Detection of quasi-periodic processes in complex systems: how do we quantitatively describe their properties?

Raoul R Nigmatullin, Airat A Khamzin and J Tenreiro Machado

Abstract

It has been shown that in reality at least two general scenarios of data structuring are possible:
(a) a self-similar (SS) scenario when the measured data form an SS structure and
(b) a quasi-periodic (QP) scenario when the repeated (strongly correlated) data form random
sequences that are *almost* periodic with respect to each other. In the second case it becomes
possible to describe their behavior and express a part of their randomness *quantitatively* in
terms of the deterministic amplitude–frequency response belonging to the generalized Prony
spectrum. This possibility allows us to re-examine the conventional concept of measurements
and opens a new way for the description of a wide set of different data. In particular, it
concerns different complex systems when the 'best-fit' model pretending to be the description
of the data measured is *absent* but the barest necessity of description of these data in terms of
the reduced number of quantitative parameters exists. The possibilities of the proposed
approach and detection algorithm of the QP processes were demonstrated on actual data:
spectroscopic data recorded for pure water and acoustic data for a test hole. The suggested
methodology allows revising the accepted classification of different incommensurable and
self-affine spatial structures and finding accurate interpretation of the generalized Prony
spectroscopy that includes the Fourier spectroscopy as a partial case.

Keywords: complex systems, random data processing, quasi-periodic process, the generalized
Prony spectrum

(Some figures may appear in colour only in the online journal)

List of acronyms

AFR	amplitude–frequency response
GMV-function	generalized mean value function
GPCF	generalized Pearson correlation function
GPS	generalized Prony spectrum
CC-factor	complete correlation factor
LLSM	linear least-squares method
QP-process	quasi-periodic process
QT-property	quasi-translational property
SRA	sequence of the ranged amplitudes
SS-process	self-similar process

1. Introduction and formulation of the problem

Scientific research pursues one important purpose: to
find the relationships existing between two (or more)
measured variables and express them in terms of fundamental
mathematical relationships. Actually, all these fundamental
relationships (so-called laws) constitute the basis of natural
sciences and this paradigm is emerging in social sciences
also. This is why researchers are trying to establish unknown
regularities (rules) existing in nature between different
variables. Nevertheless, it became evident that with the
increasing complexity of a system at different stages

of its evolution, the fundamental and simple (from the mathematical point of view) rules that had been established earlier were difficult to find and then (if they were found) to justify. It implies, particularly, the cases where the organization of the matter forms some complex structure. In many cases these relationships are *diffusive* and covered by the interruption of other uncontrollable factors known in measurements as an influence of random fluctuations or 'noise'. These uncontrollable factors can hide completely the desired relationship and, therefore, in many cases these factors play a destructive role. Besides these hidden (in most cases) relationships, there are other general rules that can characterize the degree of correlation between self-similar (SS) temporal sequences. These correlated relationships can serve as a specific indication of the existence of a relevant law and, on a certain stage, help in the understanding of a specific behavior of the complex system studied. The development of the fractal geometry [1, 2] and its impact in the development of the fractional calculus [3, 4] allows one to establish new mathematical relationships expressed in terms of the non-integer operators of differentiation and integration. In [5] the first author solved the inverse problem that existed for SS structures developing on different time scales. In other words, the original algorithm was suggested, which (a) tested the presence of an SS structure in the random sequence considered; (b) allowed the corresponding functional equation that describes the SS structure to be restored and (c) provided a good fit of the random data considered to the analytical function that follows from the solution of the corresponding functional equation. Nevertheless, we think that this SS scenario, in spite of its attractiveness and generality, is *not* unique. In this paper we want to show that there is another possibility of the existence of the almost periodic processes which many complex systems can reveal in their temporal evolution. We define this type of temporal evolution as the quasi-periodic (QP) process and describe its general properties. So, the problem that is considered in this work can be outlined as follows.

1. Formulate some general conditions that help to identify the presence of some QP process in the random data considered.
2. Find a functional equation and its solution that yields the description of the identified QP process.
3. Suggest some computing procedure for the fitting of the QP data to the analytical function that follows from the solution of the corresponding functional equation.

The content of this paper is organized as follows. In section 2 we formulate the answers to the problem posed above. Furthermore, it contains the mathematical description of the QP process and the interpretation of the meaning of the generalized Prony spectrum (GPS) that includes the conventional Fourier decomposition as a *partial* case. Section 3 contains the basic steps of the mathematical algorithm that is tested on actual data. Finally, in section 4 we summarize the results and outline the perspectives of this approach for the quantitative description of time-dependent random data that are registered in different complex systems.

We want to stress here the following idea. The detection of the QP processes helps to suggest a new scheme

for the classification of the disordered (self-affine) and incommensurable systems. They have a specific memory and this spatial memory can be expressed also in terms of spatial GPS. These possibilities are also discussed in the final section.

Here and below, under the *complex system* we imply a system when a conventional model is *absent* [6]. Under *simplicity* of the acceptable model we imply the proper hypothesis ('best fit' model) containing a minimal number of the fitting parameters (FP) that describe the behavior of the system considered quantitatively. The different approaches that exist nowadays for the description of these systems are collected in a recent review [7].

2. Description of a quasi-periodic process in terms of Prony's spectrum

It is well-known that a pure periodic process with the given period T satisfies the following functional equation:

$$\Pr(t \pm T) = \Pr(t). \tag{1}$$

The general solution of this functional equation is known and it can be expressed in the form of the Fourier series (if the initial function is defined on the discrete set of the given points $[t_j] j = 1, 2, \ldots, N$)

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

Instead of equation (1) we consider a somewhat generalized temporal process

$$F(t + T) = aF(t) + b, \tag{3}$$

where the parameters a and b determine some real constants. This functional equation means that the temporal evolution of some process taking place on the interval $t > T$ is based on events that took place presumably in the nearest past $(t < T)$. This functional equation was considered for the first time in [8] but in the present paper we want to expand it beyond the scope of the previous consideration. The solution of this equation can be written in the following form [8]:

$$a \neq 1 : F(t) = \exp(\lambda t) \Pr(t) + c_0, \lambda = \frac{\ln(a)}{T}, c_0 = \frac{b}{1-a},$$

$$a = 1 : F(t) = \Pr(t) + b\frac{t}{T}. \tag{4}$$

If $a > 1$ then we have the *increasing* exponential factor $(\lambda > 0)$. For this situation the influence of the past events on the present event is essential. For $a < 1$ we have the effect of the exponential decay $(\lambda < 0)$ and in this case the influence of the past events (that took place for $t < T$) on the present event $(t > T)$ is negligible. For $a = 1$ $(b \neq 0)$ we have alongside with periodic oscillations the appearance of a linear temporal trend and, finally, for $a = 1$ and $b = 0$ the solution (4) is reduced to conventional solution (2). It is natural that equation (3) admits the following generalization:

$$F(t + LT) = \sum_{s=0}^{L-1} a_s F(t + sT) + b. \tag{5}$$

This functional equation describes mathematically a wide class of the QP processes and can be interpreted as follows. The process that takes place during the interval $[(L-1)T, LT]$ partly depends on the processes that have happened during the previous temporal intervals $[sT, (s+1)T]$ with $s = 0, 1, \ldots, L-2$. The set of constants $[a_s]$ $(s = 0, 1, \ldots, L-1)$ can be quantitatively interpreted as the presence of a memory in the system considered. In comparison with the functional equation (5), equation (3) can be interpreted as a system having the *shortest* memory. The solution of the generalized functional equation (5) can be presented in two forms (A) and (B):

$$(A) \quad \sum_{s=0}^{L-1} a_s \neq 1: \quad F(t) = \sum_{r=1}^{L} \exp\left(\ln(\lambda_r)\frac{t}{T}\right)$$

$$\times \Pr_r(t) + c_0, \quad c_0 = \frac{b}{1 - \sum_{s=0}^{L-1} a_s},$$

$$(B) \quad \sum_{s=0}^{L-1} a_s = 1: \quad F(t) = \sum_{r=1}^{L} \exp\left(\ln(\lambda_r)\frac{t}{T}\right) \quad (6)$$

$$\times \Pr_r(t) + c_1\frac{t}{T}, \quad c_1 = \frac{b}{L - \sum_{s=0}^{L-1} s a_s}.$$

