

PAPER • OPEN ACCESS

# Empirical anomaly measure for finite-variance processes


To cite this article: Katarzyna Maraj *et al* 2021 *J. Phys. A: Math. Theor.* **54** 024001

View the [article online](#) for updates and enhancements.

## You may also like

- [A grand canonical genetic algorithm for the prediction of multi-component phase diagrams and testing of empirical potentials](#)  
William W Tipton and Richard G Hennig
- [A ternary EAM interatomic potential for U–Mo alloys with xenon](#)  
D E Smirnova, A Yu Kuksin, S V Starikov et al.
- [Ab initio local energy and local stress: application to tilt and twist grain boundaries in Cu and Al](#)  
Hao Wang, Masanori Kohyama, Shingo Tanaka et al.

# Empirical anomaly measure for finite-variance processes

Katarzyna Maraj, Dawid Szarek, Grzegorz Sikora and  
Agnieszka Wyłomańska\* 

Faculty of Pure and Applied Mathematics, Hugo Steinhaus Center, Wrocław  
University of Science and Technology, Wyspińskiego 27, 50-370 Wrocław, Poland

E-mail: [agnieszka.wylomanska@pwr.edu.pl](mailto:agnieszka.wylomanska@pwr.edu.pl)

Received 2 September 2020, revised 9 November 2020

Accepted for publication 20 November 2020

Published 15 December 2020



CrossMark

## Abstract

Anomalous diffusion phenomena are observed in many areas of interest. They manifest themselves in deviations from the laws of Brownian motion (BM), e.g. in the non-linear growth (mostly power-law) in time of the ensemble average mean squared displacement (MSD). When we analyze the real-life data in the context of anomalous diffusion, the primary problem is the proper identification of the type of the anomaly. In this paper, we introduce a new statistic, called empirical anomaly measure (EAM), that can be useful for this purpose. This statistic is the sum of the off-diagonal elements of the sample autocovariance matrix for the increments process. On the other hand, it can be represented as the convolution of the empirical autocovariance function with time lags. The idea of the EAM is intuitive. It measures dependence between the ensemble-averaged MSD of a given process from the ensemble-averaged MSD of the classical BM. Thus, it can be used to measure the distance between the anomalous diffusion process and normal diffusion. In this article, we prove the main probabilistic characteristics of the EAM statistic and construct the formal test for the recognition of the anomaly type. The advantage of the EAM is the fact that it can be applied to any data trajectories without the model specification. The only assumption is the stationarity of the increments process. The complementary summary of the paper constitutes of Monte Carlo simulations illustrating the effectiveness of the proposed test and properties of EAM for selected processes.

\* Author to whom any correspondence should be addressed.



Original content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](https://creativecommons.org/licenses/by/4.0/). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

Keywords: anomalous diffusion, anomaly type testing, Monte Carlo simulations, fractional Brownian motion

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

## 1. Introduction

Anomalous diffusion phenomena are observed in many areas of interest. They manifest themselves in deviations from the laws of Brownian motion (BM), e.g. in the non-linear growth (mostly power-law) in time of the ensemble average mean squared displacement (MSD), namely  $E[X^2(t)] \sim t^\alpha$ , where  $\alpha$  is called the anomalous diffusion exponent. Depending on the  $\alpha$  parameter one can distinguish between sub-diffusive ( $\alpha < 1$ ), normal ( $\alpha = 1$ ), and super-diffusive ( $\alpha > 1$ ) behavior. Although the anomalous diffusion is mostly considered in the means of ensemble average MSD, this phenomena influences also the probability density function (PDF) of the corresponding process. The classical diffusion is composed by both linear growth of the second moment of the process as well as its Gaussian distribution. Any deviations from these properties are related to the anomalous diffusion, see for instance [1–4].

The classical anomalous diffusion models are fractional Brownian motion (FBM) [5, 6], Lévy stable motion [7] and continuous-time random walk [8, 9]. To the class of anomalous diffusion processes, we include also the subordinated processes (called also time-changed processes) [10–13]. The most popular is the time-changed BM driven by the so-called inverse to the strictly increasing Lévy stable subordinator [14–17] for which the PDF is described by the fractional diffusion equation [8]. The family of anomalous diffusion models contains also the processes with time- or position-dependent diffusion coefficients such as scaled BM [18, 19] or heterogeneous diffusion models [20]. It is worth to mention also the superstatistical process where the diffusion coefficient is a random variable [21] or diffusive diffusivity model, called also Brownian yet non-Gaussian diffusion process, where the diffusion coefficient is described by another process (like the Ornstein–Uhlenbeck one) [22]. In the literature, one can also find the processes with resetting [18, 23] that also belong to the large class of the anomalous diffusion models, see also [24–27] and references therein.

The anomalous diffusion processes have found many practical applications, including physical phenomenon [28, 29], finance [30, 31], ecology [32], hydrology [33], biology [34] as well as meteorology and geophysics [6, 35].

When we analyze the real-life data in the context of anomalous diffusion, the primary problem is the proper identification of the type of anomaly. Through the identification of the type of anomaly, we understand the recognition between sub- or super-diffusive regime. In the statistical and application-oriented literature, one can find various methods used in this analysis. One of the most classical approaches is based on the statistics calculated for real-life data. The known approaches applied in the problem of anomalous diffusion behavior recognition and parametrization are based on simple statistics that exhibit specific behavior for different anomaly types (i.e. for sub- and super-diffusive processes). For instance, the time-averaged MSD is one of the classical statistics used to the anomalous diffusive behavior identification as well as the base for the estimation of the anomalous diffusive exponent [36–42]. The detrended fluctuation analysis (DFA)-based algorithms are popular for the detection of long-range dependence (strongly related to the anomalous diffusion) in nonstationary time series, [43–48]. Moreover, the detrended moving-average (DMA)-based methods were used for the so-called scaling exponent analysis [49–53] and to test FBM identification based on real-life

data [54]. It should be mentioned, the literature devoted to the statistical tools for anomalous diffusion processes analysis is very rich, see for instance the papers [55–62].

In the case of second-order processes (i.e. finite-variance) the universal and simplest statistics, that can be applied in the considered problem, is the sample autocovariance function (ACVF). For the zero-mean Gaussian processes the ACVF fully characterizes the distribution of the given model. Moreover, the ACVF exhibits different behavior for processes with different anomaly types thus it is a natural candidate for the analysis of the second-order anomalous diffusion models. The sample ACVF is the classical tool for the detection of the long-range dependence strictly related to the anomalous diffusion behavior [63, 64] as well as for testing the FBM [65]. However, most of the methods mentioned above based on the ACVF, utilize this statistic in one specific time lag. For instance, the test for FBM proposed by the authors in [65] requires selecting the specific time lag and the results are based on the value of the statistic only at this point. This approach seems to be effective however much of the information of the model included in ACVF for all arguments is not used.

In this paper, we take a step forward and propose to use the sample ACVF for all available time points (determined by the available trajectory length). We introduce the new statistic, called the empirical anomaly measure (EAM), which is the sum of the off-diagonal elements of the sample autocovariance matrix for the increments process. On the other hand, the EAM can be represented as the convolution of the empirical ACVF with time lags. The idea of the EAM is intuitive. It measures difference between the ensemble-averaged MSD of a given process from ensemble-averaged MSD of the classical BM. Thus, it can be used to measure the distance between the anomalous and normal diffusion. In this article, we prove the main probabilistic characteristics of the EAM statistic and construct the test for the recognition of the anomaly type. The advantage of the proposed approach is the fact that it can be applied to any real-life data without the model specification. The complementary summary of the paper constitutes of Monte Carlo simulations illustrating the effectiveness of the introduced test and properties of EAM for the exemplary process, namely the FBM.

The paper is organized as follows: in section 2 we introduce the idea of the EAM statistic and prove its main probabilistic properties for the general class of second-order processes. Next, in section 3 we examine the FBM and the behavior of EAM for this process. We indicate the different behavior of the considered statistic for different anomaly types. This section is the starting point for the introduction of the test statistic for the anomalous diffusion behavior recognition. In section 4 we describe the testing procedure and by the Monte Carlo simulations, we check its effectiveness for the FBM. The last section concludes the paper and gives a general overview of future research.

## 2. Empirical anomaly measure

Let us consider the zero-mean second-order process  $\{X(n)\} = \{X(n), n = 0, 1, \dots\}$ , starting at zero with the ACVF  $\{\gamma_X(i, j) = \mathbb{E}[X(i)X(j)], i, j = 0, 1, \dots\}$ . Moreover, we consider also the corresponding increments process  $\{Y(n)\} = \{Y(n), n = 0, 1, \dots\}$  defined as

$$Y(n) = X(n+1) - X(n), \quad n = 0, 1, \dots \quad (1)$$

with the ACVF  $\{\gamma_Y(i, j) = \mathbb{E}[Y(i)Y(j)], i, j = 0, 1, \dots\}$ . The crucial quantity reflecting the dynamics of the process  $\{X(n)\}$  is the ensemble average MSD that is defined as follows for any  $n$

$$\text{MSD}_X(\tau) = \mathbb{E}[(X(n+\tau) - X(n))^2], \quad \tau = 0, 1, \dots \quad (2)$$

The MSD measures an average squared displacement of the process over a time period  $\tau$ . The displacement  $X(n + \tau) - X(n)$  consists of  $\tau$  unit-time increments, i.e.

$$X(n + \tau) - X(n) = \sum_{i=n}^{n+\tau-1} Y(i). \quad (3)$$

Therefore, one can obtain

$$\text{MSD}_X(\tau) = \mathbb{E}[(X(n + \tau) - X(n))^2] = \mathbb{E} \left[ \left( \sum_{i=n}^{n+\tau-1} Y(i) \right)^2 \right] = \sum_{i,j=n}^{n+\tau-1} \gamma_Y(i, j), \quad (4)$$

which simply shows that the MSD is fully determined by the whole dependence structure of the increments process. Precisely, it is a sum of all elements of covariance matrix  $\{\gamma_Y(i, j), i, j = n, n + 1, \dots, n + \tau - 1\}$  of the increments process  $\{Y(n)\}$ . The special role in this covariance matrix plays the main diagonal  $\{\gamma_Y(i, i), i = n, n + 1, \dots, n + \tau - 1\}$  with the constant entry—variance of unit-time increment. The equation (4) can be rewritten as

$$\mathbb{E}[(X(n + \tau) - X(n))^2] = \sum_{i=n}^{n+\tau-1} \gamma_Y(i, i) + 2 \sum_{n \leq i < j \leq n+\tau-1} \gamma_Y(i, j). \quad (5)$$

