Identifying ergodicity breaking for fractional anomalous diffusion: Criteria for minimal trajectory length

Hanna Loch-Olszewska, Grzegorz Sikora,* Joanna Janczura, and Aleksander Weron
Faculty of Pure and Applied Mathematics, Hugo Steinhaus Center, Wrocław University of Science and Technology, 50-370 Wrocław, Poland
(Received 2 August 2016; published 22 November 2016)

In this paper, we study ergodic properties of α-stable autoregressive fractionally integrated moving average (ARFIMA) processes which form a large class of anomalous diffusions. A crucial practical question is how long trajectories one needs to observe in an experiment in order to claim that the analyzed data are ergodic or not. This will be solved by checking the asymptotic convergence to 0 of the empirical estimator $F(n)$ for the dynamical functional $D(n)$ defined as a Fourier transform of the n-lag increments of the ARFIMA process. Moreover, we introduce more flexible concept of the ϵ-ergodicity.

DOI: 10.1103/PhysRevE.94.052136

I. INTRODUCTION

Recently, the careful data analysis helped to identify various physical origins of the anomalous diffusion with the mean-squared displacement being nonlinear in time, [1–4]. Especially, the field of life sciences has lately seen an immense increase in single-particle tracking (SPT) techniques and experimental results [5–8].

Modern fluorescence microscopy that probes the stochastic motion of individual labeled tracer particles shows some significant, uncovered deviations from the laws of Brownian motion in a variety of physical and biological systems [1,9,10]. In SPT experiments, the tracer's position is tracked over time and the corresponding dependence $x(t)$ is then treated as a random process [11–13]. The advantage of SPT is that it provides the full information about the tracer's motion. Unfortunately, in many cases the data acquisition procedure is complicated, so that only few trajectories of considerable length are available [2,9].

From one side, fractional diffusion is traditionally related to the concept of fractional dynamic equations, especially the fractional Fokker-Planck equation (FFPE) and the popular continuous time random walk (CTRW) model [10]. This is a very active field of study in physics, mechanics, chemistry, and biology. From the other side, there has been a great interest in long-range dependent and self-similar processes, in particular, fractional Brownian motion (FBM) [14] and autoregressive fractionally integrated moving average time series, called ARFIMA processes [15,16]. These processes, originally introduced in econometrics, have been recently effectively used to describe fractional anomalous diffusion in physical and biological systems [17,18]. From the physical point of view, it is known that ARFIMA(p,d,q) is a discrete time representation of the fractional Langevin equation (FLE) that takes into account the memory parameter d. It also includes other popular models of fractional dynamics such as FBM, which is the limiting case of aggregated ARFIMA process. Modeling physical or biological phenomena with ARFIMA process has promising potential [14,17,19–25].

In many practical applications of the ARFIMA(p,d,q) model it is sufficient to describe data with p and q equal to 0 or 1 [26,27]. ARFIMA(1,d,1) can be considered as first order approximation of the arbitrary model taking into account only the first lag.

The problem of ergodicity of the process is an essential issue for the real-life processes. Verification of the Boltzmann ergodic hypothesis for a given system, i.e., asymptotic equality of time and ensemble averages [20], is one of the most fundamental problems in statistical mechanics [1,2]. Recently, ergodicity breaking of systems exhibiting anomalous behavior has been observed in various fields of physics, biology, and related sciences [6,22–25,28–33].

The use of highly photostable fluorescent probes, in particular quantum dots (QDs), has opened the possibility of recording long single-particle trajectories allowing better detection of anomalous diffusive behavior and thorough investigation of its causes [24,25,28,34]. The occurrence of ergodicity breaking implies that the molecules under investigation, although chemically identical, show somehow different dynamic properties [35]. The first example of ergodicity breaking on the cell membrane was provided in [25], where the authors studied the lateral diffusion of the voltage-gated potassium channel Kv2.1 on the membrane of human embryonic kidney cells. A second example by studying the diffusion of the pathogen recognition receptor DC-SIGN on the plasma membrane of stably transfected CHO cells was described in [28]. Ergodicity breaking has been also observed experimentally for the intracellular diffusion [6,30].

This paper is structured as follows. In Sec. II, our considerations start with a short introduction to the dynamical functional (DF) approach to ergodicity testing methodology [36–39]. Basic properties of α-stable ARFIMA processes and time series description of ARFIMA model are given in Sec. III. This representation allows us in Sec. IV to present an analytical form of the important statistic for our study: dynamical functional $D(n)$ for an ARFIMA(0,d,0) model with symmetric α-stable noise. Next, we study asymptotic behavior of the dynamical functional $D(n)$ and the corresponding test functional $F(n)$, which asymptotic behavior determines ergodicity of the process. Since exact formula for $F(n)$ of an ergodic process is only asymptotically equal to 0, we will find a minimal $n_0(\varepsilon)$ for which with a given accuracy $\varepsilon > 0$ the inequality $|F(n)| < \varepsilon$ holds. Not fulfilling this condition for any natural n will indicate the so called ε-ergodicity breaking, which from the experimental point of view provides a practical

*grzegorz.sikora@pwr.edu.pl

criterion for minimal sample length of the studied trajectory. Obtained theoretical and numerical results are illustrated also by simulations for some special cases. Section V concludes this work.

II. ERGODICITY TESTING: DYNAMICAL FUNCTIONAL APPROACH

It is well known that for every stationary and ergodic process $Y_k, k \in \mathbb{N}$, the Boltzmann ergodic hypothesis holds, that is, the temporal and ensemble averages coincide:

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} g(Y_k) = \langle g(Y_0) \rangle,$$

where g is an arbitrary integrable function (deterministic) and $\langle \cdot \rangle$ denotes the ensemble average [40]. A straightforward consequence of ergodicity is that instead of an ensemble average, which requires repeating an experiment many times, one can as well analyze a time average of only one long trajectory. This property is the main reason for considering ergodicity in the context of real-life stochastic processes [36].

