ISSN 0005-1055, Automatic Documentation and Mathematical Linguistics, 2025, Vol. 59, Suppl. 1, pp. S15–S28. © Allerton Press, Inc., 2025.
Russian Text © The Author(s), 2025, published in Elektronnye Biblioteki, 2025, Vol. 28, No. 1, pp. 41–75.

New Method of Description of Eddy-Covariance Ecologic Data

R. R. Nigmatullin[a], * (ORCID: 0000-0003-2931-4428), A. A. Litvinov[b], ** (ORCID: 0009-0000-3901-3704), and S. I. Osokin[b], * (ORCID: 0000-0002-0699-5390)**

[a] *Kazan National Research Technical University named after A.N. Tupolev, Radioelectronics and Informative Measurements Technics Department, Kazan, 420111 Russia*

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

**e-mail: renigmat@gmail.com*

***e-mail: litvinov85@gmail.com*

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

Received December 15, 2024

Abstract—In this paper, the authors propose the foundations of an original theory of quasi-reproducible experiments (QREs) based on the testable hypothesis that there exists an essential correlation (memory) between successive measurements. From this hypothesis, which the authors call for brevity as the verified partial correlation principle (VPCP), it can be proved that there exists a universal fitting function (UFF) for quasi-reproducible (QR) measurements. In other words, there is some common platform or bridge across which, figuratively speaking, a true theory (claiming to describe data from first principles or verifiable models) and an experiment offering this theory for verification of measured data, maximally cleaned from the influence of uncontrollable factors and apparatus/software function, meet. Actually, the proposed theory provides the potential researcher with a method of the purification of initial data, finally suggesting a curve that describes the data, is periodic, and is cleaned from a set of uncontrollable factors. The final curve corresponds to an ideal experiment. The proposed theory has been tested on eddy-covariance ecological data related to the content of CH_4, CO_2, and 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 CH_4, CO_2, and H_2O vapor in the 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 most reliable quantitative description of this kind of data that belongs to the tested complex system. We should also note that the final fitting function that is removed from uncontrollable factors becomes purely periodic and corresponds to an ideal experiment. Applications of this theory to practical applications, its place among other alternative approaches (especially touching the professional interests of ecologists), and its further development are discussed in the paper.

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

DOI: 10.3103/S000510552570089X

1. INTRODUCTION AND STATEMENT OF THE PROBLEM

Is it possible to construct a universal fitting function to describe non-stationary quasi-reproducible experiments (QREs) for a wide class of complex systems? The question sounds absurd and meaningless for any experienced specialist. Everyone knows how the traditional interaction of theory and experiment occurs. A theory offers models, hypotheses based on certain assumptions, and postulates. The experiment tests these hypotheses, trying to exclude the influence of uncontrolled factors and distortions (noise) introduced by the measuring equipment (it is usually defined as an instrumental function) as much as possible. What that is fundamentally new can be introduced into this traditional scheme? Can we imagine that it is possible to find a certain verifiable or testable principle to which almost all measurements are subject? If such a principle is found, then from its mathematical formulation it will be possible to derive a certain universal fitting function that will allow a description of all measurements. The authors call this the verified partial correlation principle (VPCP). On the basis of the VPCP, it will be possible to obtain a general mathematical model (defined as an intermediate model, IM), which all measurements satisfying this principle will have to obey. What is this principle that most measurements satisfy? If we expand the VPCP, we can formulate this as follows: successive measurements retain partial correlation (memory) with each

other and remain fully or partially correlated as a result of a series of successive measurements. Naturally, this makes it necessary to clarify the nature and type of these measurements and to translate this imprecise verbal formulation into strict mathematical language. Therefore, it is necessary to first provide several concepts and then translate the principle into the language of mathematical formulas. An ideal experiment (IE) is understood by the authors as one in which a sequence of measurements m ($m = 0, 1, 2, ..., M-1$), carried out over a certain average period T with respect to the control external variable x leads to the same value of the measured response function F(x). In this sense, all measurements that correspond to the IE are fully or absolutely correlated. Mathematically, this statement looks like this:

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

Here, the controlled (manipulated) variable x can coincide with the time variable (t), frequency (ω), wavelength (λ), etc. Because the experiment conducted on this set of variables is single-factor, it is assumed that other controlled variables that affect the response function remain unchanged over a certain range of their values in a single-factor experiment. The solution to functional equation (1) is a periodic segment of the corresponding Fourier series. For discrete data, the segment of the Fourier series is usually written as

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

Parameter T defines some average measurement period relative to the input variable x. From simple equation (1), it follows that expression (2) can be used as a fitting function for the response function in the IM. In this idealized case, the IM coincides with a segment of the Fourier series, and the coefficients of this expansion can act as fitting parameters that correspond to the IE. In fact, the fitting parameters form the sought-after amplitude-frequency response (AFR). It is quite obvious that the requirement of IE (1) is not implemented in reality, and analysis of various data shows that instead of equation (1), a more general functional equation should be written

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

Here, the function $a_l(x)$ takes into account the influence of a set of uncontrolled factors. Most of the experiments follow relation (3), and hence, these experiments can be defined as QREs. It is possible to find analytical solutions of equation (3) for a wide class of functions. Then it is possible to substantially eliminate the influence of uncontrolled factors and obtain a pure periodic function corresponding to IE (1) as a result. It is possible to outline the basics of this more general theory based on a series of successive measurements applicable to the description of QREs.

Of course, any theory will be incomplete if it is not verified by experiment. The experiments that we have conducted concerned the quantitative analysis of gases such as CH_4, CO_2, and H_2O vapor contained in the atmosphere and, hence, the proposed theory provides novel and more reliable capabilities in the analysis of eddy-covariance data. Moreover, the proposed algorithms are sufficiently general and can be applied to the analysis of many similar QREs.

2. BASICS OF QRE THEORY

2.1. Self-Consistent Solutions of Equation (3)

It is possible to obtain solutions of the functional equation (3), provided that the length L characterizing the memory between measurements is known. Thus, let us assume that all successive measurements satisfy the equation

$$F_{L+m}(x) = \sum_{l=0}^{L-1} a_l(x)F_{l+m}(x), \quad m = 0,1,\ldots,M-1. \tag{4}$$

To find unknown functions $a_l(x)$ ($l = 0, 1, ..., L$; $L < M$), we can generalize the linear least squares method (LLSM) and require that *functional* variance take a minimum value

$$\sigma(x) = \frac{1}{M-L} \times \sum_{m=0}^{M-L-1}\left[F_{L+m}(x) - \sum_{l=0}^{L-1} a_l(x)F_{l+m}(x)\right]^2 = \min. \tag{5}$$

To obtain the desired solution, it is necessary to take the average value for the remaining measurements ($l = 0, 1, ..., M-L-1, L$; $L < M$). Taking the functional derivatives of (5) with respect to the unknown functions $a_l(x)$, we obtain

$$-\frac{\delta\sigma(x)}{\delta a_l(x)} = \frac{1}{M-L} \times \sum_{m=0}^{M-L-1}\left[F_{l+m}(x)\left(F_{L+m}(x - \sum_{s=0}^{L-1} a_s(x)F_{s+m}(x)\right)\right] = 0. \tag{6}$$

Here we also applied the averaging procedure over the entire set of admissible measurements, assuming that the set functions $a_l(x)$ ($l = 0, 1, ..., L$; $L < M$) does not depend on the index m. Introducing the definitions of the pair correlation functions (7)

$$K_{L,l} = \frac{1}{M-L}\sum_{m=0}^{M-L-1} F_{L+m}(x)F_{l+m}(x),$$
$$K_{s,l} = \frac{1}{M-L}\sum_{m=0}^{M-L-1} F_{s+m}(x)F_{l+m}(x), \quad (7)$$
$$s,l = 0,1,\ldots,L-1,$$

we can obtain a system of linear equations for calculating unknown functions $a_l(x)$

$$\sum_{s=0}^{L-1} K_{s,l}(x)a_s(x) = K_{L,l}(x). \quad (8)$$

It makes sense to define this procedure as a functional linear least squares method (FLLSM), which includes the usual LLSM as a special case.