In the particular case, when the process  $\{X(n)\}$  has stationary increments, and therefore  $\gamma_Y(i, j) = \gamma_Y(i - j)$ , the equation (5) can be rewritten as

$$\mathbb{E}[X^2(\tau)] = \tau \gamma_Y(0) + 2 \sum_{i=1}^{\tau-1} (\tau - i) \gamma_Y(i). \quad (6)$$

Formulas equations (4)–(6) were well studied in the physical literature in terms of particles diffusion and velocity correlation [66–68]. However in this paper, we concentrate more on statistical properties and applicability potential of quantities in equations (4)–(6). From equation (6) one can conclude that the main diagonal of increments covariance matrix builds inside MSD the linear function of time period  $\tau$ . When the ACVFs  $\gamma_Y(\cdot)$  are summable, the dominated convergence theorem yields [69]

$$\lim_{\tau \rightarrow \infty} \frac{\mathbb{E}[X^2(\tau)]}{\tau} = \gamma_Y(0) + 2 \sum_{i=1}^{\infty} \gamma_Y(i). \quad (7)$$

Thus, the ensemble average MSD of  $\{X(n)\}$  decays linearly fast. Once, the ACVFs  $\gamma_Y(\cdot)$  stop being summable (super-diffusion), the ensemble average MSD of  $\{X(n)\}$  can grow faster than linearly and the actual rate of increase of  $\mathbb{E}[X^2(\tau)]$  is related to the rate of decay of the ACVF  $\gamma_Y(\cdot)$ . When the ACVF is summable (the sub-diffusive case), the ensemble average MSD of  $\{X(n)\}$  grows slower than linear function. The rate of increase of  $\mathbb{E}[X^2(\tau)]$  depends on the parameter which characterizes the asymptotic behavior of the ACVF of the increment process. In the case of the anomalous diffusion models, this parameter is equal to the anomalous diffusion exponent  $\alpha$ .

The second summand in equation (6)—the sum of all off-diagonal entries of increments covariance matrix—gives the information about the deviation of MSD  $\mathbb{E}[X^2(\tau)]$  from the linear function of time  $\tau \gamma_Y(0)$ . Thus, the following quantity can be considered as a diffusion

anomaly measure (AM) of the process  $\{X(n)\}$

$$\text{AM}_X(\tau) := 2 \sum_{i=1}^{\tau-1} (\tau - i) \gamma_Y(i) = \mathbb{E}[(X(\tau))^2] - \tau \gamma_Y(0). \quad (8)$$

Now, let us assume that we possess a finite trajectory  $\{X(n), n = 1, 2, \dots, N\}$  of the process  $\{X(n)\}$  and the corresponding trajectory of the increments  $\{Y(n), n = 1, \dots, N - 1\}$ . We assume the process  $\{X(n)\}$  has stationary increments. The natural candidate for the estimator of anomaly measure is the statistic that we call the EAM defined as follows

$$\widehat{\text{AM}}_X(\tau) = 2 \sum_{i=1}^{\tau-1} (\tau - i) \hat{\gamma}_Y(i), \quad (9)$$

for any  $\tau = 1, 2, \dots, N - 1$ . In the above equation  $\hat{\gamma}_Y(i)$  is the empirical ACVF of  $\{Y(n), n = 1, \dots, N - 1\}$  defined as follows

$$\hat{\gamma}_Y(i) = \frac{1}{N-i} \sum_{j=1}^{N-i} Y(j)Y(j+i), \quad i = 1, 2, \dots, N - 1.$$

In the following lemma, we will show that the EAM statistic defined in (9) is the unbiased estimator of the anomaly measure defined in (8).

**Lemma 2.1.** *Let us consider the finite trajectory  $\{X(n), n = 1, 2, \dots, N\}$  of the zero-mean second-order process  $\{X(n)\}$  with stationary increments. In that case, the corresponding EAM defined in (9) is the unbiased estimator of the anomaly measure given in (8).*

**Proof.** One can show that for the trajectory of the increments  $\{Y(n), n = 1, \dots, N - 1\}$  the following holds

$$\begin{aligned} \mathbb{E}[\widehat{\text{AM}}_X(\tau)] &= 2 \sum_{i=1}^{\tau-1} \frac{\tau-i}{N-i} \sum_{j=1}^{N-i} \mathbb{E}[Y(j)Y(j+i)] = 2 \sum_{i=1}^{\tau-1} \frac{\tau-i}{N-i} \sum_{j=1}^{N-i} \gamma_Y(i) \\ &= 2 \sum_{i=1}^{\tau-1} (\tau-i) \gamma_Y(i) = \text{AM}_X(\tau). \end{aligned}$$

That indicates the EAM ( $\widehat{\text{AM}}_X$  statistic defined in equation (9)) is the unbiased estimator of the anomaly measure given in equation (8).  $\square$

From lemma 2.1, formula (8) and the fact that  $\gamma_Y(0) = E[X(1)^2]$  is constant, one can see that the expected value of EAM is negative and decreasing with respect to  $\tau$  for the sub-diffusive processes while it is positive and increasing function for the super-diffusive case.

In the following lemma, we prove the formula for the variance of the EAM for the general zero-mean Gaussian process with stationary increments.

**Lemma 2.2.** *Let us consider the finite trajectory  $\{X(n), n = 1, 2, \dots, N\}$  of the zero-mean Gaussian process  $\{X(n)\}$  with the stationary increments. The variance of the corresponding EAM defined in equation (9) is given by*

$$\begin{aligned} \text{Var} [\widehat{\text{AM}}_X(\tau)] &= 4 \sum_{i=1}^{\tau-1} \sum_{k=1}^{\tau-1} \frac{(\tau-i)(\tau-k)}{(N-i)(N-k)} \sum_{j=1}^{N-i} \sum_{l=1}^{N-k} (\gamma_Y(j-l)\gamma_Y(j+i-l-k) \\ &\quad + \gamma_Y(j-l-k)\gamma_Y(j+i-l)), \end{aligned} \quad (10)$$

where  $\{Y(n), n = 1, \dots, N-1\}$  are increments.

**Proof.** To calculate the variance of the statistic defined in equation (9), first the second moment is calculated

$$\begin{aligned} \mathbb{E} [\widehat{\text{AM}}_X(\tau)^2] &= \mathbb{E} \left[ 4 \left( \sum_{i=1}^{\tau-1} \frac{\tau-i}{N-i} \sum_{j=1}^{N-i} Y_j Y_{j+i} \right) \left( \sum_{k=1}^{\tau-1} \frac{\tau-k}{N-k} \sum_{l=1}^{N-k} Y_l Y_{l+k} \right) \right] \\ &= 4 \sum_{i=1}^{\tau-1} \sum_{k=1}^{\tau-1} \frac{(\tau-i)(\tau-k)}{(N-i)(N-k)} \sum_{j=1}^{N-i} \sum_{l=1}^{N-k} \mathbb{E} [Y_j Y_{j+i} Y_l Y_{l+k}]. \end{aligned} \quad (11)$$

Let us highlight, the above formula (11) is true for any second-order process, without the assumption of the Gaussianity. Let us note, according to the Isserlis' theorem [70], we have that, if  $\{Z(1), \dots, Z(n)\}$  is zero-mean multivariate Gaussian random vector, then the following holds

$$\mathbb{E} \left[ \prod_{i=1}^n Z(i) \right] = \sum_{p \in P_n^2} \prod_{\{i,j\} \in p} \mathbb{E}[Z(i)Z(j)],$$

where  $P_n^2$  is a set of all the distinct pairings of  $\{1, 2, \dots, n\}$ . Using Isserlis' theorem, we get the following

$$\begin{aligned} \mathbb{E} [\widehat{\text{AM}}_X(\tau)^2] &= 4 \sum_{i=1}^{\tau-1} \sum_{k=1}^{\tau-1} \frac{(\tau-i)(\tau-k)}{(N-i)(N-k)} \sum_{j=1}^{N-i} \sum_{l=1}^{N-k} (\mathbb{E}[Y(j)Y(j+i)] \\ &\quad \times \mathbb{E}[Y(l)Y(l+k)] + \mathbb{E}[Y(j)Y(l)]\mathbb{E}[Y_{j+i}Y_{l+k}] \\ &\quad + \mathbb{E}[Y_j Y_{l+k}]\mathbb{E}[Y_{j+i}Y_l]) \\ &= 4 \sum_{i=1}^{\tau-1} \sum_{k=1}^{\tau-1} \frac{(\tau-i)(\tau-k)}{(N-i)(N-k)} \sum_{j=1}^{N-i} \sum_{l=1}^{N-k} (\gamma_Y(i)\gamma_Y(k) \\ &\quad + \gamma_Y(j-l)\gamma_Y(j+i-l-k) + \gamma_Y(j-l-k)\gamma_Y(j+i-l)) \\ &= \left[ 2 \sum_{i=1}^{\tau-1} (\tau-i)\gamma_Y(i) \right]^2 + 4 \sum_{i=1}^{\tau-1} \sum_{k=1}^{\tau-1} \frac{(\tau-i)(\tau-k)}{(N-i)(N-k)} \sum_{j=1}^{N-i} \sum_{l=1}^{N-k} \\ &\quad \times (\gamma_Y(j-l)\gamma_Y(j+i-l-k) + \gamma_Y(j-l-k)\gamma_Y(j+i-l)) \\ &= \mathbb{E}[\widehat{\text{AM}}_X(\tau)]^2 + 4 \sum_{i=1}^{\tau-1} \sum_{k=1}^{\tau-1} \frac{(\tau-i)(\tau-k)}{(N-i)(N-k)} \sum_{j=1}^{N-i} \sum_{l=1}^{N-k} \\ &\quad \times (\gamma_Y(j-l)\gamma_Y(j+i-l-k) + \gamma_Y(j-l-k)\gamma_Y(j+i-l)). \end{aligned}$$

Finally, we obtain that the variance of the estimator is given by formula (10).  $\square$

Lemma 2.2 can be generalized for any vector of random variables (without the assumption of multivariate Gaussian distribution). In the following Lemma, we derive the formula for the variance for EAM for any random vector for that the cumulant-generating function [71] exists.