In general, ergodicity of the process can be verified using, e.g., the dynamical functional, described below (and that approach will be used further in the paper), or by a comparison of the time averaged and ensemble averaged mean-square displacements which is the classical approach [38].

The dynamical functional (DF) $D(k)$ corresponding to the process Y_k was defined by Podgorski and Weron (see [37]) as

$$D(k) = \langle \exp\{i[Y_k - Y_0]\} \rangle, \tag{1}$$

that is a Fourier transform of $Y_k - Y_0$ evaluated for the Fourier-space variable $\omega = 1$. It describes the dynamical behavior of the underlying process via observing its increments and can be used as a measure of the process interdependence, especially if the correlation function does not exist. Recently, it was used in empirical data analysis (see, for example, in [36,38,39]).

The DF test for ergodicity was proposed in [36]. This test is based on the analysis of the asymptotic behavior of the test functional $F(n)$:

$$F(n) = \frac{1}{n} \sum_{k=0}^{n-1} D(k) - |\langle \exp\{iY_0\} \rangle|^2, \tag{2}$$

and the verification of its convergence to 0, i.e.,

$$\lim_{n\to\infty} F(n) = 0 \tag{3}$$

for an ergodic process.

Since the convergence in (3) is asymptotic, the crucial question is how long should trajectories be to claim that the analyzed data are ergodic or not. This will be analyzed in details in Sec. IV.

What is very interesting and especially useful in the empirical data analysis is that the DF test can be also applied to the only one-trajectory empirical data, provided that data are observed long enough (see [36] and its modification [39]), which is also shown as a by-product in this article. The ensemble average is then exchanged with the time average. Again, the question of the sufficient trajectory length is crucial here.

III. ARFIMA MODEL

In this article, we will work on a class of ARFIMA models [15,16] which encompass discrete time versions of many celebrated diffusion processes. It can be used for modeling fractional dynamics and long memory by fractionally integrated (FI) part, but on the other hand also short-time dependencies by moving average (MA) part and autoregressive (AR) part. Moreover, when properly scaled and aggregated, ARFIMA models converge to continuous time anomalous diffusion processes with such prominent examples as the fractional Brownian motion [ARFIMA(0,d,0) with Gaussian (or α-stable) noise] and the Ornstein-Uhlenbeck process [ARFIMA(1,0,0)]. Since the experimental data are usually recorded in discrete time, the ARFIMA process seems to be a natural tool for modeling anomalous dynamics. Moreover, it was recently shown that the MA part of the ARFIMA model can be successfully used for modeling data with measurement errors [41]. We focus our interests on processes ARFIMA(0,d,0) and ARFIMA(1,d,1), as they are commonly fitted to empirical data [41].

The ARFIMA(p,d,q) model, originally introduced by Grander and Joyeux in [15] and Hosking in [16], is defined as a discrete time process $\{Y_t\}$, $t = 0,1,2,\ldots$, that satisfies the equation

$$\Phi(B)(1-B)^d Y_t = \Theta(B)\varepsilon_t, \tag{4}$$

where $\Phi(z) = 1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p$ and $\Theta(z) = 1 + \theta_1 z + \theta_2 z^2 + \cdots + \theta_q z^q$ are polynomials of, respectively, autoregressive and moving average parts, B is the backward shift operator ($BY_t = Y_{t-1}$), d is the fractional differencing order, and ε_t is the noise process. In the classical definition ε_t is a Gaussian noise process, however, in this work we consider a more general case with ε_t's being i.i.d. symmetric α-stable random variables (with $0 < \alpha < 2$) (see [42–44]) defined by the characteristic function

$$\varphi_{\varepsilon_t}(s) = \exp(-|\sigma s|^\alpha), \tag{5}$$

where σ is a scaling parameter. The ARFIMA(p,d,q) model is fractionally integrated of order d with

$$(1-B)^d = \sum_{k=0}^{\infty} \binom{d}{k} (-B)^k.$$

When d is a positive noninteger, such operation introduces long memory, or long-range dependence, of the process, used, e.g., for modeling superdiffusion [42]. Hence, the ARFIMA(p,d,q) process is a composition of the fractional part FI(d) and ARMA(p,q) model.

IV. MINIMAL TRAJECTORY LENGTH

A. Analytical results

For α-stable ARFIMA process, it was shown recently in [45] that

$$D(n) = \exp\left(\sigma^\alpha \sum_{j=0}^{\infty} |c_{j+n}|^\alpha - |c_j|^\alpha - |c_j - c_{j+n}|^\alpha \right), \tag{6}$$

where the coefficients c_j come from the moving average representation of the ARFIMA model:

$$Y_t = \sum_{j=0}^{\infty} c_j \varepsilon_{t-j}. \tag{7}$$

The explicit form of the c_j coefficients of moving averages can be calculated explicitly via simple series transformations (see, e.g., [42]), for example, for FI(d), i.e., ARFIMA(0,d,0), the coefficients are

$$c_0 = 1, \qquad c_j = \frac{\Gamma(j+d)}{\Gamma(d)\Gamma(j+1)} \quad \text{for } j \neq 0 \tag{8}$$

and for ARFIMA(1,d,1)

$$\begin{aligned} c_0 &= \left(\frac{\phi-1}{\phi}\right)^{-d} - \frac{\Gamma(1+d)}{\phi\Gamma(2)\Gamma(d)}\,{}_2F_1(1+d,1,2,1/\phi), \\ c_j &= \left(\frac{\phi-1}{\phi}\right)^{-d}\phi^{j-1}(\phi+\theta) - \frac{\Gamma(d+j)}{\phi\Gamma(j+1)\Gamma(d)} \\ &\quad \times \left[\frac{j+d}{j+1}\,{}_2F_1(1+d+j,1,2+j,1/\phi)\right. \\ &\quad \left. + \theta\,{}_2F_1(d+j,1,1+j,1/\phi)\right] \text{ for } j \neq 0, \end{aligned}$$

where ${}_2F_1$ is the hypergeometric function [46]. The complexity of the formula for those coefficients increases with growing rank of both AR and MA parts, that is, p and q of ARFIMA(p,d,q).