Here the functions $\Pr_r(t)$ define a set of periodic functions from expressions (2), the values λ_r $(r = 1, 2, \ldots, L)$ coincide with the roots of the characteristic polynomial

$$P(\lambda) = \lambda^L - \sum_{s=0}^{L-1} a_s \lambda^s = 0. \quad (7)$$

In general these roots can be positive, negative, g-fold degenerated (with the value of the degeneracy g) and complex conjugated. We should note also that for the case (B) in (6) one of the roots λ_r coincides with the unit value ($\lambda_1 = 1$) that leads to the pure periodic solution. As before, the finite set of the unknown periodic functions $\Pr_r^{(T)}(t)(r = 1, 2, \ldots, L)$ is determined by their decomposition coefficients $Ac_k^{(r)}, As_k^{(r)}, r = 1, 2, \ldots, L; \; k = 1, 2, \ldots, K$:

$$\Pr_r^{(T)}(t) = A_0^{(r)} + \sum_{k=1}^{K \gg 1} \left[Ac_k^{(r)} \cos\left(2\pi k \frac{t}{T}\right) \right.$$

$$\left. + As_k^{(r)} \sin\left(2\pi k \frac{t}{T}\right) \right]. \quad (8)$$

We want to stress here the following fact. The conventional Prony decomposition [9–11] implies that exponential multipliers figuring before periodic functions have only real and decaying values. But solution (6) implies that other roots from algebraic equation (7) are also possible. These roots can change completely the structure of the periodic part (8). In particular, it is instructive to reproduce the corresponding decompositions [8] for the case of negative, degenerated and complex-conjugated roots also. When a root λ from (7) accepts a *negative* value the function in (8) becomes *anti-periodic* (see definition (9) below) and the

solution for this case is written as

$$F(t) = \exp\left(\ln|\lambda|\frac{t}{T}\right) \Pr^{(a)}(t), \quad \Pr^{(a)}(t+T) = -\Pr^{(a)}(t),$$

$$\Pr^{(a)}(t) = \sum_{k=1}^{K \gg 1} \left[Ac_k \cos\left((2k-1)\pi\frac{t}{T}\right) \right. \quad (9)$$

$$\left. + As_k \sin\left((2k-1)\pi\frac{t}{T}\right) \right].$$

If the found root λ_g $(\lambda_g > 0)$ from (7) is g-fold degenerated with the degree of the degeneracy equaling g then the solution for $F(t)$ is written as

$$F(t) = \exp\left(\ln(\lambda_g)\frac{t}{T}\right) \sum_{r=0}^{g-1} \left(t^r \Pr_r(t)\right),$$

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

Finally, we write the solution for the complex-conjugated roots $\lambda_c = |\lambda_c| \exp(\pm j\varphi)$

$$F(t) = \exp\left(\ln|\lambda_c|\frac{t}{T}\right) \Pr_c(t),$$

$$\Pr_c(t) = \sum_{k=0}^{K \gg 1} \left[Ac_k \cos\left((2k\pi + \varphi)\frac{t}{T}\right) \right. \quad (11)$$

$$\left. + As_k \sin\left((2k\pi + \varphi)\frac{t}{T}\right) \right].$$

Taking into account these decompositions one can say that they can be referred to as the amplitude (Ac_k, As_k)-frequency ($\omega_k = (2k\pi + \varphi)/T$) response (AFR) corresponding to the GPS because they take into account the different types of roots that can appear in general solution (6). These decompositions (8)–(11) have clear meaning and correspond to the linear recording of memory effects that can exist in the random sequences considered. The memory effect (considered for the discrete set of data) is expressed quantitatively by the enumerable set of real constants $\{a_s\}$ figuring in equation (5). In this sense the process without memory (Markovian process) corresponds to the following set of constants: $a_0 = 0$, $a_1 = 0, \ldots, a_{L-1} = 1$ and its solution coincides with a pure periodic function (2). So, in this sense the Fourier series can be interpreted as the process without memory. From another side, the conventional expression for the mean function when in expression (5) $\langle F(t) \rangle = F(t + (N+1)T)$ and $a_s = 1/N$, $b = 0$ can be interpreted as the process with uniform memory.

3. Description of the general algorithm

In this section we propose an algorithm that detects the QP process from the real data and performs the corresponding fit to general solution (6). The algorithm can be divided into two fundamental steps.

Step 1: *finding the value of T*. The first important step is related to finding the period T from *a priori* information.

If this parameter can be found then in the general fitting procedure it becomes easier to reduce the nonlinear fit to the well-known linear least-squares method (LLSM). The development of the method when this *a priori* information is absent merits special research. In order to find the value T it is necessary to have data repeated many times with some fixed period of T. If some necessary conditions for repeating the measuring process are fulfilled then one can link this time of recycling with T.

What is it necessary to undertake if this period T is known approximately?

In this case one can use the statistics of the fractional moments [12, 13] and suggest a procedure that can separate statistically homogeneous segments of the length T from general sampling having the length $T_L \gg T$. In order to find these statistically homogeneous segments it is necessary to suggest at least two criteria.

Firstly, the criterion based on external correlations.

For the sequence having N number of points within the temporal segment of length T_L we calculate the generalized mean value (GMV) function

$$\mathrm{GMV}_p(s) = \left(\frac{1}{N} \sum_{j=1}^{N} \left| \mathrm{nrm}_j(s) \right|^{\mathrm{mom}_p} \right)^{1/\mathrm{mom}_p},$$

$$\mathrm{mom}_p = \exp\left(Ln_p\right), \quad Ln_p = mn + \left(\frac{p}{P}\right)(mx - mn), \quad (12)$$

$$p = 0, 1, \ldots, P.$$

Here the value s fixes the given segment. The normalized sequence located in the interval $0 < \mathrm{nrm}(y) < 1$ is determined below by an expression (13). The value mom_p determines the current moment from the interval $[0, P]$. The value P determines the final value of the linear function Ln_p located in the interval $[mn, mx]$. The values mn and mx define correspondingly the limits of the moments in the uniform logarithmic scale. In many practical cases these values are chosen as $mn = -15$, $mx = 15$ and P is chosen as an integer value located in the interval [50–100]. This choice is related to the fact that the transition region of the random sequences considered and expressed in the form of the GMV-functions is concentrated usually in the interval $Ln_p \in [-5, 5]$. The extended interval $[-15, 15]$ is taken usually for the calculation of the limiting values of this function in the space of the fractional moments. The initial sequences are chosen in that way: the minimum of the GMV-function coincides with zero value while the maximal value of this function coincides with $\max(\mathrm{nrm}_j(y))$. In formula (12) the random sequence is normalized to the unit value in accordance with expressions (A) and (B):

$$(A) \quad \mathrm{nrm}_j(y) = \frac{y_j^{(+)}}{\max(y_j^{(+)})} - \frac{y_j^{(-)}}{\min(y_j^{(-)})},$$

$$y_j^{(\pm)} = \tfrac{1}{2}\left(y_j \pm |y_j|\right),$$

$$(B) \quad \mathrm{nrm}_j(y) = \frac{\Delta y_j}{\max(\Delta y_j)}, \quad \Delta y_j = y_j - \min(y_j). \tag{13}$$

