as "dn") contains measurements with slopes $(1-\Delta, \min(Sl_m))$. The value Δ for each set of the QR measurements is defined independently in each specific case. This ordered curve Sl_m is important and reflects the quality of the performed measurements and the quality of used equipment. How to find Δ based on expression (28)? One can divide the ordered curve Sl_m after subtraction the unit value on two parts—positive part $(0, \max(Sl_m - 1)$ and the negative part $(\min(Sl_m - 1), 0)$. In each part one takes the halves of each selected part, i.e., $\Delta_1 = \max(Sl_m - 1)/2$ and $\Delta_2 = \min(Sl_m - 1)/2$. These values can be used for division on the three desired parts/clusters.

The first group is formed as: (a) the distance from the initial point of the BLC (0,0) to the first point of intersection $(m_1, 1+\Delta_1)$ determines the number of measurements $Nup(m = 1, 2, \ldots, m_1 = Nup)$ falling into the first (upper) group, they are characterized by the average $Yup(x)$ curve; (b) the distance between the two points $(m_1, 1+\Delta_1)$, $(m_2, 1+\Delta_2)$ of intersection of the straight line with the BLC determines on the axis OX the number of measurements $Nmn(m_1 + 1, m_2 - 1)$ falling into the middle group "mn" with slopes close to unity; finally, (c) $(m_2, 1+\Delta_2), (M-1, 0)$ the last group of measurements equaled Ndn falls into the lower group "dn" and is characterized by the average curve $Ydn(x)$. If the number of measurements $Nmn > Nup + Ndn$, then such a cycle of measurements is rated as "good" and is relatively stable. In the case when $Nmn \approx Ndn \approx Nup$, then such measurements are rated as "acceptable" (with a score of "satisfactory") and finally, the case when $Nmn < Nup + Ndn$ is rated poorly and such measurements are generally rated as unsatisfactory. Quantitatively, all three cases can be evaluated using a ratio:

$$Rt = \left(\frac{Nmn}{Nup + Ndn + Nmn}\right) \cdot 100\% = \left(\frac{Nmn}{M}\right) \cdot 100\%. \tag{29}$$

In expression (29), parameter M determines the full number of measurements. Based on this evaluation, the following criteria can be introduced: "excellent" and "good" grades are given to an experiment when $60\% < Rt < 100\%$; an "acceptable" or "satisfactory" grade is given to an experiment when $30\% < Rt < 60\%$; and finally, a "poor" grade is given when $0 < Rt < 30\%$. Therefore, by creating this triad from the original measurements, one can introduce the following definitions:

$$F_0(x) \cong \langle a_1(x)\rangle F_1(x) + \langle a_2(x)\rangle F_2,$$

$$F_0(x) \equiv Yup(x) = \frac{1}{Nup}\sum_{m=0}^{Nup-1} y_m^{(up)}(x), \quad 1 + \Delta_l < Sl_m < \max(Sl_m),$$

$$F_2(x) \equiv Ydn(x) = \frac{1}{Ndn}\sum_{m=0}^{Ndn-1} y_m^{(dn)}(x), \quad \min(Sl_m) < Sl_m < 1 - \Delta_2,$$

$$F_1(x) \equiv Ymn(x) = \frac{1}{Nmn}\sum_{m=0}^{Nmn-1} y_m^{(mn)}(x), \quad 1 - \Delta_2 < Sl_m < 1 + \Delta_1. \tag{30}$$

Here the Sl_m function defines a distribution of slopes arranged in decreasing order; the parameters $\Delta_{1,2}$, related to the value of the confidence interval, are chosen independently for each series of measurements. We have here added three "artificially" created measurements $F_{2,1,0}(x)$ to the previous set $y_m(x)$. As a result of this procedure, the $\langle a_{1,2}(x)\rangle$ functions independent of the index m remain almost the same (for sufficiently large values of M) compared to the case where no such clustering procedure was applied to the original measurements. We also assume that the averaged function $Ymn(x)$ is identified with the initial measurement $F_1(x)$, while the other two measurements $F_{0,2}(x)$ coincide with the functions $Yup(x), Ydn(x)$, respectively. The solution of equation (30) is determined by expressions (26) and (27). This clustering procedure turns out to be very efficient and can be applied to a wide range of cases. The details of this procedure are described for a non-trivial example discussed below.