Now let us return to equation (3). We seek the solution of this equation in the form

$$F_0(x) = [\kappa(x)]^{x/T}\,\mathrm{Pr}(x),$$
$$F_m(x) = [\kappa(x)]^{m+x/T}\,\mathrm{Pr}(x). \quad (9)$$

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

$$\Phi(x) = A_0 + \sum_{k=1}^{K\gg 1}\left[Ac_k \cos\left(2\pi k\frac{x}{T}\right) + As_k \sin\left(2\pi k\frac{x}{T}\right)\right]. \quad (10)$$

Obviously, the expansion coefficients Ac_k, As_k ($k = 1, 2, \ldots, K$) in (10) depend on the specific type of the expanded function. Substituting the trial solution (6) into equation (4), we obtain an equation for calculating the unknown functions $\kappa(x)$ appearing in solution (9):

$$[\kappa(x)]^L - \sum_{l=0}^{L-1} a_l(x)[\kappa(x)]^l = 0. \quad (11)$$

If the roots of the functional equation (11) $\kappa_q(x), q = 1,2,\ldots,L$ can be found, the general solution for the function $F_m(x)$ can be written in the form

$$F_0(x) = \sum_{q=1}^{L}[\kappa_q(x)]^{x/T}\,\mathrm{Pr}_q(x),$$
$$F_m(x) = \sum_{q=1}^{L}[\kappa_q(x)]^{m+(x/T)}\,\mathrm{Pr}_q(x), \quad (12)$$
$$m = 0,1,\ldots,M-1.$$

The number of periodic functions $\mathrm{Pr}_q(x)$ must coincide with the number of functions $\kappa_q(x)$, $q = 1,2,\ldots,L$ included in the last expression in (12). This expression can be considered as the general solution of the functional equation (8). This solution can be interpreted as follows: if successive measurements are partially correlated with each other (can "remember" each other) and change during the average measurement period T, then the fitting function for describing these measurements is self-consistent and is determined by the entire set of random measurements that participate in this process. Obviously, this new result generalizes the previous results [1–5] obtained for the case when the functions $\langle a_l(x)\rangle$. can be approximated by constants a_l. It is desirable to obtain solutions for equation (4) for the case when the functions $\langle a_l(x)\rangle$ are not completely periodic or are obtained a priori from other conditions. However, to the best of the authors' knowledge, the mathematical theory of solutions of functional equations is almost undeveloped [6] relative, for example, to a full-fledged theory of solutions of differential or integral equations. Thus, in our opinion, the proposed theory defines a new direction for mathematicians working in the field of functional analysis intending to find applications of their results for physics, chemistry, and engineering. For practical applications, it makes sense to consider the case of short memory with ($L = 2$) in more detail, as the number of approximating parameters for this case is minimal. As shown below, the case of a long memory $L > 2$ can also be reduced to the case of short memory. The results for this case are needed to describe real measurements, which are given in the next section.

For the case of short memory with $L = 2$ we get

$$F_{2+m}(x) = a_1(x)F_{1+m} + a_0(x)F_m,$$
$$m = 0,1,\ldots,M-1. \quad (13)$$

Equation (8) for this case takes the form (14)

$$K_{00}(x)a_0(x) + K_{10}(x)a_1(x) = K_{20}(x),$$
$$K_{10}(x)a_0(x) + K_{11}(x)a_1(x) = K_{21}(x). \quad (14)$$

We write the solution to equation (13) in the form

$$F_0(x) = [\kappa_1(x)]^{x/T}\,\mathrm{Pr}_1(x) + [\kappa_2(x)]^{x/T}\,\mathrm{Pr}_2(x),$$
$$\kappa_{1,2}(x) = \frac{a_1(x)}{2} \pm \sqrt{\left(\frac{a_1(x)}{2}\right)^2 + a_0(x)}. \quad (15)$$

If one of the roots in (15) becomes negative (e.g., $\kappa_2(x) < 0$), then the general solution for this case can be written as

$$F_0(x) = [\kappa_1(x)]^{x/T}\,\mathrm{Pr}_1(x) + [|\kappa_2(x)|]^{x/T}\cos\left(\pi\frac{x}{T}\right)\mathrm{Pr}_2(x). \quad (16)$$

If the order of measurements for assessing the influence of non-stationarity of the process as a whole is significant, the proposed theory will allow the entire non-stationary sequence to be reconstructed with the relations

$$F_m(x) = [\kappa_1(x)]^{m+(x/T)}\,\mathrm{Pr}_1(x) + [\kappa_2(x)]^{m+(x/T)}\cos\left(\pi\left(\frac{x}{T}+m\right)\right)\mathrm{Pr}_2(x),$$
$$\mathrm{Pr}_{1,2}(x) = A_0 + \sum_{k=1}^{K_m}\left[Ac^{(m)}_{k,1,2}\cos\left(2\pi k\frac{x}{T}\right) + As^{(m)}_{k,1,2}\sin\left(2\pi k\frac{x}{T}\right)\right], \quad (17)$$
$$m = 0,1,\ldots,M-1.$$

In this case, the functions $\mathrm{Pr}_{1,2}(x \pm T) = \mathrm{Pr}_{1,2}(x)$ retain their periodicity during the average period T, but the expansion coefficients $\mathrm{Ac}^{(m)}_{k,1,2}$, $\mathrm{As}^{(m)}_{k,1,2}$ ($k = 1, 2, \ldots, K_m$) appearing in (17) may differ from the case of $m = 0$ and reflect the influence of possible instability across the entire measurement process. If the true sequence of measurements is not essential, and the measurement results remain invariant with respect to the permutations of all measurements with each other, then all measurements can be grouped into three independent groups (forming a specific triad), and the long memory case will again be reduced to the short memory case considered above. This simple idea makes it possible to significantly reduce the number of fitting parameters and again obtain a fitting function with a minimum number of parameters. The procedure that is associated with the formation of the necessary triad is described in the next subsection.

2.2. Clustering Procedure and Reduction to IE

As was emphasized earlier [5], the evaluation of the "true' value of L based on a general criterion or principle is still an *unsolved* problem." If the assumption of permutations of measurements with each other can be justified and seems quite reasonable, then the following procedure of clustering and dividing all measurements into three groups (forming a specific triad) can be proposed. To do this, consider the distribution of slopes (angular tangents) of each measurement with respect to their mean, whose tangent is equal to or close to unity:

$$\mathrm{Sl}_m = \mathrm{slope}(y, y_m) \equiv \frac{(y_m y)}{(yy)},$$
$$y = \left(\frac{1}{M}\right)\sum_{m=0}^{M-1} y_m, \quad (AB) = \sum_{j=1}^{N} A_j B_j. \quad (18)$$

The round bold brackets in (18) represent the scalar product between two functions with $j = 1, 2, \ldots, N$ points of measurement of the initial data. We assume that random measurements $y_m(x)$ for $m = 0, 1, \ldots, M-1$ approximate the functions $F_m(x)$ ($y_m(x) \cong F_m(x)$) that appears in equation (17). If we construct a random distribution of the slope function Sl_m depending on the number of measurements m, then it is convenient to arrange them on the graph in the descending order $\mathrm{Sl}_0 > \mathrm{Sl}_1 > \ldots > \mathrm{Sl}_{M-1}$. These distribution slope functions can be divided into three groups. The upper group of measurements (up) has slopes localized in the interval $(1 + \Delta, \max(\mathrm{Sl}_m))$; the middle group (mn) contains measurements with slopes in the interval $(1 - \Delta, 1 + \Delta)$; and finally the lower group (dn) contains measurements with slopes in the interval $(1 - \Delta, \min(\mathrm{Sl}_m))$. The value of Δ for each set of QR measurements is determined independently for each specific case. This ordered curve Sl_m is important, and it reflects the quality of the measurements made and the equipment used.

How can we find Δ based on expression (18)? The ordered curve Sl_m can be divided (after the unit value is subtracted) into two parts—the positive part $(0, \max(\mathrm{Sl}_m)$ and the negative part $(0, \min(\mathrm{Sl}_m))$. For each part, we remove half of each selected part, i.e., $\Delta_1 = \max(\mathrm{Sl}_m - 1)/2$ and $\Delta_2 = \min(\mathrm{Sl}_m - 1)/2$. These values can be used to divide into three parts/clusters selected in this way. To understand the details of the clustering procedure, we write this procedure in more detail.