For the α-stable ARFIMA the test functional $F(n)$, introduced in the previous section, takes the following form:

$$F(n) = \frac{1}{n}\sum_{k=0}^{n-1}\left(\exp\left\{\sigma^\alpha \sum_{j=0}^{\infty} |c_{j+n}|^\alpha - |c_j|^\alpha - |c_j - c_{j+n}|^\alpha\right\}\right. \\ \left. - \exp\left\{-2\sigma^\alpha \sum_{j=0}^{\infty}|c_j|^\alpha\right\}\right), \tag{9}$$

as $|\langle\exp\{iY_0\}\rangle|^2 = \exp\left(-2\sigma^\alpha\sum_{j=0}^{\infty}|c_j|^\alpha\right)$.

In Fig. 1, the exemplary trajectories of test functional $F(n)$ for ARFIMA(1,d,1) model are depicted. The blue lines are the values obtained from 1000 simulated sample paths, each of the length 500, while the red ones are the theoretical ones; the green dashed lines are 95% confidence intervals, introduced by Janczura and Weron [38].

In the real world, sometimes it could be challenging to obtain very long sample paths of the process, which could reveal the information about ergodicity. Because exact formula (9) for $F(n)$ for an ergodic process is only asymptotically equal to 0, we intend to find such n_0 that

$$n_0(\varepsilon) = \min\{n \in \mathbb{N} : |F(n)| < \varepsilon\} \tag{10}$$

for a given accuracy $\varepsilon > 0$. The underlying assumption (for this time being) is that the process we are considering is ergodic. Our aim is to verify how long it takes to observe this characteristic and later generalize it as a tool for testing ergodicity breaking. As mentioned before, from now on we focus on two exemplary variants of the ARFIMA process: ARFIMA(0,d,0) and ARFIMA(1,d,1).

TABLE I. The minimal trajectory length of α-stable ARFIMA(0,d,0) process to claim ε-ergodicity: comparison of three proposed methods for $\varepsilon = 0.05$. The first two columns were calculated using the analytical formulas, while the third one is obtained by calculating mean of 100 $n_0(\varepsilon)$ values, each derived for the test functional based on 1000 sample paths of the length 12 000. The most outlying values are those for Gaussian model with high d value; our studies show that the speed of $F(n)$ convergence declines with the growth of d (which is well explained by superdiffusion). The fall is even more rapid for the high values of d.

		Eq. (9)	Eqs. (14) and (19)	Simulation
$\alpha = 2$	$d = 0.43$	121	158	1276 (443.66)
ARFIMA	$d = 0.06$	18	5	18 (3.53)
(0,d,0)	$d = -0.39$	16	16	11 (2.41)
$\alpha = 1$	$d = -0.07$	10	620	48 (13.68)
ARFIMA	$d = -0.44$	9	4	24 (3.15)
(0,d,0)	$d = -0.89$	16	2	19 (2.49)

What we would like to stress is that in general ARFIMA processes are ergodic (see [47]) and the testing functional showing ergodicity breaking for such process is only pointing to incorrect model choice. Hence, one can use the ε-ergodicity breaking similarly to the statistical test, with the H_0 hypothesis concerning the model chosen to the data. The procedure of calculation of $n_0(\varepsilon)$ for the empirical data is the following: first, the proper model should be chosen [26,27]; second, its parameters should be estimated [42,48], and, finally, the value of $n_0(\varepsilon)$ should be calculated.

Of course, the values of $n_0(\varepsilon)$ differ with the underlying model. Minimal $n_0(0.05)$ values for the ARFIMA(0,d,0) process, for which the exact formula (9) fulfils the inequality $|F(n)| < 0.05$, are given in the first column of Table I. They were obtained by MATLAB computations of the exact formula, replacing infinite sum in the formula (9) with the finite one, up to $j = 3000$. That causes the numerical bias, which in the case of such high level of truncation is well controlled (see [49]). For simplicity, we assume $\sigma = 1$.

The values of α and d were chosen in the following manner: the α values are corresponding to the cases that are analytically derived in Secs. IV B and IV C, that is, $\alpha = 1$ and 2, respectively. For every α there are three different values of d, with similar location within the proper interval [it is needed to choose $d \in (-1/\alpha, 1 - 1/\alpha)$] (see, for example, [43]).

As was shown by many experiments (and can be partially visible in the Fig. 2), the trajectory length sufficient to identify ergodicity breaking appears to be linear with respect to the accepted ε. Generally, the higher the positive d parameter, the longer it takes for $F(n)$ to converge; this is the expected behavior, as $d > 0$ is responsible for the long-range dependence. Nevertheless, our studies show that such a case is observed only for $d > 0.4$.