The next issue to be considered in this section is related to the reduction of real measurements to IE. According to the definition given in [5] (see also definition (1)), by IE we mean the situation when

$$F_m(x) \equiv F(x + mT) = F_{m+1}(x) \equiv F(x + (m+1)T), \tag{31}$$

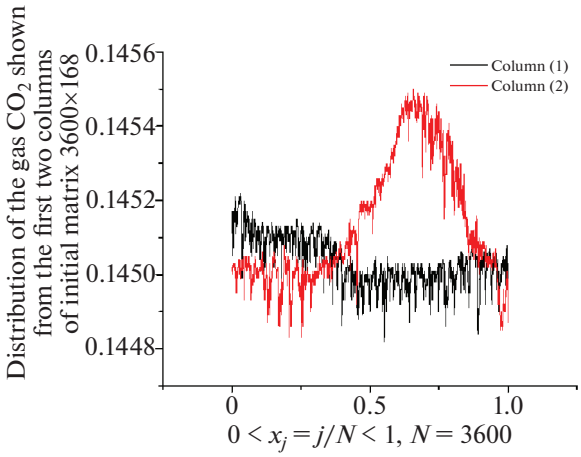

Fig. 1. Initial data containing second data recorded during the first hour (black lines) and second hour (red lines), correspondingly. In order to smooth these data and eliminate the HF fluctuations one can use expression (36). The result is shown in the next figure.

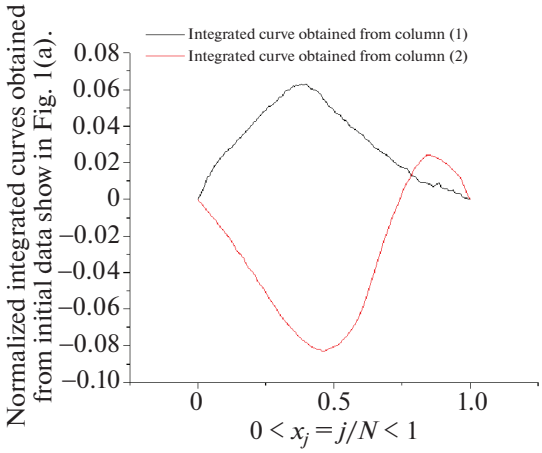

Fig. 2. Based on expression (36) one can integrate data and receive the smoothed curves that are suitable for the further analysis. One can notice that integrated data demonstrate their oscillating character.

the response function (measurement result) remains the same for the whole series of measurements included in one cycle. As mentioned above, in this case the IE coincides with the segment of the Fourier series (2). Therefore, the question arises: is it possible to extract the purely periodic F-components $Pr_q(x)$ $(q = 1, 2, \ldots, L)$ from the general solution (22) and to present to theorists for comparison the purified function, which should be compared with the hypothesis claiming to quantitatively describe the experimental results from a microscopic point of view? It makes sense to show this procedure in detail for the case of a "short" memory $(L = 2)$, keeping in mind this situation as the most likely one. As it has been shown above the case of a large number of measurements $2 < L < M$ at some reasonable suppositions is reduced to the short memory case also.

1. $L = 2$, case when $\kappa_{1,2}(x) > 0$

$$F_0(x) = \Big[\kappa_1(x)\Big]^{x/T} \mathrm{Pr}_1(x) + \Big[\kappa_2(x)\Big]^{x/T} \mathrm{Pr}_2(x),$$
$$F_1(x) = \kappa_1(x)^{1+(x/T)} \mathrm{Pr}_1(x) + \kappa_2(x)^{1+(x/T)} \mathrm{Pr}_2(x). \tag{32}$$

From this system of equations, we can easily find the desired periodic function $\mathrm{Pr}(x)$, which will