We form groups as follows:

(a) distance from the initial point of the *XOY* axis (0, 0) to the first intersection point $(m_1, 1 + \Delta_1)$ determines the number of measurements *N*up ($m = 1, 2, \ldots, m_1 = N\mathrm{up}$) that fall into the upper group, characterized by the average curve Yup(x);

(b) distance between two points $(m_1, 1 + \Delta_1)$, $(m_2, 1 + \Delta_2)$ of the intersection of a straight line with an ordered distribution of slopes determines the number of measurements Nmn $(m_1 + 1, m_2 - 1)$ that fall into the middle group mn, with slopes close to one along the *OX* axis; and, finally,

(c) $(m_2, 1 + \Delta_2)$, $(M - 1, 0)$—the last group of measurements that is equal to *N*dn—falls into the lower group dn and is characterized by the last part of the curve *Y*dn(x).

If the number of measurements *N*mn > *N*up + *N*dn, then this measurement cycle is assessed as good and is relatively stable. Where *N*mn ≈ *N*dn ≈ *N*up, such measurements are considered acceptable (with a rating of satisfactory), and, finally, the case when *N*mn < *N*up + *N*dn is considered bad, and such measurements are generally assessed as unsatisfactory. All three cases can be quantitatively assessed using the ratios

$$Rt = \left(\frac{N\mathrm{mn}}{N\mathrm{up} + N\mathrm{dn} + N\mathrm{mn}}\right)\times 100\% = \left(\frac{N\mathrm{mn}}{M}\right)\times 100\%. \quad (19)$$

In expression (19), parameter M determines the total number of measurements. From this assessment, the following criteria can be introduced: an experiment can be rated as excellent or good when $60\% < Rt < 100\%$; an experiment is rated as satisfactory when $30\% < Rt < 60\%$; and finally, a bad assessment is given when $Rt < 30\%$. Therefore, having created this triad

from the initial measurements, the following definitions can be introduced:

$$
\begin{gathered}
F_0(x) \cong a_1(x)F_1(x) + a_2(x)F_2(x), \\
F_0(x) \equiv Y\mathrm{up}(x) = \frac{1}{N\mathrm{up}} \sum_{m=0}^{N\mathrm{up}-1} y_m^{(\mathrm{up})}(x), \\
1 + \Delta_1 < Sl_m < \max(Sl_m), \\
F_2(x) \equiv Y\mathrm{dn}(x) = \frac{1}{N\mathrm{dn}} \sum_{m=0}^{N\mathrm{dn}-1} y_m^{(\mathrm{dn})}(x), \\
\min(Sl_m) < Sl_m < 1 - \Delta_2, \\
F_1(x) \equiv Y\mathrm{mn}(x) = \frac{1}{N\mathrm{mn}} \sum_{m=0}^{N\mathrm{mn}1} y_m^{(\mathrm{mn})}(x), \\
1 - \Delta_2 < Sl_m < 1 + \Delta_1.
\end{gathered}
\tag{20}
$$

Here, the function Sl_m defines the distribution of slopes, arranged in descending order; the parameters $\Delta_{1,2}$ that are associated with the values of the confidence interval are chosen independently for each series of measurements. Here we have added three "artificially" created measurements $F_{2,1,0}(x)$ to the previous set $y_m(x)$. As a result, the m-independent functions $a_{1,2}(x)$ remain almost unchanged (for sufficiently large values of M) relative to when such a clustering procedure was not applied to the original measurements. We also assume that the averaged function $Y\mathrm{mn}(x)$ is identified with the original measurement $F_1(x)$, and the other two measurements $F_{0,2}(x)$ coincide with the functions $Y\mathrm{up}(x)$ and $Y\mathrm{dn}(x)$, respectively. The solution of equation (20) is determined by the expressions (16) and (17). This clustering procedure is very effective and can be applied to a wide range of cases. The details of this procedure are described in the 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)), IE is understood by us as the situation when

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

and the response function (measurement result) remains unchanged for the entire series of measurements that are included in one cycle. As noted, in this case, the IE coincides with the segment of the Fourier series (2). In this connection, the question arises: is it possible to isolate purely periodic Fourier components $\mathrm{Pr}_q(x)$ ($q = 1, 2, \ldots, L$) from the general solution (12) and present the purified function to theoreticians for comparison, comparing it 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 having a short memory ($L = 2$), keeping in mind this situation as the most probable. As was shown above, the case of a large number of measurements $2 < L < M$ can be reduced to the case of short memory under some reasonable assumptions.

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

$$
\begin{aligned}
F_0(x) &= \left[\kappa_1(x)\right]^{x/T} \mathrm{Pr}_1(x) + \left[\kappa_2(x)\right]^{x/T} \mathrm{Pr}_2(x), \\
F_1(x) &= \kappa_1(x)^{1+(x/T)} \mathrm{Pr}_1(x) + \kappa_2(x)^{1+(x/T)} \mathrm{Pr}_2(x).
\end{aligned}
$$

From this system of equations, we can easily find the desired periodic function $\mathrm{Pr}(x)$, which is represented as a linear combination of functions $\mathrm{Pr}_{1,2}(x)$

$$
\begin{gathered}
\mathrm{Pr}_1(x) = \left[\kappa_1(x)\right]^{-(x/T)} \frac{F_0(x)\kappa_2(x) - F_1(x)}{\kappa_2(x) - \kappa_1(x)}, \\
\mathrm{Pr}_2(x) = \left[\kappa_2(x)\right]^{-(x/T)} \frac{F_1(x) - F_0(x)\kappa_1(x)}{\kappa_2(x) - \kappa_1(x)}, \\
\mathrm{Pr}(x) = w_1 \mathrm{Pr}_1(x) + w_2 \mathrm{Pr}_2(x).
\end{gathered}
\tag{21}
$$

Here, for insurance, we introduced the unknown weight constants $w_{1,2}$ for use as fitting parameters in the final stage of comparing the IE fitting function with the hypothesis that is obtained from a competing model or microscopic theory. Obviously, the zeros in the functions $\varkappa_{1,2}$ in (21) do not determine the desired periodic functions, and the degenerate case, they must be considered separately.

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

$$
\begin{aligned}
F_0(x) &= \left[\kappa_1(x)\right]^{x/T} \mathrm{Pr}_1(x) \\
&+ \left[\left|\kappa_2(x)\right|\right]^{x/T} \cos\left(\pi \frac{x}{T}\right) \mathrm{Pr}_2(x), \\
F_1(x) &= \kappa_1(x)^{1+(x/T)} \mathrm{Pr}_1(x) \\
&- \left|\kappa_2(x)\right|^{1+(x/T)} \cos\left(\pi \frac{x}{T}\right) \mathrm{Pr}_2(x),
\end{aligned}
$$

The solution in this case takes the form

$$
\begin{gathered}
\mathrm{Pr}_1(x) = \left[\kappa_1(x)\right]^{-(x/T)} \frac{F_1(x) + \left|\kappa_2(x)\right| F_0(x)}{\kappa_1(x) + \left|\kappa_2(x)\right|}, \\
\mathrm{Pr}_2(x) = \left[\left|\kappa_2(x)\right|\right]^{-(x/T)} \frac{F_0(x)\kappa_1(x) - F_1(x)}{\kappa_1(x) + \left|\kappa_2(x)\right|}, \\
\mathrm{Pr}(x) = w_1 \mathrm{Pr}_1(x) + w_2 \mathrm{Pr}_2(x).
\end{gathered}
$$

The cases when the degenerate roots coincide with each other equally $\kappa_1(x) \equiv \kappa_2(x)$ and the case of complex conjugate roots" $\kappa_{1,2}(x) = \mathrm{Re}\left(\kappa(x)\right) \pm i\,\mathrm{Im}\left(\kappa(x)\right))$ are omitted. The authors suggest that the inquisitive and advanced reader obtain these as an exercise.

A thorough analysis of the presented theory shows that it makes it possible to go beyond the permissible values of the controlled variable x. However, a detailed description of this possibility goes beyond the scope of this article, so this detailed consideration is omitted.