**Lemma 2.3.** *Let us consider the finite trajectory  $\{X(n), n = 1, 2, \dots, N\}$  of the process  $\{X(n)\}$  with the stationary increments. If the cumulant-generating functions [71]*

$$K_{i,j,k,l}(t_1, t_2, t_3, t_4) = \log \mathbb{E}[e^{t_1 Y(i) + t_2 Y(j) + t_3 Y(k) + t_4 Y(l)}], \quad (i, j, k, l) \in \{1, \dots, N-1\}^4 \quad (12)$$

exist, where  $\{1, \dots, N-1\}^4$  is the 4-ary Cartesian power of set  $\{1, \dots, N-1\}$ . Then the variance of the EAM defined in equation (9) is given by

$$\begin{aligned} \text{Var} [\widehat{\text{AM}}_X(\tau)] &= 4 \sum_{i=1}^{\tau-1} \sum_{k=1}^{\tau-1} \frac{(\tau-i)(\tau-k)}{(N-i)(N-k)} \sum_{j=1}^{N-i} \sum_{l=1}^{N-k} \sum_{p \in P_4} \\ &\times \prod_{b \in p} \kappa(\{Y_m^{ijkl}\}_{m \in b}) - \text{AM}_X^2(\tau), \end{aligned} \quad (13)$$

where  $\{Y(n), n = 1, \dots, N-1\}$  is the trajectory of the corresponding increments and  $\kappa(\{Y_m^{ijkl}\}_{m \in b})$  is the joint cumulant [71] of  $\{Y_m^{ijkl}\}_{m \in b}$  (for the definition of  $Y_m^{ijkl}$  see the equation (15)) with  $b$  being the element from the set  $p$  taken from  $P_4$ —the partition of  $\{1, 2, 3, 4\}$ .

**Proof.** In the considered case the second moment of the statistic defined in equation (9) is given in equation (11). Here we will use the generalization of Isserlis' theorem [70]—the moment-cumulants formula [72], which states that if for a random vector  $\{Z(1), \dots, Z(n)\}$  the cumulant-generating function exists (12), then the following holds

$$\mathbb{E} \left[ \prod_{i=1}^n Z(i) \right] = \sum_{p \in P_n} \prod_{b \in p} \kappa(\{Z(m)\}_{m \in b}), \quad (14)$$

where  $P_n$  is a partition of  $\{1, 2, \dots, n\}$ . In order to utilize the formula from equation (14), the random variables need to have an (arbitrary but fixed) order. This way we define the bijection from the set of the random variables  $\{Y(j), Y(j+i), Y(l), Y(l+k)\}$  to the set  $\{1, 2, 3, 4\}$ , hence the partition  $P_4$  is well defined. Thus, we introduce a new notation

$$Y_1^{ijkl} := Y(j), \quad Y_2^{ijkl} := Y(j+i), \quad Y_3^{ijkl} := Y(l), \quad Y_4^{ijkl} := Y(l+k). \quad (15)$$

Then, using the formula (14), we get the following

$$\mathbb{E} [\widehat{\text{AM}}_X(\tau)^2] = 4 \sum_{i=1}^{\tau-1} \sum_{k=1}^{\tau-1} \frac{(\tau-i)(\tau-k)}{(N-i)(N-k)} \sum_{j=1}^{N-i} \sum_{l=1}^{N-k} \sum_{p \in P_4} \prod_{b \in p} \kappa(\{Y_m^{ijkl}\}_{m \in b}).$$

Therefore, taking into consideration the formula for the expected value of EAM given in lemma 2.1, we conclude that the variance of the estimator is given by equation (13).  $\square$



### 3. Empirical anomaly measure for fractional Brownian motion

The FBM  $\{X_H(t), t \geq 0\}$  with Hurst index  $H \in (0, 1)$  is a continuous and centered Gaussian process that starts at zero (almost surely) with ACVF [5, 73–75]

$$E[X_H(t)X_H(s)] = \frac{1}{2}D(t^{2H} + s^{2H} - |t - s|^{2H}), \quad \text{where } t, s \geq 0.$$

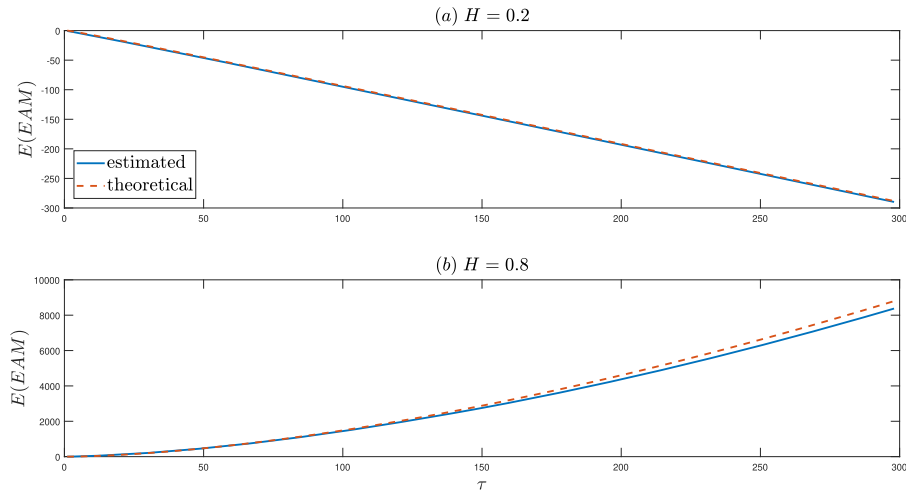
The parameter  $D$  in the above equation is called the diffusion coefficient. For given  $t \geq 0$ , the random variable  $X_H(t) \sim N(0, Dt^{2H})$ . The FBM has stationary increments and is self-similar. Moreover, it is considered as one of the classical process used to describe the anomalous diffusive phenomena. Indeed, for  $H < 1/2$  it exhibits the sub-diffusive behavior while for  $H > 1/2$ —super-diffusive one. Moreover, the FBM with  $H > 1/2$  exhibits also the so-called long-range dependence. For  $H = 1/2$ , the FBM reduces to the standard BM. Thus, for FBM the anomalous diffusive exponent is equal to  $\alpha = 2H$ .

In this paper, we consider the discrete-time FBM, i.e. the process  $\{X_H(n)\} = \{X_H(n), n = 0, 1, \dots\}$ . Through  $\{Y_H(n)\}$  we denote the increment process defined in equation (1). Thus, in the examined case, the following holds

$$\gamma_{Y_H}(i) = \frac{D}{2}(|i+1|^{2H} + |i-1|^{2H} - 2|i|^{2H}), \quad i \in \mathbb{Z}.$$

Using lemmas 2.1 and 2.2 we can calculate the expected value and the variance of the EAM statistic for FBM. Namely, we have the following

$$\begin{aligned} \mathbb{E}[\widehat{\text{AM}}_{X_H}(\tau)] &= 2 \sum_{i=1}^{\tau-1} (\tau-i) \gamma_{Y_H}(i) = D \sum_{i=1}^{\tau-1} (\tau-i) ((i+1)^{2H} + (i-1)^{2H} - 2i^{2H}) \\ &= D \left[ \sum_{i=1}^{\tau-1} (\tau-i) ((i+1)^{2H} - i^{2H}) - \sum_{i=1}^{\tau-1} (\tau-i) (i^{2H} - (i-1)^{2H}) \right] \\ &= D \left[ \sum_{i=1}^{\tau-1} (\tau-i) ((i+1)^{2H} - i^{2H}) - \sum_{i=0}^{\tau-2} (\tau-i-1) ((i+1)^{2H} - i^{2H}) \right] \\ &= D \left[ \sum_{i=1}^{\tau-1} (\tau-i) ((i+1)^{2H} - i^{2H}) - \sum_{i=0}^{\tau-2} (\tau-i) ((i+1)^{2H} - i^{2H}) \right. \\ &\quad \left. + \sum_{i=1}^{\tau-1} i^{2H} - \sum_{i=0}^{\tau-2} i^{2H} \right] \\ &= D [\tau^{2H} - (\tau-1)^{2H} - \tau + (\tau-1)^{2H}] = D (\tau^{2H} - \tau) \\ &= D \tau \left( \tau^{2(H-\frac{1}{2})} - 1 \right). \end{aligned}$$



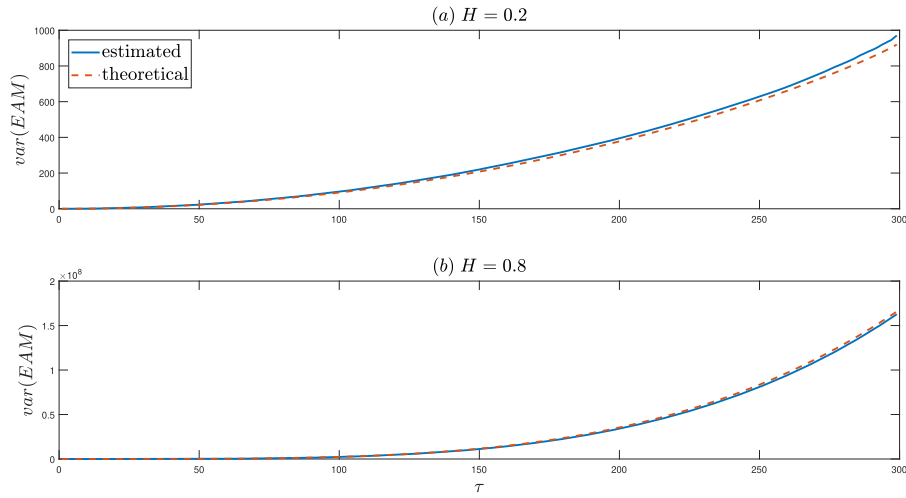
**Figure 1.** The comparison of the theoretical and empirical expected value of EAM for  $H = 0.2$  [panel (a)] and  $H = 0.8$  [panel (b)]. In order to calculate the empirical expected value we simulated 1000 trajectories of length  $N = 300$  for FBM with the corresponding  $H$  parameter.