$$j = 1, 2, \ldots, N, \quad 0 < \mathrm{nrm}(y) < 1.$$

Here, as it was done above, the set y_j defines the initial random sequence that can contain a trend or can be compared with another trendless sequence. The symbol $|\ldots|$ and index j ($j = 1, 2, \ldots, N$) determine the absolute value and number of the measured points, correspondingly. The second case (B) in (13) corresponds to the case when the initial sequence is positive. For detection of the external correlations between two sequences s_1 and s_2 formed from initial sequence T_L it is necessary to calculate $\mathrm{GMV}_p(s_1)$ and $\mathrm{GMV}_p(s_2)$ and to plot the function $\mathrm{GMV}_p(s_2)$ (having initially the same chosen length $T = T_L$) with respect to the previous function $\mathrm{GMV}_p(s_1)$. If this dependence forms a curve close to the straight line

$$\mathrm{GMV}_p(s_2) \cong \mathrm{Sl}\,\mathrm{GMV}_p(s_1) + \mathrm{It}, \tag{14}$$

(where Sl defines the slope and It corresponds to interception) then the initial segments compared are similar and are strongly correlated. Then, using this simple test expression for other sequences, one can select all that are contained in the given sampling having the length $T_L \gg T$. So, after this selection one can form the desired relationship (5) and evaluate the unknown set of constants $[a_s]$ ($s = 0, 1, \ldots, L - 1$) by LLSM.

Secondly, the criterion based on internal correlations.

Besides this criterion, based on the *external* correlations, one can suggest a more delicate selection of similar sequences based on *internal* correlations. For this we introduce the generalized Pearson correlation function (GPCF) [13]

$$\mathrm{GPCF}_p = \frac{\mathrm{GMV}_p(s_1, s_2)}{\sqrt{\mathrm{GMV}_p(s_1, s_1)\mathrm{GMV}_p(s_2, s_2)}}, \tag{15}$$

where

$$\mathrm{GMV}_p(s_1, s_2, \ldots, s_K) =$$

$$\left(\frac{1}{N} \sum_{j=1}^{N} \left| \mathrm{nrm}_j(s_1) \cdot \mathrm{nrm}_j(s_2) \cdot \ldots \cdot \mathrm{nrm}_j(s_K) \right|^{\mathrm{mom}_p} \right)^{1/\mathrm{mom}_p} \tag{16}$$

determines the GMV-function of the Kth order. The GPCF_p determined by expression (15) coincides with the conventional definition of the Pearson correlation coefficient at $\mathrm{mom}_p = 2$. Other values are determined by expressions (12) and (13). If the limits mn and mx in (12) have the opposite signs and accept sufficiently large values, then the GPCF function has two plateaus (equaled unit at small numbers of mn (i.e. $\mathrm{GPCF}_{mn} = 1$)) and another limiting value GPCF_{mx} depends on the degree of internal correlation between two random sequences compared. This right-hand limit (defined as Lm) is located between two values

$$M \equiv \min(\mathrm{GPCF}_p) \leqslant Lm \equiv \mathrm{GPCF}_{mx} \leqslant 1. \tag{17}$$

The appearance of two plateaus implies that all information about possible correlations is complete and further increase of the limiting numbers (mx, mn) in (12) is *useless*. Numerous tests showed that the high degree of correlation between two random sequences is achieved when $Lm = 1$, while the lowest correlation is observed when $Lm = M$. This simple observation, having a general character for all random

sequences, allows us to introduce a new correlation parameter, the complete correlation (CC) factor, which is determined as

$$CC = M \left(\frac{Lm - M}{1 - M} \right). \qquad (18)$$

We would like to stress here that this factor is determined on the *total* set of the fractional moments located between $\exp(mn)$ and $\exp(mx)$ (see expression (12)). As mentioned above, in practical calculations for many cases it is sufficient to put $mn = -15$ and $mx = +15$. The CC factor is equal to unit value when the degree of correlation is high and the case $Lm = M$ corresponds to the lowest (remnant) degree of correlation that can be observed between the compared random sequences. In addition, we also want to stress the following fact. This CC factor does *not* depend on the amplitude of the random sequences. The pair random sequences compared should be normalized to the interval $0 \leqslant |y_j| \leqslant 1$. It reflects the *internal* structure of correlations of the compared random sequences based presumably on the similarity of their probability distribution functions that are *not* known in many cases. A recent example related to the application of the statistics of the fractional moments was considered in paper [14].

These two criteria based on statistical proximity of the random sequences compared help to form the basic functional equation (5) and, hence, to find the desired set of the constants $[a_s]$ $(s = 0, 1, \ldots, L - 1)$ (by the LLSM) forming a specific discrete memory of the complex system considered.

Step 2: the final fit of the statistically close sequences to the fitting function (6). In practical calculations it is desirable to have similar sequences with small value of the parameter $L(= 2, 3, 4)$ in (5) that corresponds to the short memory. This requirement is related to the following observation. If we limit the decompositions in (8)–(11) by the fixed number of modes $(k = 1, 2, \ldots, K)$ then the total number of FPs are located in the interval

$$2KL \leqslant FP \leqslant (2K + 1) L. \qquad (19)$$

From the last expression it follows that all this approach has a sense only in the case when the number of points N that forms the chosen segment of length T exceeds the value $(2K + 1)L$. In other words, we imply that the condition of reduction of initial data to a small number of calculated parameters is fulfilled when

$$N > L (2K + 1). \qquad (20)$$

The value of L is calculated by the LLSM from the functional equation (5) while the value K in decompositions (8)–(11) is found from the expression

$$\mathrm{Re\,lErr}(K) = \left(\frac{\mathrm{stdev}\left(y(t_j) - F(t_j; K) \right)}{\mathrm{mean}\left| y(t_j) \right|} \right) \times 100\% \leqslant 10\%. \qquad (21)$$

Here $y(t_j)$ is an initial random sequence having the length T and $F(t; K)$ is the fitting function from (6). It contains initially the number of FPs that satisfy conditions (19) and (20), where the parameter L is supposed to be known from expression (5). Other k parameters (Ac_k, As_k) $(k = 1, 2, \ldots, K)$ are linear and evaluated again with the help

of the LLSM when the upper infinite limit of the periodic function defined by expressions (8)–(12) is replaced by the finite value K. This value is calculated from minimization of the fitting function to initial expression and the value of the relative error from (21) should not exceed 10%. In order to decrease the number of the reduced parameters that enter (20) one can apply the detection of the QP processes for the sequences Jy_j that are more correlated (in comparison with initial signals) and obtained from the initial sequences with the help of integration

$$Dy_j = y_j - \mathrm{mean}(y),$$

$$Jy_j = Jy_{j-1} + 0.5 \left(x_j - x_{j-1} \right) \left(Dy_j + Dy_{j-1} \right), \quad Jy_0 = 0,$$

$$j = 1, 2, \ldots, N. \qquad (22)$$

As an integration operation of a discrete sequence, one can use the simplest trapezoid method. This simple procedure was applied successfully for the smoothing of the initial signals [14] and the detection of the hidden traps [15], and so it can be applied in the same manner for the treatment of other data. Finally, we obtain

$$y_j \rightarrow \mathrm{QP}(T, \lambda_r, Ac_k, As_k), \qquad (23)$$

the description of the initial signal in terms of AFR of the identified QP process that contains the period of the process T, the set of exponential factors λ_r $(r = 1, 2, \ldots, L)$ and linear combination of amplitudes (Ac_k, As_k) $(k = 1, 2, \ldots, K)$. The set of frequencies ω_k is obtained easily from the known period T. We suppose that this recognition procedure can contain additional and valuable information proving the existence of the QP processes in different complex systems.