3. TESTING THE PROPOSED THEORY ON REAL DATA

3.1. Description of Real Data

Here we add a few words about real data and their features. As real data, we took eddy-covariance ecologic data related to the content of CH_4, CO_2, and vapor H_2O in the atmosphere, where the corresponding detectors for measuring the content of the sought gases are located. Thus, while in our previous work [7], we processed data that are only directly related only to the concentration of carbon dioxide, methane, and water vapor, in this work we consider the balance of carbon dioxide, methane, and water vapor, i.e., the product of the corresponding concentration according to the value of the vertical velocity.

3.2. Data Processing Procedure

To reduce the influence of high-frequency (HF) random fluctuations, we integrate the original data using a trapezoidal method, pre-normalizing them in relation to the following expressions:

$$Y_j = \frac{y_j - \mathrm{mean}(y)}{\mathrm{Range}(y)}, \quad j = 1,2, \ldots, N,$$
$$J_0 = 0, \;\; J_j = J_{j-1} + \frac{1}{2}(x_j - x_{j-1})(Y_j + Y_{j-1}), \qquad (22)$$
$$\mathrm{Range}(y) = \max(y) - \min(y).$$

These expressions make the original data y_j (a) dimensionless and (b) filtered. The integration procedure eliminates high-frequency fluctuations and smooths possible outliers. After this pre-processing, three rectangular matrices can be prepared for each selected gas: CH_4, CO_2, and H_2O atmospheric vapor. Each matrix contains a total of N rows × M columns, where N contains 3600 1-second points in each column (1-hour measurement), and $M = 24 \times 7 = 168$ h per week. Preliminary analysis showed that this procedure for preparing data for 1 week is the most acceptable and optimal.

Assuming that the said prepared data can be classified as QR experiments, three normalized integral curves Jr (r = up, mn, and dn) can be obtained according to the clustering procedure described in the previous section. Our further goal is to describe these nine curves for the three original matrices with the minimum number of fitting parameters. Careful analysis shows that, without violation of the validity of the proposed general theory, we can assume in expression (15) that $\mathrm{Pr}_1(x) = \mathrm{Pr}_2(x) \approx \mathrm{Pr}(x)$. The functions $\kappa_{1,2}(x)$ are defined in the same expression in the line below. The unknown functions $a_{1,2}(x)$, which calculate the roots $\kappa_{1,2}(x)$, are derived from expressions (7) and (8). The unknown nonlinear parameter in the form of the maximum period T that appears in the definition of the periodic function (2) can be calculated through minimizing the relative error

$$\mathrm{RelErr}(T_{mx}, K_{\min}) = \min\left(\frac{\mathrm{stdev}(J_r(x) - F_0(x, T_{mx}, K_{\min}))}{\mathrm{mean}|J_r(x)|}\right) \times 100\%. \qquad (23)$$

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

$$F_0(x, T_{mx}, K_{\min}) = A_0 + G(x, T_{mx})Ac_0 + G(x, T_{mx}) \times \sum_{k=1}^{K_{\min}}\left[Ac_k \cos\left(2\pi k \frac{x}{T_{mx}}\right) + As_k \sin\left(2\pi k \frac{x}{T_{mx}}\right)\right], \qquad (24)$$
$$G(x, T_{mx}) = \left(\kappa_1(x)^{x/T_{mx}} + \kappa_2(x)^{x/T_{mx}}\right).$$

In the last expression, it is convenient to use the normalized input variable $x_j = j/N$, localized in the interval [0, 1]. Therefore, the most probable interval on which the value of T_{mx} can be found is determined as (0.5, 1.75). This assumption was confirmed by numerical estimates using formula (23). From the simplified fitting function (24), it is easy to find the periodic function (2) corresponding to the IE

$$\mathrm{Pr}(x) \cong A_0 + \sum_{k=1}^{K_{\min}}\left[\mathrm{Ac}_k \cos\left(2\pi k \frac{x}{T_{mx}}\right) + \mathrm{As}_k \sin\left(2\pi k \frac{x}{T_{mx}}\right)\right] \equiv A_0 + \sum_{k=1}^{K_{\min}}[\mathrm{Amd}_k \cos(\omega_k x - \mathrm{Ph}_k)], \qquad (25)$$
$$\mathrm{Amd}_k = \sqrt{Ac_k^2 + As_k^2}, \;\; \mathrm{Ph}_k = \tan^{-1}\left(\frac{As_k}{Ac_k}\right).$$

The simplified fitting function (25) (containing the minimum number of approximation parameters $P_{\mathrm{rm}} = [(T_{\mathrm{mx}}, A_0) + 2K_{\min}]$) allows us to obtain the desired amplitude–frequency response that contains the leading minimum frequency $\omega_{\min} = 2\pi/\mathrm{T}_{\mathrm{mx}}$ and the total frequency segment $\omega_k = \omega_{\min}\mathrm{k}$ located in the interval $[1, K_{\min}]$.

The confidence tube (CT) allows us to select three characteristic curves: two extremal ones and one average one, which are invariant with respect to permutations of all remaining points localized within each column.

To save space for showing similar figures for all data, we present the details of data processing only for methane CH_4. The remaining data are processed in a similar way. To simplify the preliminary data analysis, it is useful to use the concept of the CT. which is formed from three values: maximum, average, and minimum, respectively, taken from each column of the original rectangular matrix. In fact, all other measured data in each column are located inside the CT. The three distributions Mx(m), Mn(m), and Min(m) calculated for all columns m = 1, 2, ..., M = 168 are shown in Fig. 1a.