The variance of EAM for FBM has the following form

$$\begin{aligned}
 \text{Var} \left[ \widehat{\text{AM}}_{X_H}(\tau) \right] &= 4 \sum_{i=1}^{\tau-1} \sum_{k=1}^{\tau-1} \frac{(\tau-i)(\tau-k)}{(N-i)(N-k)} \sum_{j=1}^{N-i} \sum_{l=1}^{N-k} (\gamma_{Y_H}(j-l) \gamma_{Y_H} \\
 &\quad \times (j+i-l-k) + \gamma_{Y_H}(j-l-k) \gamma_{Y_H}(j+i-l)) \\
 &= D^2 \sum_{i=1}^{\tau-1} \sum_{k=1}^{\tau-1} \frac{(\tau-i)(\tau-k)}{(N-i)(N-k)} \sum_{j=1}^{N-i} \sum_{l=1}^{N-k} \{ [|j-l+1|^{2H} \\
 &\quad + |j-l-1|^{2H} - 2|j-l|^{2H}] \\
 &\quad \times [|j+i-l-k+1|^{2H} + |j+i-l-k-1|^{2H} - 2|j+i-l-k|^{2H}] \\
 &\quad + [|j-l-k+1|^{2H} + |j-l-k-1|^{2H} - 2|j-l-k|^{2H}] \\
 &\quad \times [|j+i-l+1|^{2H} + |j+i-l-1|^{2H} - 2|j+i-l|^{2H}] \}.
 \end{aligned}$$

In figure 1 we present a comparison of the theoretical and empirical expected value of EAM for FBM. To calculate the empirical expected value we simulated 1000 trajectories of length  $N = 300$  for FBM. In panel (a) we present the comparison for the sub-diffusive case, namely for the FBM with  $H = 0.2$ . As one can see, the theoretical and empirical expected values are less than zero for all  $\tau$ s. In this case, the statistic notably decreases with increasing  $\tau$ . Also, one can observe that theoretical and empirical values coincide. In panel (b) we demonstrate the results for the super-diffusive case, namely for the FBM with  $H = 0.8$ . In this case, the statistic increases with growing  $\tau$  and is always higher than zero. The theoretical and empirical values are almost the same.

In figure 2 we present the comparison of the theoretical and empirical variance of EAM for FBM in log-log scale. To calculate the empirical variance we simulated 1000 trajectories

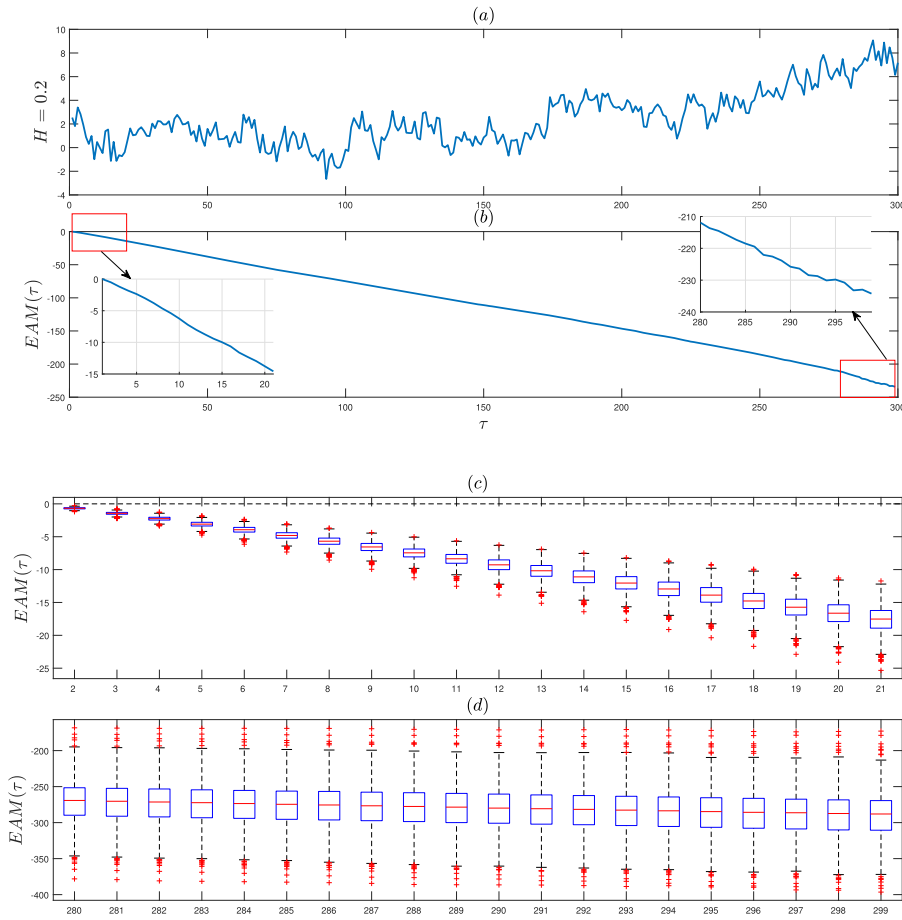


**Figure 2.** The comparison of the theoretical and empirical variance of EAM for  $H = 0.2$  [panel (a)] and  $H = 0.8$  [panel (b)] (in log–log scale). In order to calculate the empirical variance we simulated 1000 trajectories of length  $N = 300$  for FBM with the corresponding  $H$  parameter.

of length  $N = 300$  of FBM. In the panel (a) we present the results for FBM with  $H = 0.2$  and in panel (b)—with  $H = 0.8$ . As one can notice, the theoretical and empirical variances coincide. In both cases, the statistic increases with growing  $\tau$ , but for the super-diffusive case, the variance takes higher values than for the sub-diffusive process [notice the higher scale in panel (b)].

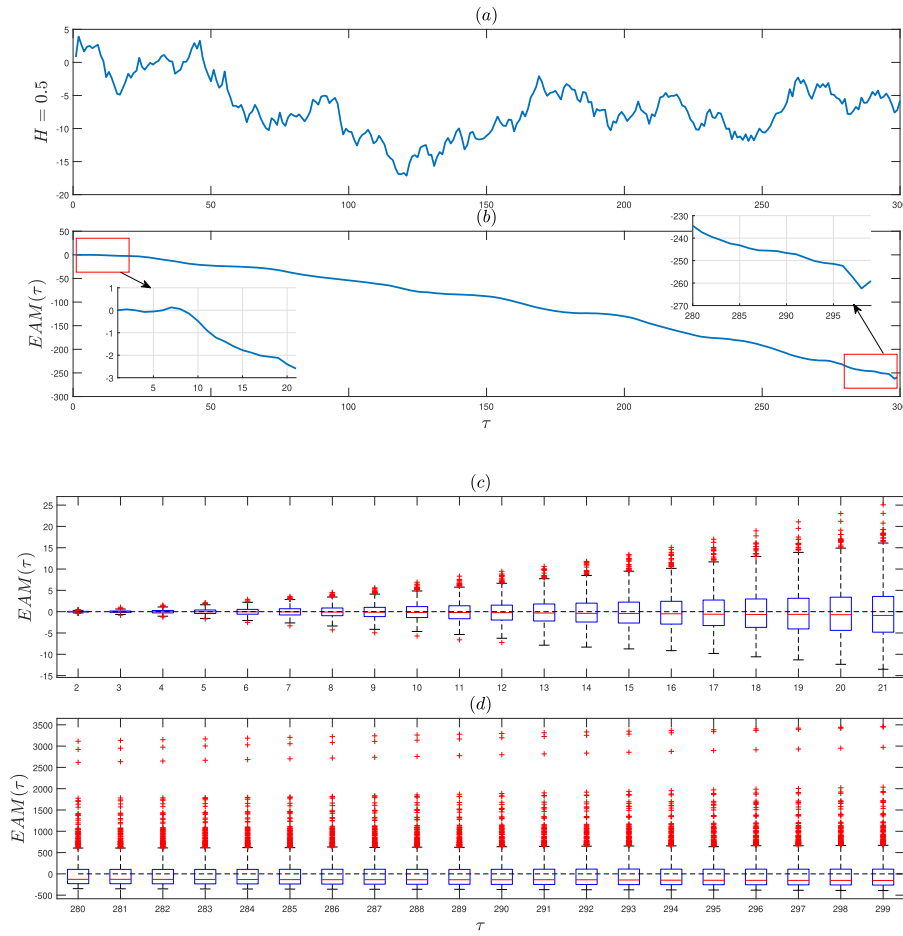
In figure 3 we present the exemplary trajectory of the FBM in the sub-diffusive case ( $H = 0.2$ ) with  $D = 1$ , see the panel (a), and the corresponding EAM statistic for the whole range of  $\tau$  parameter calculated according to the equation (9), see the panel (b). The length of the trajectory is  $N = 300$ . The similar trajectory lengths we observe in real-life data, see for instance [19]. As one can see the values of the statistic for the simulated trajectory are less than zero for all considered  $\tau$ s. Moreover, we have also zoomed the beginning and the end of the plot to see the specific behavior of the EAM statistic for small and large values of the arguments. As one can see, for both considered ranges the statistic takes values smaller than zero and it decreases. However, to demonstrate the overall pattern and the specific behavior of the EAM statistic we have made the  $M = 1000$  Monte Carlo simulations of the FBM with  $H = 0.2$ . The length of each simulated trajectory is  $N = 300$ . Then, for each trajectory, we calculate the EAM statistic for small (first 20) and large (last 20) values of  $\tau$  parameter. Finally, on panels (c) and (d) in figure 3 we demonstrate the box-plots of the obtained values. One can see, for all simulated trajectories of FBM with  $H = 0.2$  the value of the statistic is smaller than zero and especially for small values of  $\tau$  the statistic noticeably decreases for increasing  $\tau$ . Moreover, the variance of the statistic is smaller than for the large values of  $\tau$ . For large values of the arguments, the EAM statistic also does not exceed the zero value, however, it is more stabilized (but it still decreases) in contrast to the small range of the  $\tau$  parameter. The range of the statistic is notably larger than for the small values of the arguments.