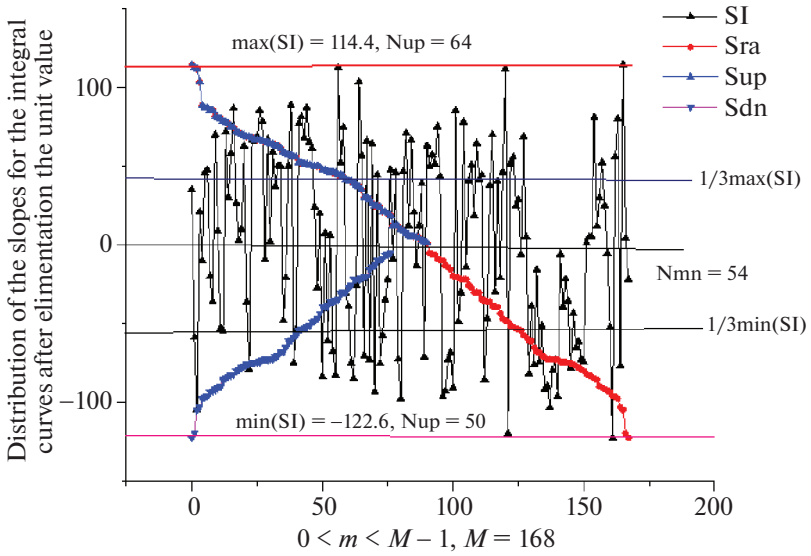

Fig. 3. The further data reduction becomes possible with the help of the distribution slopes curve Slm $(0 < m < M-1, M = 168)$. This curve defined by expression (28) helps to evaluate the number of slopes forming the "up" of the slopes distribution $(Nup = 64)$ and the slopes referring to the "dn" $(Ndn = 50)$ and the middle of the curve $(Nmn = 54)$. Using expressions (30) one can use the corresponding analysis to consideration only three integrated curves only. They are shown in the figure below.

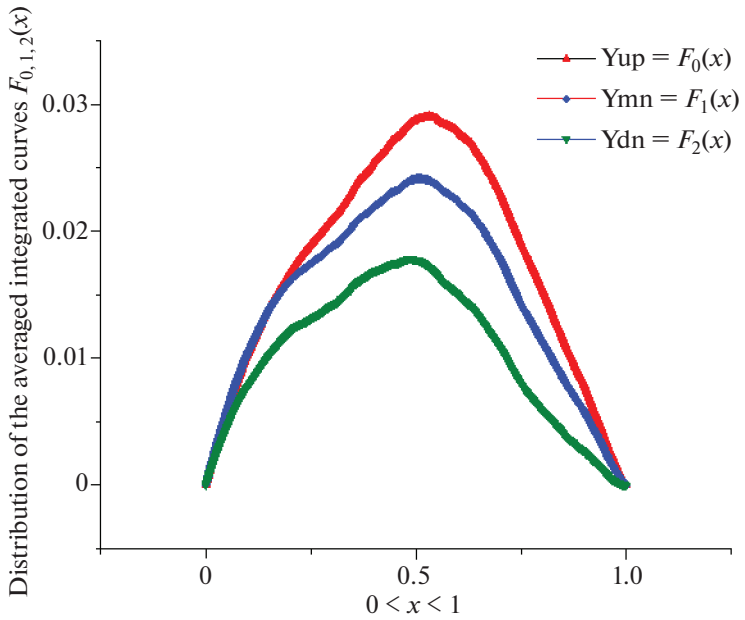

Fig. 4. Distribution of the integrated averaged curves obtained from expression (30). Namely, these curves should be fitted in order to receive the desired AFR and the "purified" curves from random uncontrollable factors and corresponding to an "ideal" experiment.

be represented as a linear combination of the functions

$$\mathrm{Pr}_1(x) = \Big[\kappa_1(x)\Big]^{-(x/T)} \frac{F_0(x)\kappa_2(x) - F_1(x)}{\kappa_2(x) - \kappa_1(x)},$$

$$\mathrm{Pr}_2(x) = \Big[\kappa_2(x)\Big]^{-(x/T)} \frac{F_1(x) - F_0(x)\kappa_1(x)}{\kappa_2(x) - \kappa_1(x)},$$

$$\mathrm{Pr}(x) = w_1\mathrm{Pr}_1(x) + w_2\mathrm{Pr}_2(x). \tag{33}$$

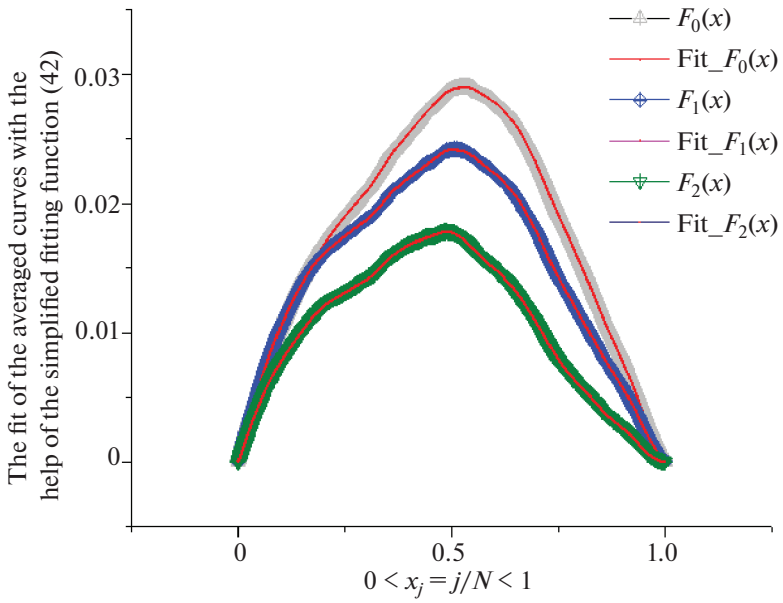

Fig. 5. The fit of the averaged integral curves $F_{0,1,2}(x)$ obtained with the help of the fitting function (38). The fitting function (38) is expressed in the form of the red solid line.