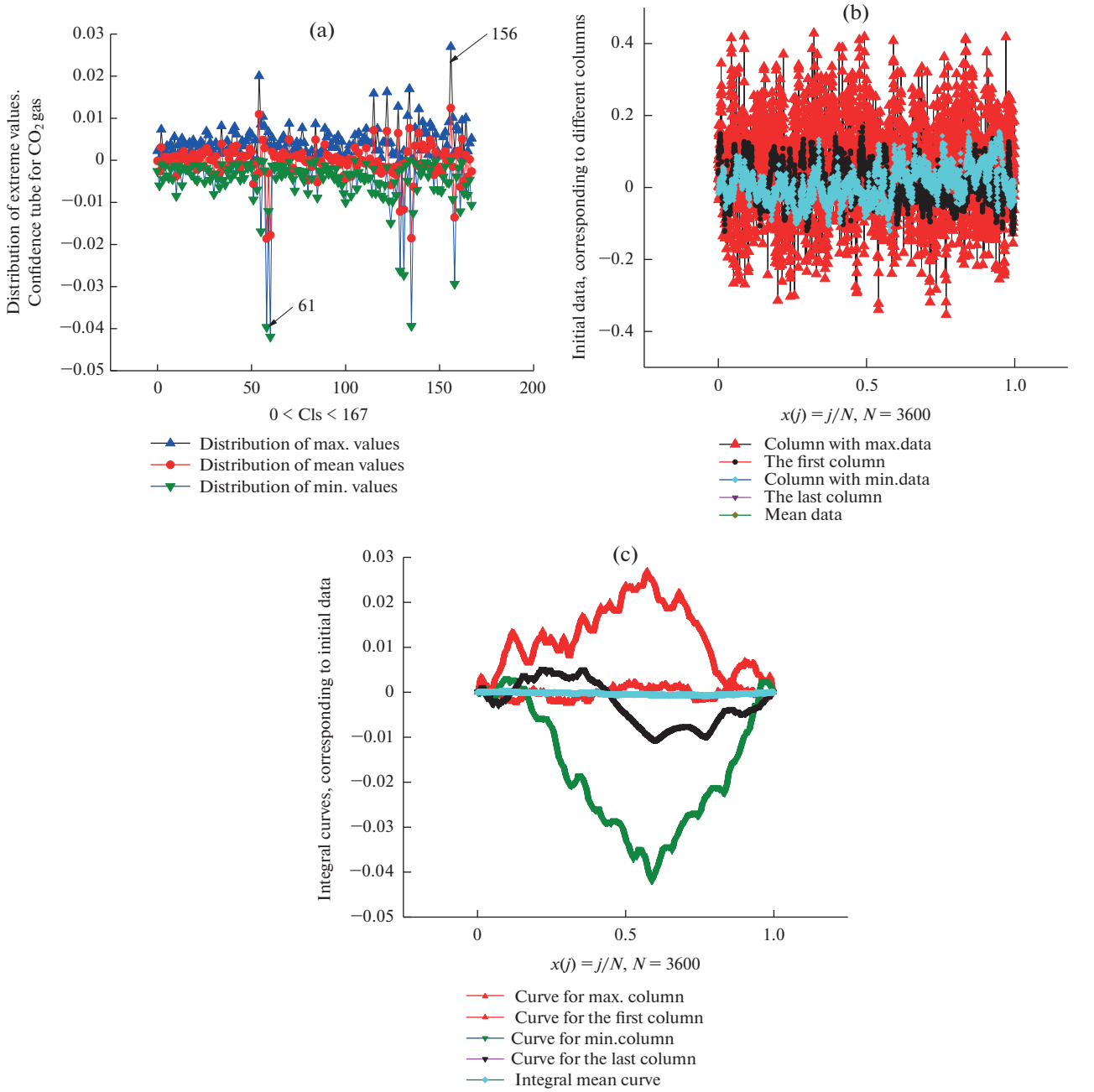

Fig. 1. (a) Confidence tube (distribution of extreme values) for CO_2 gas. Extreme values $M(\max) = 156$, $M(\min) = 61$ are shown by arrows. (b) Some examples of initial data corresponding to different columns, minimum value $m = 61$, maximum value $m = 156$, initial and final columns $m = 0.167$, including average data. (c) Integrated data obtained according to expression (22).

Figure 1b shows the original data corresponding to the extreme values $m = 65$ (minimum value) and $m = 101$ (maximum value), respectively.

All of the original data (Fig. 1b) appear as trendless sequences. They do not demonstrate any specific features relative to their initially prepared cumulative data, as shown in Fig. 1c.

The integration procedure performed according to expression (22) is shown in Fig. 1c.

A completely different picture emerges after the original data are summed (integrated). It is noticeable that the normalized data exhibit oscillatory properties that are aggravated by small fluctuations. Some of the data increase their value, while others tend to decrease their value. Therefore, it is necessary to focus on the analysis of the integrated data presented in this figure (Fig. 1c).

As mentioned above, Fig. 1c demonstrates the effectiveness of the integration. It eliminates high-frequency fluctuations and allows smoothed curves to be obtained for further analysis. Figure 2a shows the distribution of slopes (following the elimination of the unit value).

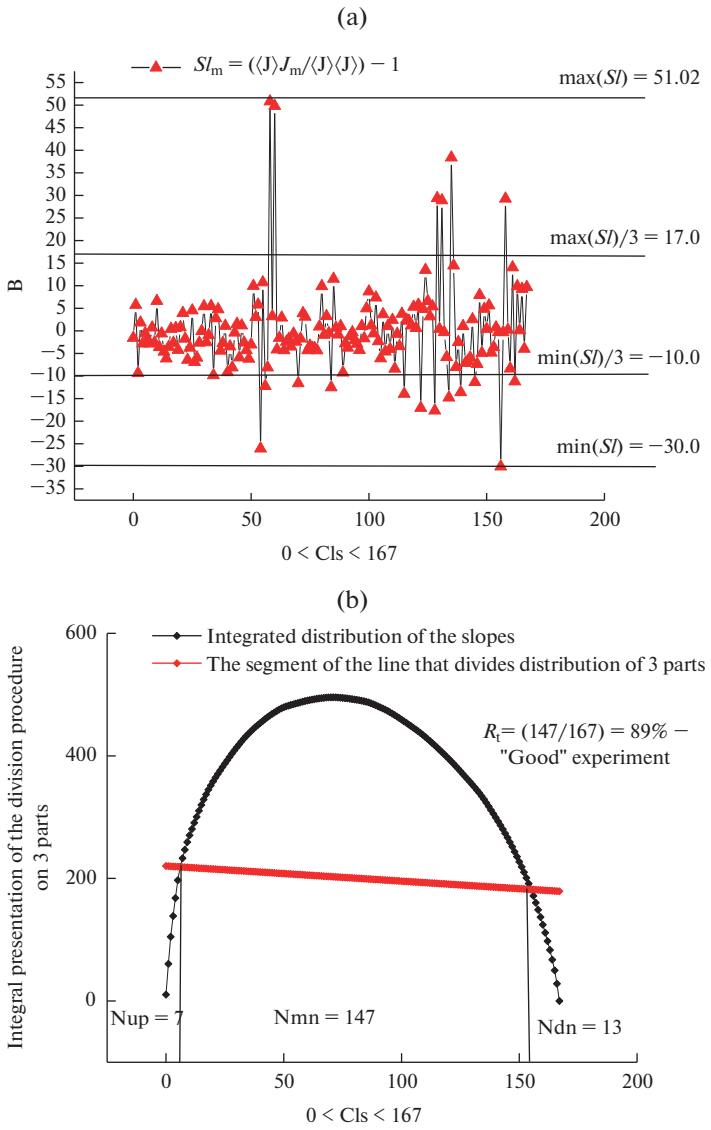

Fig. 2. (a) Distribution of integral slopes $Sl_m = [(J_m \cdot \langle J\rangle)/(\langle J\rangle\langle J\rangle)] - 1$ and dividing them into three parts to obtain averaged integral curves according to expression (20). In this case, $\Delta_{1,2} = 1/3$. (b) For a clear presentation of the results shown in the previous figure, it is convenient to arrange all the slopes and integrate them to obtain a sequence of ranked amplitudes.

The entire segment with boundaries (min(Sl), max(Sl)) can be divided into three almost equal segments: (1/3max(Sl), max(Sl) for $F_0(x)$), (1/3min(Sl),1/3max(Sl) for $F_1(x)$), and (min(Sl), 1/3min(Sl) for $F_2(x)$). The dividing lines for each of the three parts of the measurements are shown in Fig. 2a. The number of measurements that are included in each selected segment is shown in Fig. 2b.

In this representation, the values Nup = 7, Nmn = 147, and Ndn = 13 can easily be found, and an estimate of the quality of this first week experiment Rt = 89% for the data on CO_2 gas can be made. For another gas CH_4, other Rt values can be expected. They are given in Table 1 (see below).

This is the most convenient figure for demonstrating the proposed algorithm using a bell-shaped curve. This information helps the desired averaged integral functions $F_{0,1,2}$ to be obtained, based on expression (20). Figure 3a shows the final result for CH_4 gas.

Relative to pure concentration data, flux data show their oscillatory nature.

These three curves can then be approximated with the simplified function (24). Only one nonlinear approximation parameter, T_{mx}, can be found from minimizing the relative error (23), under the assumption that this parameter is in the interval [0.5T, 1.75T]. The other approximation parameters (A_0, Ac_0, Ac_k, As_k (k = 1, 2, ..., K_{min})) are found using the LLSM.

Table 1. Calculation of the main parameters *N*up, *N*mn, *N*dn, and Rt for CH_4, CO_2, and H_2O for four weeks related to February 2024. For details, see expression (20)

Gas/week	*N*up	*N*mn	*N*dn	*Rt*, %
CH_4/week 1 (0–7)	35	121	11	72.455
CH_4/week 2 (7–14)	21	116	30	69.461
CH_4/week 3 (14–21)	16	141	10	84.431
CH_4/week 4 (21–28)	18	137	12	82.035
CO_2/week 1 (0–7)	7	152	8	91.018
CO_2/week 2 (7–14)	5	157	5	94.012
CO_2/week 3 (14–21)	10	140	17	83.832
CO_2/week 4 (21–28)	17	142	8	85.031
H_2O/week 1 (0–7)	4	159	4	95.211
H_2O/week 2 (7–14)	1	164	2	98.204
H_2O/week 3 (14–21)	14	149	4	89.222
H_2O /week 4 (21–28)	10	154	3	92.216

The approximation of these three curves is shown in Fig. 3b.

The amplitude–frequency response distributions taken as modules and phases $\mathrm{Amd}_k = \left(\mathrm{Ac}_k^2 + \mathrm{As}_k^2\right)^{1/2}$, $\mathrm{Ph}_k = \tan^{-1}\left(\dfrac{\mathrm{As}_k}{\mathrm{Ac}_k}\right)$ for three averaged integral functions $F_{0,1,2}$ are shown in Figs. 4a and 4b, respectively.

In Fig. 4c, we demonstrate three periodic curves that correspond to the ideal experiment.

