Fractional process as a unified model for subdiffusive dynamics in experimental data

Krzysztof Burnecki,* Grzegorz Sikora,† and Aleksander Weron‡
Hugo Steinhaus Center, Institute of Mathematics and Computer Science, Wroclaw University of Technology, Wyspianskiego 27, 50-370 Wroclaw, Poland
(Received 18 April 2012; revised manuscript received 14 September 2012; published 17 October 2012)

We show how to use a fractional autoregressive integrated moving average (FARIMA) model to a statistical analysis of the subdiffusive dynamics. The discrete time FARIMA(1,d,1) model is applied in this paper to the random motion of an individual fluorescently labeled mRNA molecule inside live *E. coli* cells in the experiment described in detail by Golding and Cox [Phys. Rev. Lett. **96**, 098102 (2006)] as well as to the motion of fluorescently labeled telomeres in the nucleus of live human cells (U2OS cancer) in the experiment performed by Bronstein *et al*. [Phys. Rev. Lett. **103**, 018102 (2009)]. It is found that only the memory parameter d of the FARIMA model completely detects an anomalous dynamics of the experimental data in both cases independently of the observed distribution of random noises.

DOI: 10.1103/PhysRevE.86.041912 PACS number(s): 87.16.A−, 05.40.Fb, 02.50.Ey, 02.70.−c

I. INTRODUCTION

A phenomenon observed in recent nanoscale single-molecule biophysics experiments is subdiffusion, which largely departs from the classical Brownian diffusion theory since the mean-squared displacement (MSD) is sublinear. Determining anomalous diffusion in crowded fluids, e.g., in the cytoplasm of living cells [1–4], or in more controlled *in vitro* experiments [5–7], is a challenging problem. Subdiffusion can have three different physical origins: obstruction, binding, and viscoelasticity [8,9].

The most popular theoretical models that are commonly employed are continuous-time random walk (CTRW) [3,8], obstructed diffusion (OD) [4,10], fractional Brownian motion (FBM) [6,11], fractional Lévy α-stable motion (FLSM) [12], and fractional Langevin equation (FLE) [7,11,13]. These models can be divided into two categories: with short memory (CTRW, OD) and fractional with long (power-law) memory (FBM, FLSM, and FLE).

In this paper we propose a FARIMA model which unifies the latter category. The FARIMA(1,d,1) process $X(t)$ is represented by the following fractional difference equation with constant coefficients:

$$(1-B)^d[X(t)-\phi_1 X(t-1)] = Z(t)-\psi_1 Z(t-1), \quad (1)$$

where $t = 0,\pm1,\ldots$.

The fractional difference operator $(1-B)^d$ is defined in the supplemental material [14], $d < 1-1/\alpha$, and $0 < \alpha \leqslant 2$ is the index of stability. The noise defined by $Z(t)$ consists of independent and identically distributed (i.i.d.) random variables belonging to the domain of attraction of the Lévy α-stable law [15]. Hence, the noise may be either Gaussian or non-Gaussian with finite variance or it may have infinite variance. For the finite variance case, apart from the Gaussian law, the typical choices are log-normal and normal-inverse Gaussian distributions [16,17]. For infinite variance variables one may consider, for example, symmetric and skewed Lévy stable and Pareto distributions [18]. For the description of general FARIMA(p,d,q) processes, where p and q are positive integers, see the supplemental material [14].

First, from the physical point of view, it is known that FARIMA is a discrete time analog of FLE that takes into account the memory parameter d [19]. It also includes other popular models of subdiffusive dynamics as FBM and FLSM, which are limiting cases of aggregated FARIMA. In contrast to FBM and FLSM, it can also describe different light- and heavy-tailed noises and a short (exponential) dependence. FARIMA, similarly as increments of FBM, exhibits power-law (fractional) long-time correlations. This results in an anomalous diffusion exponent MSD. Therefore, FARIMA can be associated with the third origin of anomalous diffusion, i.e., viscoelasticity [4].

In Ref. [20] the p-variation test was applied to the Golding-Cox experimental data [3]. The authors analyzed six two-dimensional (2D) sample paths (all those having more than $2^9 = 512$ points, which is reasonable for the p-variation test) from their set of 27 trajectories. The test clearly demonstrated that the subdiffusion cannot stem from the CTRW model. Moreover, the tests in Refs. [20,21] and [12] show that there is no reason to reject the hypothesis that the data follow FBM or FLSM. This resolved a controversy over the underlying reason for the Golding-Cox subdiffusion [22–24]. Here we look for a unified approach to fractional subdiffusive dynamics.

As a candidate suitable for extensive statistical analysis of the subdiffusive dynamics in biological cells we propose here the FARIMA model [25–28]. It is a discrete time analog of FLE [19] that allows for the non-Gaussian law (Lévy α-stable) and the long-term dependence (long memory). The structure of the paper is as follows. Basic building blocks of FARIMA process are discussed in in Sec. II. Moreover, a close relation to FBM and its α-stable extension FLSM is explained in detail. An original estimation procedure for parameters including the memory parameter d of the FARIMA model is also provided. Next, in Sec. III we present the sample MSD concept. Fitting FARIMA parameters to the mRNA and telomere data with the help of the introduced estimation procedure and the standard ITSM package [28] is presented in Secs. IV and V, respectively. The contents of the output windows of the ITSM are included in the supplemental material [14]. The goodness

*krzysztof.burnecki@pwr.wroc.pl
†grzegorz.sikora@pwr.wroc.pl
‡aleksander.weron@pwr.wroc.pl

of fit is illustrated by plotting sample MSD for all analyzed trajectories along with the 95% confidence intervals calculated assuming the FARIMA model. Discussion of the results is given in Sec. VI. Conclusions are presented in the last section.