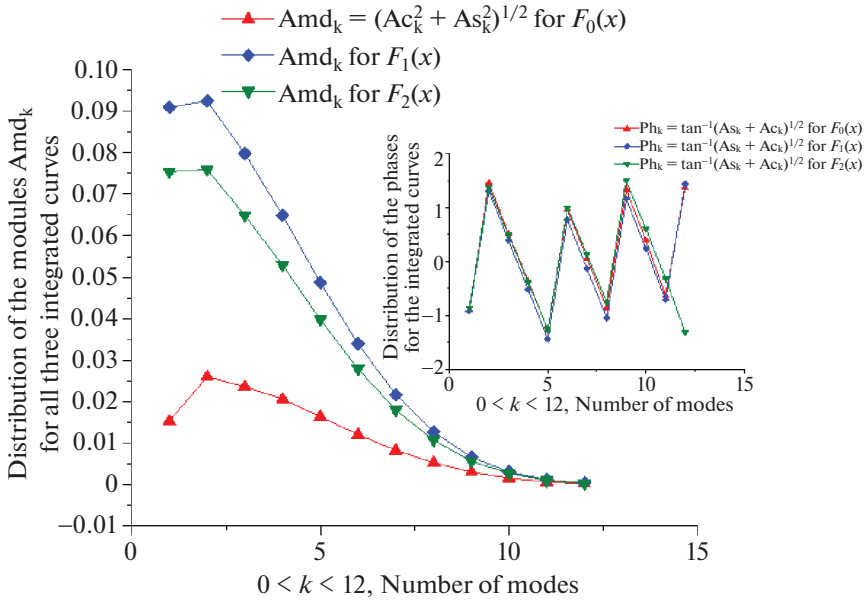

Fig. 6. The amplitude-frequency responses shown relatively the number of modes for all integrated curves shown in the previous figure 5.

Here, for "insurance", we have introduced unknown weight constants w_1 and w_2 to use them as fitting parameters at the final stage of comparison of the fitting function of the IE with the hypothesis derived from a competing model or microscopic theory. Obviously, the zeros of the functions κ_1 and κ_2 in (33) do not define the sought periodic functions, and the degenerated case must be treated separately.

2. $L = 2$, the case when $\kappa_1(x) > 0$ and $\kappa_2(x) < 0$

$$F_0(x) = \Big[\kappa_1(x)\Big]^{x/T} \mathrm{Pr}_1(x) + \Big[|\kappa_2(x)|\Big]^{x/T} \cos\left(\pi \frac{x}{T}\right) \mathrm{Pr}_2(x),$$

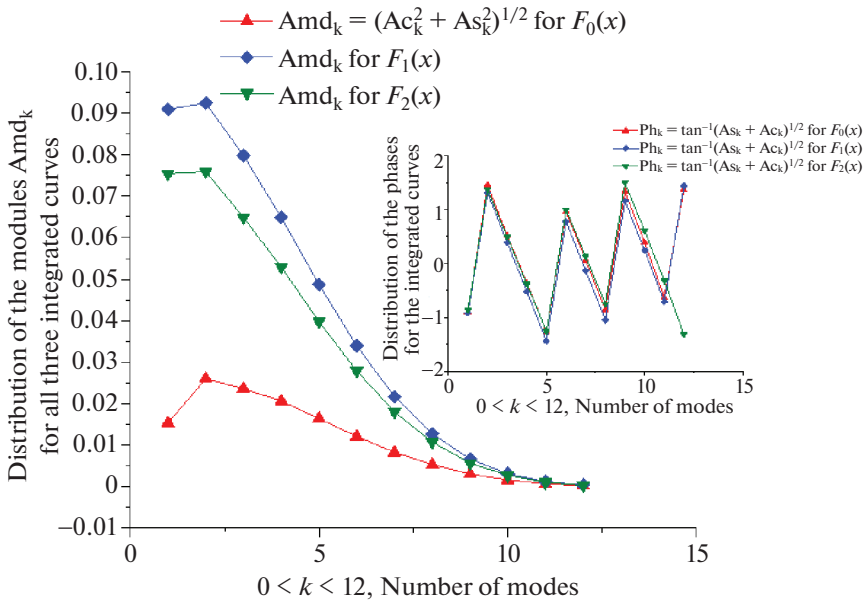

Fig. 7. This significant figure demonstrates clearly the integrated curves that are "cleaned" from the influence of the uncontrollable factors. These curves can be served as a specific "bridge" connecting the specific theory with experiment. The optimal period Tmx as earlier is found from the criterion corresponding to the minimal error requirement (see expression (37)).