4. Detection of the QP processes on real data

In this section we are going to show *how* to detect the QP processes from real data. The basic requirements are the following:

(a) The same measurements should be repeated during period T many times. The interval between measurements (time-break) should be approximately the same.

(b) It is desirable that the interval T is controlled during measurements and accepts the same value.

But if these requirements are out of the accurate control then one can try to find other sets of the strongly correlated variables for detection of the QP processes. Two randomly taken sets of data (obtained from different sources) should demonstrate to the skeptically tuned reader that QP processes exist in many periodic processes studied and it is necessary only to keep some conditions for their detection.

4.1. The recognition of QP processes from Raman spectra recorded for pure water

As the first example we consider the Raman spectra of pure (double degasified) water. The Raman spectroscopy (http://en.wikipedia.org/wiki/Raman_spectroscopy) is a well-known
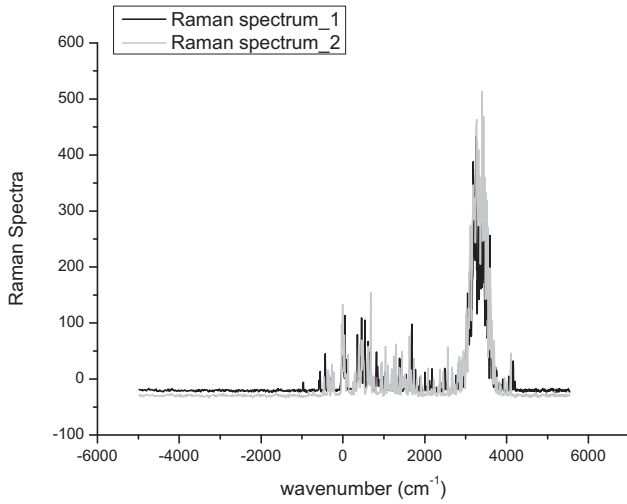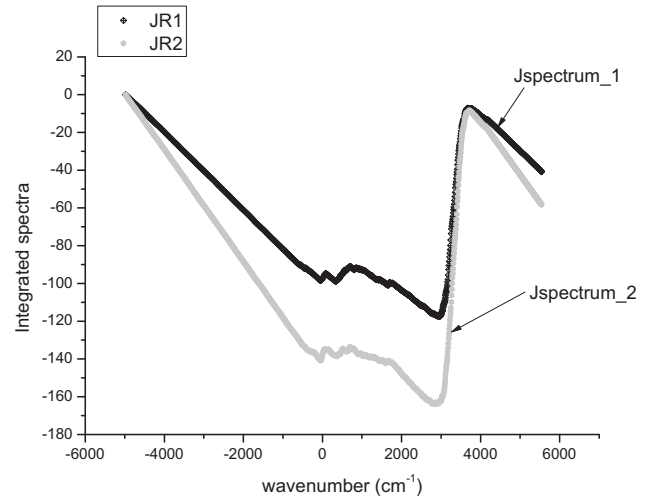
**Figure 1.** Two similar Raman spectra belonging to two different passages for double-gas-free water are shown.
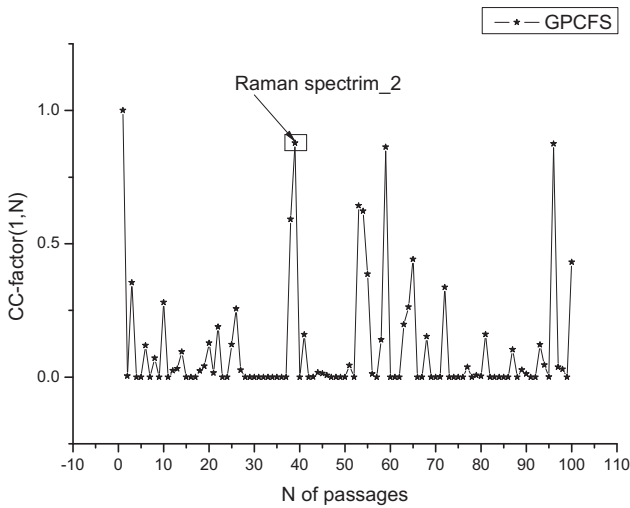
Figure 2. The comparison of the different Raman spectra with respect to the first spectrum (Raman spectrum_1) expressed in terms of the CC factor (expression (18)). One can notice that this sampling containing 103 passages is rather heterogeneous. As one can see from this figure only four spectra are located in the interval of correlations [0.8–1.0] and seven have correlations between [0.5–1.0].


Figure 3. Comparison of the integrated (cumulative) spectra obtained from expression (22). One can note that in comparison with figure 1, the high-frequency fluctuations are essentially decreased and all possible similarities (differences) are becoming noticeable.
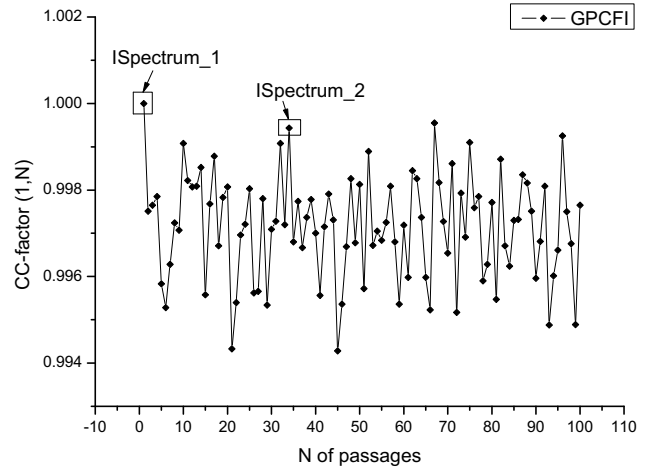
Figure 4. The comparison of the different integrated spectra with respect to the first spectrum (JSpectrum_1) expressed in terms of the CC factor (expression (18)). One can note that integrated spectra containing again 103 passages become strongly correlated and homogeneous to each other. As one can see from this figure all correlations (1,N) are located in the interval of correlations [0.994–1.0]. The same statement is valid for all pair correlations compared.

spectroscopic method for the investigation of vibration modes of different substances and so we can omit its general description. For us it is important only for the following specific information [16, 17] associated with these available data: the type of laser that defines the basic pumping frequency and the period of time when one passage (containing 1024 measured points) is obtained. For the data treated below one passage is performed during 10^{-8} s. One experiment containing approximately 100–110 passages with time-breaks is finished during 100 s. The typical Raman spectrum corresponding to one passage for double-gas-free water was shown in figure 1. The wavenumbers (in cm^{-1}) are counted off from the laser pumping frequency equaling 532 nm and all wavenumbers are classified as Stokes(+)–anti-Stokes(−) regions, correspondingly. For comparison in figure 1 we give only one similar spectrum