In figure 4 we demonstrate the exemplary trajectory of FBM with  $H = 0.5$  and  $D = 1$ , i.e. the diffusive case and the corresponding EAM statistic, calculated according to formula (9),



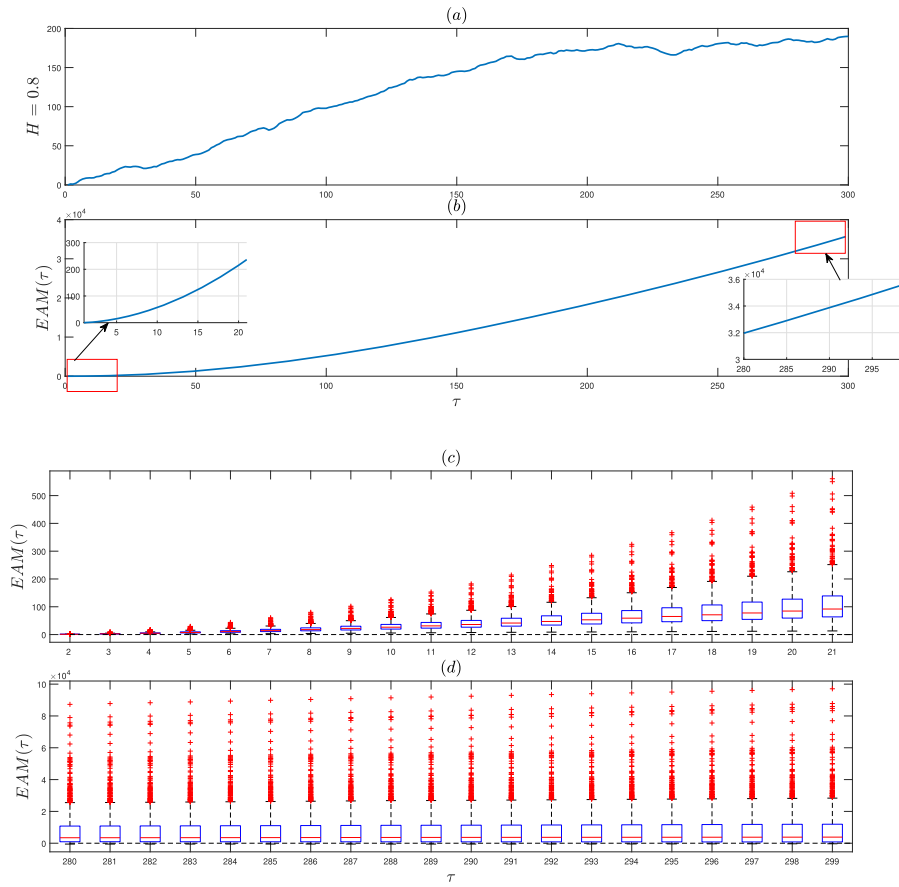
**Figure 3.** The exemplary trajectory of length  $N = 300$  of FBM with  $H = 0.2$  [panel (a)] and the EAM statistic for this trajectory for the whole range of  $\tau$  [panel (b)]. The panels (c) and (d) present the box-plots of the EAM statistic values for small [panel (c)] and large [panel (d)]  $\tau$  parameters. The box-plots are calculated based on  $M = 1000$  Monte Carlo simulations of the FBM with  $H = 0.2$ .

see panels (a) and (b), respectively. As one can see, the statistic for the small values of the  $\tau$  parameters seems to be close to zero, however when we zoomed the plot, one can observe it exceeds a little bit the zero value. When we analyze the large values of the  $\tau$  parameters, the EAM statistic is smaller than zero and it decreases with increasing  $\tau$ . Similar as for the sub-diffusive case (i.e. for  $H = 0.2$ ), we made  $M = 1000$  Monte Carlo simulations of FBM trajectories with  $H = 0.5$  of length  $N = 300$ . For each trajectory, we calculate the EAM statistic for all possible  $\tau$  values (i.e. for  $\tau = 2, 3, \dots, 299$ ) and finally in panels (c) and (d) of figure 4 we demonstrate the box-plots of the obtained values for small and large values of the arguments. The medians of the EAM values are close to zero and the variance of the statistic increases along with the  $\tau$  parameter. This is especially visible for the small values of the  $\tau$  parameters. However, comparing the sub-diffusive case, we can conclude the statistic is more stabilized.



**Figure 4.** The exemplary trajectory of length  $N = 300$  of FBM with  $H = 0.5$  and  $D = 1$  [panel (a)] and the EAM statistic for this trajectory for the whole range of  $\tau$  [panel (b)]. The panels (c) and (d) present the box-plots of the EAM statistic values for small [panel (c)] and large [panel (d)]  $\tau$  parameters. The box-plots are calculated based on  $M = 1000$  Monte Carlo simulations of the FBM with  $H = 0.5$ .

In figure 5 we present the exemplary trajectory of length  $N = 300$  for FBM with  $H = 0.8$  (super-diffusive case) and  $D = 1$ , see panel (a), and the corresponding EAM statistic, see panel (b). One can clearly see that the statistic increases with respect to the  $\tau$  parameter and is larger than zero both for small and large values of the arguments. Similar to previous cases, we made  $M = 1000$  Monte Carlo simulations of the sample trajectory of FBM with  $H = 0.8$  of length  $N = 300$  and for each trajectory we calculated the EAM statistic. Finally, we created the box-plots of the obtained values, see panels (c) and (d) of figure 5. The following conclusions can be drawn: for small values of the  $\tau$  parameters all values of the EAM statistic are higher than zero, this function increases with respect to its arguments and the variance of the statistic also increases. For large values of the  $\tau$  parameters the median of the EAM statistic is larger than zero, this function is more stabilized than for the small values of the arguments, however the variance of the statistic is larger than for the small values of the  $\tau$  parameter. One can see the



**Figure 5.** The exemplary trajectory of length  $N = 300$  of FBM with  $H = 0.8$  [panel (a)] and the EAM statistic for this trajectory for the whole range of  $\tau$  [panel (b)]. The panels (c) and (d) present the box-plots of the EAM statistic values for small [panel (c)] and large [panel (d)]  $\tau$  parameters. The box-plots are calculated based on  $M = 1000$  Monte Carlo simulations of the FBM with  $H = 0.8$ .

noticeably different behavior of EAM for a super-diffusive case than we have for sub- and diffusive processes.

The presented simulation results clearly indicate the differences of behavior of the EAM statistic for sub-diffusive, diffusive, and super-diffusive cases. The main differences for the three cases are summarized in table 1 and became the starting point for proposing a simple test for detecting anomalous behavior in real-life data. The test is precisely described in the next section where we also present its effectiveness for the simulated trajectories of the FBM. Here we present the version for the super-diffusive, behavior testing, but the simple modification of the algorithm allows for testing also the sub-diffusive regime. The main advantage is that the test is model-free and thus it can be used for any processes with a finite second moment.

**Table 1.** The behavior of the EAM statistic for sub-diffusive, diffusive, and super-diffusive cases.

$\tau$	Sub-diffusive case		Diffusive case		Super-diffusive case	
	EAM values	EAM monotonicity	EAM values	EAM monotonicity	EAM values	EAM monotonicity
Small	$< 0$	Decreases	$\approx 0$	Stabilizes	$> 0$	Increases
Large	$< 0$	Decreases	$\approx 0$	Stabilizes	$> 0$	Increases

#### 4. Simple test for anomaly detection based on EAM

In this section, we present a simple and intuitive application of the EAM statistic defined in equation (9). As it was mentioned above, the specific behavior of the statistic summarized in table 1 for sub-diffusive, diffusive, and super-diffusive cases can be a starting point for the test for anomalous diffusive behavior detection for real-life data. This test can be applied without the assumption of the specific model behind the data. However, in this section, we demonstrate its effectiveness for simulated trajectories of FBM for different values of the Hurst exponent and three selected trajectory lengths. As one can see in table 1 the EAM statistic can be considered as a detector for the discrimination between sub- and super-diffusive behavior.

Based on our results we can formulate the test with the following hypotheses

$H_0$  : the data exhibit super-diffusive behavior,

$H_1$  : the data do not exhibit super-diffusive behavior. (16)

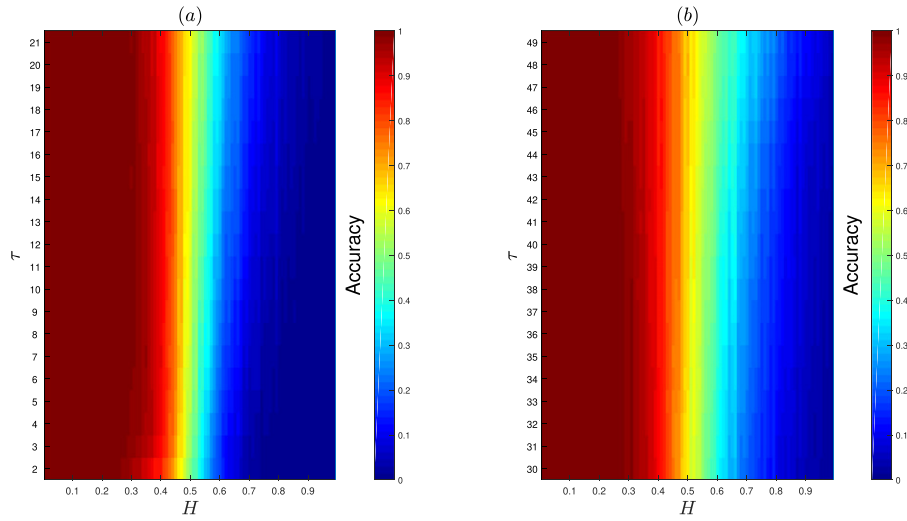
The procedure of testing is straightforward. Namely, for the real-life data set  $x(1), x(2), \dots, x(N)$  of length  $N$  first we estimate the EAM according to the formula (9) for selected values of  $\tau$  parameters. Next, we check the behavior of the statistic along the set of  $\tau$  values. If the statistic increases and takes values greater than zero we can suspect the super-diffusive regime. In other cases we, reject the  $H_0$  hypothesis.

When the number of trajectories from the same experiment is available, then we can calculate the empirical probability that the estimated value of the EAM statistic is greater than zero. We remind, zero is the borderline case discriminating between sub- and super-diffusion. More precisely, if the number of the available trajectories from the same experiment is  $M$ , each trajectory is of length  $N$ , then for each  $\tau$  the probability that the EAM is greater than zero can be calculated as follows

$$\hat{p}(\tau) = \frac{1}{M} \# \{ \widehat{AM}_i(\tau) > 0, \quad i = 1, 2, 3, \dots, M \}, \quad (17)$$

where the  $\widehat{AM}_i(\tau)$  is the empirical value of the AM statistic calculated for  $i$ th trajectory in point  $\tau$  (according to equation (9)). If the calculated  $\hat{p}(\tau)$  is small (i.e. smaller than the given confidence level) for selected values of the  $\tau$  parameters one can expect the data do not exhibit super-diffusive behavior. For a large value of  $\hat{p}(\tau)$  we do not reject the  $H_0$  hypothesis. One can enhance the information obtained for one specific value of  $\tau$  and analyze the  $\hat{p}(\cdot)$  for a set of  $\tau$  parameters.