It is interesting to note that these have different scales and, therefore, they cannot all be shown in one figure. The main fitting parameters that are associated with the periodic functions (25) for all three selected substances are collected in Tables 1–4.

Comments on Tables 1–4. In these tables, we list the main fitting parameters that form the periodic functions (25) that are calculated for all three substances CH_4, CO_2, and H_2O. All of these parameters are related to the first week of February 2024. In Table 4, it should be noted that the parameters related to the averaged integral Jup and Jmn curves coincide with each other (we fix the degenerate case).

The other data related to CO_2 gas and H_2O vapor can be processed in a similar way. Let us show the fitting of the average integral curves for CO_2 gas (Fig. 5a) and H_2O vapor (Fig. 5b).

The outliers that are clearly visible in this figure cannot be explained. In our opinion, more additional modes are needed for a more accurate fit. To compare all the analyzed gases on one scale, we fixed $K = 8$ for all gases.

Here, we showed the fit of only two curves, *J*up and *J*dn. These outline the limits of the other curves, including the averaged *J*mn curve. As noted, to compare all the analyzed gases on one scale, we set $K = 8$ for all of them. If we compare these curves with each other, we observe that the integral curves exhibit oscillatory properties that are not noticeable in the original data.

Below, we only show the main and most expressive figures. In Figs. 6a–6c, we justify the optimal choice of rectangular matrices corresponding to a periodicity of 1 week.

In these figures, one can notice clearly defined and noticeable oscillations of CH_4 gas and pronounced trends in the change of these oscillations when moving to the next possible and independent oscillation. Similar behavior is observed for CO_2 gas and H_2O vapor. In general, this observation is important; it completely

Table 2. Set of basic parameters associated with periodic functions and related to averaged integral curves for CH_4 gas

k	Ac_k (*J*up)	As_k (*J*up)	Ac_k (*J*mn)	As_k (*J*mn)	Ac_k (*J*dn)	As_k (*J*dn)
	T_{mx}	A_0	T_{mx}	A_0	T_{mx}	A_0
T_{mx}, A_0	1.0635	4.60467E–6	1.9591	1.26201E–5	1.9365	2.48197E–5
$k = 1$	–2.88078E–4	–5.51339E–5	5.19784	17.2207	–0.01551	0.66897
$k = 2$	–7.46133E–5	4.28281E–5	11.3271	–7.52796	0.52892	0.02386
$k = 3$	–2.62737E–5	8.27381E–6	–6.51193	–5.37605	0.02362	–0.35579
$k = 4$	–1.19733E–5	1.03999E–5	–1.62614	3.90184	–0.19997	–0.01698
$k = 5$	–6.33079E–6	1.27707E–5	1.64533	0.16105	–0.00907	0.09133
$k = 6$	1.66724E–7	7.38361E–6	–0.09653	–0.46822	0.03229	0.00344
$k = 7$	4.66842E–6	3.3709E–6	–0.0801	0.04488	8.36735E–4	–0.00801
$k = 8$	4.53143E–6	1.54078E–6	0.00657	0.00599	–0.00109	–8.27201E–5

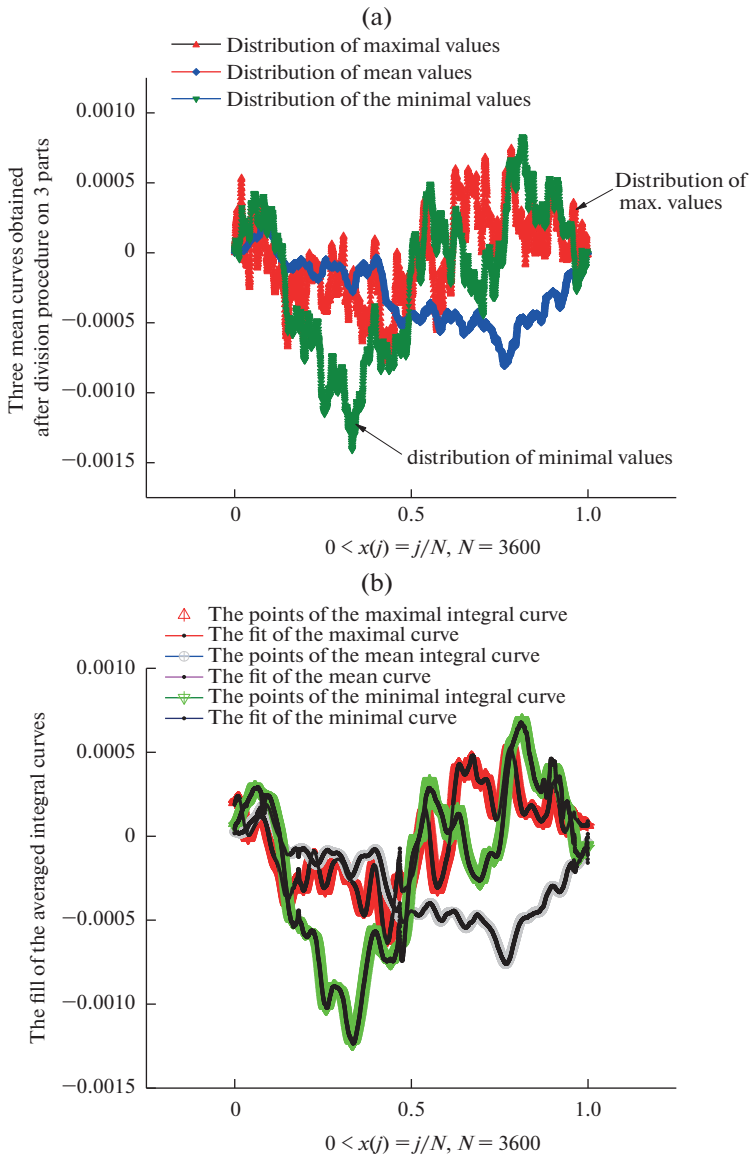

Fig. 3. (a) Three averaged integral curves (normalized and calculated as integral data) for CO_2 gas, which were finally obtained for the procedure of fitting with function (24) and reduction to an ideal experiment. (b) Fitting the averaged integral curves Jup, Jmn, and Jdn to the simplified fitting curve (24).

changes all previous and generally accepted algorithms for analyzing environmental (eddy-covariance) data.

4. RESULTS AND DISCUSSION

We proposed a theory for the QREs, based on a given sample of successive measurements in a single-factor experiment. These measurements must be specified as a rectangular matrix $N \times M$, the columns of which ($m = 1, 2, ..., M$) include data points ($j = 1, 2, ..., N$) that correspond to one experiment implemented during one period T. All columns ($m = 1, 2, ..., M$) define the entire cycle of successive/repeated measurements that are implemented across the entire observation period $M{\cdot}T$. Thanks to the described algorithm, it is possible to obtain only three key averaged curves (see expression (20)), which can be described by a simplified fitting function (24). This function plays the dual role of (a) fitting the averaged measurements and (b) extracting a pure periodic function (25) corresponding to an ideal experiment. The other two important results can be formulated as follows.

1. Creation of fully computerized laboratories, where the input coincides with the initial data presented in the form of a matrix $N \times M$, and the output coincides with a limited number of adjustable parameters $Prm = [(T_{mx}, A_0, Ac_0) + 2K_{min}]$ following from expression (25).

