Computational Statistics and Data Analysis 168 (2022) 107401

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

www.elsevier.com/locate/csda

# Statistical test for anomalous diffusion based on empirical anomaly measure for Gaussian processes

Dawid Szarek [a], Katarzyna Maraj-Zygmąt [a], Grzegorz Sikora [a,*], Diego Krapf [b], Agnieszka Wyłomańska [a]

[a] Faculty of Pure and Applied Mathematics, Hugo Steinhaus Center, Wroclaw University of Science and Technology, Wyspianskiego 27, 50-370 Wroclaw, Poland
[b] Department of Electrical and Computer Engineering, Colorado State University, Fort Collins, CO 80523, USA

ARTICLE INFO

*Article history:*
Received 11 May 2021
Received in revised form 11 November 2021
Accepted 12 November 2021
Available online 1 December 2021

*Keywords:*
Autocovariance function
Fractional Brownian motion
Monte Carlo simulations
Biological data

ABSTRACT

Gaussian processes with anomalous diffusion behavior are considered. A new statistical test for the model identification that is based on the empirical anomaly measure (EAM) is introduced. This measure is considered as the distance between the anomalous and normal diffusion. In particular, the main properties of the EAM based on the quadratic form representation of Gaussian processes are investigated. The effectiveness of the test is evaluated for the fractional Brownian motion. Theoretical results and simulation studies are supported by the analysis of experimental data describing the sub-diffusive motion of microspheres in agarose hydrogels.



## 1. Introduction

Gaussian processes are crucial in statistical analysis. These processes are fully characterized by the covariance and expected value. In recent years various reports were devoted to the investigation of different Gaussian processes, see e.g. Stroud et al. (2017); Coeurjolly and Porcu (2018); Davis et al. (2021); Gerber and Nychka (2021); Wang and Xu (2019); Mariñas-Collado et al. (2019); Ernst et al. (2017); Zhou and Xiao (2018); Bachoc et al. (2019). In general, Gaussian systems have found various interesting applications, see e.g. Li et al. (2020); Lee et al. (2020); Arias Velásquez and Mejía Lara (2020).

Gaussian anomalous diffusion processes are a special case of Gaussian models, where the anomalous diffusion behavior of a given process is manifested in a non-linear second moment. More precisely, the process $\{X(t),\ t \in \mathbb{R}\}$ is classified as anomalous diffusive if its second moment scales asymptotically as a power function $\mathbb{E}\left[X^2(t)\right] \sim t^{\beta}$, where $\beta \neq 1$ is called the anomalous diffusion exponent. Depending on the $\beta$ parameter one can distinguish between sub-diffusive ($\beta < 1$) and super-diffusive ($\beta > 1$) regimes. In the case $\beta = 1$, the process exhibits diffusive behavior (also called normal diffusion). The origin of normal diffusion is the tacit assumption that Brownian particle moves in an infinite structureless medium acting as a heat bath. This assumption is incorrect when the motion takes place in a complex medium, (Oliveira et al., 2019), as recently observed in various physical phenomena, see e.g. Vilk et al. (2021); Weigel et al. (2011); Sabri et al. (2020).

The classical example of Gaussian anomalous diffusion is fractional Brownian motion (FBM). This process was introduced in Kolmogorov (1940) and it is considered as the generalization of the classical Brownian motion (BM). The FBM is the only Gaussian self-similar process with stationary increments, (Beran, 1994). The corresponding increment process, called frac-

* Corresponding author.
*E-mail address:* gregor.sikora@gmail.com (G. Sikora).

https://doi.org/10.1016/j.csda.2021.107401
0167-9473/

tional Gaussian noise, is the classical example of a long-memory process. The FBM $\{X(t)\}$, $t \geq 0$, characterized by the Hurst index $H \in (0, 1)$ and the diffusion coefficient $D > 0$, is a continuous and centered Gaussian process with the autocovariance function (ACVF)

$$\mathbb{E}\left[X(t)X(s)\right] = D\left(t^{2H} + s^{2H} - |t-s|^{2H}\right), \tag{1}$$

where $t, s \geq 0$. It should be highlighted that, the FBM itself is not stationary process, thus the ACVF depends on two arguments. For $H = 1/2$, FBM reduces to Brownian motion. In addition to FBM, the class of anomalous diffusion processes is very rich, (Metzler et al., 2014a; Krapf and Metzler, 2019). It includes, for instance fractional Lévy stable motion, (Samorodnitsky and Taqqu, 1994), continuous-time random walk, (Montroll and Weiss, 2017; Scher, 2017) and subordinated processes (called also time-changed processes), (Magdziarz and Weron, 2007; Krapf, 2015). See also these publications Bodrova et al. (2019); Thapa et al. (2021); Fuliński (2011); Beck and Cohen (2003); Chechkin et al. (2017); Kusmierz et al. (2014); Sokolov (2012); Klages et al. (2008); Höfling and Franosch (2013); Metzler et al. (2014b); Weigel et al. (2012); Krapf et al. (2016).

The so-called statistics of time averages play a crucial role in the analysis of the anomalous diffusion processes. Three statistics of time averages have been proposed for Gaussian anomalous diffusion processes, namely, the sample autocovariance function (ACVF), (Balcerek and Burnecki, 2020), time-averaged mean squared displacement (TAMSD), (Sikora et al., 2017a) and detrended moving average (DMA), (Sikora, 2018). For interesting properties and applications of these statistics we refer the readers to references (Kepten et al., 2015; Gal et al., 2013; Michalet, 2010; Michalet and Berglund, 2012; Ernst and Köhler, 2013; Sikora et al., 2017b). These three statistics can be represented as a quadratic form. However, in the testing procedures, the appropriate test statistics are considered for one specific time lag, thus some information about the process is not taken into account.

Our original contribution to the statistical inference theory for Gaussian anomalous diffusion processes in this article is the new statistical test for the proper model identification of anomalous diffusion phenomena. The test is based on a novel statistic of time averages called the empirical anomaly measure (EAM), which is the sum of the off-diagonal elements of the sample autocovariance matrix of the increment process. The EAM measures the distance between the second moment of a given process and the second moment of the classical Brownian motion. Thus, in the case of anomalous diffusion, it can be considered as the distance between the anomalous and normal diffusion. The EAM was introduced in Maraj et al. (2021), where the general properties and possible applications were presented. In this paper, we study the probabilistic properties of this statistic using its quadratic form representation for Gaussian processes (Mathai and Provost, 1992). This representation gives the possibility to recognize the theoretical foundations and thus we propose a strict statistical test for anomalous diffusion behavior identification. The testing procedure consists of two steps. At the first stage, the type of anomalous diffusion for real trajectories can be discriminated. At the second stage, one can test the theoretical Gaussian model for data description. While the proposed testing procedure is based on strong mathematical foundations, it can be considered as a guide for experimentalists who deal with anomalous diffusive data.