II. FARIMA MODELS

FARIMA series are fractionally differenced autoregressive moving average (ARMA) series [25]. The model generalizes three broad classes of time series, namely, the autoregressive (AR) models, the integrated (I) models, and the moving average (MA) models [26,28]. In the Gaussian case they posses the property of exponentially decaying correlation function. The generalization allows for a power-law correlation structure which leads to the notion of long-term dependence. Such an approach is widely used in modeling the dynamical behavior of various complex physical systems. Consequently, the FARIMA models (called also ARFIMA) have already appeared in the physical literature [16,17,29–33].

In many applications a FARIMA$(1,d,1)$ model is sufficient to describe the data well; see, e.g., Ref. [16]. So, for simplicity of presentation, we concentrate on this case only. The basic building blocks of a FARIMA$(1,d,1)$ model are the AR(1) process, $X(t) = \phi_1 X(t-1) + Z(t)$, and MA(1) process, $X(t) = Z(t) - \psi_1 Z(t-1)$. AR(1) is a causal or future-independent function of noise and stands for the regression. The explanatory variable is the observation immediately prior to our current observation. In order to get an idea about the role of MA part let us concentrate on the case when $Z(t)$ is a white noise. It appears that if $X(t)$ is a stationary 1-correlated time series, i.e., $X(s)$ and $X(t)$ are independent whenever $|t-s| > 1$ (in contrast to an i.i.d. sequence, which is zero-dependent), then it can be represented as the MA(1) process [28]. The dependence (correlation in the white noise case) is only one lag long and its intensity is fully controlled by the parameter ψ_1. Hence, ARMA models introduce short memory of the process.

The time series $X(t)$ is an ARMA(1,1) process if it is stationary and satisfies (for every t) linear difference equation with constant coefficients:

$$X(t) - \phi_1 X(t-1) = Z(t) - \psi_1 Z(t-1), \tag{2}$$

where $t = 0, \pm 1, \ldots$.

A stationary solution of ARMA(1,1) equation exists if and only if $\phi_1 \neq \pm 1$. If $|\phi_1| < 1$, then the unique stationary solution exists and is causal, since $X(t)$ can be expressed in terms of the current and past values $Z(s)$, $s \leqslant t$. Otherwise, if $|\phi_1| > 1$, then the solution is not causal since $X(t)$ is then a function of Z_s, $s \geqslant t$ [28].

ARMA models provide a general framework for studying stationary processes and simple prediction algorithms to describe future behavior of the dynamics. For example, linear prediction of the causal ARMA(1,1) process looks as follows:

$$\hat{X}(t+1) = \phi_1 X(t) + \theta_t(\phi_1,\psi_1)[X(t) - \hat{X}(t)], \tag{3}$$

where $t = 0, \pm 1, \ldots$.

A generalization of this class, which incorporates a wide range of nonstationary series, is provided by ARIMA processes, i.e., processes that reduce to ARMA processes when differenced finitely many times.

Let us observe that ARMA models are short memory stationary processes since their autocorrelation function ρ at lag h converges exponentially to zero as $h \to \infty$.

Finally, stationary processes with much more slowly decreasing autocorrelation function, i.e., with long (power-law) memory are called FARIMA processes and are given by Eq. (1). Such processes are asymptotically self-similar with the parameter $H = d + 1/\alpha$. The process $X(t)$ is well defined for $-\infty < d < 1 - 1/\alpha$. In the Gaussian case, i.e., when $\alpha = 2$, the rate of decay of the autocovariance function $\mathrm{Cov}(n) = \langle X(n)X(0)\rangle - \langle X(n)\rangle\langle X(0)\rangle$ for the FARIMA model is n^{2d-1}. Moreover, for $d > 0$ we have $\sum_{n=0}^{\infty} |\mathrm{Cov}(n)| = \infty$. This serves as a classical definition of long memory [26]. However, in this paper we call all FARIMA processes with $d \neq 0$ long-term dependent to distinguish between ARMA models with exponentially decaying autocorrelation function and FARIMA models with the power-law decay. For $\alpha < 2$ the covariance does not exist, and one has to replace it, e.g., with the codifference (see Ref. [15]). The codifference of the FARIMA process was studied in Ref. [34], where it was proved that for $d > 1 - 2/\alpha$ FARIMA possesses long-term dependence in the classical sense.

Modeling with FARIMA time series with infinite variance allows one to take into account heavy tails. The FARIMA processes offer also a lot of flexibility in modeling long power-law and short exponential dependencies by choosing the memory parameter d and appropriate autoregressive and moving average coefficients in Eq. (1). Hence, they can be better tailored to empirical data. Furthermore, Brownian motion (BM) and Lévy stable motion (LSM) correspond, in the limit sense [35], to FARIMA$(0,0,0)$ with Gaussian and α-stable noise, respectively. Similarly, FLSM corresponds to FARIMA$(0,d,0)$ with $d = H - 1/\alpha$, where H is the self-similarity parameter. For $\alpha = 2$ we recognize the FBM model.

In this context a basic question arises whether the α-stable distribution is Gaussian or not. There exists a simple and computationally convenient method that can be used in practice. Let $\{X_1, X_2, \ldots, X_N\}$ be a sample from an experiment described by an α-stable distribution. For each $1 \leqslant i \leqslant N$ form a statistics based on the first i observations $S_i^2 = 1/(i-1) \sum_{k=1}^{i} (X_k - \bar{X}_i)^2$, where $\bar{X}_i = 1/i \sum_{k=1}^{i} X_k$. Then plot S_i^2 against i. If the empirical distribution has a finite variance, then S_i^2 should converge to a finite value; otherwise S_i^2 should diverge [36]. A similar idea based on empirical fourth moment was used for an algorithm of recognition of stable distribution with Lévy index α close to 2, which was proposed in Ref. [37].