On the other hand, as one can see in Fig. 2, treating $F(n)$ as a function of α only (that implies different α ranges for every value of d) suggests that the longest data sample is required in the case of the smallest α value. The values depicted are the mean values of the set of 100 values of $n_0(\varepsilon)$, each calculated on the base of 1000 simulated sample paths. The surprisingly high and not perfectly aligning values for $\varepsilon = 0.01$ are the result of

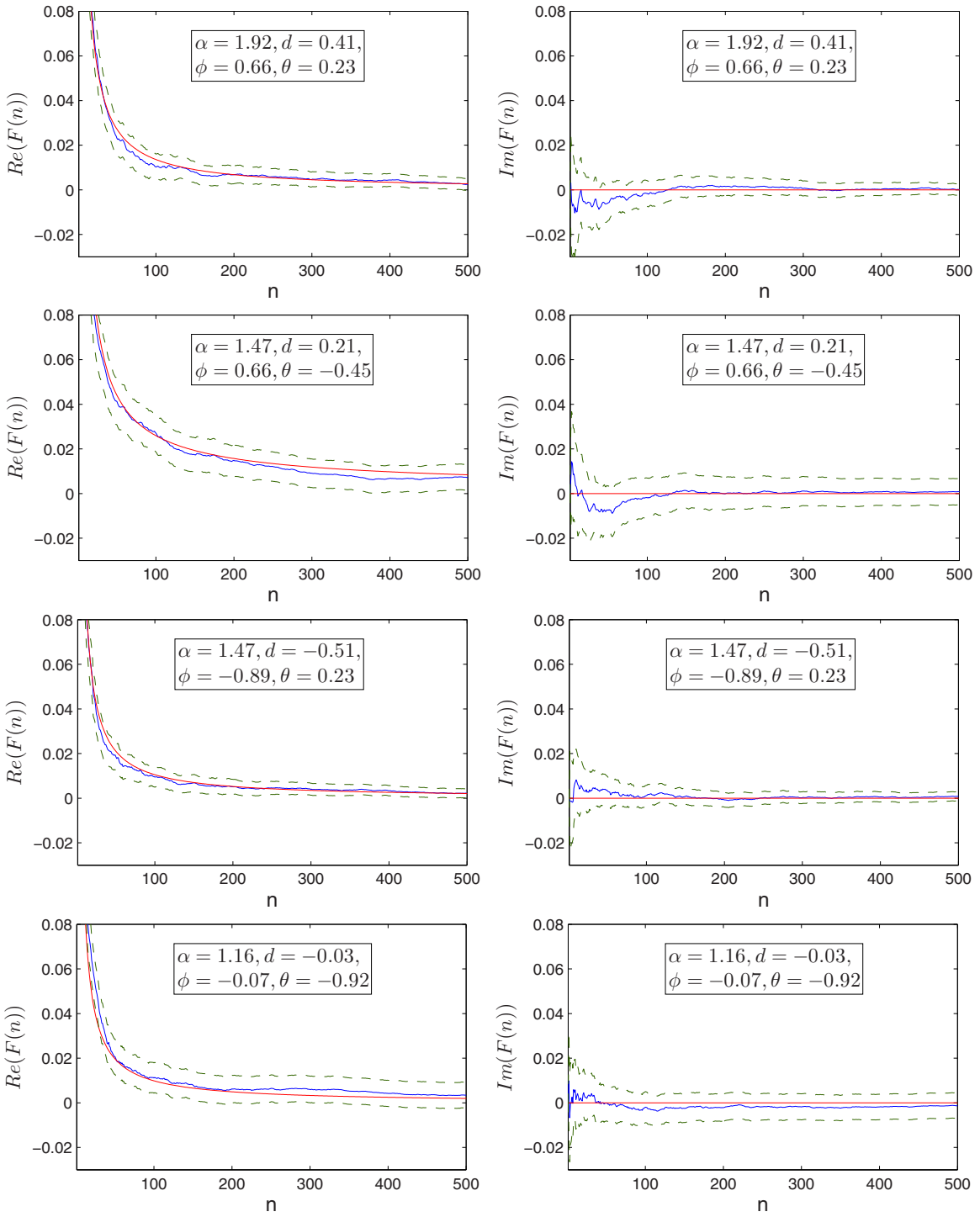

FIG. 1. The test functional $F_\alpha(n)$ for ARFIMA(1,d,1) model with various values of α, d, ϕ, and θ. The blue lines are the values obtained from 1000 simulated sample paths, each of the length 500, while the red ones are the theoretical ones; the green dashed lines are 95% confidence intervals. Please note that the imaginary part is theoretically equal to 0, as according to the formula (9); during the simulations some numerical noise occurs.

numerical perturbations and experiment constraints (the same truncation has been used in all three cases; for smaller values of ε, more c_j coefficients need to be used).

Because calculations of the exact formula (9) require large computational burden, in the following we find an analytical formula for the bound of $|F(n)|$. To this end, we concentrate on two noteworthy cases of α, namely $\alpha = 1$ and 2, corresponding to two well-known distributions. They can allow us to make an estimation for the rest of two reasonable intervals, respectively, $\alpha \in (0,1]$ and $\alpha \in (1,2]$.

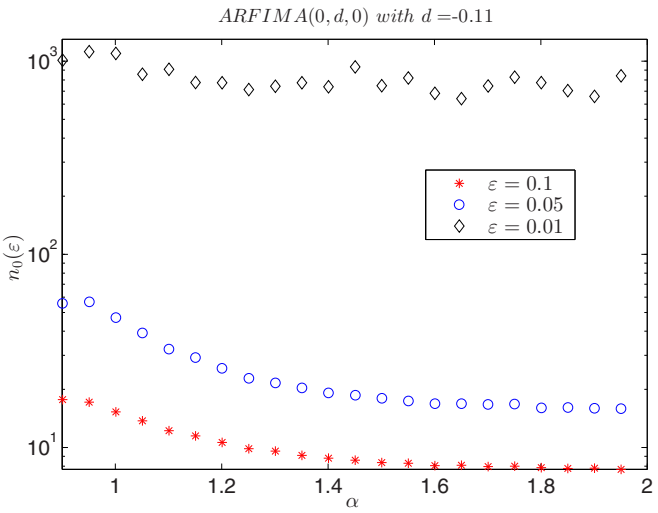

FIG. 2. The minimal sample length $n_0(\varepsilon)$ of ARFIMA(0, -0.11,0) process for different α-stable innovations. Please note that the OY axis is in log scale to better present the common behavior.