identified with the help of CC-factor (see expressions (15) and (18) above). Comparing the first spectral record with the other 100 passages is given in figure 2. As one can note on this plot these spectra (in spite of apparent similarity) are strongly deviated from each other. In order to make them more similar one can integrate them in accordance with expression (22). The numerical integration suppresses high-frequency fluctuations and makes the spectra more correlated to each other. The pair of the integrated spectra is shown in figure 3. The variation of the CC factor of the first integrated spectra with others is shown in figure 4. In comparison with figure 2 the CC factor is located in the interval [0.994–1.0] that proves their close statistical proximity to each other. Having these smoothed spectra one can try to detect the QP-processes
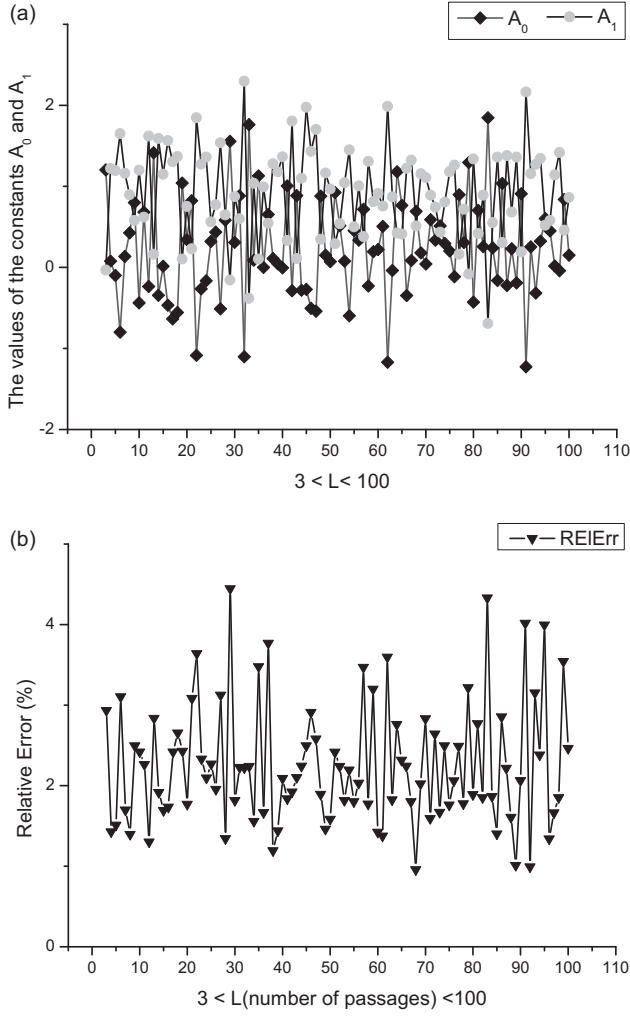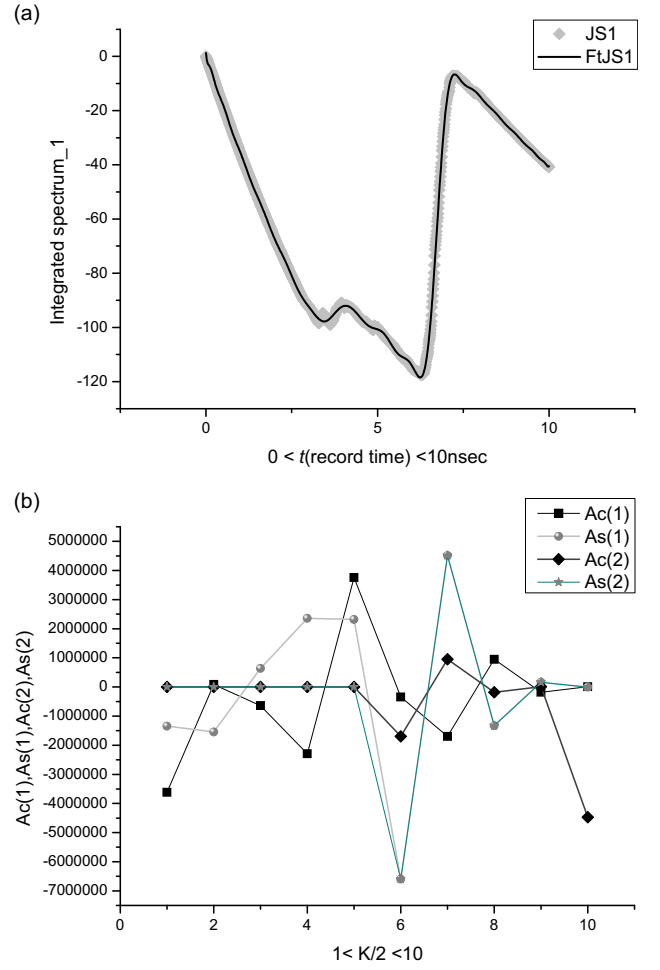
Figure 5. (a) The variations of the constants found by the LLSM from equation (24). One can notice that all of them located in the interval [−2, 3] and it proves again that this sampling expressed in terms of the integrated spectra is homogeneous. Because of saving place for other figures the calculated roots are not shown. We want to mark here that the roots can be negative and complex-conjugated. (b) Here we show the variations of the relative error defined by (21). One can note that all these values are located in the interval [0.75, 4.5](%) and this plot proves that hypothesis (24) is realized with high accuracy.


Figure 6. (a) Integrated spectrum_1 expressed in terms of the generalized Prony's decomposition (24). In order to receive the accurate fit in the limits of the relative error less than 1.5% (= 1.08% if one can be more precise) only 20 modes in decomposition (25) for periodic functions are necessary. We note that in comparison with the previous figure 3 we realized the fit for the time of recording (and *not* for the corresponding wave numbers as it is accepted). (b) The values of the amplitudes (Ac_k, As_k) ($k = 1, 2, \ldots, 10$) calculated with the help of the LLSM that enter solution (25). These amplitudes form the desired AFR that represents itself the reduced description of the integrated spectrum. Similar values one can calculate for other integrated spectra that are connected with each other by means of relationship (24).

expressed in terms of expression (5) for the short memory case ($L = 2$). We test the simplest hypothesis

$$F(t + LT) \cong A_1(L)F(t + T) + A_0(L)F(t). \qquad (24)$$

Here the function $F(t)$ determines the initial integrated spectrum and the constants $A_{1,0}(L)$ are calculated approximately in the frame of LLSM. The index $L = 3, 4, \ldots 100$ counts the number of passages. The influence of the passages located between one and L is supposed to be negligible. The recording time t for registration of one passage is located in the interval [0,10 ns]. We omit experimental time-breaks and suppose that T coincides with the duration of one passage recording 10 ns. But we want to note here that the result qualitatively is *not* changed when we take into account the time-breaks and consider $T = 1$ s, having in mind that the recording time together with

time-breaks occupies the period of 1 s. The dependence of the coefficients $A_{0,1}(L)$ with respect to a number of passage L is shown in figure 5(a). Figure 5(b) shows the variation of the relative error that provides an approximate fit of the left part to the pair of the previous measurements in expression (24). The solution of the functional equation (24), in accordance with general expression (6), is written in the form

$$\sum_{r=0}^{1} A_r(L) \neq 1 : F(t) = \sum_{r=1}^{2} \exp\left(\ln\left(\lambda_r(L)\right)\frac{t}{T}\right)\mathrm{Pr}_r(t),$$
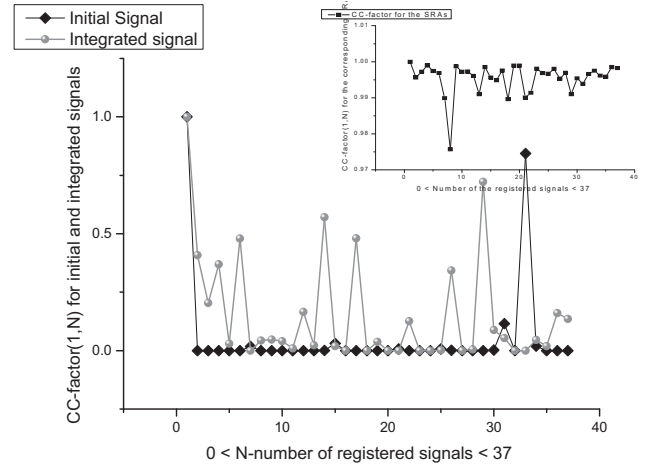$$(25)$$

where

$$\lambda_{1,2}(L) = -\frac{A_1(L)}{2} \pm \sqrt{\left(\frac{A_1(L)}{2}\right)^2 + A_0(L)}, \qquad (26)$$