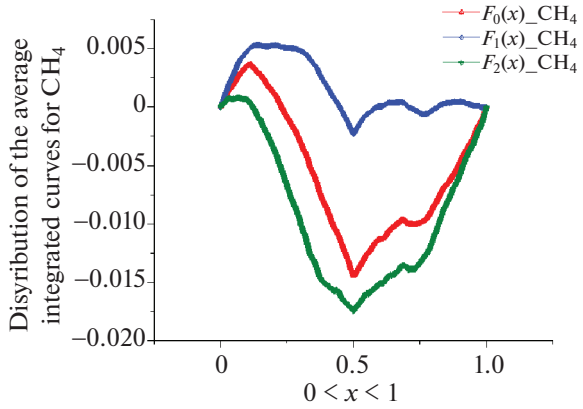

Fig. 8. The averaged integrated curves obtained for the gas CH_4.

$$F_1(x) = \kappa_1(x)^{1+(x/T)}\mathrm{Pr}_1(x) - |\kappa_2(x)|^{1+(x/T)}\cos\left(\pi\frac{x}{T}\right)\mathrm{Pr}_2(x). \tag{34}$$

Solution in this case accepts the form

$$\mathrm{Pr}_1(x) = \Big[\kappa_1(x)\Big]^{-(x/T)}\frac{F_1(x) + |\kappa_2(x)|F_0(x)}{\kappa_1(x) + |\kappa_2(x)|},$$
$$\mathrm{Pr}_2(x) = \Big[|\kappa_2(x)|\Big]^{-(x/T)}\frac{F_0(x)\kappa_1(x) - F_1(x)}{\kappa_1(x) + |\kappa_2(x)|},$$
$$\mathrm{Pr}(x) = w_1\mathrm{Pr}_1(x) + w_2\mathrm{Pr}_2(x). \tag{35}$$

The cases when degenerate "roots" coincide identically with each other $\kappa_1(x) \equiv \kappa_2(x)$ and the case of complex-conjugate "roots" ($\kappa_{1,2}(x) = \mathrm{Re}(\kappa(x)) \pm i\mathrm{Im}(\kappa(x))$) are omitted. The authors suggest that the inquisitive reader obtain them as an exercise. A careful analysis of this theory shows that it allows going beyond the allowed values of the controlled variable x. This possibility is considered in Mathematical Appendix 2.

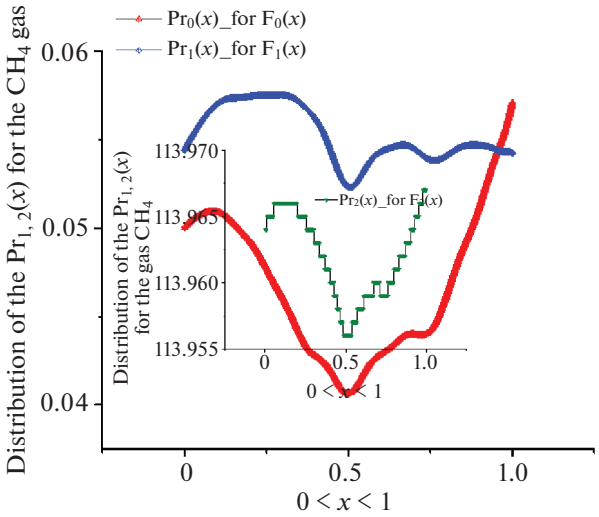

Fig. 9. This figure shows the purified pure periodic curves $\mathrm{Pr}_{0,1,2}(x)$ obtained from initial averaged curves for the CH_4 gas shown in Fig. 8. One notes that the periodic function $\mathrm{Pr}_2(x)$ has significant values and, therefore, it is depicted on the central figure separately.