Modifying work in Ref. [38], we estimate the vector $\beta_0 = (\phi_1, \psi_1, d)$. For a sample $\{X_1, X_2, \ldots, X_N\}$ we denote the normalized periodogram by

$$\tilde{I}_N(\lambda) = \frac{\left|\sum_{t=1}^{N} X_t e^{-i\lambda t}\right|^2}{\sum_{t=1}^{N} X_t^2}, \quad -\pi \leqslant \lambda \leqslant \pi. \tag{4}$$

The estimator of the true parameter vector β_0 is well defined as the vector argument β for which the following function attains its minimum value:

$$g(\beta) = \int_0^{\pi} \tilde{I}_N(\lambda) W(\lambda,\phi_1,\psi_1)(2 - 2\cos\lambda)^d \, d\lambda, \tag{5}$$

where

$$W(\lambda,\phi_1,\psi_1) = \frac{\left(1-2\phi_1\cos\lambda+\phi_1^2\right)}{\left(1+2\psi_1\cos\lambda+\psi_1^2\right)}. \tag{6}$$

For details, see Ref. [39].

III. SAMPLE MSD

Let $\{X_i,\ i=1,\ldots,N\}$ be a sample of length N. The sample MSD was introduced in Ref. [12] as

$$M_N(\tau) = \frac{1}{N-\tau}\sum_{k=1}^{N-\tau}(X_{k+\tau}-X_k)^2. \tag{7}$$

The sample MSD is a time-average MSD on a finite sample regarded as a function of difference τ between observations. It is a random variable in contrast to the ensemble average, which is deterministic.

If the sample comes from an H-self-similar process with stationary increments belonging to the domain of attraction of Lévy α-stable law, then for large N

$$M_N(\tau) \sim \tau^{2d+1}, \tag{8}$$

where $d = H - 1/\alpha$ and $\sim$ means similarity in distribution.

Precisely:

(i) If $\alpha = 2$, then for large N and small τ $M_N(\tau) \sim \tau^{2d+1}\langle X_1^2\rangle$, where $d = H - 1/2$.

(ii) If $\alpha < 2$, then for large N/τ, $M_N(\tau) \sim C(N)\tau^{2d+1}S_{\alpha/2}$, where $C(N) = N^{2/\alpha}$, $d = H - 1/\alpha$, and $S_{\alpha/2}$ is a Lévy $\alpha/2$-stable random variable with the skewness parameter $\beta = 1$ [12].

In particular, for a FBM we obtain the well-known result that $M_N(\tau) \sim \tau^{2H}$, and for both Brownian and Lévy stable motions we arrive at the diffusion case, namely, $M_N(\tau) \sim \tau$ since $d = 0$; see also Ref. [40]. As a consequence, we see that the memory parameter d controls the type of anomalous diffusion. If $d < 0$ ($H < 1/\alpha$), in the negative dependence case, the process follows the subdiffusive dynamics; if $d > 0$ ($H > 1/\alpha$), the character of the process changes to superdiffusive. The subdiffusion pattern arises when the dependence is negative, so possible large positive jumps are quickly compensated by large negative jumps, and on average the process in crowded fluids travels shorter distances than the light-tailed Brownian motion.

Let us now define the partial sum process: $\{Y_n = \sum_{i=1}^{n} X_i : n = 1,2,\ldots,N\}$, where $X(t)$ is FARIMA. For FARIMA processes with noise in the domain of attraction of the Lévy stable law with $\alpha \leqslant 2$, sample MSD calculated for the partial sum process behaves like in Eq. (8). This stems from the fact that aggregated FARIMA time series are asymptotically $d + 1/\alpha$-self-similar and they converge to FBM for $\alpha = 2$ and to FLSM for $\alpha < 2$ [35].

Finally, as a by-product, the sample MSD can serve as a method of estimating d [17]. The method is well defined for the general Lévy stable case, and the estimator, denoted by d_{MSD}, has a very small variance. In order to apply it, first, we calculate the sample MSD for the partial sum process. Next, applying Eq. (8), we fit the linear regression line according to

$$\ln[M_N(\tau)] = \ln(C) + (2d_{\mathrm{MSD}}+1)\ln(\tau), \tag{9}$$

where C is assumed to be constant, $\tau = 1,\ldots,\tau_{\max}$, and $\tau_{\max}$ has to be chosen in such a way to ensure a large value of $N/\tau_{\max}$ [12].

IV. FITTING FARIMA PARAMETERS TO THE MRNA DATA

We study here the data of Golding and Cox describing the motion of individual fluorescently labeled mRNA molecules inside live *E. coli* cells [3]. The data clearly follow the subdiffusive character and consist of 27 2D trajectories [20]. We identified four Gaussian and nine α-stable trajectories among them. Here we analyze one Gaussian trajectory (no. 12x), the longest trajectory (no. 4y), which follows the non-Gaussian Lévy stable law, and the trajectory (no. 2y), which is neither Gaussian nor Lévy stable with $\alpha < 2$. Since these trajectories have missing data at some time points, which would influence the FARIMA estimation procedure, we concentrate on their

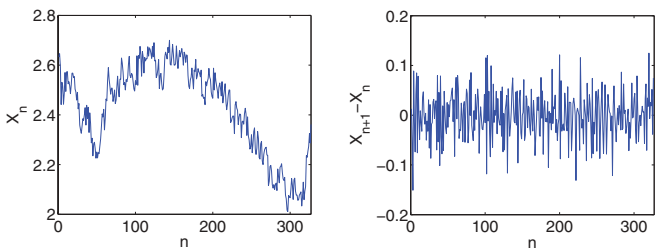

FIG. 1. (Color online) A plot of the x coordinate of particle position during a motion of a tagged RNA molecule inside an *E. coli* cell (left panel) and its increments (right panel). Trajectory no. 12 with Gaussian ($\alpha = 2$) noise.

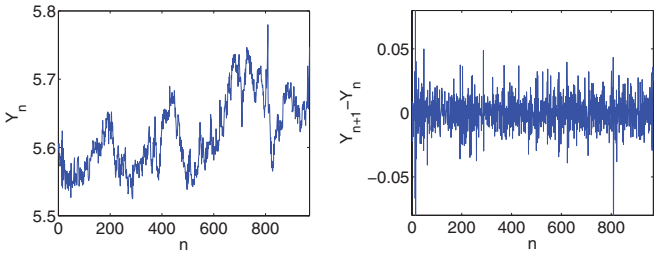

FIG. 2. (Color online) A plot of the y coordinate of particle position during a motion of a tagged RNA molecule inside an *E. coli* cell (left panel) and its increments (right panel). Trajectory no. 4 with Lévy stable noise with $\alpha = 1.81$.