As it was mentioned, the EAM is defined as the convolution of the sample ACVF for all possible arguments (determined by the trajectory length) of the increment process and appropriate lag values. Thus, intuitively, it contains more information about the process than the sample ACVF for one selected value of lag. In the proposed holistic procedure, all values of EAM are taken into account, thus the whole information about the process is used for testing. In the literature, one can find different approaches used for testing various Gaussian processes. Most of the methods are dedicated to the specific Gaussian model, (Beran et al., 2016; Wei et al., 2017), however, there are also approaches, similar to the EAM-based method, that can be applicable for the general case, see e.g. (Zhou, 2011).

The rest of the paper is organized as follows: In Section 2, we study the probabilistic properties of the EAM for Gaussian processes based on the theoretical foundations of quadratic form representation. Based on this representation we provide the formulas for the moments of the considered statistic as well as the main characteristics, such as the probability density function (PDF) and cumulative distribution function (CDF). The theoretical properties give the possibility to design a statistical test for anomalous diffusion based on EAM. In Section 3, we describe the testing procedure that consists of the two steps, described above. In Section 4, we present the simulation study for the selected Gaussian anomalous diffusion process, namely the FBM. We demonstrate how the two-step procedure can be applied in this case. However, there is a need to highlight that although the simulation study is performed for FBM, the proposed testing methodology is universal and can be applied to any Gaussian process with stationary increments. As it was mentioned, the EAM statistic arises as a linear combination of the sample ACVF for all possible arguments and appropriate lag values, thus it can be considered as the generalization (or extension) of the sample ACVF. In Section 5, we compare the effectiveness of the new approach with the methodology based on the sample ACVF. The study is performed for FBM. In Section 6, we describe the holistic testing procedure based on EAM that can be considered as the practical guide for experimentalists. In this procedure we take into account the EAM values for all possible arguments, thus one can infer the whole information about the process being tested. In this section, we also demonstrate the effectiveness of the holistic procedure for simulated trajectories of FBM. In Section 7, crucial from an applicability point of view, we analyze real data by using the introduced methodology. We study the sub-diffusive motion of particles embedded in a viscoelastic media, namely polystyrene microspheres within agarose hydrogels. The last section contains conclusions.

## 2. Probabilistic properties of empirical anomaly measure for Gaussian processes

In this section first, we review the definition of the anomaly measure (AM) and its empirical version, called empirical anomaly measure, recently introduced in Maraj et al. (2021).

We consider the zero-mean Gaussian process (starting at zero) with stationary increments indexed by natural numbers $\{X(n)\} = \{X(n),\ n = 0, 1, \ldots\}$. Its autocovariance function (ACVF) is denoted as $\{\gamma_X(i, j) = \mathbb{E}[X(i)X(j)],\ i, j = 0, 1, \ldots\}$. Moreover, as $\{Y(n)\} = \{Y(n),\ n = 0, 1, \ldots\}$ we denote the corresponding increment process defined as follows

$$Y(n) = X(n+1) - X(n), \quad n = 0, 1, \ldots \tag{2}$$

with the ACVF $\{\gamma_Y(i, j) = \mathbb{E}[Y(i)Y(j)],\ i, j = 0, 1, \ldots\}$. In this paper we assume the process under consideration has stationary increments. In case $\gamma_Y(i, j)$ depends only on one argument and $\{\gamma_Y(i) = \mathbb{E}[Y(i)Y(0)],\ i = 0, 1, \ldots\}$.

**Definition 2.1.** (Maraj et al., 2021) Let us consider the zero-mean Gaussian process (starting at zero) with stationary increments $\{X(n)\} = \{X(n),\ n = 0, 1, \ldots\}$ and the corresponding increment process $\{Y(n)\} = \{Y(n),\ n = 0, 1, \ldots\}$. The corresponding AM is defined as follows

$$AM_X(\tau) \coloneqq 2\sum_{i=1}^{\tau-1}(\tau - i)\gamma_Y(i),\ \ \tau = 2, 3, \ldots \tag{3}$$

It was shown that AM defined in (3) can be considered as the quantity indicating the anomalous diffusive behavior of the corresponding process. More precisely, it was demonstrated in Maraj et al. (2021) that AM can be also represented as

$$AM_X(\tau) = \mathbb{E}[(X(\tau))^2] - \tau\gamma_Y(0),\ \tau = 2, 3, \ldots \tag{4}$$

where $\mathbb{E}[(X(\tau))^2]$ is the second moment of a given process. Thus, anomaly measure gives the information about the deviation of the second moment from the linear function $\tau\gamma_Y(0)$. Therefore, if the process $\{X(n)\}$ is a diffusive one, the corresponding AM equals zero. When it is sub-diffusive, the AM is negative and it is decreasing function of $\tau$ while, in the super-diffusive case, the corresponding AM takes positive values and increases with respect to $\tau$. More details related to the interpretation of the anomaly measure can be found in Maraj et al. (2021).

In the next definition, we introduce the empirical anomaly measure that is a natural candidate for the estimator of the AM.

**Definition 2.2.** (Maraj et al., 2021) Let us consider the finite trajectory $\boldsymbol{X}_{N+1} = \{X(n),\ n = 1, 2, \ldots, N+1\}$ of the zero-mean Gaussian process with stationary increments and the trajectory of the increments $\boldsymbol{Y}_N = \{Y(n),\ n = 1, \ldots, N\}$. The corresponding EAM is defined as follows

$$\widehat{AM}_X(\tau) = 2\sum_{i=1}^{\tau-1}(\tau - i)\hat{\gamma}_Y(i),\ \tau = 2, 3, \ldots, N. \tag{5}$$

In the above equation $\hat{\gamma}_Y(i)$ is the empirical autocovariance function of $\boldsymbol{Y}_N$ defined as follows

$$\hat{\gamma}_Y(i) = \frac{1}{N-i}\sum_{j=1}^{N-i} Y(j)Y(j+i),\ i = 1, 2, \ldots, N-1. \tag{6}$$

In Theorem 2.1 we prove that EAM for the zero-mean Gaussian process with stationary increments can be represented as the quadratic form of $\boldsymbol{Y}_N$. However, let us first recall the general idea of the quadratic forms for Gaussian processes, (Mathai and Provost, 1992). For the finite trajectory $\boldsymbol{X}_N$ of any zero-mean Gaussian process any statistic $S_N(\tau)$ depending on parameter $\tau$ that takes the following form

$$S_N(\tau) = \boldsymbol{X}_N^{\mathsf{T}}\boldsymbol{A}(\tau)\boldsymbol{X}_N \tag{7}$$

is called the quadratic form of $\boldsymbol{X}_N$. In Eq. (7) $\boldsymbol{X}_N^{\mathsf{T}}$ denotes a transposition of the vector $\boldsymbol{X}_N$ and $\boldsymbol{A}(\tau)$ is a symmetric matrix, which depends on the $\tau$ parameter.