2. Creation of a new metrological standard based on a comparison of the template detector/device with those tested according to the optimal number of

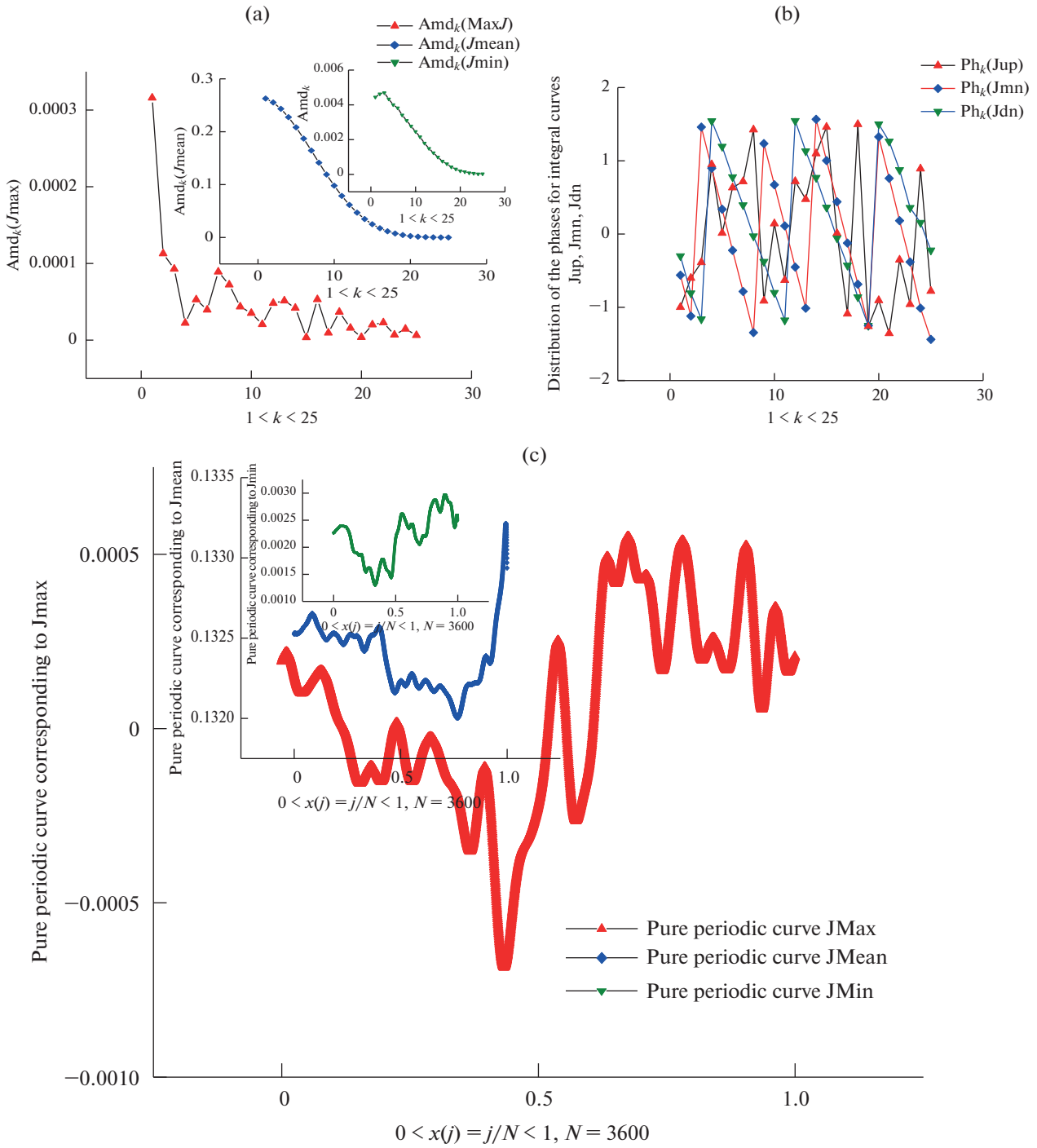

Fig. 4. (a) Distributions of the moduli Amd_k for all three integral curves Jmn, Jmn, and Jdn shown in Fig. 3 (b). It is interesting to note that it is impossible to place them in one figure, since they have different amplitudes. (b). Distributions of phases $Ph_k = \tan^{-1}(As_k/Ac_k)$ for three curves shown in Fig. 3(b). (c) Modified purely periodic curves Pr(x) obtained according to expression (25) from simplified fitting curve (24). The periodic parameters are shown in Figs. 4a and 4b.

adjustable parameters Prm obtained from the registered fluctuations.

Let us note the following fact: the periodicity we discovered in the previous data [7] (data on the concentrations of carbon dioxide, methane, and water vapor), disappears in the presented data (balance/amount of carbon dioxide, methane, and water vapor). It is destroyed in wind flows.

A problem requiring further study follows from the solution of (19) based on the assumption $a_l(x)$,

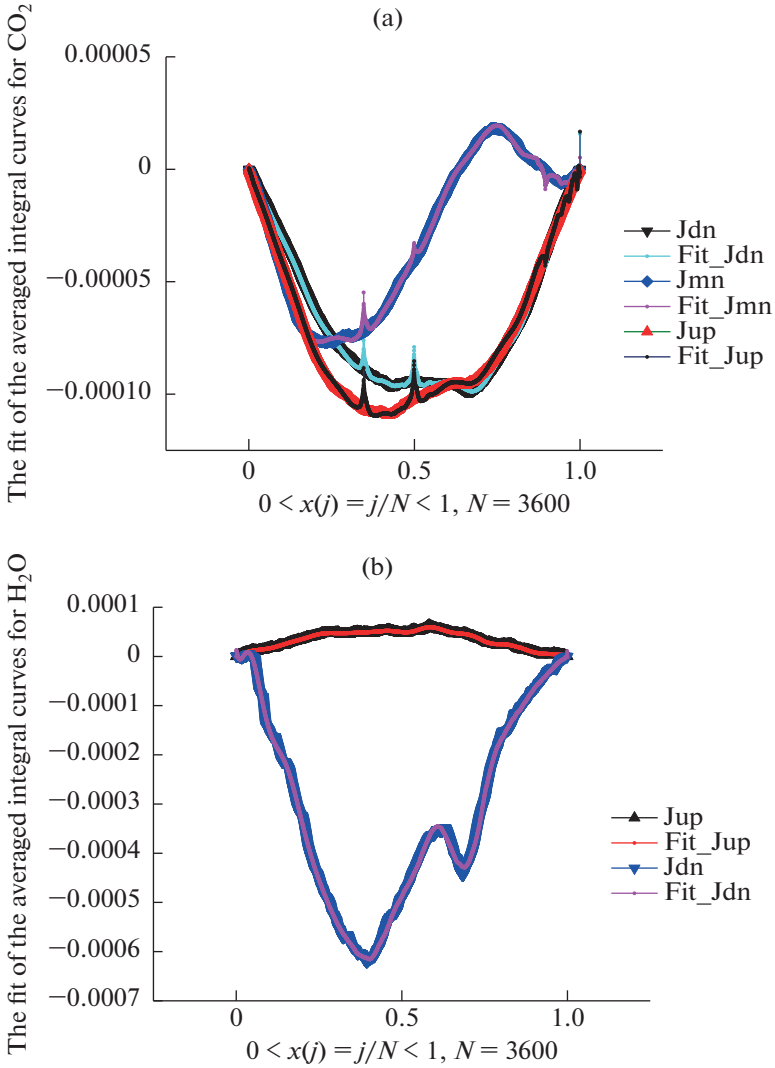

Fig. 5. (a) Fitting averaged integral curves for CO_2 gas. (b) Fitting averaged integral curves for H_2O vapor.

Table 3. Set of basic parameters associated with periodic functions and related to averaged integral curves for CO_2 gas

k	Ac_k (Jup)	As_k (Jup)	Ac_k (Jmn)	As_k (Jmn)	Ac_k (Jdn)	As_k (Jdn)
	T_{mx}	A_0	T_{mx}	A_0	T_{mx}	A_0
T_{mx}, A_0	2.4994	−4.87956E−5	2.4994	−1.89231E−5	2.4994	−5.15067E−5
$k = 1$	5.739	35.1052	8.32128	48.8561	7.0547	44.1951
$k = 2$	27.9061	−9.51763	38.6716	−13.8051	35.3632	−11.8506
$k = 3$	−10.336	−18.7419	−14.9711	−25.7217	−13.0844	−23.9188
$k = 4$	−10.3636	8.57309	−13.9556	12.3532	−13.2663	11.0218
$k = 5$	5.59923	4.51429	7.98282	5.86663	7.27551	5.74028
$k = 6$	1.42352	−2.85784	1.71977	−4.0054	1.76143	−3.72706
$k = 7$	−1.10406	−0.25837	−1.51022	−0.24596	−1.43388	−0.28867
$k = 8$	0.00563	0.30348	0.03973	0.40179	0.02329	0.38935
$k = 9$	0.05261	−0.01533	0.06669	−0.02622	0.06612	−0.02288
$k = 10$	−0.00271	−0.00426	−0.004	−0.00509	−0.00372	−0.00519