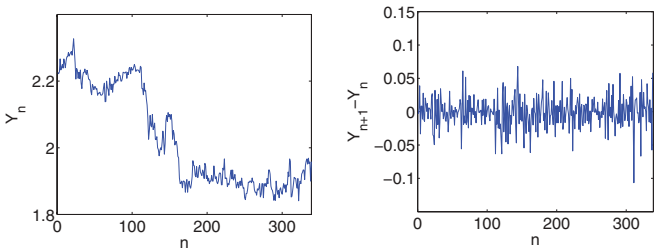

FIG. 3. (Color online) A plot of the y coordinate of particle position during a motion of a tagged RNA molecule inside an *E. coli* cell (left panel) and its increments (right panel). Trajectory no. 2 with unknown noise.

TABLE I. Results of the proposed estimation procedure for three sample trajectories of the Golding and Cox data.

Traj.	Model	ϕ_1	ψ_1	α	d_{MSD}
12x	FARIMA(0, −0.23,1)	0	0.12	2	−0.21
4y	FARIMA(0, −0.14,1)	0	0.16	1.81	−0.17
2y	FARIMA(1, −0.08,1)	0.29	0.48	?	−0.16

longest continuous parts: $\{(X_n, Y_n) : n = 1,2,\ldots,N\}$ with $N = 327$, $N = 970$, and $N = 339$, respectively. The data and their increments are presented in Figs. 1–3. First, we estimated FARIMA parameters by applying the estimation procedure given by Eq. (5). The results along with MSD estimates (d_{MSD}) are presented in Table I. We also compared them with the estimates calculated with the standard ITSM package, which was introduced for light-tailed distributions [28] (for details see the supplemental material [14]). The ITSM algorithm leads to different results. This is due to the fact that the procedures use different statistics for estimation, but since the MSD estimates (d_{MSD}) are closer to the memory parameter values obtained by the former procedure and most of the data are not Gaussian, we choose the values depicted in Table I as the FARIMA model parameters.

The goodness of fit of FARIMA processes is illustrated in Figs. 4–6. We plotted sample MSD for three analyzed trajectories along with the 95% confidence intervals calculated under the assumption of the FARIMA model. We can see the empirical trajectiories lie within the bounds except for several values, which indicates the model is well fitted. We also note that the $M_N(\tau)$ statistic (the last column in Table I) is reliable only for large N/τ [12]. We set similar minimal values for N/τ for different trajectories.

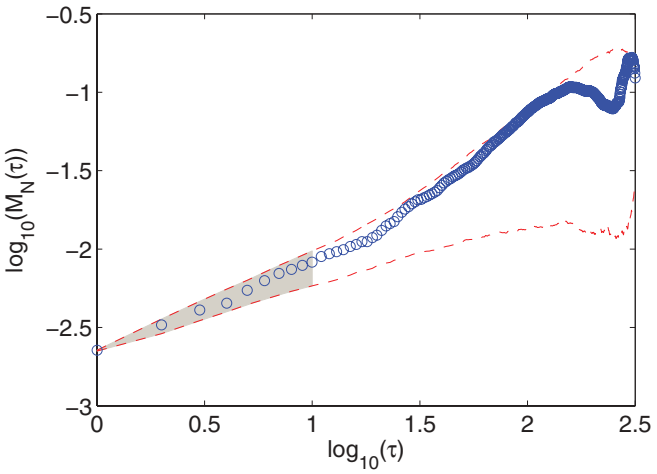

FIG. 4. (Color online) Sample MSD (blue circles) for the x coordinate of the trajectory no. 12 with Gaussian ($\alpha = 2$) noise and estimated 95% confidence intervals obtained via Monte Carlo simulations under the assumption of the Gaussian FARIMA model (red dashed lines) in double logarithmic scale. The values of the statistic lie within the confidence interval for the FARIMA model supporting its goodness of fit. The memory parameter $d_{\mathrm{MSD}} = -0.21$ was obtained by fitting the line to the values in the gray region.

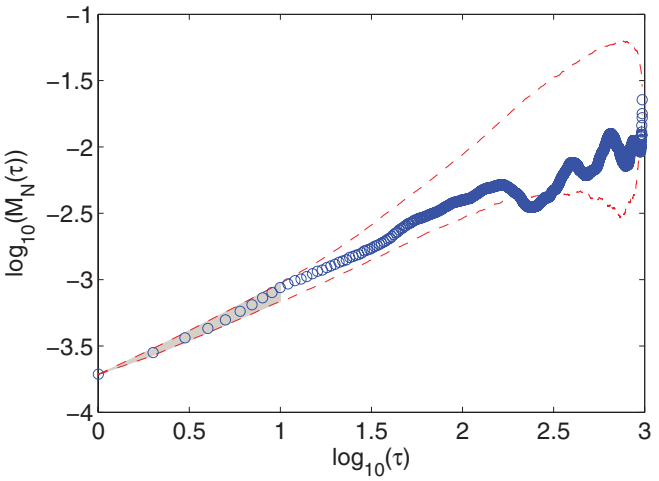

FIG. 5. (Color online) Sample MSD (blue circles) for the y coordinate of the trajectory no. 4 with Lévy stable noise with $\alpha = 1.81$ and estimated 95% confidence intervals obtained via Monte Carlo simulations under the assumption of the Lévy stable FARIMA model (red dashed lines) in double logarithmic scale. The values of the statistic lie within the confidence interval for the FARIMA model supporting its goodness of fit. The memory parameter $d_{\mathrm{MSD}} = -0.17$ was obtained by fitting the line to the values in the gray region.

V. FITTING FARIMA PARAMETERS TO THE TELOMERE DATA

We investigate here the data describing transient anomalous diffusion of telomeres in the nucleus of living human cells (U2OS cancer) presented in Ref. [11], where the diffusion properties of telomeres in a broad time range of almost six orders of magnitude by combining different imaging set-ups on the same microscope were examined.