**Theorem 2.1.** *For the finite trajectory $\boldsymbol{X}_{N+1}$ of the zero-mean Gaussian process with stationary increments, the EAM defined in Eq. (5) can be represented as the quadratic form of the increments vector $\boldsymbol{Y}_N$, namely*

$$\widehat{AM}_X(\tau) = \boldsymbol{Y}_N^{\mathsf{T}}\boldsymbol{T}(\tau)\boldsymbol{Y}_N, \tag{8}$$

*where $\boldsymbol{T}(\tau) = \left[t_{|i-j|}\right]$, $i, j = 1, 2, \cdots, N$ is the indefinite symmetric Toeplitz matrix of the bandwidth $\tau - 1$ ($\tau > 1$) with $t_{|i-j|} = \frac{\tau-|i-j|}{N-|i-j|}\mathbb{1}_{0<|i-j|<\tau}$, where $\mathbb{1}_{x\in A}$ is the indicator function that is equal to 1 if $x \in A$ and 0 otherwise.*

**Proof.** To prove the theorem we need to transform Eq. (5) into the form given by Eq. (7). First, let us plug (6) into Eq. (5). Then we obtain $\widehat{AM}_X(\tau) = 2\sum_{i=1}^{\tau-1} \frac{\tau-i}{N-i} \sum_{j=1}^{N-i} Y(j)Y(j+i)$. Then, using the indicator function we obtain the equivalent form

$$\begin{aligned}\widehat{AM}_X(\tau) &= 2\sum_{i=1}^{\tau-1} \frac{\tau-i}{N-i} \sum_{j=1}^{N-i} Y(j)Y(j+i)\mathbb{1}_{j\le N-i} \\ &= 2\sum_{i=1}^{\tau-1} \frac{\tau-i}{N-i} \sum_{j=1}^{N-1} Y(j)Y(j+i)\mathbb{1}_{j\le N-i} = 2\sum_{j=1}^{N-1} Y(j) \sum_{i=1}^{\tau-1} \left(\frac{\tau-i}{N-i}\mathbb{1}_{j\le N-i}\right) Y(j+i).\end{aligned}$$

Finally we obtain

$$\begin{aligned}\widehat{AM}_X(\tau) &= 2\sum_{j=1}^{N-1} Y(j) \sum_{i=1}^{\tau-1} \left(\frac{\tau-i}{N-i}\mathbb{1}_{i\le N-j}\right) Y(j+i) \overset{j+i=l}{=} 2\sum_{j=1}^{N-1} Y(j) \sum_{l=1+j}^{\tau-1+j} \left(\frac{\tau-(l-j)}{N-(l-j)}\mathbb{1}_{l-j\le N-j}\right) Y(l) \\ &\overset{l\to i}{=} 2\sum_{j=1}^{N-1} Y(j) \sum_{i=1+j}^{\tau-1+j} \left(\frac{\tau-i+j}{N-i+j}\mathbb{1}_{i\le N}\right) Y(i) = 2\sum_{i=1}^{N}\sum_{j=1}^{N} Y(i)Y(j) \left(\frac{\tau-i+j}{N-i+j}\mathbb{1}_{j+1\le i\le \tau-1+j}\right) \\ &= \sum_{i=1}^{N}\sum_{j=1}^{N} Y(i)Y(j) \left(\frac{\tau-i+j}{N-i+j}\mathbb{1}_{j<i<j+\tau}\right) + \sum_{j=1}^{N}\sum_{i=1}^{N} Y(j)Y(i) \left(\frac{\tau-j+i}{N-j+i}\mathbb{1}_{i<j<i+\tau}\right) \\ &= \sum_{i=1}^{N}\sum_{j=1}^{N} Y(i)Y(j) \left(\frac{\tau-(i-j)}{N-(i-j)}\mathbb{1}_{j<i<j+\tau}\right) + \sum_{i=1}^{N}\sum_{j=1}^{N} Y(i)Y(j) \left(\frac{\tau+i-j}{N+i-j}\mathbb{1}_{j-\tau<i<j}\right) \\ &= \sum_{i=1}^{N}\sum_{j=1}^{N} Y(i)Y(j) \left(\frac{\tau-|i-j|}{N-|i-j|}\mathbb{1}_{0<|i-j|<\tau}\right).\end{aligned}$$

The last equation can be written as $\widehat{AM}_X(\tau) = \boldsymbol{Y}_N^{\mathsf{T}} \boldsymbol{T}(\tau) \boldsymbol{Y}_N$, where $\boldsymbol{T}(\tau)$ is the square matrix of size $\tau-1$ with the elements $t_{|i-j|} = \frac{\tau-|i-j|}{N-|i-j|}\mathbb{1}_{0<|i-j|<\tau}$, i.e.

$$\boldsymbol{T}(\tau) = \begin{bmatrix} 0 & \frac{\tau-1}{N-1}\mathbb{1}_{1<\tau} & \frac{\tau-2}{N-2}\mathbb{1}_{2<\tau} & \cdots & \frac{\tau-N+1}{1}\mathbb{1}_{N-1<\tau} \\ \frac{\tau-1}{N-1}\mathbb{1}_{1<\tau} & 0 & \frac{\tau-1}{N-1}\mathbb{1}_{1<\tau} & \cdots & \frac{\tau-N+2}{2}\mathbb{1}_{N-2<\tau} \\ \frac{\tau-2}{N-2}\mathbb{1}_{2<\tau} & \frac{\tau-1}{N-1}\mathbb{1}_{1<\tau} & 0 & \cdots & \frac{\tau-N+3}{3}\mathbb{1}_{N-3<\tau} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{\tau-N+1}{1}\mathbb{1}_{N-1<\tau} & \frac{\tau-N+2}{2}\mathbb{1}_{N-2<\tau} & \frac{\tau-N+3}{3}\mathbb{1}_{N-3<\tau} & \cdots & 0 \end{bmatrix}. \tag{9}$$