(a)

(b)
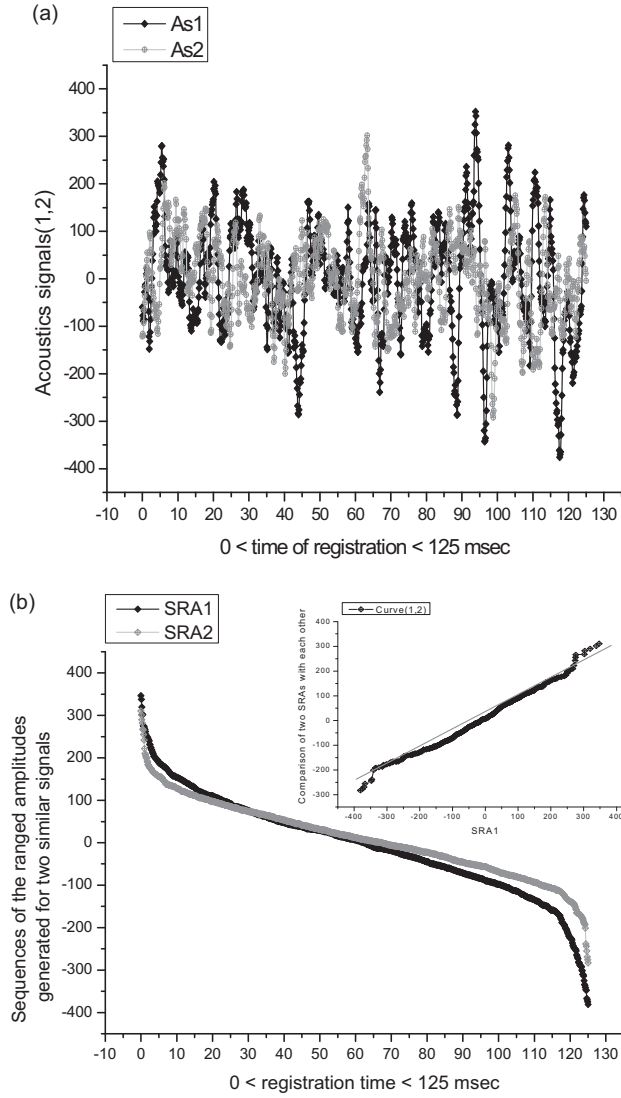
Figure 7. (a) Two similar reflected acoustic signals recorded on the same depth. (b) Two SRAs generated from two acoustic signals depicted on the previous figure. The plot given inside a small frame shows that these two SRAs are strongly correlated.

Figure 8. Here we demonstrate the variations of CC-factor (expression (18)) between the first signals and other signals registered in late time. Even for integrated sequences these signals remain weakly correlated. In order to increase their correlations one can generate the corresponding SRAs and compare their correlations. As one can see from the figure in a small frame these sequences are strongly correlated. All correlations are located in the interval [0.97–1.0].

the roots of quadratic equations associated with equation (24). The periodic functions $Pr_{1,2}(t)$ are defined by expression (8)–(11) and depend essentially on the character of roots (positive, negative, complex-conjugated) of expression (26). In order to provide the acceptable fit (with the values of the relative error less than 1%) only 20 modes ($K = 20$) in the corresponding decompositions (8)–(11) are necessary. This fit for $L = 2$ is shown in figure 6(a). We want to stress here that in comparison with figure 3 this dependence is given with respect of the recording time t from interval [0,10 ns] and *not* to the respect of the wavenumbers as in figure 3. The values of the decomposition coefficients Ac_k, As_k with respect to the number of a current mode k ($k = 0(1), 2, \ldots, 10$) are shown in figure 6(b) for $L = 3$. The coefficients Ac_k, As_k for other values of L look similar and so they are not shown. From analysis of this simple example some important conclusions are following:

1. The repetition of experimental measurements (at the *same* external conditions) *cannot* reproduce similar

results as initially one might expect. So, for the reliable detection of the QP-processes the high-correlated random sequences calculated from initial measurements are *necessary*.

2. The fit of these functions is realized in respect of the observation/registration time, but not in respect of the actual external parameter (as the wavenumber in this case).

3. The calculation of the decomposition coefficients Ac_k, As_k for the highly correlated sequences allows one to reduce the initial information (i.e. contained in 1024 measured points) to 42–44 parameters (in this case) when the actual model pretending on description of the integrated spectra is *absent*.

4.2. The recognition of QP processes from random acoustic signals (geophysics data)

Many companies working in the region of geophysical industry/prospecting register the reflected acoustic signals emitted from the acoustic generator located inside a test hole on the certain depth. For further analysis only the following details are important. The time of recording of the reflected acoustic signal containing 1024 measured points occupies the interval [120–150 ms]. The interval between the registered signals is in the limits [1 s–1 min] and depends on the character of the rocks surrounded. After the registration of 20–40 signals the receiver changes its depth of location inside the test hole. The author (RRN) received these reflected acoustic signals for analysis from the company (http://tgtoil.com). Based only on these data (without other supplied geophysical information) we try to find an answer to the following question: Do these data belong to some QP-process or not? How to present data in a more correlated form in order to detect the QP-process with high reliability? In this example the period of time between
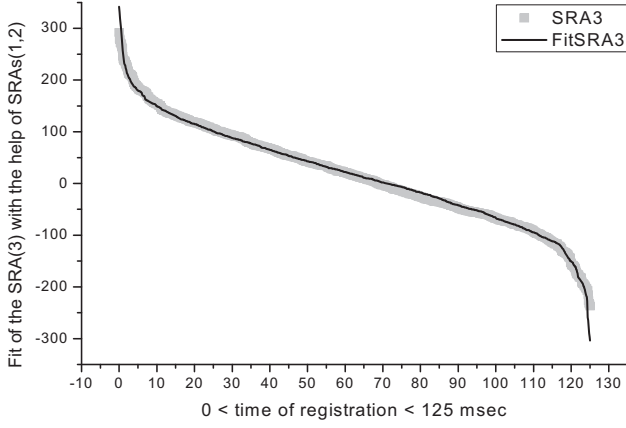
Figure 9. The verification of the hypothesis (27). The fit with the help of two previous SRAs looks acceptable. The relative fitting error does not exceed 9%. The same fit can be realized for other sequences presented in the form of the SRAs. The values of the fitting constants ($A_0(L)$, $A_1(L)$, $C_0(L)$) and the value of the fitting error are shown on figures 10.