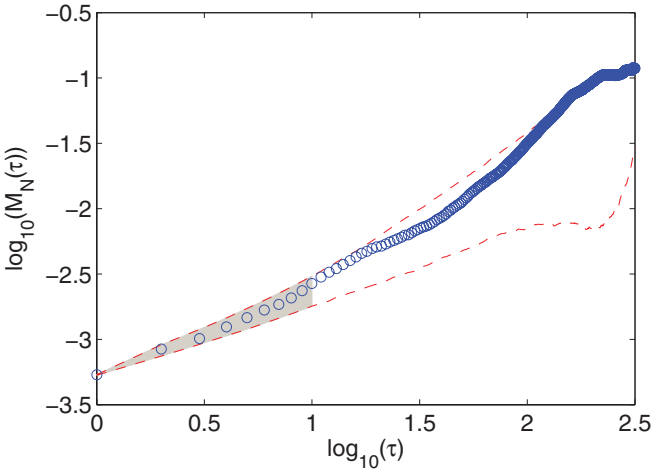

FIG. 6. (Color online) Sample MSD (blue circles) for the y coordinate of the trajectory no. 2 with unknown noise and estimated 95% confidence intervals obtained via Monte Carlo simulations under the assumption of the FARIMA model with noise generated from empirical distribution function (red dashed lines) in double logarithmic scale. The values of the statistic lie within the confidence interval for the FARIMA model supporting its goodness of fit. The memory parameter $d_{\mathrm{MSD}} = -0.16$ was obtained by fitting the line to the values in the gray region.

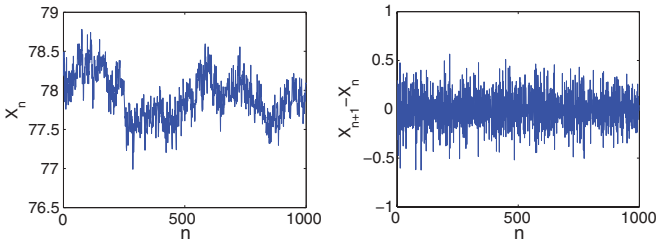

FIG. 7. (Color online) A plot of the x coordinate of telomere data (left panel) and its increments (right panel). Trajectory no. 0707 with Gaussian noise.

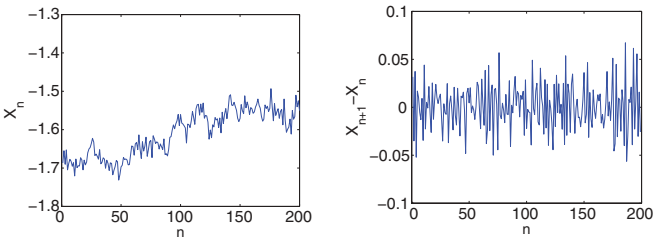

FIG. 9. (Color online) A plot of the x coordinate of telomere data (left panel) and its increments (right panel). Trajectory no. 47 with Gaussian noise.

We concentrate on the data, which clearly follow the subdiffusive character and consist of 180 trajectories of length 1000 observed over time range 10^{-2}–10 [s] (CCD) and 2D confocal data set of 94 trajectories of length 200 measured over 1–10^2 [s] (2D confocal) [11]. All of them follow the Gaussian distribution. We study in detail four different trajectories: no. 0707x and no. 0707y from the CCD data set as well as no. 47y and no. 3y from the 2D confocal data set. The data and their increments are presented in Figs. 7–10. Since the data are Gaussian, we estimated FARIMA parameters with the standard ITSM package, which was proposed for light-tailed distributions [28]. The estimated parameters along with MSD estimates (d_{MSD}) are presented in Table II (for the output windows of ITSM see the supplemental material [14]). We also compared them with the estimates calculated by the estimation procedure given by Eq. (5). It appears that the results are similar. Hence, we choose the values depicted in Table II as the FARIMA model parameters.

The goodness of fit of FARIMA processes is illustrated in Figs. 11–14. We plotted the sample MSD function for four analyzed trajectories along with the 95% confidence intervals calculated under the assumption of the FARIMA model. We can see the empirical trajectiories lie within the bounds except for several values, which indicates the model is well fitted. We also note that the $M_N(\tau)$ statistic is reliable only for large N/τ [12]. In all cases we set similar minimal values for N/τ.

VI. DISCUSSION

Recently, due to the advance of single-particle tracking experimental techniques, the diffusive behaviors of various intracellular biomolecules have been identified. In general, stochastic motions of in vivo biomolecules exhibit very complex subdiffusive behaviors. Several anomalous diffusion models such as CTRW, OD, FBM, FLSM, and FLE were proposed to explain those observed behaviors. It seems that stochastic properties of the intracellular diffusions are beyond that of the proposed diffusion models. Underlying stochastic property is a theoretically challenging and timely problem.

Five extensively studied in the literature models of subdiffusion dynamics can be divided into two categories: short memory (CTRW, OD) and long memory, namely, with power-law dependence (FBM, FLSM, and FLE). We additionally proposed here a FARIMA model, which extends the latter category. First, from the physical point of view, it is known that FARIMA is a discrete-time analog of FLE that takes into account the memory parameter d [12]. It also includes other popular models of subdiffusive dynamics as FBM and FLSM, which are limiting cases of aggregated FARIMA. In contrast to FBM and FLSM, it can also describe different light- and heavy-tailed noises and a short (exponential) memory. The complete understanding of the differences between these types of fractional motion is to date somewhat fragmentary. Nevertheless, FLE can be distinguished from FBM since its displacement correlation has a positive value at short times [8]. We would like to point out that the FARIMA modeling is a technique of the analysis of the time series. The FARIMA model as a such does not take into account any microscopic dynamics.