In our simulation study, we made the following experiment. For each of the analyzed values of  $N \in \{50, 100, 300\}$  we simulate  $M = 1000$  trajectories of the FBM for  $H \in \{0.01, 0.02, \dots, 0.99\}$ . For each  $H$  and each  $\tau$  parameter, we calculate the empirical probability that the EAM is below zero, which is exactly equal to  $1 - \hat{p}(\tau)$ , where  $\hat{p}(\tau)$  is defined



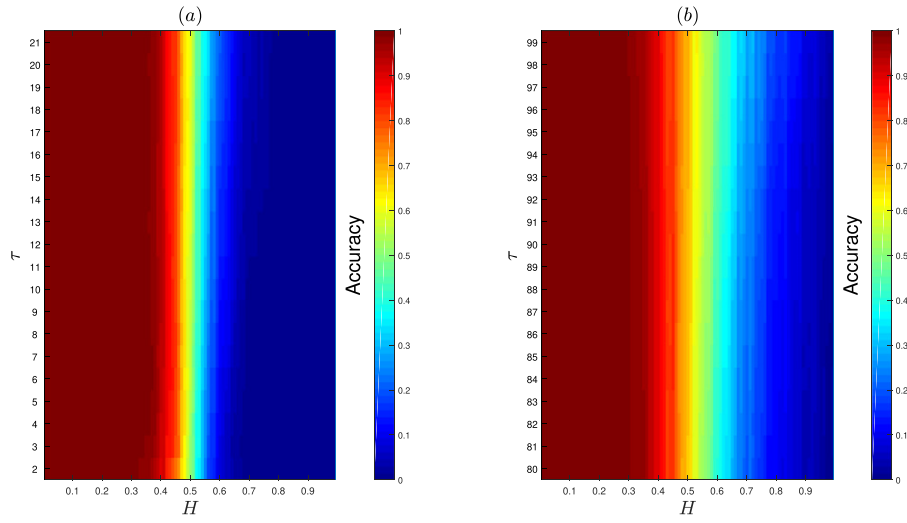
**Figure 6.** The power of the test for super-diffusive behavior detection based on the EAM statistic for  $H \in \{0.01, 0.02, \dots, 0.99\}$  for small [panel (a)] and large [panel (b)] values of  $\tau$  parameters. In order to calculate the power of the test, we simulated 1000 trajectories of length  $N = 50$  for FBM with the corresponding  $H$  parameter.

as in equation (17). For the test with hypotheses defined as in (16) the empirical probability  $1 - \hat{p}(\tau)$  is equal to the power of the test. We remind that the power of the test is the probability that the test rejects the  $H_0$  hypothesis when a specific alternative hypothesis  $H_1$  is true. The statistical power ranges from 0 to 1, and as statistical power increases, the probability of making a type II error (wrongly failing to reject the null hypothesis) decreases.

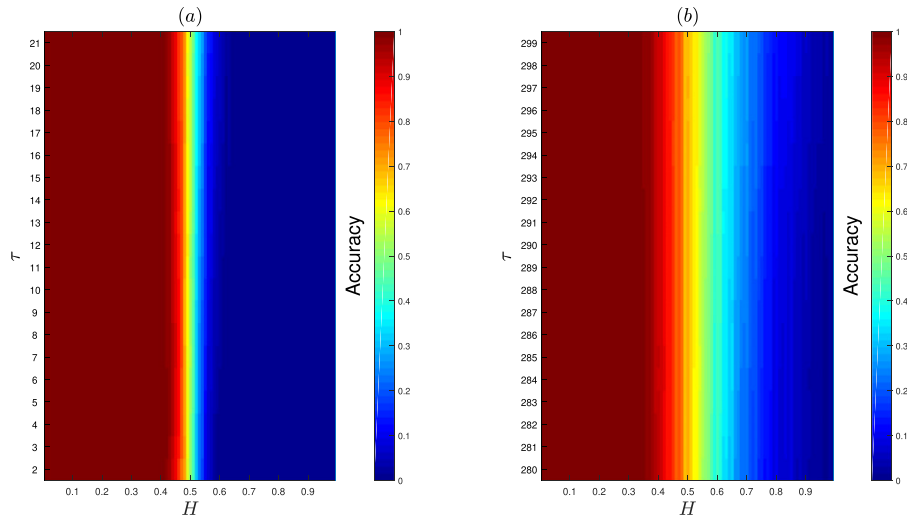
In figures 6–8, we present the power of the test for all considered  $H$  values for small (first 20 values) and large (last 20 values)  $\tau$  parameters for three considered values of  $N$ . One can see, for small trajectory length  $N = 50$  (see figure 6), the power for  $H \leq 0.3$  is equal to one for small and large values of  $\tau$  parameter, it decreases when  $H$  tends to 0.5 however in the neighborhood of 0.5 it is still higher than 0.5. For small  $\tau$  the test seems to be more effective. For  $H > 0.6$  the power is equal to zero. When we analyze the case with  $N = 100$ , see figure 7, the message is more clear. For  $H < 0.4$  the power of the test for small and large values of  $\tau$  is close to 1 while it is equal to 0 for small  $\tau$  and  $H > 0.55$ . For large values of  $\tau$  parameter, the test is not so effective as in the case of small values of the arguments of EAM statistic. As we expected, for a larger value of  $N$ , namely for  $N = 300$ , see figure 8, the power of the test seems to be not effective only in the close neighborhood of  $H = 0.5$  parameter and a small value of  $\tau$ . For large values of  $\tau$  the trajectory length has no influence on the power of the test, see panels (b) of figures 6–8.

In figures 9–11 in the [appendix](#), we also present the fraction of values of the EAM statistic that are below zero for  $H = \{0.01, 0.02, \dots, 0.5\}$  and above zero for  $H = \{0.51, 0.52, \dots, 0.99\}$  for small [panel (a)] and large [panel (b)] values of  $\tau$  parameter and three considered trajectories lengths. More precisely, for  $H = \{0.01, 0.02, \dots, 0.5\}$  we present  $1 - \hat{p}(\tau)$  and for  $H = \{0.51, 0.52, \dots, 0.99\}$ —the  $\hat{p}(\tau)$ , where  $\hat{p}(\tau)$  is defined in equation (17). The fractions are calculated based on  $M = 1000$  Monte Carlo simulations of the FBM with different values of  $H$  parameter. One can see that for the strict sub-diffusive case ( $H \ll 0.5$ ) and





**Figure 7.** The power of the test for super-diffusive behavior detection based on the EAM statistic for  $H \in \{0.01, 0.02, \dots, 0.99\}$  for small [panel (a)] and large [panel (b)] values of  $\tau$  parameters. In order to calculate the power of the test, we simulated 1000 trajectories of length  $N = 100$  for FBM with the corresponding  $H$  parameter.



**Figure 8.** The power of the test for super-diffusive behavior detection based on the EAM statistic for  $H \in \{0.01, 0.02, \dots, 0.99\}$  for small [panel (a)] and large [panel (b)] values of  $\tau$  parameters. In order to calculate the power of the test, we simulated 1000 trajectories of length  $N = 300$  for FBM with the corresponding  $H$  parameter.

super-diffusive case ( $H \gg 0.5$ ) and small values of  $\tau$  parameter the calculated empirical probability is equal to zero. However, for the large  $N$ , only in the close neighborhood of  $H = 0.5$  the calculated probability is different than 1, however it is still large ( $\sim 0.6$ ). The situation is different for large values of  $\tau$  parameter. Here the calculated probabilities are close to 1 for

**Table 2.** The means of the computational times (in seconds) of testing algorithm based on EAM statistic for  $H = 0.2$  and three values of trajectory lengths  $N = 50, 100$  and  $200$ . To calculate the means we considered 1000 trajectories for each length. The calculations were provided for  $\tau = 2$  and  $\tau = N - 1$ .

$N$	$\tau = 2$	$\tau = N - 1$
50	$1.41798 \times 10^{-5}$	$1.46983 \times 10^{-4}$
100	$1.41899 \times 10^{-5}$	$2.92116 \times 10^{-4}$
300	$1.48452 \times 10^{-5}$	$8.35019 \times 10^{-4}$

small values of  $H$  parameter. However, when we analyze the case  $H \gg 0.5$  the calculated values increase but they never meet 1. The presented results clearly indicate the test based on the EAM statistic for anomaly detection is more effective for small values of  $\tau$  parameter. Obviously, it is also more effective for larger trajectories.

In order to demonstrate the advantages of the proposed approach, we have calculated the computational time of the used testing procedure for the exemplary case. More precisely, for 1000 simulated trajectories of FBM with  $H = 0.2$  with length  $N$  we calculated the computational time of the algorithm. Finally, we obtained the mean of the computational times for all trajectories. For the comparison, we have analyzed three trajectory's lengths, namely  $N = 50, 100$  and  $300$ . In table 2 we demonstrate the means of the computational times for two values of  $\tau$  parameter, namely  $\tau = 2$  and  $\tau = N - 1$ . One can observe the differences of the computational times which follow directly from the definition of the EAM statistic. For larger  $\tau$ , the number of components used in EAM calculation is larger than for small  $\tau$ . However, one can see that even for a large value of  $\tau$  the proposed algorithm is relatively fast.

For the implementation of the test and the simulation study, we used MATLAB R2017a. Simulations were performed on Intel(R) Core(TM) i7-7700HQ CPU @ 2.80 GHz.

## 5. Conclusions and future work

In this paper, we have introduced a new statistic, called the EAM, which is useful in the problem of anomalous diffusion behavior recognition for the second-order processes. Its construction requires the whole information of sample ACVF of the given process (i.e. the ACVF for all arguments) and it is defined as the convolution of the ACVF with appropriate time lags. Thus, the proposed approach utilizes the whole information about the process in contrast to the classical approaches, where the methods are based on the ACVF in a specific time lag. The idea of the EAM is intuitive. It measures the deviation of the ensemble-averaged MSD of the considered process from the ensemble-averaged MSD for the classical diffusion model, namely BM. Thus, the EAM is a natural candidate for the detection of the anomaly type. The proven probabilistic characteristics for the considered statistic indicate this statement. By using the Monte Carlo simulations we have shown that EAM exhibits different behavior for different anomaly types. Thus, we proposed a simple test for the super-diffusive behavior recognition based on real-life data. For the exemplary anomalous diffusion process, namely FBM, we have demonstrated the effectiveness of the proposed approach. The proposed methodology is relatively simple. It utilizes the EAM statistic which is based on the sample ACVF and takes under consideration the fact that the EAM is decreasing with respect to  $\tau$  and negative for sub-diffusive processes while in the super-diffusive case it increases and takes positive values. Moreover, the EAM-testing method is computationally fast.