B. Dynamical functional for Cauchy distribution

The Cauchy distribution is a special case of the family of α-stable distributions when $\alpha = 1$. In this case, we know explicitly its density and it is related for example to bulk-mediated diffusion [29,50]. First, let us note that for $\alpha = 1$ we have

$$D(n) = \exp\left(-\sigma \sum_{j=0}^{n-1} |c_j| - \sigma \sum_{j=0}^{\infty} |c_j - c_{j+n}|\right), \tag{11}$$

where σ is a scale parameter of innovation process ε_n in the ARFIMA model. As for ARFIMA(0,d,0) in case of $\alpha \leqslant 1$ (equivalently $d < 0$) $c_j = \frac{\Gamma(j+d)}{\Gamma(d)\Gamma(j+1)} < 0$ for $j > 0$,

$$\sum_{j=0}^{\infty} |c_j| = 2. \tag{12}$$

That is indeed a very interesting observation, as it is independent of the memory parameter d. Second, $c(0) = 1$ and c_n for $n > 0$ are negative and increasing to 0. Hence, we can remove the absolute value in the reasonable way:

$$D(n) = \exp\left\{-\sigma\left(1 - \sum_{j=1}^{n-1} c_j\right) - \sigma\left[1 - c_n + \sum_{j=1}^{\infty}(c_{j+n} - c_j)\right]\right\} = \exp\left(-2\sigma - \sigma \sum_{j=1}^{\infty} c_{j+n} + \sigma \sum_{j=1}^{n} c_j + \sigma \sum_{j=1}^{\infty} c_j\right)$$
$$= \exp\left(-2\sigma - \sigma \sum_{j=n+1}^{\infty} c_j + \sigma \sum_{j=1}^{n} c_j + \sigma \sum_{j=1}^{\infty} c_j\right).$$

That leads us to the following form of $D(n)$:

$$D(n) = \exp\left(-2\sigma + 2\sigma \sum_{j=0}^{n} c_j\right)$$
$$= \exp\left\{-2\sigma + 2\sigma\left[\frac{\Gamma(n+1+d)}{d\Gamma(d)\Gamma(n+1)} - 1\right]\right\}$$
$$= \exp\left[-2\sigma + 2\sigma\left(\frac{n+d}{d} c_n - 1\right)\right], \tag{13}$$

that is, the dynamical functional in this case can be expressed via a single coefficient of moving average form of the process. That observation points very strongly on the fact that $D(n)$ is in fact some measure of the process dependence. Hence, the test functional $F(n)$ we would like to examine has the form

$$F(n) = \frac{1}{n} \sum_{k=0}^{n-1} \exp\left(-4\sigma + 2\sigma \frac{k+d}{d} c_k\right) - \exp(-4\sigma)$$
$$= \exp(-4\sigma) \frac{1}{n} \sum_{k=0}^{n-1} \left[\exp\left(2\sigma \frac{k+d}{d} c_k\right) - 1\right]. \tag{14}$$

A useful remark is that $2\sigma \frac{k+1+d}{d} c_k > 0$ for every $n > 0$ and d, hence, the functional is clearly non-negative. Moreover, one can notice that $\frac{\Gamma(n+1+d)}{d\Gamma(d)\Gamma(n+1)}$ is in fact the codifference of the process and as it was shown by Kokoszka and Taqqu in [44], it is positive for every $\alpha \in (0,1]$.

C. Dynamical functional for Gaussian distribution

The Gaussian distribution can be considered as a special case of the family of α-stable distributions when $\alpha = 2$. However, many of its properties are in a sharp contrast to other α-stable distributions for $\alpha < 2$ which have, for example, infinite second moment. The Gaussian distribution is important for such basic stochastic processes in physics and biology as Brownian motion and FBM [6].

For $\alpha = 2$,

$$D(n) = \exp\left(\sigma^2 \sum_{j=0}^{\infty} 2c_{j+n} c_j - 2c_j^2\right). \tag{15}$$

Provided that c_j are as in (8), after a few transformations one obtains

$$\sum_{j=0}^{\infty} c_j^2 = \frac{\Gamma(1-2d)}{\Gamma(1-d)^2} \tag{16}$$

and

$$\sum_{j=0}^{\infty} 2c_{j+n} c_j = \frac{\Gamma(1-2d)\Gamma(d+n)}{\Gamma(1-d)\Gamma(d)\Gamma(-d+n+1)}. \tag{17}$$

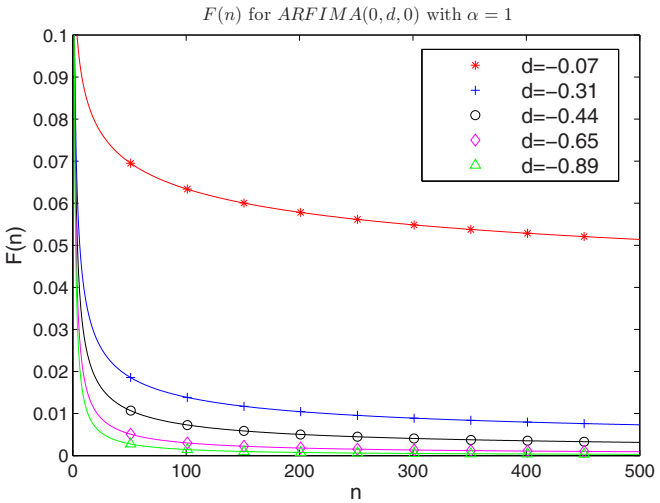

FIG. 3. The test functional $F_\alpha(n)$ for $\alpha = 1$ with respect to n for few d values.