We proposed the FARIMA model for the description of anomalous diffusions of mRNA and telomeres inside cells as a unified model. In both cases, the experimentally obtained time series were successfully fitted by the FARIMA$(1,d,1)$ model with the memory parameter $d < 0$. The meaning of estimated values of ϕ_1, ψ_1, d, and d_{MSD} presented in Tables I and II is as follows. AR(p) is a causal or future-independent function of noise and stands for the regression. In the case of $p = 1$ the explanatory variable is the observation immediately

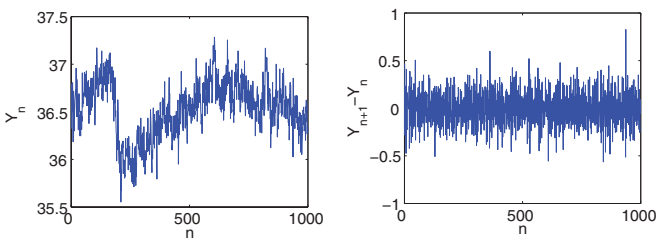

FIG. 8. (Color online) A plot of the y coordinate of telomere data (left panel) and its increments (right panel). Trajectory no. 0707 with Gaussian noise.

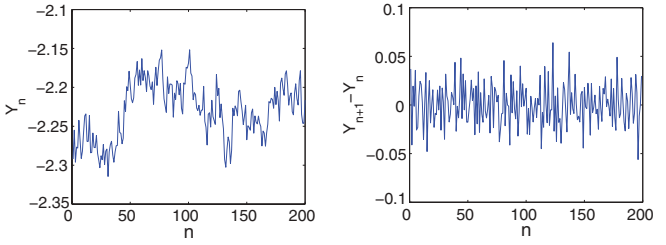

FIG. 10. (Color online) A plot of the y coordinate of telomere data (left panel) and its increments (right panel). Trajectory no. 3 with Gaussian noise.

TABLE II. Results of the ITSM estimation procedure for four sample trajectories of the telomere data.

Traj.	Model	ϕ_1	ψ_1	α	d_{MSD}
0707x	FARIMA(0, −0.36,1)	0	0.30	2	−0.39
0707y	FARIMA(0, −0.36,1)	0	0.26	2	−0.37
47x	FARIMA(1, −0.31,0)	−0.28	0	2	−0.29
3y	FARIMA(0, −0.36,0)	0	0	2	−0.27

prior to our current observation. The dependence is fully controlled by the parameter ϕ_1. On the other hand, in the finite variance case, if $X(t)$ is a stationary q-correlated time series, i.e., $X(s)$ and $X(t)$ are independent whenever $|t-s|>q$ (in contrast an i.i.d. sequence is zero dependent), then it can be represented as the MA(q) process [28]. In particular, if $q=1$, then the dependence is only one lag long, and its intensity is fully controlled by the parameter ψ_1. Hence, ARMA models introduce short memory of the process. The class of linear time series models, which includes ARMA, provides a general framework for studying stationary processes. We concentrate on stationary processes because only for this class do there exist simple prediction algorithms to describe future behavior of the dynamics. Finally, the FARIMA model generalizes the ARMA process to allow for long-term (power-law) dependencies defined by the memory parameter d. In this paper d is estimated by Eq. (5) in Table I and by the ITSM package in Table II. This is a consequence of the fact that the mRNA data follow also non-Gaussian laws. Moreover, as a double check of our statistical procedure, we estimated the memory parameter by a natural and simple estimator d_{MSD} based on the sample mean-squared displacement methodology [12].

It appears that mRNA and telomore motions have different stochastic properties depending on the trajectories.

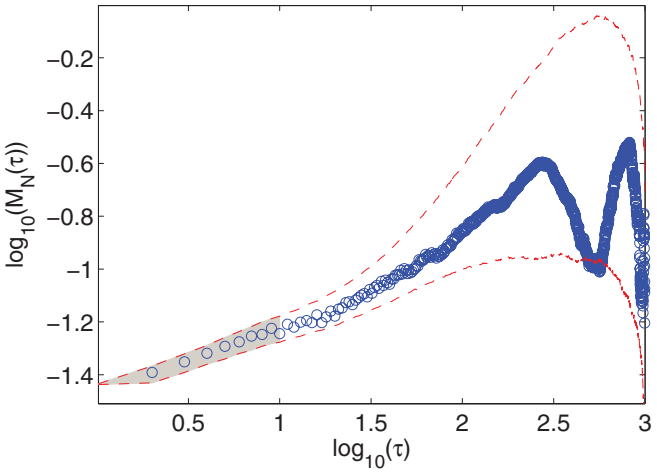

FIG. 11. (Color online) Sample MSD (blue circles) for the x coordinate of the trajectory no. 0707 from telomere data with Gaussian noise and estimated 95% confidence intervals obtained via Monte Carlo simulations under the assumption of the Gaussian FARIMA model (red dashed lines) in double logarithmic scale. The values of the statistic lie within the confidence interval for the FARIMA model supporting its goodness of fit. The memory parameter $d_{\mathrm{MSD}}=-0.39$ was obtained by fitting the line to the values in the gray region.

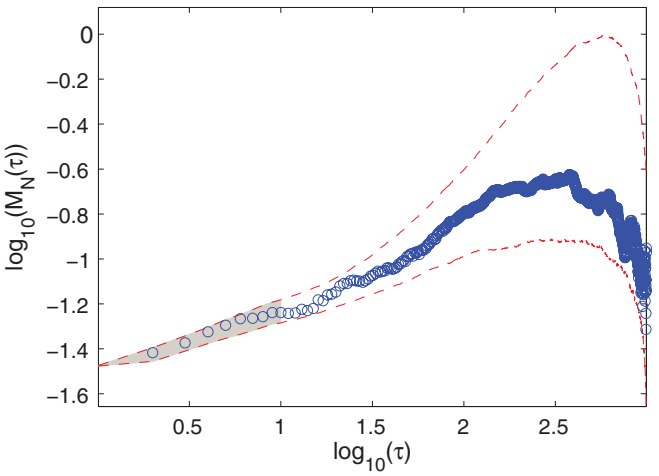

FIG. 12. (Color online) Sample MSD (blue circles) for the y coordinate of the trajectory no. 0707 from telomere data with Gaussian noise and estimated 95% confidence intervals obtained via Monte Carlo simulations under the assumption of the Gaussian FARIMA model (red dashed lines) in double logarithmic scale. The values of the statistic lie within the confidence interval for the FARIMA model supporting its goodness of fit. The memory parameter $d_{\mathrm{MSD}}=-0.37$ was obtained by fitting the line to the values in the gray region.