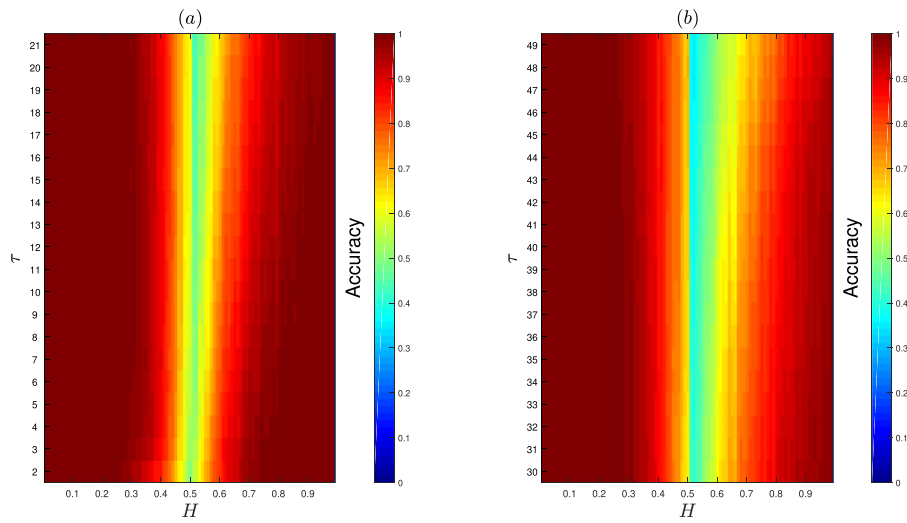
Although in this paper we have shown the application of the EAM in the problem of super-diffusive behavior identification, the considered statistic can be also used for testing the

sub-diffusive regime (or the general anomalous diffusive one). The simple modification of the described in section 4 procedure allows the recognition of the anomalous diffusive regime of any type. The universality of the proposed approach comes from the specific behavior of the statistic for different anomaly types. The advantage of the introduced methodology is related to the fact it is model-free and can be applied to real-life data without the knowledge of the theoretical process behind the data. However, the simplifying assumption is the stationarity of the increments. This paper can be useful for practitioners who require the simple intuitive methods and algorithms for the real-life data investigation without the preliminary knowledge about the theoretical foundations related to the testing assumptions.

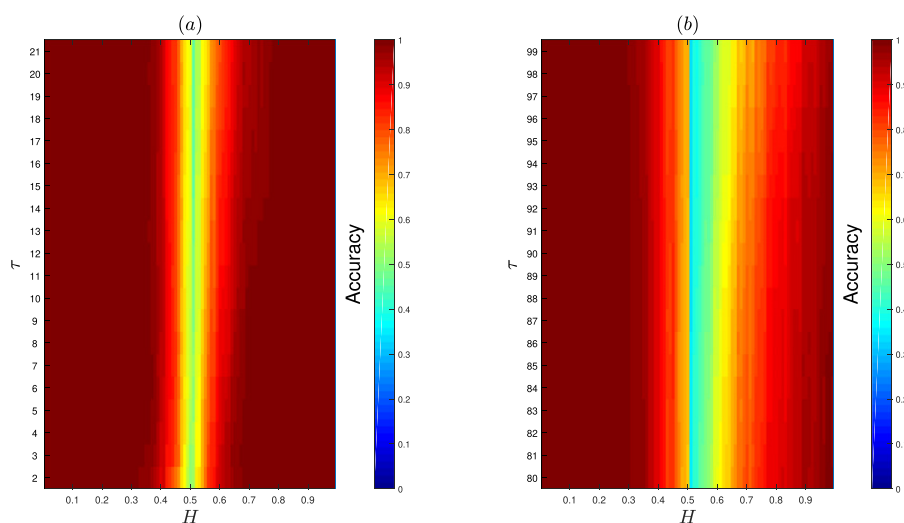
This paper is the preliminary one in the investigation of the EAM statistic in different directions. On one hand, the probabilistic properties (like distribution) of the EAM should be explored for instance in order to introduce a more strict statistical test for anomaly regime recognition. On the other hand, the considered statistic could be also used for the anomalous diffusion parameter estimation without preliminary knowledge about the model. The described testing procedure could be also enhanced by an introduction to the testing schema the intelligent methods, similar as it is used for other statistics useful in the anomalous diffusion behavior analysis known from the literature, see for instance [61, 76–78]. Moreover, the EAM statistic could be also useful in the problem of the transient diffusion recognition. Thus, based on this methodology one can detect the so-called structure breakpoint for the processes with time-varying anomalous diffusion parameter, see e.g. [56, 79].

## Appendix

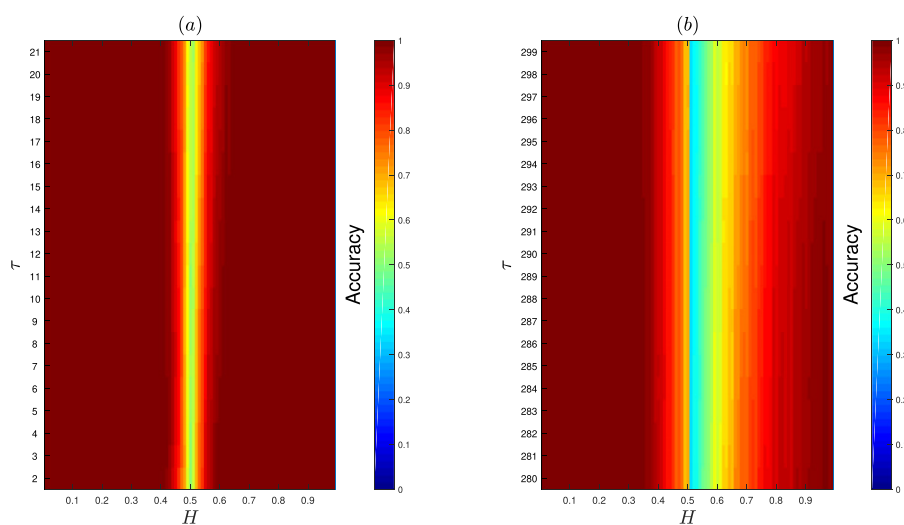
See (figures 9–11).



**Figure 9.** The fraction of values of the EAM statistic that are below zero for  $H = \{0.01, 0.02, \dots, 0.5\}$  and above zero for  $H = \{0.51, 0.52, \dots, 0.99\}$  for small [panel (a)] and large [panel (b)] values of  $\tau$  parameters. In order to calculate the empirical probability of anomalous diffusive regime, we considered 1000 trajectories of length  $N = 50$  for each  $H$ .



**Figure 10.** The fraction of values of the EAM statistic that are below zero for  $H = \{0.01, 0.02, \dots, 0.5\}$  and above zero for  $H = \{0.51, 0.52, \dots, 0.99\}$  for small [panel (a)] and large [panel (b)] values of  $\tau$  parameters. In order to calculate the empirical probability of anomalous diffusive regime, we considered 1000 trajectories of length  $N = 100$  for each  $H$ .



**Figure 11.** The fraction of values of the EAM statistic that are below zero for  $H = \{0.01, 0.02, \dots, 0.5\}$  and above zero for  $H = \{0.51, 0.52, \dots, 0.99\}$  for small [panel (a)] and large [panel (b)] values of  $\tau$  parameters. In order to calculate the empirical probability of anomalous diffusive regime, we considered 1000 trajectories of length  $N = 300$  for each  $H$ .

## ORCID iDs

Agnieszka Wyłomańska  <https://orcid.org/0000-0001-9750-1351>

## References

- [1] Zeng C, Chen Y and Yang Q 2012 The fbm-driven Ornstein–Uhlenbeck process: probability density function and anomalous diffusion *Fractional Calculus Appl. Anal.* **15** 479–92
- [2] Chen W, Sun H, Zhang X and Korošak D 2010 Anomalous diffusion modeling by fractal and fractional derivatives *Comput. Math. Appl.* **59** 1754–8
- [3] Mura A and Pagnini G 2008 Characterizations and simulations of a class of stochastic processes to model anomalous diffusion *J. Phys. A: Math. Theor.* **41** 285003
- [4] Piryatinska A, Saichev A I and Woyczynski W A 2005 Models of anomalous diffusion: the subdiffusive case *Physica A* **349** 375–420
- [5] Beran J 1994 *Statistics for Long-Memory Processes* (London: Chapman and Hall)
- [6] Mandelbrot B B and Van Ness J W 1968 Fractional Brownian motions, fractional noises and applications *SIAM Rev.* **10** 422–37
- [7] Samorodnitsky G and Taqqu M 1994 *Stable Non-Gaussian Random Processes* (London: Chapman and Hall)
- [8] Metzler R and Klafter J 2000 The random walk’s guide to anomalous diffusion: a fractional dynamics approach *Phys. Rep.* **339** 1–77
- [9] Metzler R and Klafter J 2004 The restaurant at the end of the random walk: recent developments in the description of anomalous transport by fractional dynamics *J. Phys. A: Math. Gen.* **37** R161–208
- [10] Sato K-I 1999 *Lévy Processes and Infinitely Divisible Distributions* (Cambridge: Cambridge University Press)
- [11] Wylomańska A 2012 Arithmetic Brownian motion subordinated by tempered stable and inverse tempered stable processes *Physica A* **391** 5685–96
- [12] Wylomańska A 2013 Tempered stable process with infinitely divisible inverse subordinators *J. Stat. Mech.* P10011
- [13] Wylomańska A and Gajda J 2016 Stable continuous-time autoregressive process driven by stable subordinator *Physica A* **444** 1012–26
- [14] Magdziarz M and Weron A 2007 Fractional Fokker–Planck dynamics: stochastic representation and computer simulation *Phys. Rev. E* **75** 056702
- [15] Magdziarz M 2009 Langevin picture of subdiffusion with infinitely divisible waiting times *J. Stat. Phys.* **135** 763–72
- [16] Gajda J and Magdziarz M 2010 Fractional Fokker–Planck equation with tempered alpha-stable waiting times: Langevin picture and computer simulation *Phys. Rev. E* **82** 011117
- [17] Gajda J and Wylomańska A 2014 Fokker–Planck type equations associated with fractional Brownian motion controlled by infinitely divisible processes *Physica A* **405** 104–13
- [18] Bodrova A S, Chechkin A V and Sokolov I M 2019 Scaled Brownian motion with renewal resetting *Phys. Rev. E* **100** 012120
- [19] Thalpa S, Wylomańska A, Sikora G, Wagner C E, Krapf D, Kantz H, Chechkin A V and Metzler R 2020 Leveraging large-deviation statistics to decipher the stochastic properties of measured trajectories (arXiv:2005.02099)
- [20] Fulinski A 2011 Anomalous diffusion and weak nonergodicity *Phys. Rev. E* **83** 061140
- [21] Beck C and Cohen E G D 2003 Superstatistics *Physica A* **322** 267–75
- [22] Chechkin A V, Seno F, Metzler R and Sokolov I M 2017 Brownian yet non-Gaussian diffusion: from superstatistics to subordination of diffusing diffusivities *Phys. Rev. X* **7** 021002
- [23] Kusmierz L, Majumdar S N, Sabhapandit S and Schehr G 2014 First order transition for the optimal search time of Lévy flights with resetting *Phys. Rev. Lett.* **113** 220602
- [24] Sokolov I M 2012 Models of anomalous diffusion in crowded environments *Soft Matter* **8** 9043–52
- [25] Klages R, Radons G and Sokolov I M 2008 *Anomalous Transport: Foundations and Applications* (New York: Wiley)
- [26] Hoefling F and Franosch T 2013 Anomalous transport in the crowded world of biological cells *Rep. Prog. Phys.* **76** 046602
- [27] Metzler R, Jeon J-H, Cherstvy A G and Barkai E 2014 Anomalous diffusion models and their properties: non-stationarity, non-ergodicity, and ageing at the centenary of single particle tracking *Phys. Chem. Chem. Phys.* **16** 24128–64
- [28] Nezhadhighi M G, Rajabpour M A and Rouhan S 2011 First-passage-time processes and subordinated Schramm–Loewner evolution *Phys. Rev. E* **84** 011134