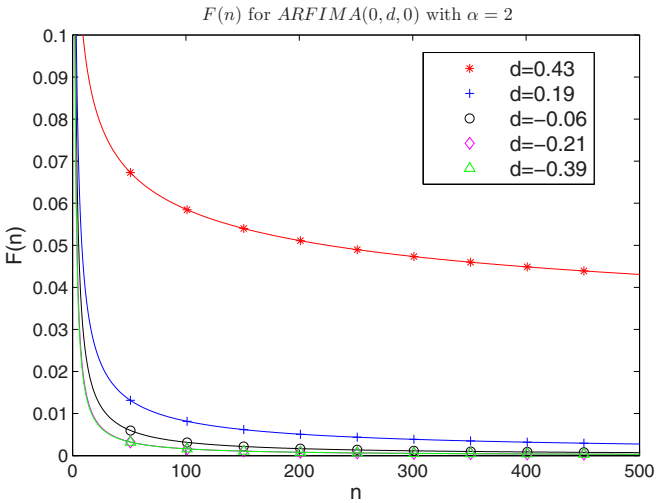

FIG. 4. The test functional $F_\alpha(n)$ for $\alpha = 2$ with respect to n for few d values.

Hence,

$$D(n) = \exp\left\{\sigma^2\left[\frac{\Gamma(1-2d)\Gamma(d+n)}{\Gamma(1-d)\Gamma(d)\Gamma(-d+n+1)} - 2\frac{\Gamma(1-2d)}{\Gamma(1-d)^2}\right]\right\}. \tag{18}$$

That leads to the following formula for $F(n)$:

$$\begin{aligned} F(n) =& \frac{1}{n}\sum_{k=0}^{n-1}\exp\left\{\sigma^2\left[\frac{\Gamma(1-2d)\Gamma(d+k)}{\Gamma(1-d)\Gamma(d)\Gamma(-d+k+1)} - 2\frac{\Gamma(1-2d)}{\Gamma(1-d)^2}\right]\right\} - \exp\left[-2\sigma^2\frac{\Gamma(1-2d)}{\Gamma(1-d)^2}\right] \\ =& \exp\left[-2\sigma^2\frac{\Gamma(1-2d)}{\Gamma(1-d)^2}\right] \\ & \frac{1}{n}\sum_{k=0}^{n-1}\left\{\exp\left[\sigma^2\frac{\Gamma(1-2d)\Gamma(d+k)}{\Gamma(1-d)\Gamma(d)\Gamma(-d+k+1)}\right] - 1\right\}. \end{aligned} \tag{19}$$

As it was shown in [44], $\sigma^2\frac{\Gamma(1-2d)\Gamma(d+k)}{\Gamma(1-d)\Gamma(d)\Gamma(-d+k+1)}$ being the covariance of the process is not necessarily positive; in this case of $\alpha = 2$ it is greater than 0 for $d > 0$, but not for $d < 0$.

Those formulas, although not reversible with respect to n (in case of arbitrary d), are simple enough to be easily computed and the relevant value n can be read from the plot. Such graphs are presented in Figs. 3 and 4, respectively, for $\alpha = 1$ and 2.

In the second column of Table I there are $n_0(0.05)$ values for ARFIMA(0,d,0) calculated using the formula from this and previous section.

D. Simulation results

In order to show that analytical results are adequate, the first two columns of Table I present analytical results for the ARFIMA(0,d,0) model to find n_0 satisfying (10). Now, we turn to a more experimentally grounded analysis. In the third column of Table I, such $n_0(0.05)$ are obtained via simulating sample paths of the process are given. For each case of α and d we simulated 1000 sample paths, calculated the empirical values of $F(n)$, and searched for the mentioned $n_0(\varepsilon)$ value. The results are then averaged over 100 repetitions. Note that in parentheses we also give the standard deviations. What can concern the reader is the particularly high value obtained for $\alpha = 2$ and $d = 0.43$.

As in experiments one usually has a single trajectory, which is needed to examine this case in details. To this end, we simulate sample paths of ARFIMA model, calculate the empirical values of $F(n)$ [the ensemble average in (1) is now replaced with a time average], and search for the mentioned $n_0(\varepsilon)$ value. The results were then averaged over 100 repetitions. The obtained mean results, together with the standard deviations, are given in Table II. Note that this can be done only for an ergodic process, so we would obtain only a necessary (but not sufficient) condition for ergodicity (for details see [36]), that means based on those results we can only claim lack of ε-ergodicity, or rather in this case incorrect model choice, as we know the process is ergodic. The two bolded cases, i.e., ARFIMA(0,d,0) with $\alpha = 2$ and ARFIMA(1,0,0) with $\alpha = 2$, when properly aggregated, are discrete time versions of the popular physical diffusion models: increments of the fractional Brownian motion and the Ornstein-Uhlenbeck process, respectively. Recall that parameter d in the ARFIMA(0,d,0) model is linked with the Hurst exponent H with the relation $H = d + 1/\alpha$. For the ARFIMA(1,0,0) model we use three different values of the ϕ parameter, describing the speed of mean reversion (note that the memory parameter d is now equal to 0). What is particularly interesting from the experimentalist point of view is that those values are not very high, which means the sample length required to state something about ergodicity is possible to obtain in most of the experiments.

Finally, we would like to demonstrate how our methodology of the ε-ergodicity can be easily applied to real molecular biology experimental data. We consider the single trajectory studied in Fig. 10 in [45], which is related to the analysis of fluorescently labeled telomeres in the nucleus of living human cell originating from the U2OS cancer cell line. To

TABLE II. The minimal trajectory length of α-stable ARFIMA process for the necessary condition to claim ε-ergodicity in the one trajectory case. The values in parentheses are the standard deviations. Please note that for one trajectory one obtains only a necessary (but not sufficient) condition for ergodicity, which is the reason of the smaller $n_0(\varepsilon)$ values in case of high d values in comparison with Table I. The two bolded cases are discrete time versions of the popular physical diffusion models.