Our approach is based on a thorough statistical analysis of single-particle tracking experiments and is not a mean field approximation. We believe that cytoplasm forming a crowded fluid (complex system) does not always have to follow the same distribution with fixed parameters. For example, in the telomere data, PDFs calculated for all sample trajectories resemble a Gaussian PDF, but their standard deviation differs for different trajectories and time regimes. On the other hand, for the mRNA data consisting of 27 2D trajectories

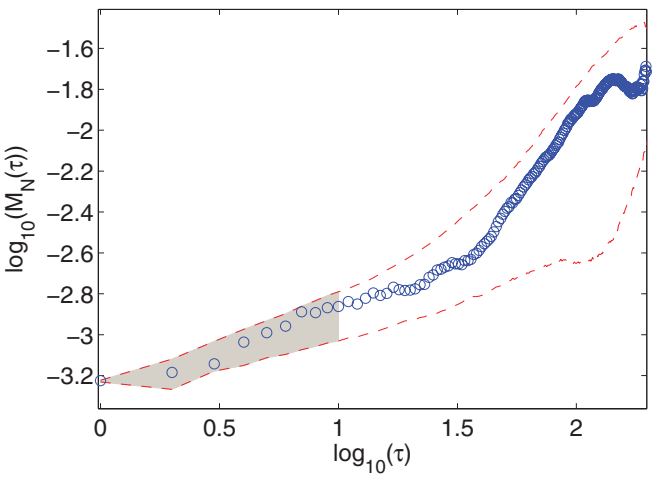

FIG. 13. (Color online) Sample MSD (blue circles) for the x coordinate of the trajectory no. 47 from telomere data with Gaussian noise and estimated 95% confidence intervals obtained via Monte Carlo simulations under the assumption of the Gaussian FARIMA model (red dashed lines) in double logarithmic scale. The values of the statistic lie within the confidence interval for the FARIMA model supporting its goodness of fit. The memory parameter $d_{\mathrm{MSD}}=-0.29$ was obtained by fitting the line to the values in the gray region.

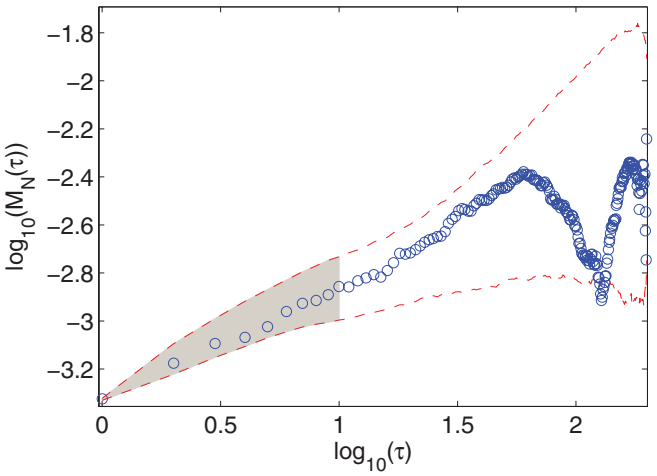

FIG. 14. (Color online) Sample MSD (blue circles) for the y coordinate of the trajectory no. 3 from telomere data with Gaussian noise and estimated 95% confidence intervals obtained via Monte Carlo simulations under the assumption of the Gaussian FARIMA model(red dashed lines) in double logarithmic scale. The values of the statistic lie within the confidence interval for the FARIMA model supporting its goodness of fit. The memory parameter $d_{\mathrm{MSD}} = -0.27$ was obtained by fitting the line to the values in the gray region.

we identified four Gaussian and nine α-stable trajectories. Statistical tests show that the others are neither Gaussian nor α-stable with $\alpha < 2$.

VII. CONCLUSIONS

In this paper we demonstrated that one can apply FARIMA$(1,d,1)$ methodology to long experimental data. Moreover, we found that only a negative memory parameter d of the FARIMA model completely determines subdiffusive dynamics of the experimental data.

For this we extended FARIMA modeling and estimation to the case of a negative memory parameter d which is tailored for subdiffusion. In comparison to earlier intensive studies [10–12,20,22–24], this allows us to propose here a new unified methodology to detect a type of subdiffusion in crowded fluids.

We investigated the applicability of the FARIMA methodology for experimental mRNA data; see Table I. The goodness of fit of the FARIMA model with the parameters estimated by Eq. (5) was justified by plots of the 95% confidence intervals; see Figs. 4–6.

Next, we also observed a similar effect for more recent data describing transient anomalous diffusion of telomeres in the nucleus of living human cells presented in Ref. [11]. Here we analyzed four randomly chosen trajectories from two categories of data: the CCD data and the 2D confocal data describing the motion of individual trajectories of fluorescently labeled telomeres (U2OS osteosarcoma cellline). To this end we employed the ITSM estimation procedure, which works well for the light-tailed distributions; see Table II. The goodness of fit of the FARIMA model was checked by plots of the 95% confidence intervals; see Figs. 11–14.

Recently, in Ref. [16] we analyzed 2D sample path extracted by image segmentation method from an original mRNA data. We examined x and y coordinates as well as the 2D trajectories. Sample MSD presented there showed clearly the subdiffusive dynamics but with slightly different parameters. A similar situation has been also observed for telomere data. Hence, we studied here only x and y coordinates of the original 2D trajectories.