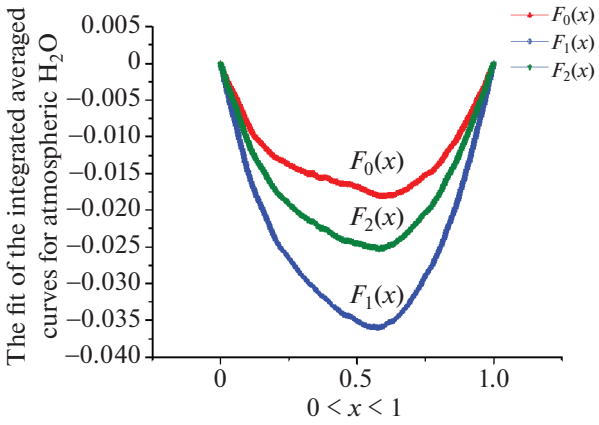

Fig. 10. The averaged integrated curves calculated for the vapors of H_2O.

3. VERIFICATION OF THE PROPOSED THEORY ON REAL DATA

In order to diminish the influence of strong variations of these high-frequency (HF), random factors we integrate the initial data by means of trapezoid method preliminary normalizing them in accordance with the following expressions

$$Y_j = \frac{y_j - mean(y)}{stdev(y)}, \quad j = 1, 2, \ldots, N,$$
$$J_j = J_{j-1} + \frac{1}{2}(x_j - x_{j-1})(Y_j + Y_{j-1}), \quad J_0 = 0. \tag{36}$$

These expressions make the initial data y_j: (a) dimensionless and (b) filtered. Integration procedure eliminates the high-frequency fluctuations and smoothes possible outlies. After this preliminary manipulation one can prepare the 3 rectangle matrices for each selected gas as CH_4, CO_2 and atmospheric humidity H_2O. Each matrix contains in total N rows $\times M$ columns, where N contains 3600 data points in each column (coinciding with 1 hour measurement) and $M = 24 \times 7 = 168$ hours per week.

Supposing that these prepared data can be classified as QR experiments one can obtain three normalized and integrated curves J_r, $(r = up, mn, dn)$ in accordance with clusterization procedure described in the previous section. Our further goal is to describe these 9 curves for the three original

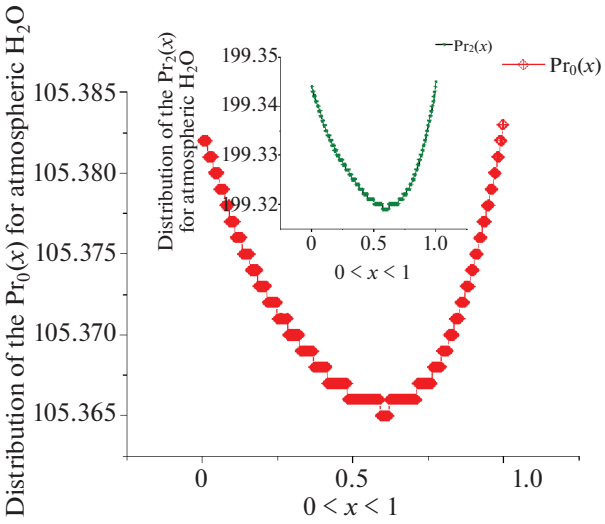

Fig. 11. This figure shows the purified pure periodic curves $\mathrm{Pr}_{0,2}(x)$ obtained from initial averaged curves for the H_2O shown in Fig. 10. One notes that the periodic function $\mathrm{Pr}_2(x)$ has another scale and, therefore, it is depicted on the central figure separately. We do not show a similar curve for $\mathrm{Pr}_1(x)$ because it has location in very narrow scale [199.3–199.5] and it cannot be shown in this picture.

matrices using a minimum number of fitting parameters. Attentive analysis shows that not disturbing the validity of the proposed general theory, in expression (25) one can put simply $\mathrm{Pr}_1(x) = \mathrm{Pr}_2(x) = \mathrm{Pr}(x)$. The functions $\kappa_1(x)$ and $\kappa_2(x)$ are defined in the same expression by the line below. The unknown functions $\langle a_1(x)\rangle$ and $\langle a_2(x)\rangle$ serving for evaluation the desired roots $\kappa_{1,2}(x)$ are derived from expressions (17) and (18). The unknown nonlinear parameter as the maximal period T value that figures in definition of periodic function (2) can be calculated from the minimization of the relative error value

$$\mathrm{Re}lErr(T_{mx}, R_{\min}) = \min\left(\frac{stdev(J_r(x) - F_0(x, T_{mx}, R_{\min}))}{mean|J_r(x)|}\right)\cdot 100\%. \tag{37}$$