		$\varepsilon = 0.1$	$\varepsilon = 0.05$	$\varepsilon = 0.01$
$\alpha = 2$	$d = 0.43$	36 (26.03)	67 (150.77)	443 (855.82)
ARFIMA	$d = 0.06$	25 (8.28)	34 (29.40)	190 (398.07)
$(1,d,1)$	$d = -0.39$	22 (3.96)	28 (12.51)	128 (163.72)
$\alpha = 1$	$d = -0.07$	22 (9.81)	34 (54.03)	234 (786.82)
ARFIMA	$d = -0.44$	21 (2.57)	25 (9.88)	61 (37.16)
$(1,d,1)$	$d = -0.89$	21 (2.49)	25 (7.81)	60 (34.83)
$\alpha = 2$	$d = 0.43$	45 (55.93)	79 (162.44)	419 (841.53)
ARFIMA	$d = 0.06$	25 (8.28)	40 (37.10)	166 (306.70)
$(0,d,0)$	$d = -0.39$	23 (4.218)	32 (16.73)	101 (118.77)
$\alpha = 1$	$d = -0.07$	23 (3.44)	33 (33.21)	152 (513.37)
ARFIMA	$d = -0.44$	21 (1.59)	26 (8.16)	118 (456.37)
$(0,d,0)$	$d = -0.89$	22 (2.38)	26 (8.35)	63 (44.79)
$\alpha = 2$	$\phi = 0.84$	34 (17.63)	43 (23.69)	142 (119.00)
ARFIMA	$\phi = 0.23$	25 (7.80)	32 (14.95)	127 (155.24)
$(1,0,0)$	$\phi = -0.79$	22 (3.27)	29 (10.78)	95 (103.66)

the stationary increments of this trajectory the ARFIMA$(0, -0.32, 1)$ model has been fitted, with $\theta = -0.65$ and $\sigma = 0.01$. If we calculate the n_0 for the mentioned data (using the time average, as it is the one trajectory case), we get that $n_0(0.05) = 20$. It proves that as we have very long trajectory (1000 observations, 999 increments), it can be clearly stated that the chosen model is well fitted (since we know from [45] that it is ergodic).

V. CONCLUSIONS

In this paper, we have proposed, investigated, and validated reliable statistical analysis that can be applied to study ε-ergodicity breaking for fractional anomalous diffusion data, for example, representing single-particle tracking in living cells. In many biological experiments the data acquisition procedure is complicated, so that only few such trajectories of considerable length are available. In such a situation we propose to use the ε-ergodicity. In the telomeres example presented in Sec. IV we considered $\varepsilon = 0.05$, but our methodology leaves to experimentalists a crucial decision on the level of accuracy ε, which can depend, for example, on the known level of the calibrated noise in the experiment.

In Table I is presented $n_0(0.05)$, the minimal n for which $|F(n)| < 0.05$ for three different methods (general analytical formula, closed formula for $\alpha = 2$, and $\alpha = 1$ and simulating sample paths) for ARFIMA$(0,d,0)$ process. The biggest $n_0(0.05)$ for for $\alpha = 2$ and $d = 0.43$ is suggested via simulations, which is biased by numerical burden. The more exact methods, that is the use of formulas (14) and (19), give the smaller bounds, but on the other hand, such strict criterion can lead to the false ε-ergodicity breaking identification. For the empirical use, we suggest taking into account the quality of the data while applying particular thresholds for $n_0(\varepsilon)$; nevertheless, it is worth to remember that decreasing value of ε can result in claiming ε-ergodicity breaking even for ergodic samples due to numerical burden.

We foresee two future steps that can be done in a further study: (i) calculation of the closed forms for the mentioned two cases of $\alpha = 1$ and 2 while increasing the rank of MA(p) and AR(q) polynomials and (ii) verifying the behavior (and the monotonicity, if it occurs) of $F_\alpha(n)$ as the function of α parameter [with d set in case of ARFIMA$(0,d,0)$]. The latter one will allow stating some inequalities concerning the minimal trajectory length for $\alpha \in (0,2]$ and from the practical point of view it should be sufficient for most of the applications. The initial work upon the behavior of $F_\alpha(n)$ as one-variable function (of α or d) has been already done. The simulations show that parameter d influences strongly the convergence speed of $F_\alpha(n)$, hence the values of $n_0(\varepsilon)$. Moreover, also the estimators $\hat{F}(n)$ and $\hat{n}_0(\varepsilon)$ can be interesting objects of study.

ACKNOWLEDGMENTS

The research of J.J was cofinanced by the Foundation for Polish Science under the START programme. The research of H.L.-O., G.S., and A.W. was partially supported by NCN Maestro Grant No. 2012/06/A/ST1/00258.