For the FARIMA model there exists a rich collection of efficient computer packages, e.g., ITSM [28], which allows one to easily analyze the statistical behavior of FARIMA(p,d,q) processes; see the supplemental material [14]. We focus on stationary FARIMA processes since for this class of fractional models there exist simple prediction algorithms to describe future behavior of the dynamics. We strongly believe that our approach provides a simple unified way, missing up to now, to look deeper into processes leading to anomalous diffusion in many single-particle tracking experiments.

ACKNOWLEDGMENTS

The authors would like to thank Ido Golding for providing mRNA data and Eldad Kesten for providing telomere data. G.S. would like to acknowledge a partial support of the Fellowship co-financed by European Union within the European Social Fund.

[1] M. Weiss, M. Elsner, F. Kartberg, and T. Nilsson, Biophys. J. **87**, 3518 (2004).
[2] I. M. Tolić-Norrelykke, E.-L. Munteanu, G. Thon, L. Oddershede, and K. Berg-Sorensen, Phys. Rev. Lett. **93**, 078102 (2004).
[3] I. Golding and E. C. Cox, Phys. Rev. Lett. **96**, 098102 (2006).
[4] G. Guigas, C. Kalla, and M. Weiss, Biophys. J. **93**, 316 (2007).
[5] W. Pan, L. Filobelo, N. D. Q. Pham, O. Galkin, V. V. Uzunova, and P. G. Vekilov, Phys. Rev. Lett. **102**, 058101 (2009).
[6] J. Szymanski and M. Weiss, Phys. Rev. Lett. **103**, 038102 (2009).
[7] J.-H. Jeon, V. Tejedor, S. Burov, E. Barkai, C. Selhuber-Unkel, K. Berg-Sørensen, L. Oddershede, and R. Metzler, Phys. Rev. Lett. **106**, 048103 (2011).
[8] J.-H. Jeon and R. Metzler, Phys. Rev. E **81**, 021103 (2010).
[9] C. C. Fritsch and J. Langowski, J. Chem. Phys. **137**, 064114 (2012).
[10] M. Hellmann, J. Klafter, D. W. Heermann, and M. Weiss, J. Phys. Condens. Matter **23**, 234113 (2011).
[11] E. Kepten, I. Bronshtein, and Y. Garini, Phys. Rev. E **83**, 041919 (2011).
[12] K. Burnecki and A. Weron, Phys. Rev. E **82**, 021130 (2010).
[13] I. Bronstein, Y. Israel, E. Kepten, S. Mai, Y. Shav-Tal, E. Barkai, and Y. Garini, Phys Rev. Lett. **103**, 018102 (2009).
[14] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevE.86.041912 for details of the process and results.
[15] G. Samorodnitsky and M. S. Taqqu, *Stable Non-Gaussian Random Processes* (Chapman & Hall, New York, 1994).

[16] K. Burnecki, M. Muszkieta, G. Sikora, and A. Weron, Europhys. Lett. **98**, 1004 (2012).
[17] K. Burnecki, J. Gajda, and G. Sikora, Phys. A **390**, 3136 (2011).
[18] A. Stanislavsky, K. Burnecki, M. Magdziarz, A. Weron, and K. Weron, Astrophys. J. **693**, 1877 (2009).
[19] M. Magdziarz and A. Weron, Studia Math. **181**, 47 (2007).
[20] M. Magdziarz, A. Weron, K. Burnecki, and J. Klafter, Phys. Rev. Lett. **103**, 180602 (2009).
[21] M. Magdziarz and J. Klafter, Phys. Rev. E **82**, 011129 (2010).
[22] A. Lubelski, I. M. Sokolov, and J. Klafter, Phys. Rev. Lett. **100**, 250602 (2008).
[23] Y. He, S. Burov, R. Metzler, and E. Barkai, Phys. Rev. Lett. **101**, 058101 (2008).
[24] T. Neusius, I. M. Sokolov, and J. C. Smith, Phys. Rev. E **80**, 011109 (2009).
[25] C. W. J. Granger and R. Joyeux, J. Time Series Anal. **1**, 15 (1980).
[26] J. Beran, *Statistics for Long-Memory Processes* (Chapman & Hall, New York, 1994).
[27] M. S. Taqqu and V. Teverovsky, in *A Practical Guide To Heavy Tails: Statistical Techniques and Applications*, edited by R. J. Adler *et al*. (Birkhäuser, Boston, 1998).
[28] P. J. Brockwell and R. A. Davis, *Introduction to Time Series and Forecasting* (Springer-Verlag, New York, 2002); *ITSM for Windows: A User's Guide to Time Series Modelling and Forecasting* (Springer-Verlag, New York, 1994).
[29] A. Weron, K. Burnecki, Sz. Mercik, and K. Weron, Phys. Rev. E **71**, 016113 (2005).
[30] B. Podobnik, P. C. Ivanov, K. Biljakovic, D. Horvatic, H. E. Stanley, and I. Grosse, Phys. Rev. E **72**, 026121 (2005).
[31] K. Burnecki, J. Klafter, M. Magdziarz, and A. Weron, Phys. A **387**, 1077 (2008).
[32] B. Podobnik, I. Grosse, D. Horvatic, S. Ilic, P. Ch. Ivanov, and H. E. Stanley, Eur. Phys. J. B **71**, 243 (2009).
[33] K. Burnecki, J. Stat. Mech. (2012) P05015.
[34] P. Kokoszka and M. S. Taqqu, Stoch. Proc. Appl. **60**, 19 (1995).
[35] S. Stoev and M. S. Taqqu, Fractals **12**, 95 (2004).
[36] A. Weron and R. Weron, Lect. Notes Phys. **457**, 379 (1995).
[37] K. Burnecki, A. Wyłomańska, A. Beletskii, V. Gonchar, and A. Chechkin, Phys. Rev. E **85**, 056711 (2012).
[38] P. Kokoszka and M. S. Taqqu, Ann. Statist. **24**, 1880 (1996).
[39] K. Burnecki and G. Sikora (unpublished).
[40] B. Dybiec and E. Gudowska-Nowak, Phys. Rev. E **80**, 061122 (2009).