recordings is uncertain and acoustic signals registered over the same time interval are deviated from each other. Figures 7(a) and (b) depict two similar acoustic signals taken without any filtering (figure 7(a)) and after transformation of the initial signal to the sequence of the ranged amplitudes (SRAs), when all registered amplitudes are located in the descending order ($y_1 > y_2 > \cdots > y_N$) (central plots in figure 7(b)). In the small frame of figure 7(b) these SRAs being plotted with respect to each other exhibit their strong correlations. Figure 8 gives the variations of the CC factor of the first acoustic signal with others that are registered in late moments of time. The integration of initial signals with respect to their mean value does not facilitate the situation. Figure 8 shows also the variations of CC factors for the integrated signals. Only a few of them are strongly correlated. In order to increase their correlations and to make them more or less homogeneous we form, from the initial acoustic signal, the SRAs. It is interesting to note that the corresponding SRAs, generated from initial acoustic signals, are strongly correlated. This statement is confirmed by figure 8, where the CC factor comparing different SRAs with the first one is shown in the small frame. The appearance of strong correlations between different SRAs can be explained by the following way. Any initial acoustic signal belongs to the general sampling having the volume $N!$, where a set of amplitudes is not changed essentially and forms random sequences only by means of various *permutations*. Any SRA represents itself the *limiting permutation* when the amplitudes are located in descending order. Another limiting case is formed when the same amplitudes are located in ascending order. If the given set of amplitudes generating different samplings is not changed essentially then their SRAs should be close to each other. Being plotted to respect each other two SRAs should give a curve close to a straight line (14). See figure 7(b) in a small frame. Therefore, this independent test confirms the existence of the sampling with volume $N!$ generated by random permutations and all these random sequences belonging to this sampling are strongly correlated. Consequently, one can try to confirm the existence of the
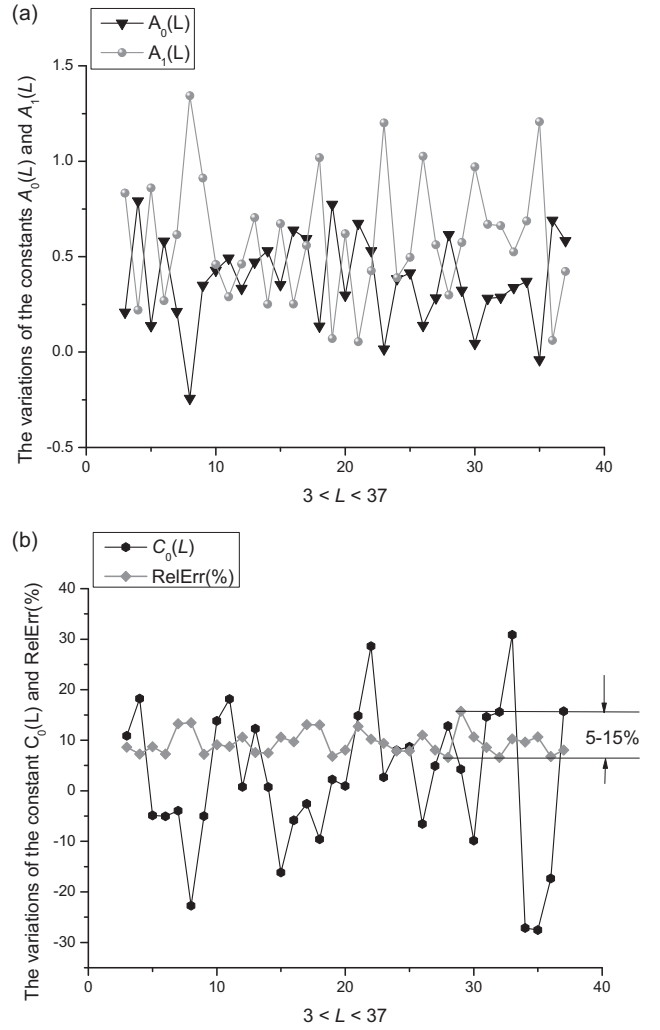

Figure 10. (a) The variations of the constants $A_0(L)$ and $A_1(L)$ found by the LLSM from equation (27). One can note that all of them located in the interval [−0.5, 1.5]. It proves again that this sampling expressed in terms of the SRAs is homogeneous. Because of space restraints, the calculated roots are not shown. We want to mark here that other calculated roots can be only positive or negative. The complex-conjugated roots are absent. (b) The variations of the constant $C_0(L)$ and the value of the relative error found by the LLSM from equation (27). The value of the relative error lies in the interval (5–15%) in spite of the fact that initial acoustic signals are weakly correlated.

QP-process that can be observed between different SRAs. Again, in order not to increase the number of FPs we consider the case with the short memory. We suppose that the following hypothesis is valid:

$$F(t + LT) \cong A_1(L)F(t + T) + A_0(L)F(t) + C_0(L). \quad (27)$$

The free constant $C_0(L)$ is added to hypothesis (27) in order to decrease the value of the relative error when the approximate equality between the right and left sides of equation (27) is verified with the help of the LLSM. The fit of the SRA for $L = 3$ by means of sequences belonging to $L = 1, 2$ is shown in figure 9. The values of the fitting constants ($A_0(L)$, $A_1(L)$, $C_0(L)$) ($L = 3, 4, \ldots, 37$) and the value of the fitting error are depicted in figures 10(a) and (b). Finally, we show the fit of the SRA ($L = 1$) to solution of the
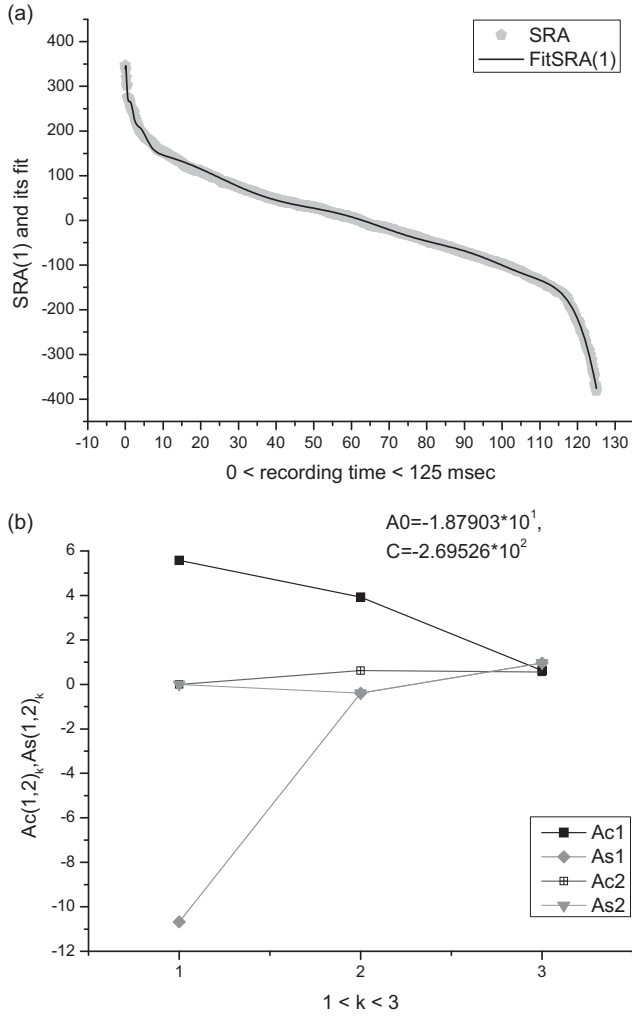
Figure 11. (a) The fit of the function $F(t)$ (coinciding with SRA $L = 1$) by its expression that follows from solution (28). In order to have acceptable fit with the value RelErr = 1.37% only 14 FP are necessary. The value of K is equaled to 6. (b) The values of the FP that enter to decomposition (28). Besides the six parameters $Ac(1, 2)_k$, $As(1, 2)_k$ ($k = 1, 2, 3$), we give the values of the constant C (at $L = 1$) and the constant A_0 that presents for the roots having positive values. In order to place all these values in one plot the actual value of the constant $Ac(2)_1$ should be increased in 10^8 times.

functional equation (27) in figure 11(a)

$$\sum_{r=0}^{1} A_r(L) \neq 1 :$$

$$F(t) = \sum_{r=1}^{2} \times \exp\left(\ln\left(\lambda_r(L)\right) \frac{t}{T}\right) \Pr_r(t) + C(L), \quad (28)$$

$$C(L) = \frac{C_0(L)}{1 - \sum_{r=0}^{1} A_r(L)}.$$

The AFR containing 14 parameters, and reproducing the fit of the function $F(t)$ (and coinciding with SRA ($L = 1$)), is shown in figure 11(b).