[1] R. Metzler, J. Jeon, A. G. Cherstvy, and E. Barkai, Phys. Chem. Chem. Phys. **16**, 24128 (2014).
[2] Y. Meroz and I. M. Sokolov, Phys. Rep. **573**, 1 (2015).
[3] A. G. Cherstvy and R. Metzler, J. Chem. Phys. **142**, 144105 (2015).
[4] D. Krapf, Curr. Top. Membr. **75**, 167 (2015).
[5] I. Golding and E. C. Cox, Phys. Rev. Lett. **96**, 098102 (2006).
[6] I. Bronstein, Y. Israel, E. Kepten, S. Mai, Y. Shav-Tal, E. Barkai, and Y. Garini, Phys. Rev. Lett. **103**, 018102 (2009).
[7] M. Magdziarz, A. Weron, K. Burnecki, and J. Klafter, Phys. Rev. Lett. **103**, 180602 (2009).
[8] C. Manzo and M. F. Garcia-Parajo, Rep. Prog. Phys. **78**, 124601 (2015).
[9] M. P. Backlund, R. Joyner, and W. E. Moerner, Phys. Rev. E **91**, 062716 (2015).
[10] R. Metzler and J. Klafter, Phys. Rep. **339**, 1 (2000).
[11] M. J. Saxton, Nat. Methods **5**, 671 (2008).
[12] K. Burnecki, M. Muszkieta, G. Sikora, and A. Weron, Europhys. Lett. **98**, 10004 (2012).
[13] E. Barkai, Y. Garini, and R. Metzler, Phys. Today **65**(8), 29 (2012).
[14] W. Deng and E. Barkai, Phys. Rev. E **79**, 011112 (2009).
[15] C. W. J. Granger and R. Joyeux, J. Time Ser. Anal. **1**, 15 (1980).
[16] H. R. M. Hosking, Biometrika **68**, 165 (1981).
[17] K. Burnecki, G. Sikora, and A. Weron, Phys. Rev. E **86**, 041912 (2012).

[18] J. Slezak and A. Weron, Phys. Rev. E **91**, 053302 (2015).
[19] J. Slezak, S. Drobczynski, K. Weron, and J. Masajada, Appl. Opt. **53**, B254 (2014).
[20] M. Magdziarz and A. Weron, Ann. Phys. (NY) **326**, 2431 (2013).
[21] A. Weron and M. Magdziarz, Phys. Rev. Lett. **105**, 260603 (2010).
[22] D. Krapf, Phys. Chem. Chem. Phys. **15**, 459 (2013).
[23] A. Fulinski, J. Chem. Phys. **138**, 021101 (2013).
[24] A. V. Weigel, M. M. Tamkun, and D. Krapf, Proc. Natl. Acad. Sci. USA **110**, E4591 (2013).
[25] A. V. Weigel, B. Simon, M. M. Tamkun, and D. Krapf, Proc. Natl. Acad. Sci. USA **108**, 6438 (2011).
[26] P. J. Brockwell and R. A. Davis, *Introduction to Time Series and Forecasting* (Springer, New York, 2002).
[27] P. J. Brockwell and R. A. Davis, *ITSM for Windows. A User's Guide to Time Series Modelling and Forecasting* (Springer, New York, 1994).
[28] C. Manzo, J. A. Torreno-Pina, P. Massignan, G. J. Lapeyre Jr., M. Lewenstein, and M. F. Garcia Parajo, Phys. Rev. X **5**, 011021 (2015).
[29] G. Campagnola, K. Nepal, B. W. Schroeder, O. E. Peersen, and D. Krapf, Sci. Rep. **5**, 17721 (2015).
[30] S. M. Ali Tabei, S. Burov, H. Y. Kim, A. Kuznetzov, T. Huynh, J. Jureller, L. H. Philipson, A. R. Dinner, and N. F. Scherer, Proc. Natl. Acad. Sci. USA **110**, 4911 (2013).
[31] X. Brokmann, J.-P. Hermier, G. Messin, P. Desbiolles, J.-P. Bouchaud, and M. Dahan, Phys. Rev. Lett. **90**, 120601 (2003).
[32] J.-H. Jeon, A. V. Chechkin, and R. Metzler, Phys. Chem. Chem. Phys. **16**, 15811 (2014).
[33] A. G. Cherstvy, A. V. Chechkin, and R. Metzler, New J. Phys. **15**, 083039 (2013).
[34] J. A. Torreno-Pina, B. M. Castro, C. Manzo, S. I. Buschow, A. Cambi, and M. F. Garcia-Parajo, Proc. Natl. Acad. Sci. USA **111**, 11037 (2014).
[35] J. A. Torreno-Pina, C. Manzo, and M. F. Garcia-Parajo, J. Phys. D: Appl. Phys. **49**, 104002 (2016).
[36] M. Magdziarz and A. Weron, Phys. Rev. E **84**, 051138 (2011).
[37] S. Cambanis, K. Podgórski, and A. Weron, Studia Math. **115**, 109 (1995).
[38] J. Janczura and A. Weron, J. Chem. Phys. **142**, 144103 (2015).
[39] Y. Lanoiselée and D. S. Grebenkov, Phys. Rev. E **93**, 052146 (2016).
[40] B. O. Koopman and J. von Neumann, Proc. Natl. Acad. Sci. USA **18**, 255 (1932).
[41] K. Burnecki, E. Kepten, Y. Garini, G. Sikora, and A. Weron, Sci. Rep. **5**, 11306 (2015).
[42] K. Burnecki and A. Weron, J. Stat. Mech. (2014) P10036.
[43] P. S. Kokoszka and M. S. Taqqu, Stoch. Proc. Appl. **60**, 19 (1995).
[44] P. S. Kokoszka and M. S. Taqqu, J. Time Ser. Anal. **15**, 203 (1994).
[45] H. Loch, J. Janczura, and A. Weron, Phys. Rev. E **93**, 043317 (2016).
[46] J. W. Pearson, *Computation of Hypergeometric Functions* (Oxford University Press, Oxford, 2009).
[47] S. Ling and W. K. Li, J. Am. Stat. Assoc. **92**, 1184 (1997).
[48] K. Burnecki, J. Stat. Mech. (2012) P05015.
[49] S. Stoev and M. S. Taqqu, Fractals **12**, 95 (2004).
[50] A. V. Chechkin, I. M. Zaid, M. A. Lomholt, I. M. Sokolov, and R. Metzler, Phys. Rev. E **86**, 041101 (2012).