- [29] Failla R, Grigolini P, Ignaccolo M and Schwettmann A 2004 Random growth of interfaces as a subordinated process *Phys. Rev. E* **70** 010101(R)
- [30] Gabaix X, Gopikrishnan P, Plerou V and Stanley H E 2003 A theory of power-law distributions in financial market fluctuations *Nature* **423** 267–70
- [31] Ivanov P C, Yuen A, Podobnik B and Lee Y 2004 Common scaling patterns in intertrade times of U.S. stocks *Phys. Rev. E* **69** 056107
- [32] Scher H, Margolin G, Metzler R and Klafter J 2002 The dynamical foundation of fractal stream chemistry: the origin of extremely long retention times *Geophys. Res. Lett.* **29** 1061
- [33] Doukhan P, Oppenheim G and Taqqu M S 2003 *Theory and Applications of Long-Range Dependence* (Basel: Birkhäuser)
- [34] Golding I and Cox E C 2006 Physical nature of bacterial cytoplasm *Phys. Rev. Lett.* **96** 098102
- [35] Stanislavsky A A, Burnecki K, Magdziarz M, Weron A and Weron K 2009 Farima modelling of solar flar activity from empirical time series of soft x-ray solar emission *Astrophys. J.* **693** 1877–82
- [36] Kepten E, Weron A, Sikora G, Burnecki K and Garini Y 2015 Guidelines for the fitting of anomalous diffusion mean square displacement graphs from single particle tracking experiments *PLOS One* **10** e0117722
- [37] Gal N, Lechtman-Goldstein D and Weihs D 2013 Particle tracking in living cells: a review of the mean square displacement method and beyond *Rheol. Acta* **52** 425–43
- [38] Michalet X 2010 Mean square displacement analysis of single-particle trajectories with localization error: Brownian motion in an isotropic medium *Phys. Rev. E* **82** 041914
- [39] Michalet X and Berglund A J 2012 Optimal diffusion coefficient estimation in single-particle tracking *Phys. Rev. E* **85** 061916
- [40] Ernst D and Köhler J 2013 How the number of fitting points for the slope of the mean-square displacement influences the experimentally determined particle size distribution from single-particle tracking *Phys. Chem. Chem. Phys.* **15** 3429–32
- [41] Sikora G, Burnecki K and Wylomańska A 2017 Mean-squared displacement statistical test for fractional Brownian motion *Phys. Rev. E* **95** 032110
- [42] Sikora G, Teuerle M, Wylomańska A and Grebenkov D 2017 Statistical properties of the anomalous scaling exponent estimator based on time averaged mean square displacement *Phys. Rev. E* **96** 022132
- [43] Taqqu M S, Teverovsky V and Willinger W 1995 Estimators for long-range dependence: an empirical study *Fractals* **03** 785–98
- [44] Movahed M and Hermanis E 2008 Fractal analysis driver flow fluctuations *Physica A* **387** 915–32
- [45] Bardet J-M and Kammoun I 2008 Asymptotic properties of the detrended fluctuation analysis of long-range-dependent processes *IEEE Trans. Inf. Theor.* **54** 2041–52
- [46] Movahed M S, Jafari G R, Ghasemi F, Rahvar S and Tabar M R R 2006 Multifractal detrended fluctuation analysis of sunspot time series *J. Stat. Mech.* **P02003**
- [47] Heneghan C and McDarby G 2000 Establishing the relation between detrended fluctuation analysis and power spectral density analysis for stochastic processes *Phys. Rev. E* **62** 6103
- [48] Kiyono K 2015 Establishing a direct connection between detrended fluctuation analysis and Fourier analysis *Phys. Rev. E* **92** 042925
- [49] Alessio E, Carbone A, Castelli G and Frappietro V 2002 Second-order moving average and scaling of stochastic time series *Eur. Phys. J. B* **27** 197
- [50] Carbone A 2007 Algorithm to estimate the hurst exponent of high-dimensional fractals *Phys. Rev. E* **76** 056703
- [51] Arianos S and Carbone A 2017 Detrending moving average algorithm: a closed-form approximation of the scaling law *Physica A* **382** 9–5
- [52] Carbone A and Kiyono K 2016 Detrending moving average algorithm: frequency response and scaling performances *Phys. Rev. E* **93** 063309
- [53] Shao Y-H, Gu G-F, Jiang Z-Q and Zhou W-X 2015 Effects of polynomial trends on detrending moving average analysis *Fractals* **23** 1550034
- [54] Sikora G 2018 Statistical test for fractional Brownian motion based on detrending moving average algorithm *Chaos Solitons Fractals* **116** 54–62
- [55] Janczura J and Weron A 2015 Ergodicity testing for anomalous diffusion: small sample statistics *J. Chem. Phys.* **142** 144103
- [56] Sikora G, Wylomańska A, Gajda J, Sole L, Akin E J, Tamkun M and Krapf D 2017 Elucidating distinct ion channel populations on the surface of hippocampal neurons via single-particle tracking recurrence analysis *Phys. Rev. E* **96** 062404

- [57] Lutz E 2003 Anomalous diffusion and Tsallis statistics in an optical lattice *Phys. Rev. A* **67** 051402
- [58] Magdziarz M and Weron A 2011 Anomalous diffusion: testing ergodicity breaking in experimental data *Phys. Rev. E* **84** 051138
- [59] Weron A, Janczura J, Boryczka E, Sungkaworn T and Calebiro D 2019 Statistical testing approach for fractional anomalous diffusion classification *Phys. Rev. E* **99** 042149
- [60] Burnecki K, Kepten E, Garini Y, Sikora G and Weron A 2015 Estimating the anomalous diffusion exponent for single particle tracking data with measurement errors—an alternative approach *Sci. Rep.* **5** 11306
- [61] Janczura J, Kowalek P, Lis-Olszewska J S H and Weron A 2020 Machine learning classification of particle trajectories in living cells: machine learning versus statistical testing hypothesis *Phys. Rev. E* **102** 032402
- [62] Wyłomańska A, Chechkin A, Sokolov I M and Gajda J 2015 Codifference as a practical tool to measure interdependence *Physica A* **421** 412–29
- [63] Magdziarz M and Weron A 2011 Ergodic properties of anomalous diffusion processes *Ann. Phys., NY* **326** 2431–43
- [64] Hassani H, Leonenko N and Patterson K 2012 The sample autocorrelation function and the detection of long-memory processes *Physica A* **391** 6367–79
- [65] Balcerek M and Burnecki K 2020 Testing of fractional Brownian motion in a noisy environment *Chaos Solitons Fractals* **140** 110097
- [66] Taylor G I 1921 Diffusion by continuous movement *Proc. Lond. Math. Soc.* **20** 196–211
- [67] Shalchi A 2011 Applicability of the Taylor–Green–Kubo formula in particle diffusion theory *Phys. Rev. E* **83** 046402
- [68] Dechant D K A, Lutz E and Barkai E 2014 Scaling Green–Kubo relation and application to three aging systems *Phys. Rev. X* **4** 011022
- [69] Magdziarz M 2006 The dependence structure of the solutions of the fractional equations with  $\alpha$ -stable noise *Doctoral Thesis* Wrocław University of Science and Technology
- [70] Isserlis L 1918 On a formula for the product-moment coefficient of any order of a normal frequency distribution in any number of variables *Biometrika* **12** 134–9
- [71] Severini T A 2005 *Elements of Distribution Theory* (Cambridge: Cambridge University Press)
- [72] Mingo J A and Speicher R 2017 *Free Probability and Random Matrices* (Berlin: Springer)
- [73] Szymanski J and Weiss M 2009 Elucidating the origin of anomalous diffusion in crowded fluids *Phys. Rev. Lett.* **103** 038102
- [74] Weiss M 2013 Single-particle tracking data reveal anticorrelated fractional Brownian motion in crowded fluids *Phys. Rev. E* **88** 010101
- [75] Krapf D *et al* 2019 Spectral content of a single non-Brownian trajectory *Phys. Rev. X* **9** 011019
- [76] Kowalek P, Loch-Olszewska H and Szwabiński J 2019 Classification of diffusion modes in single-particle tracking data: feature-based versus deep-learning approach *Phys. Rev. E* **100** 032410
- [77] Muñoz-Gil G, García-March M A, Manzo C, Martín-Guerrero J D and Lewenstein M 2020 Single trajectory characterization via machine learning *New J. Phys.* **22** 013010
- [78] Arts M, Smal I, Paul M W, Wyman C and Meijering E 2019 Particle mobility analysis using deep learning and the moment scaling spectrum *Sci. Rep.* **9** 17160
- [79] Sikora G, Wyłomańska A and Krapf D 2018 Recurrence statistics for anomalous diffusion regime change detection *Comput. Stat. Data Anal.* **128** 380–94