Because the elements of the matrix $\boldsymbol{T}(\tau)$ are dependent on the absolute difference of the position ($|i-j|$), it is the symmetric Toeplitz matrix. To show that $\boldsymbol{T}(\tau)$ is indefinite, we use Sylvester's criterion, (Horn and Johnson, 2013). This criterion implies that a matrix is indefinite if there exists a negative principal minor of even order and the determinant is different from zero, i.e. in our case $|\boldsymbol{T}(\tau)| \neq 0$ for all $\tau$. We can easily calculate the first two leading principal minors of the matrix $\boldsymbol{T}(\tau)$, namely $\Delta_1 = 0$, $\Delta_2 = -\left(\frac{\tau-1}{N-1}\mathbb{1}_{1<\tau}\right)^2$. For the non-trivial case when $\tau > 1$ we have that $\Delta_2 < 0$ thus, indeed, there exists negative principal minor of even order. Also, as $\boldsymbol{T}(\tau)$ is a symmetric Toeplitz matrix, it is non-singular and thus $|\boldsymbol{T}(\tau)| \neq 0$. This implies that $\boldsymbol{T}(\tau)$ (for $1 < \tau \le N$) is indefinite. □

Based on Theorem 2.1 and the theory of quadratic forms for Gaussian processes, one can conclude that the EAM has the generalized $\chi^2$ distribution, i.e.

$$\widehat{AM}_X(\tau) \overset{d}{=} \sum_{i=1}^{N} \lambda_i(\tau) U_i, \tag{10}$$

where $\{U_i,\ i = 1, 2, \cdots, N\}$ are independent and identically distributed (i.i.d.) $\chi^2$ random variables with one degree of freedom, and $\lambda_i(\tau)$ are the eigenvalues of the matrix $\boldsymbol{T}(\tau)\boldsymbol{\Sigma}_Y$ where $\boldsymbol{\Sigma}_Y = [\gamma_Y(i,j),\ i,j = 1, 2, \cdots, N]$ is the covariance matrix of $\boldsymbol{Y}_N$. The quadratic form representation of EAM allows to calculate the moments of the statistic $\widehat{AM}(\tau)$ for each $\tau > 1$. Namely, according to Mathai and Provost (1992), if the statistic $S_N(\tau)$ has a general quadratic form given by Eq. (7) for the finite trajectory $\boldsymbol{X}_N$ of the zero-mean Gaussian process, then its r-th moment ($r \in \mathbb{N}/\{0\}$) can be calculated using the following recursive formula

$$\mathbb{E}[S_N^r(\tau)] = \sum_{k=0}^{r-1} \frac{(r-1)!}{k!} 2^{r-1-k} tr\left[(\boldsymbol{A}(\tau)\boldsymbol{\Sigma}_X)^{r-k}\right] \mathbb{E}[S_N^k(\tau)], \tag{11}$$

where $\boldsymbol{\Sigma}_X = [\gamma_X(i,j),\ i,j = 1,2,\cdots,N+1]$ is the covariance matrix of $\boldsymbol{X}_N$. Using the properties of the trace of a symmetric matrix, (Bretscher, 2005), one can obtain $tr\left[(\boldsymbol{T}(\tau)\boldsymbol{\Sigma}_X)^k\right] = \sum_{i=1}^N \lambda_i^k(\tau)$. Using the above-mentioned facts one can formulate the following Lemma.

**Lemma 2.1.** *For the finite trajectory $\boldsymbol{X}_{N+1}$ of the zero-mean Gaussian process with stationary increments, the r-th moment ($r \in \mathbb{N}_+$) of EAM defined in Eq. (5) can be calculated using the following recursive formula*

$$\mathbb{E}\left[\widehat{AM}_X^r(\tau)\right] = \sum_{k=0}^{r-1} \frac{(r-1)!}{k!} 2^{r-1-k} \mathbb{E}\left[\widehat{AM}_X^k(\tau)\right] \sum_{i=1}^N \lambda_i^{r-k}(\tau), \tag{12}$$

*where $\lambda_i(\tau)$ are eigenvalues of the matrix $\boldsymbol{T}(\tau)\boldsymbol{\Sigma}_Y$, $\boldsymbol{\Sigma}_Y$ is the covariance matrix of the increment trajectory $\boldsymbol{Y}_N$, and $\boldsymbol{T}(\tau)$ is the matrix given in Eq. (9).*

From Lemma 2.1 one can obtain

$$\mathbb{E}\left[\widehat{AM}_X(\tau)\right] = \sum_{i=1}^N \lambda_i(\tau), \quad \mathbb{E}\left[\widehat{AM}_X^2(\tau)\right] = 2\sum_{i=1}^N \lambda_i^2(\tau) + \mathbb{E}\left[\widehat{AM}_X(\tau)\right]^2. \tag{13}$$

And thus $Var\left[\widehat{AM}_X(\tau)\right] = 2\sum_{i=1}^N \lambda_i^2(\tau)$.

As it was shown in Maraj et al. (2021), one can also calculate $\mathbb{E}[\widehat{AM}_X(\tau)]$ using the definition of EAM given in Eq. (5), namely, for each $\tau = 1,2,\cdots,N$, we have $\mathbb{E}\left[\widehat{AM}_X(\tau)\right] = AM_X(\tau)$, where $AM_X(\tau)$ is given in Eq. (3). The above indicates that $\widehat{AM}_X(\tau)$ statistic is an unbiased estimator of $AM_X(\tau)$ for each $\tau = 1,2,\cdots,N$. On the other hand, using Eq. (13), one obtains that anomaly measure can also be calculated as $AM_X(\tau) = \sum_{i=1}^N \lambda_i(\tau)$, where $\lambda_i(\tau)$ are the eigenvalues of the matrix $\boldsymbol{T}(\tau)\boldsymbol{\Sigma}_Y$. From this fact, one can conclude that $\sum_{i=1}^N \lambda_i(\tau)$ is an increasing function with respect to $\tau$, and takes positive values when the process under consideration is super-diffusive, while it is decreasing and takes negative values if the process exhibits sub-diffusive behavior.