Here the simplified fitting function $F_0(x, T_{mx}, K_{\min})$ is determined as

$$F_0(x, T_{mx}, K_{\min}) = A_0 + G(x, T_{mx})Ac_0$$
$$+ G(x, T_{mx})\sum_{k=1}^{K_{\min}}\left[Ac_k\cos\left(2\pi k\frac{x}{T_{mx}}\right) + As_k\sin\left(2\pi k\frac{x}{T_{mx}}\right)\right], \tag{38}$$

$$G(x, T_{mx}) = \left(\kappa_1(x)^{x/T_{mx}} + \kappa_2(x)^{x/T_{mx}}\right).$$

In the last expression, it is convenient to use the normalized input variable $x_j = j/N$ in the interval [0,1]. Therefore, the most probable interval, where the value T_{mx} can be located, is determined as (0.5, 1.75). This supposition was confirmed also numerically. From the simplified fitting function (38) it is easy to find the periodic function (2) from (38) corresponding to the IE

$$\mathrm{Pr}(x) \cong A_0 - Ac_0 + \frac{F_0(x, T_{mx}, K_{\min}) - A_0}{G(x, T_{mx})}. \tag{39}$$

The simplified fitting function (38) (containing minimal number of the fitting parameters $Prm = [(T_{mx}, A_0, Ac_0) + 2K_{\min}]$) allows to obtain the desired AFR (amplitude-frequency response) that contains the leading minimal frequency $\omega_{\min} = 2\pi/T_{mx}$ and the total segment of frequencies $\omega_k = \omega_{\min}k$ located in the interval $[1, K_{\min}]$.

In order to save place for demonstration of similar figures associated with all data, we demonstrate the processing details for CO_2 measured data only. Other data are processed by similar way. In Fig. 1, we demonstrate the initial data (the first two columns from the initial matrix $N = 3600 \times M = 168$). Integration procedure performed with accordance of expression (36) is shown in Fig. 2. As it has

Table 1. Collection of the basic fitting parameters related to CH_4, CO_2 and H_2O

Basic parameters	CH_4	CO_2	H_2O
RelError(%)	3.14244	0.38908	0.50246
Tmx	−0.01066	2.36204E-4	2.4294E-5
A_0	−215.377	0.00615	−0.00962
Ac_0	−57.7857	−0.00592	0.0068
Ac_k, As_k for $(k = 1, K_{\min})$	406.089	0.00418	0.00168
	339.818	$-4.08188E-5$	0.00243
	98.7791	3.71344E-4	−0.00113
	113.73	$-3.38628E-4$	7.7525E-4
	−251.434	6.52879E-4	$-8.3221E-4$
	−163.355	$-1.07774E-4$	2.86282E-4
	−104.081	1.92741E-4	$-5.94241E-4$
	−79.3113	1.08318E-5	5.39341E-5
	92.1547	4.94373E-5	$-3.20753E-4$
	44.3654	$-5.17735E-5$	$-4.8566E-5$
	51.029	1.87263E-4	$-2.9099E-4$
	27.6791	1.22959E-4	$-1.57504E-4$
	−17.7267	1.0636E-4	$-1.54041E-4$
	−5.60207	4.00495E-5	$-1.39499E-4$
	−12.4916	4.53336E-5	$-9.02014E-5$
	−4.56507	5.47098E-5	$-9.0448E-5$
	1.26814	$-6.70257E-5$	$-2.41301E-5$
	0.15116	2.87109E-6	$-1.0391E-4$
	1.28586	$-3.67108E-5$	$-3.95203E-5$
	0.25297	$-5.73328E-5$	$-7.22104E-5$
	0.01024	$-1.33468E-5$	$-3.59246E-6$
	0.00543	1.39172E-5	$-4.72613E-5$
	-0.02703	1.30144E-5	3.27143E-5

been mentioned above, Fig. 2 demonstrates the effectiveness of integration procedure. It eliminates the HF fluctuations and allows to receive the smoothed curves for the further analysis. In Fig. 3, we show the distribution of the slopes (after elimination of the unit value). One can divide the whole segment with boundaries $(\min(Sl), \max(Sl))$ on almost equal three segments: $(1/3\max(Sl), \max(Sl)$ for $F_0(x))$, $(1/3\min(Sl), 1/3\max(Sl)$ for $F_1(x))$ and $(min(Sl), 1/3\min(Sl)$ for $F_2(x))$. The number

of measurements in each selected segment is shown in Fig. 3. This information helps in obtaining the desired averaged functions $F_{0,1,2}$ from expression (30). Fig. 4 demonstrate the final result. Then, these three curves can be fitted by the simplified fitting function (38). Only one nonlinear fitting parameter as T_{mx} can be found from minimization of the relative error (38), supposing that this parameter is located in the interval $[0.5T, 1.75T]$. Other fitting parameters $(A_0, Ac_0, Ac_k, As_k, (k = 1, 2 \ldots, K_{\min}))$ are found by the LLSM. The fit of these three curves is shown in Fig. 5. The distribution of the AFRs taken in the form of modules and phases $Amd_k = (Ac_k^2 + As_k^2)^{1/2}$, $Ph_k = \tan^{-1}\left(\dfrac{As_k}{Ac_k}\right)$ for three averaged functions $F_{0,1,2}$ are shown in Fig. 6.