5. Results and discussion

In this section we summarize the basic conclusions that follow from this approach.

The detection of the QP processes opens new possibilities in the description and classification of data that are registered from different complex systems. Before, all repeated data were supposed to be similar to each other and so all experimentalists in the world try to reduce a massive amount of initial data to their mean values and standard deviations. The QP processes allow one to find a *fine* structure of the massive data analyzed and even fit them in respect of observation time. One can note, for example, that the first set of data referring to the Raman spectra are more homogeneous in comparison with the geophysics data. In the first case for detection of the QP process the simple integration with respect to their mean values was sufficient while in the second case the transformation of the initial acoustic signals to the SRAs was necessary.

Usually all researchers working with different data are accustomed to thinking that the model containing a minimal number of FPs is ideal in any case. But what we should do when this 'ideal' model is absent? The detection of the QP or SS processes [5] allows us to introduce the concept of an *intermediate* model, when any general principle (such as the concept of self-similarity [5] or quasi-periodicity) allows one to reduce the redundant information that is imbedded in the initial data to other forms (to the AFR in our case) that contain less information but allows one to reproduce the initial data containing N measured points. This reduction of information is the *basic* aim of any researcher who is trying to find additional laws that help him to reduce and classify redundant information.

The research presented in this paper allows us to generalize this approach and apply it to *spatial* objects with quasi-translation (QT) properties. As is well known, a function remains translationally invariant if

$$f\left(\mathbf{r} + \sum_{i=1}^{3} n_i \mathbf{a}_i\right) = f(\mathbf{r}). \quad (29)$$

Here $f(\mathbf{r})$ is a function that can be given in scalar or vector form, vector triple ($\mathbf{a}_i$, $i = 1, 2, 3$) determines the basic translation vectors. In complete analogy with the functional equation (5) one can write the following expression:

$$F(\mathbf{r} + L_1 \mathbf{a}_1 + L_2 \mathbf{a}_2 + L_3 \mathbf{a}_3)$$
$$= \sum_{s_1, s_2, s_3 = 0}^{L_1-1, L_2-1, L_3-1} A_{s_1, s_2, s_3} F\left(\mathbf{r} + \sum_{i=1}^{3} s_i \mathbf{a}_i\right) + b. \quad (30)$$

Here s_i ($0, 1, 2, \ldots, L_i$) ($i = 1, 2, 3$) coincides with the set of integer numbers that determine the QT-property in space. Perhaps in the near future equation (30) and its solutions can serve as a new classification scheme for the identified QT structures. This equation can be generalized easily for classification of self-affine structures [1, 2]. Let us consider the following scaling equation:

$$F(\mathbf{r}\xi_1^{L_1}\xi_2^{L_2}\xi_3^{L_3}) = \sum_{s_1, s_2, s_3 = 0}^{L_1-1, L_2-1, L_3-1} A_{s_1, s_2, s_3} F(\mathbf{r}\xi_1^{s_1}\xi_2^{s_2}\xi_3^{s_3}) + b. \quad (31)$$

Making a substitution $\ln(\mathbf{r}) \to \mathbf{r}$, $\ln(\xi_i) \to a_i$ (for each component of the corresponding vector $\mathbf{a}$) and $L_i = s_i$ we reduce equation (31) to the previous equation (30).

A partial solution of the functional equation (30) is written in the form

$$F(\mathbf{r}) = \exp(\boldsymbol{\kappa} \cdot \mathbf{r}) f(\mathbf{r}), \quad \boldsymbol{\kappa} = \boldsymbol{\kappa}(\kappa_1, \kappa_2, \kappa_3), \qquad (32)$$

where the periodic function $f(\mathbf{r})$ is determined by equation (29) and the vector $\boldsymbol{\kappa}$ remains unknown. The substitution of a partial solution to equation (30) leads finally to the following algebraic equation for finding the desired roots λ_I and unknown components of the vector $\boldsymbol{\kappa}$:

$$\lambda_1^{L_1} \lambda_2^{L_2} \lambda_3^{L_3} = \sum_{s_1,s_2,s_3=0}^{L_1-1,L_2-1,L_3-1} A_{s_1,s_2,s_3} \lambda_1^{s_1} \lambda_2^{s_2} \lambda_3^{s_3},$$

$$\lambda_i = \exp(\kappa_i a_i), \quad i = 1, 2, 3. \qquad (33)$$

Particularly, for different polymers and large molecules forming a complex structure on different space/time scales the ideal translation or scaling scheme for their classification is *not* valid. The solution of equation (33) is not a trivial procedure and can be obtained for partial cases by analogy with expression (6) when the polynomials (7) for each direction can be at least *separated* or *factorized*. The detailed consideration of these new possibilities merits special research.

Acknowledgments

References

[1] Mandelbrot B 1982 *Fractal Geometry of Nature* (San Francisco, CA: Freeman)
[2] Aharony A and Feder J (ed) 1989 Fractals in physics *Essays in Honour of B Mandelbrot* (North-Holland: Amsterdam)
[3] Samko S G, Kilbas A A and Marichev O I 1987 *Integrals and Derivatives of the Fractional Order and their Some Applications* (Minsk: Nauka and Tekhnika' Publishing House) p 688 (in Russian)
[4] Machado J A T, Kiryakova V and Mainardi F 2011 Recent history of fractional calculus *Commun. Nonlinear Sci. Numer. Simul.* **16** 1140
[5] Nigmatullin R R, Machado J A and Tand Menezes R 2013 Self-similarity principle: the reduced description of randomness *Central Eur. J. Phys.* **11**(6) 724–39
[6] Nigmatullin R R 2012 *J. Appl. Nonlinear Dyn.* **1** 207
[7] Kwapien J and Drozdz S 2012 *Phys. Rep.* **515** 115
[8] Nigmatullin R R 2008 *Phys. Wave Phenom.* **16** 119
[9] Osborne M R and Smyth G K 1991 A modified Prony algorithm for fitting functions defined by difference equations *SIAM J. Sci. Stat. Comput.* **12** 362
[10] Kahn M, Mackisack M S, Osborne M R and Smyth G K 1992 On the consistency of Prony's method and related algorithms *J. Comput. Graph. Stat.* **1** 329
[11] Osborne M R and Smyth G K 1995 A modified Prony algorithm for fitting sums of exponential functions *SIAM J. Sci. Stat. Comput.* **16** 119
[12] Nigmatullin R R 2006 The statistics of the fractional moments: is there any chance to read 'quantitatively' any randomness? *J. Signal Process.* **86** 2529
[13] Nigmatullin R R 2010 Universal distribution function for the strongly-correlated fluctuations: general way for description of random sequences *Commun. Nonlinear Sci. Numer. Simul.* **15** 637
[14] Nigmatullin R R, Ionescu C and Baleanu D 2013 NIMRAD: novel technique for respiratory data treatment *J. Signal Video Process.* doi:10.1007/s11760-012-0386-1
[15] Ciurea M L, Lozanu S, Stavaracher I, Lepadatu A M, Iancu V, Mitroi M R, Nigmatullin R and Baleanu C M 2011 Stressed induced traps in multilayed structures *J. Appl. Phys.* **109** 013717
[16] Pershin S M, Bunkin A F, Lukyanchenko V A and Nigmatullin R R 2007 Detection of the OH band fine structure in a liquid water by means of new treatment procedure based on the statistics of the fractional moments *Lazer Phys. Lett.* **4** 809
[17] Pershin S M, Bunkin A F and Lukyanchenko V A 2010 Evolution of the spectral components of ice in the OH band of water at temperatures 13–99 C *Quantum Electron.* **40** 1146