The knowledge about the distribution of the EAM statistic gives us the ability to calculate the exact forms of its cumulative distribution function and probability density function. Namely, for the generalized $\chi^2$ distributed random variable $\widehat{AM}_X(\tau)$, given in (10), the above mentioned functions are given by

$$CDF(t) = F_\tau(t) = P\left(\widehat{AM}_X(\tau) < t\right) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{it\omega}}{i\omega} \left(\prod_{k=1}^N (1 + 2i\lambda_k(\tau)\omega)^{-\frac{1}{2}}\right) d\omega, \tag{14}$$

$$PDF(t) = f_\tau(t) = \frac{d}{dt} F_\tau(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{it\omega} \left(\prod_{k=1}^N (1 + 2i\lambda_k(\tau)\omega)^{-\frac{1}{2}}\right) d\omega, \tag{15}$$

where $t \in \mathbb{R}$, $i = \sqrt{-1}$ and $\{\lambda_k(\tau)\}$ are the eigenvalues of the matrix $\boldsymbol{T}(\tau)\boldsymbol{\Sigma}_Y$.

In Fig. 1 we present the comparison of the empirical CDF of $\widehat{AM}_X(\tau)$ for FBM for three selected values of the Hurst exponent corresponding to sub-diffusive, diffusive and super-diffusive cases ($H = 0.2$, $H = 0.5$ and $H = 0.8$), calculated based on 1000 trajectories of length $N+1 = 2001$ with the theoretical CDF given in Eq. (14) for $\tau = 2$ and $\tau = 2000$. As we can see, the theoretical and empirical cumulative distribution functions are close to each other in all considered cases.

## 3. Statistical test for anomalous diffusion based on EAM

In this section, we present a possible application of the EAM for the statistical testing of the anomalous diffusion behavior. We propose a two-stage procedure: In the first step, we can test if the considered phenomenon corresponds to the anomalous diffusive regime. In the second step, we can statistically test if the given theoretical model of anomalous diffusive regime corresponds to analyzed data. Based only on the first stage, one can obtain information about the type of anomalous diffusion without the model specification. The second stage of the proposed procedure gives more precise information about the model as well as estimation of its parameters. From a practical point of view, both stages are important and can be implemented separately, depending on the problem under consideration.

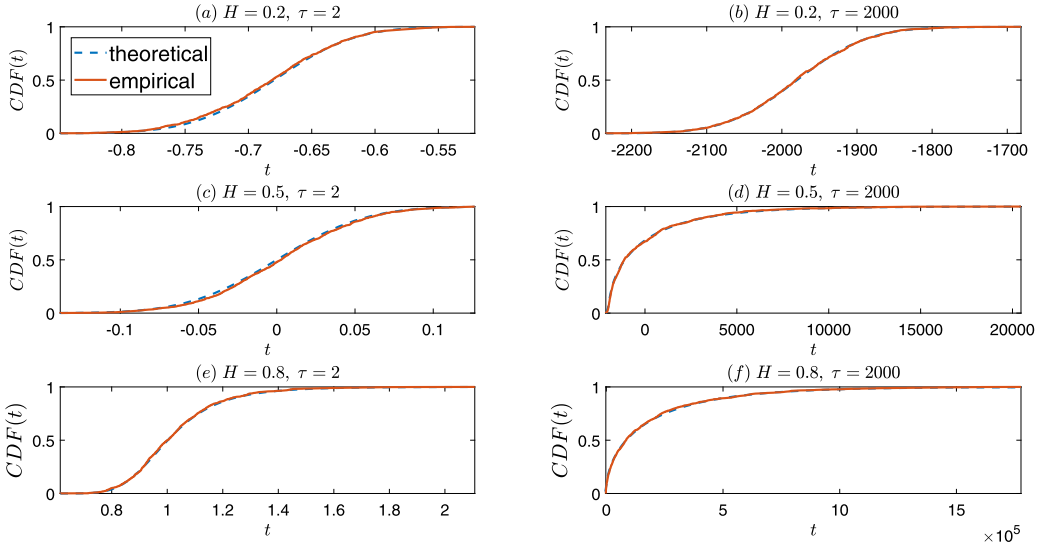


**Fig. 1.** The comparison of the empirical cumulative distribution function of $\widehat{AM}_X(\tau)$ for FBM for $H = 0.2$, $H = 0.5$ and $H = 0.8$ calculated based on 1000 trajectories of length $N + 1 = 2001$ with the theoretical CDF for small $\tau = 2$ and large $\tau = 2000$. Panels $(a)$ and $(b)$ show results for sub-diffusive case, panels $(e)$ and $(f)$ - for super-diffusive case while panels $(c)$ and $(d)$ correspond to normal diffusion.

### 3.1. First step of the testing procedure

In the first step, we assume that the $M$ trajectories of the same length $N + 1$ from the same experiment are available. We denote them as $\boldsymbol{X}^i_{N+1} = \{X_i(n),\ n = 1, 2, \ldots, N + 1\}$, $i = 1, 2, \ldots, M$. For each trajectory $\boldsymbol{X}^i_{N+1}$ we calculate the $\widehat{AM}_X(\tau)$ for $\tau = 2, \ldots, N$ according to Eq. (5). Finally, for each $\tau$ we calculate the empirical mean of the values $\widehat{AM}_X(\tau)$. The visual test proposed here indicates the type of anomaly (sub-diffusion, super-diffusion, or normal diffusion) corresponding to the examined phenomena. More precisely, if we observe the calculated empirical mean of the EAM statistic increases with respect to $\tau$ and takes positive values, one can infer the theoretical process describing the data belongs to the super-diffusive class of processes. If the empirical mean of the considered statistic is negative and decreases with $\tau$, we can assume it is a sub-diffusive case. If the considered value is close to zero and stabilizes, one can assume normal diffusion. In the next section, we analyze simulated data, and we demonstrate that even for a small number of the available trajectories ($M = 20$) we observe the behavior predicted above.

### 3.2. Second step of the testing procedure

At this stage, the testing procedure is based on one trajectory $\boldsymbol{X}_{N+1} = \{X(n),\ n = 1, 2, \ldots, N + 1\}$. The null hypothesis ($H_0$) is such that the analyzed vector of observations $\boldsymbol{X}_{N+1}$ forms a trajectory of a theoretical centered Gaussian process with stationary increments with specified values of the parameters. For a given $\tau$, the test statistic $\widehat{AM}_N(\tau)$ is calculated according to Eq. (5). We propose to apply a two-sided testing procedure. We proceed as follows: the $H_0$ hypothesis is rejected if the test statistic $\widehat{AM}_N(\tau)$ is extreme, either larger than an upper critical value or smaller than a lower critical value at a given significance level $\alpha$. The acceptance region of the test is given as follows

$$\left[Q_{\alpha/2}(\tau),\, Q_{1-\alpha/2}(\tau)\right], \tag{16}$$

where $Q_p(\tau)$ is the quantile of order $p$ of generalized $\chi^2$ distribution given in Eq. (10). If the value of the statistic $\widehat{AM}_N(\tau)$ falls into the acceptance region at a given significance level $\alpha$, we conclude that the test does not reject the $H_0$ hypothesis at significance level $\alpha$. The quantiles $Q_p(\tau)$ are calculated by using the exact formula of the CDF for the test statistic given in Eq. (14).