The most interesting result from this research is shown in Figs. 7, 9, and 11. One can eliminate the influence of uncontrolled factors and get the pure periodic functions $Pr_{0,1,2}(x)$ based on expression (39). As one notice from these figures the purified periodic functions have different scales and can be significantly different from the initial functions $F_{0,1,2}(x)$. The basic fitting parameters associated with function (38) are collected in Table 1.

4. RESULTS AND DISCUSSION

In this paper, the authors proposed the theory for the QRE(s) that based on the given sampling of successive measurements. They should be given in the form of the rectangle matrix $N \times M$, where $(j = 1, 2, \ldots, N)$ includes in itself the number of data points corresponding to one experiment realized during one period T, while index $m = 1, 2, \ldots, M$ determines the whole cycle of successive/repeatable measurements realized during the whole observation period $M \cdot T$. Thanks to algorithm described in the paper one can receive only 3 key averaged measurements, see expression (30), that can be fitted by the simplified fitting function (38). This function plays a double role (a) the fit of the averaged measurements and (b) extraction the pure periodic function (39) corresponding to an "ideal experiment". Other 2 important results can be formulated as follows.

1. Creation of the completely computerized laboratories, where "input" coincides with initial data presented in the form of the matrix $N \times M$, while the "output" coincides with limited number of the fitting parameters $Prm = [(T_{mx}, A_0, Ac_0) + 2K_{\min}])$ followed from expression (39).

2. The creation of new metrological standard based on comparison of the "pattern" detector/device with the tested ones based on the optimal number of the fitting parameters Prm that are obtained from the registered fluctuations.

The problem that is needed in further research follows from the solution (19) based on the supposition $\langle a_l(x)\rangle$, $\kappa(x \pm T) = \kappa(x)$, $\Pr(x \pm T) = \Pr(x)$. What new solutions will the rejection of this assumption lead to? The answer to this question requires further research.

5. FUNDING

This work was funded by the subsidy allocated to Kazan Federal University for the state assignment in the sphere of scientific activities, project No FZSM-2024-0004.

REFERENCES

1. R. Nigmatullin and R. Rakhmatullin, "Detection of quasi-periodic processes in repeated-measurements: New approach for the fitting and clusterization of different data," Commun. Nonlin. Sci. Numer. Simul. **19**, 4080–4093 (2014).
2. R. R. Nigmatullin, A. A. Khamzin, and J. T. Machado, "Detection of quasi-periodic processes in complex systems: How do we quantitatively describe their properties?" Phys. Scr. **89**, 015201(11pp) (2014).
3. R. R. Nigmatullin, S. I. Osokin, D. Baleanu, S. Al-Amri, A. Azam, and A. Memic, "The first observation of memory effects in the InfraRed (FT-IR) Measurements: Do successive measurements remember each other?" PLoS One **9**, e94305 (2014).
4. R. R. Nigmatullin, R. M. Rakhmatullin, and S. I. Osokin, "How to reduce reproducible measurements to an ideal experiment?" Magn. Reson. Solids **16** (2), 1–19 (2014). http://mrsej.kpfu.ru

5. R. R. Nigmatullin, W. Zhang, and D. Striccoli, "General theory of experiment containing reproducible data: The reduction to an ideal experiment," Commun. Nonlin. Sci. Numer. Simul. **27**, 175–192 (2015).
6. M. Kuczma, "A servey of the theory of functional equations," Publ. Elektroteh. Fak. Univ. Beogradu **130**, 1–64 (1964).
7. R. R. Nigmatullin, D. Striccoli, G. Boggia, and C. Ceglie, "A novel approach for characterizing multimedia 3D video streams by means of quasiperiodic processes," Signal Image Video Process. (2016). https://doi.org/10.1007/s11760-016-0866-9
8. J. Feder, *Fractals* (Plenum, New York, 1988).
9. R. Nigmatullin and D. Baleanu, "New relationships connecting a class of fractal objects and fractional integrals in space," Fract. Calc. Appl. Anal. **16** (4), 1–26 (2013).
10. A. Wald, *Sequential Analysis* (Wiley, New York, 1947).