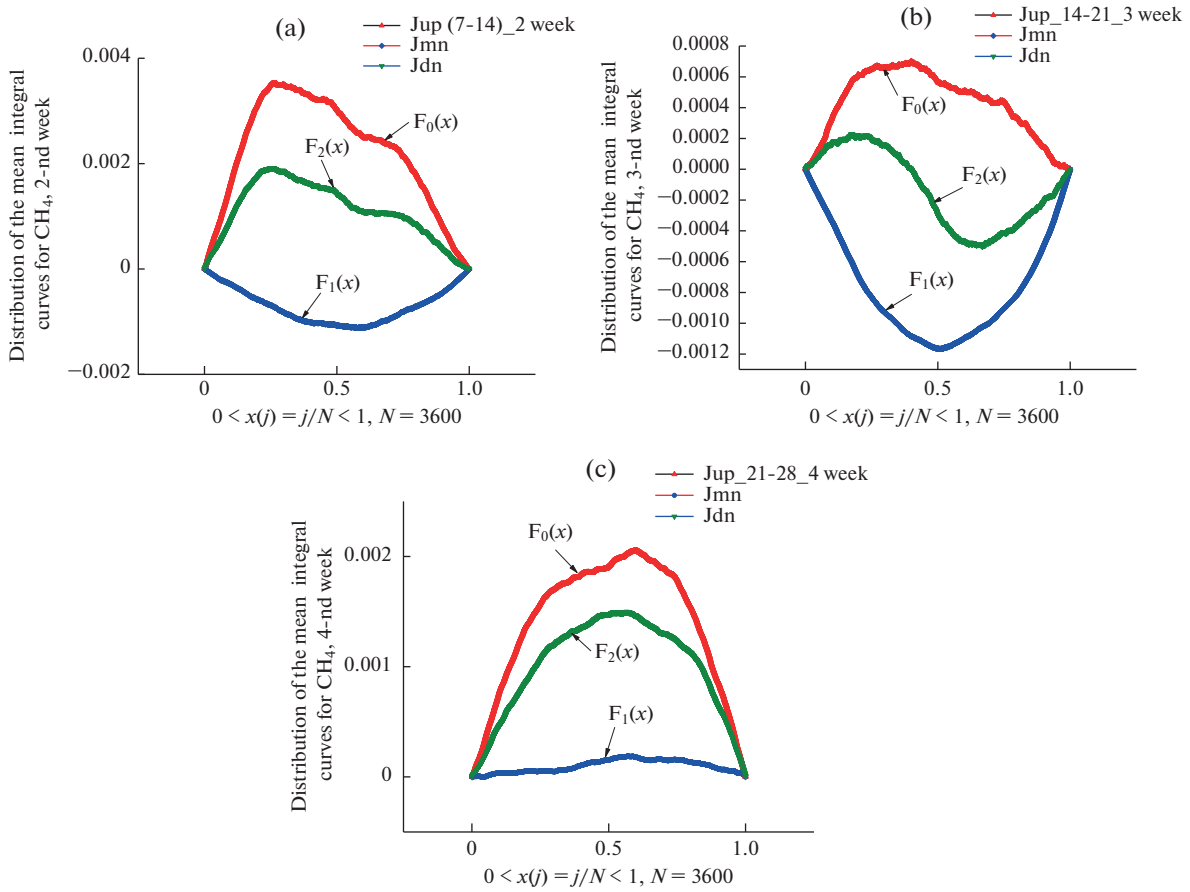

Fig. 6. (a) Evolution of three averaged integral curves during the second week. (b) Evolution of three averaged integral curves during the third week. (c) Evolution of three averaged integral curves during the last fourth week corresponding to February 2024.

Table 4. Set of basic parameters associated with periodic functions and related to averaged integral curves for H_2O vapors

k	Ac_k (Jup)	As_k (Jup)	Ac_k (Jmn)	As_k (Jmn)	Ac_k (Jdn)	As_k (Jdn)
	T_{mx}	A_0	T_{mx}	A_0	T_{mx}	A_0
T_{mx}, A_0	1.3505	−1.27131E−5	1.3505	−1.27131E−5	1.09098	3.59308E−6
$k = 1$	0.00137	−0.00128	0.00137	−0.00128	−3.27398E−5	−8.20476E−6
$k = 2$	5.30561E−5	−0.00145	5.30561E−5	−0.00145	−3.27837E−6	1.00861E−5
$k = 3$	−7.98729E−4	−9.09201E−4	−7.98729E−4	−9.09201E−4	1.17805E−6	7.3396E−7
$k = 4$	−0.00101	−1.24383E−4	−0.00101	−1.24383E−4	7.9935E−7	1.58899E−6
$k = 5$	−5.68944E−4	4.69685E−4	−5.68944E−4	4.69685E−4	1.83581E−6	−7.37466E−7
$k = 6$	−9.90357E−5	4.96939E−4	−9.90357E−5	4.96939E−4	−1.5128E−6	−6.08823E−8
$k = 7$	1.89188E−4	2.73372E−4	1.89188E−4	2.73372E−4	−1.24363E−6	8.58641E−7
$k = 8$	1.63936E−4	3.39089E−5	1.63936E−4	3.39089E−5	1.38365E−6	2.41622E−6
$k = 9$	7.96425E−5	−2.93135E−5	7.96425E−5	−2.93135E−5	1.4022E−6	−1.53316E−6
$k = 10$	3.69983E−6	−3.16791E−5	3.69983E−6	−3.16791E−5	−8.70399E−7	−1.5842E−8

$\kappa(x \pm T) = \kappa(x)$, $\Pr(x \pm T) = \Pr(x)$. What new solutions will the rejection of this assumption lead to? The answer to this question requires further study.

ABBREVIATIONS AND NOTATION

AFR	amplitude-frequency response
(F)LLSM	(functional) linear least square method
IM	intermediate model
IE	ideal experiment
QRE(s)	quasi-reproducible experiment(s)
VPCP	verified partial correlation principle
CT	confidence tube
HF	high-frequency

FUNDING

The work was carried out using funds from a subsidy allocated to the Kazan Federal University to fulfill a state assignment in the field of scientific activity, project no. FZSM-2024-0004.

CONFLICT OF INTEREST

The authors of this work declare that they have no conflicts of interest.

REFERENCES

1. Nigmatullin, R. and Rakhmatullin, R., Detection of quasi-periodic processes in repeated measurements: New approach for the fitting and clusterization of different data, *Commun. Nonlinear Sci. Numer. Simul.*, 2014, vol. 19, no. 12, pp. 4080–4093. https://doi.org/10.1016/j.cnsns.2014.04.013
2. Nigmatullin, R.R., Khamzin, A.A., and Machado, J.C., Detection of quasi-periodic processes in complex systems: How do we quantitatively describe their properties?, *Phys. Scr.*, 2013, vol. 89, no. 1, p. 15201. https://doi.org/10.1088/0031-8949/89/01/015201
3. Nigmatullin, R.R., Osokin, S.I., Baleanu, D., Al-Amri, S., Azam, A., and Memic, A., The first observation of memory effects in the infrared (FT-IR) measurements: Do successive measurements remember each other?, *PLoS One*, 2014, vol. 9, no. 4, p. e94305. https://doi.org/10.1371/journal.pone.0094305
4. Nigmatullin, R.R., Osokin, S.I., and Rakhmatullin, R.M., How to reduce reproducible measurements to an ideal experiment?, *Magn. Reson. Solids*, 2014, vol. 16, no. 2, p. 3.
5. Nigmatullin, R.R., Zhang, W., and Striccoli, D., General theory of experiment containing reproducible data: The reduction to an ideal experiment, *Commun. Nonlinear Sci. Numer. Simul.*, 2015, vol. 27, nos. 1–3, pp. 175–192. https://doi.org/10.1016/j.cnsns.2015.02.024
6. Kuczma, M., A servey of the theory of functional equations, *Publikacije Elektrotehnickog Fakulteta Univerziteta u Beogradu*, 1964, vol. 130, pp. 1–64.
7. Nigmatullin, R.R., Litvinov, A.A., and Osokin, S.I., Quasi-reproducible experiments: Universal fitting function for quantitative description of complex systems data, *Lobachevskii J. Math.*, 2024, vol. 45, no. 8, pp. 3959–3974. https://doi.org/10.1134/s1995080224604739

Translated by L. Solovyova

SPELL: OK