Using the exact formula for the CDF of the EAM statistic, one can calculate also the $p_{value}$ of the proposed test. The $p_{value}$ is defined as the largest probability, under the null hypothesis $H_0$, that the test statistic takes a value as extreme or more extreme than the value actually observed. In this paper, we consider the two-sided test based on the $\widehat{AM}_X(\tau)$ for specific value of $\tau$. Thus, in this case the $p_{value}$ is defined as

$$p_{value} = 2\min\left\{\mathrm{P}(\widehat{AM}_X(\tau) \le t|H_0), \mathrm{P}(\widehat{AM}_X(\tau) \ge t|H_0)\right\} = 2\min\left\{CDF(t), 1 - CDF(t)\right\}, \tag{17}$$

where $t$ is the observed value of the test statistic $\widehat{AM}_X(\tau)$ and $CDF(t)$ is the cumulative distribution function of $\widehat{AM}_X(\tau)$ under the null hypothesis $H_0$. In the considered case the $CDF(t)$ can be calculated using Eq. (14).

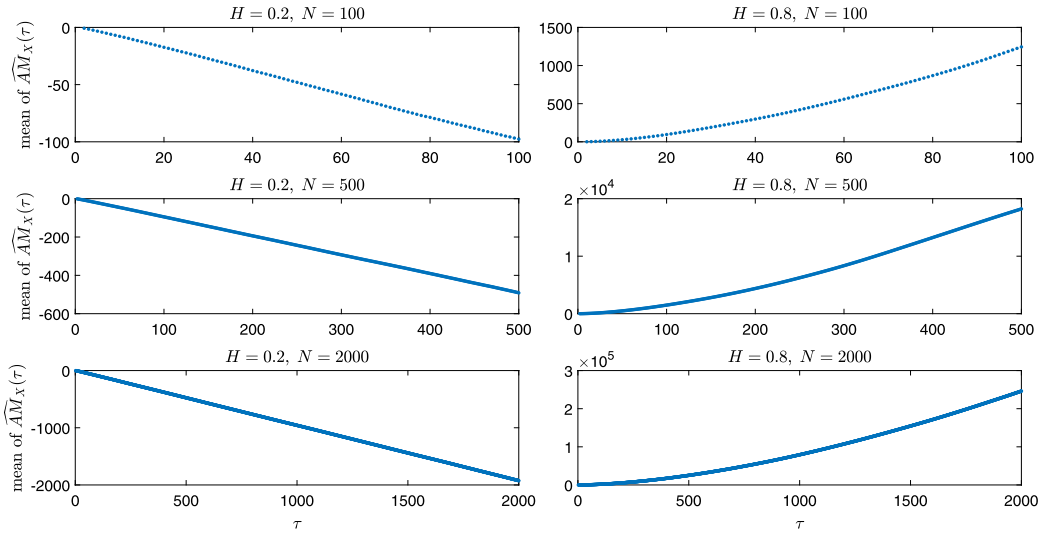


**Fig. 2.** Mean of $\widehat{AM}_X(\tau)$ for FBM with $H = 0.2$ and $H = 0.8$. We consider three lengths of trajectories $N \in \{100, 500, 2000\}$. In order to calculate the expected values for each trajectory length we simulated 20 trajectories of FBM with the corresponding $H$ parameter.

## 4. Simulation study

In this section, we demonstrate the effectiveness of the proposed testing procedure for the simulated data. In the first step of the testing method, we assume there are available $M$ trajectories of the same length corresponding to the examined phenomena. The visual test used here indicates the anomalous diffusion type (i.e. sub-diffusion, super-diffusion, or diffusion). In order to show the effectiveness of this algorithm, in Fig. 2 we show the empirical mean of EAM calculated for FBM with $H = 0.2$ and $H = 0.8$ for trajectories with length $N + 1 \in \{101, 501, 2001\}$. The empirical mean is calculated for $M = 20$ trajectories for all considered cases. The number of simulated trajectories is selected to be equal to the number of the real trajectories analyzed in Section 7.

As one can see, even for a small number of trajectories, for the sub-diffusive case (i.e. for $H = 0.2$, left panels in Fig. 2) one can see the mean of the statistic takes negative values and it is decreasing function with respect to $\tau$ while for the super-diffusive case (i.e. for $H = 0.8$, right panels of Fig. 2) we have the opposite situation, the mean of EAM is increasing function of $\tau$ and takes positive values only.

In order to demonstrate the critical cases, we consider also the FBM with Hurst index close to 0.5, namely $H \in \{0.45, 0.46, \cdots, 0.55\}$. In Fig. 3 we show the empirical mean (for $M = 20$ trajectories) of EAM calculated for the trajectory length $N + 1 \in \{101, 501, 2001\}$. It is easily seen that even for a short trajectory's length and index $H$ close to 0.5, one can distinguish the sub- and super-diffusive regimes. However, in some cases, the discrimination between anomalous diffusion regimes can be difficult. For the cases when $H < 0.5$, the mean of $\widehat{AM}_X(\tau)$ decreases and takes the negative values, while for $H > 0.5$ there is an opposite situation. For the normal diffusion (i.e. when $H = 0.5$) the mean of EAM stabilizes and is close to zero. This condition is especially visible for the long trajectories. This simple visual test will be used in the testing procedure for the preliminary knowledge about possible range of the parameters of the testing process. We have also applied it for the real data analysis examined below.

In the second step of the testing procedure, we examine separately each available trajectory. We test if the given trajectory corresponds to FBM with a specified value of the $H$ parameter. In all cases, we assume the diffusion coefficient $D = 1$. As the example of the testing $H$ parameter, we consider two cases, namely $H_0 = 0.2$ and $H_0 = 0.8$. In Fig. 4 we present the power of the test for small values of $\tau$ ($\tau = \{2, 3, \cdots, 11\}$). In the preliminary simulation studies, also the large values of the $\tau$ parameters were examined. However, the obtained results clearly indicate that for the small values of $\tau$ parameter the test based on EAM is more effective, see Maraj et al. (2021). Thus, in this part we present the results only for a small range of the $\tau$ parameter. In our considerations, we analyze the FBM for all possible values of the $H$ parameters. To get these results, we simulated 1000 trajectories of FMB with lengths $N \in \{100, 500, 2000\}$ for parameters $H \in \{0.01, 0.02, \cdots, 0.99\}$. We assume the significance level $\alpha = 0.05$. For all simulated trajectories, we have calculated the EAM statistic $\widehat{AM}_X(\tau)$ for considered values of $\tau$ parameters. Then, for a given $\tau$, we calculated the number of trajectories for which the corresponding value of EAM fails into the constructed acceptance region given in Eq. (16). In the further analysis of this number, we denote it as $M_1$. The power of the test is equal to $1 - M_1/M$. This is the value demonstrated in Fig. 4. As one can see, the longer the trajectory, the test is more restrictive. For short trajectories (panel ($a$) and ($d$)) the power of the test is equal to zero almost for all trajectories corresponding to $H \in [0.01, 0.3]$ for test with the null hypothesis $H_0 = 0.2$, and trajectories corresponding to $H \in [0.7, 0.9]$ for the null hypothesis with $H_0 = 0.8$. For longer trajectories, we get better results, the interval where we accept the null hypothesis in case it is false, is narrower. It is worth to mention, for a sub-diffusive case

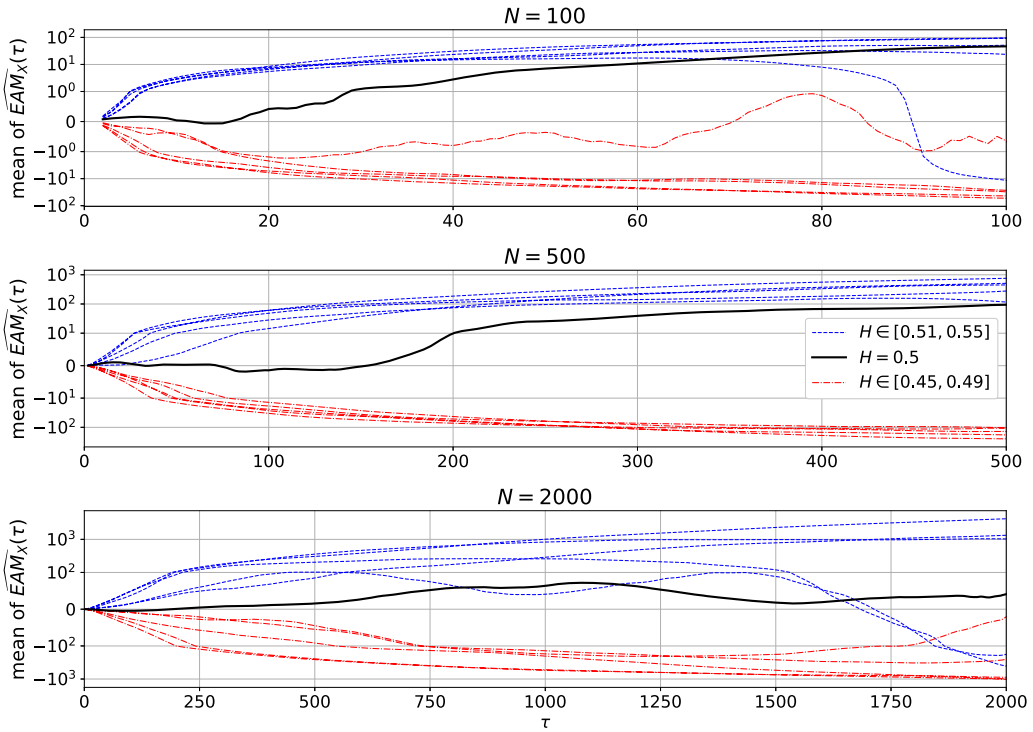


**Fig. 3.** Mean of $\widehat{AM}_X(\tau)$ for ordinary BM (i.e. FBM with $H = 0.5$) and FBM with Hurst index close to 0.5 presented in symmetric logarithmic scale. We consider three lengths of trajectories $N \in \{100, 500, 2000\}$. In order to calculate the expected values for each trajectory length we simulated 20 trajectories of FBM with the corresponding $H$ parameter.

(i.e. with $H_0 = 0.2$) we obtain better results than for $H_0 = 0.8$. For all lengths of trajectories, the interval where we accept the null hypothesis in case it is false is narrower for sub-diffusion. For $N = 100$, we can also observe that when $H_0 = 0.2$ the higher $\tau$, the better performance of the test. In the case with $H_0 = 0.8$, it is the opposite.

We should highlight that, the proposed test for FBM is only an example of using the EAM statistic. As it was mentioned, the proposed methodology can be used for any Gaussian zero-mean process with stationary increments. It seems the proposed testing procedure is very selective and sensitive for the changes of the Hurst index, even in the close neighborhood of the true value of this parameter.

## 5. Comparison with empirical autocovariance-based method

In this section, we compare the introduced testing procedure with the method proposed recently in Balcerek and Burnecki (2020), which is based on the empirical autocovariance function given in Eq. (6) for Gaussian processes. In Balcerek and Burnecki (2020), the authors analyzed the stationary zero-mean Gaussian processes. They proved that in this case the statistic (6) can be expressed as the quadratic form given in Eq. (7) of a Gaussian process with appropriate matrix $\boldsymbol{A}(\tau)$. Based on that, the authors could identify the distribution of the empirical autocovariance function and find the acceptance region for the testing problem if the given trajectory corresponds to the zero-mean stationary Gaussian process. The testing procedure is similar to the one demonstrated in this paper. Only one difference is related to different statistics used. In the paper Balcerek and Burnecki (2020) the test statistic is $\hat{\gamma}_Y(i)$ $i = 1, 2, \cdots, N-1$ defined in Eq. (6) while in the method presented in this paper this is $\widehat{AM}_X(\tau)$, $\tau = 1, 2, \cdots, N$ defined in Eq. (5). Clearly, both statistics are connected, i.e. the EAM statistic is just the linear combination of the empirical autocovariance function with appropriate weights. Moreover, for $\tau = 2$ we have $\widehat{AM}_X(2) = \hat{\gamma}_Y(1)$ where $\mathbf{Y}_N$ is the trajectory of increments of $\mathbf{X}_{N+1}$. To compare the effectiveness of both algorithms, we performed a similar simulation analysis, as we demonstrated in Section 4 (second step of the testing procedure), for the test proposed in Balcerek and Burnecki (2020). More precisely, for the same simulated trajectories corresponding to FBM analyzed in the previous section, we test FBM with two $H_0$ hypotheses, namely $H_0 = 0.2$ and $H_0 = 0.8$. In Fig. 5 we present the power of the test for small values of $i$ ($i = 1, 2, \cdots, 10$) and three considered trajectory lengths. The power is calculated in the same way as we presented in our methodology. For simulated 1000 trajectories of FMB with lengths $N \in \{100, 500, 2000\}$ for parameters $H \in \{0.01, 0.02, \cdots, 0.99\}$ we check if the calculated empirical autocovariance falls into the acceptance region calculated for the test statistic. In order to calculate the acceptance region, we used 10000 simulations of the random variables corresponding to the distribution of empirical autocovariance function for a given $i$. We assume the significance level $\alpha = 0.05$. Then, for a given $i$, we calculated the number of trajectories for which the

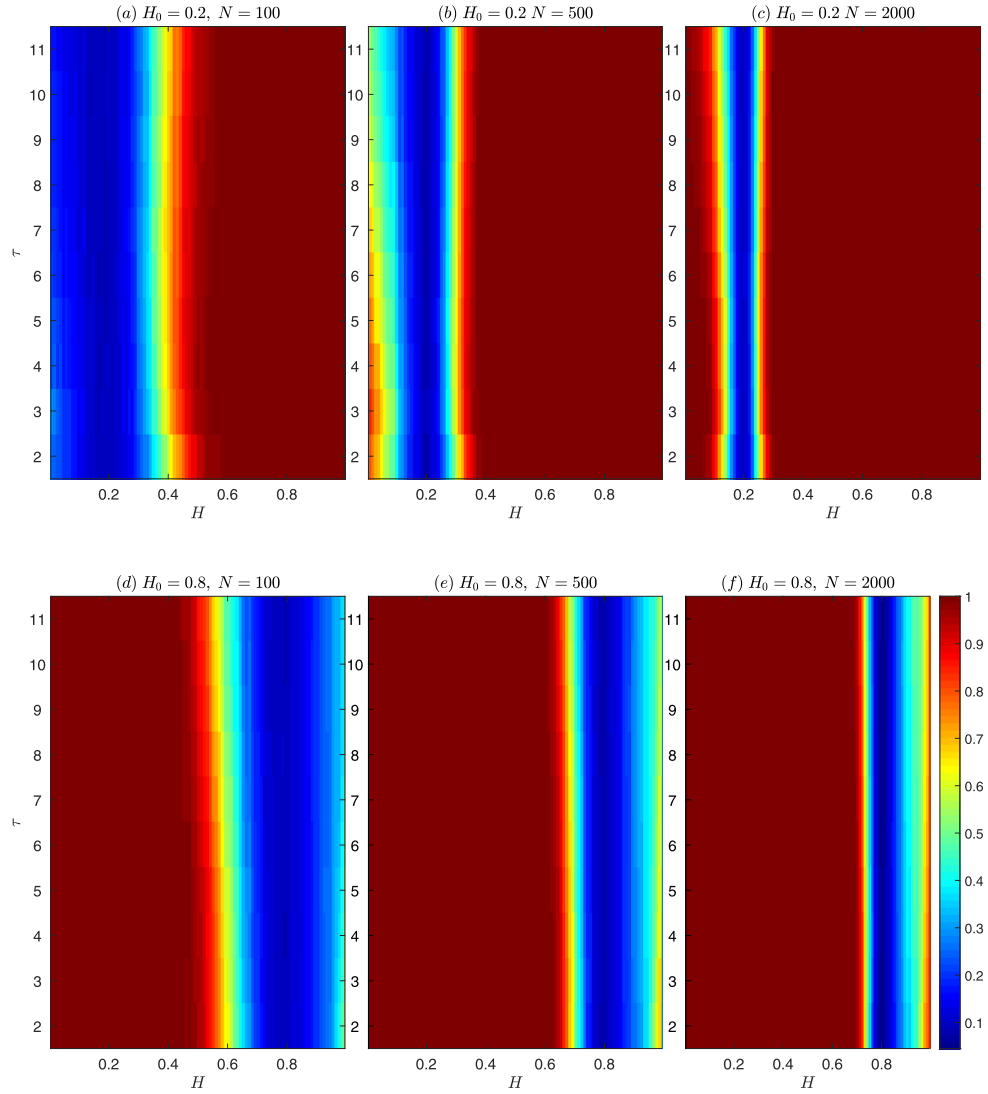


**Fig. 4.** The power of the test for two null hypotheses, namely FBM with $H_0 = 0.2$ (panels $(a)$–$(c)$) and FBM with $H_0 = 0.8$ (panels $(d)$–$(f)$) based on the $\widehat{AM}_X(\tau)$ statistic for $H \in \{0.01, 0.02, \cdots, 0.99\}$ for small values of $\tau$ parameters ($\tau = \{2, 3, \ldots, 11\}$) and three different trajectory lengths $N \in \{100, 500, 2000\}$. In order to calculate the power of the test for each trajectory length we simulated 1000 trajectories of FBM with the corresponding $H$ parameter. (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

corresponding value of the test statistic falls into the constructed acceptance region. We divide this number by the number of considered trajectories, i.e. by $M = 1000$.

Comparing Figs. 4 and 5 one can easily observe, the EAM-based test outperforms the empirical autocovariance-based test for $\tau > 2$ (or equivalently $i > 1$). As it was mentioned, for $\tau = 2$ and $i = 1$ the statistics have equal values. The test based on $\widehat{AM}_X(\tau)$ is more selective and reacts on small changes of $H$ parameter even for small trajectory lengths. This is intuitive as the EAM statistic contains the values of the empirical autocovariance function for appropriate lags and thus it contains more information about the process in contrast to $\hat{\gamma}_Y(i)$ for a specific value of $i$.

## 6. The holistic testing procedure for real trajectories

As it was described in the previous sections, the EAM statistic is useful for testing if experimental trajectories correspond to sub- or super-diffusion (first step of the testing procedure) or to test if given model with specific values of the parameters can be used to describe the real trajectory (second step of the testing procedure). In this section, we will discuss the practical aspects related to the second step of our procedure. Even if we suspect a specific theoretical model, in general, we do not know the corresponding theoretical parameters. Thus, in the second step of the testing procedure applied to real trajectories, we propose to test a given model for all possible parameters (for instance, FBM with any $H$ from the discretized

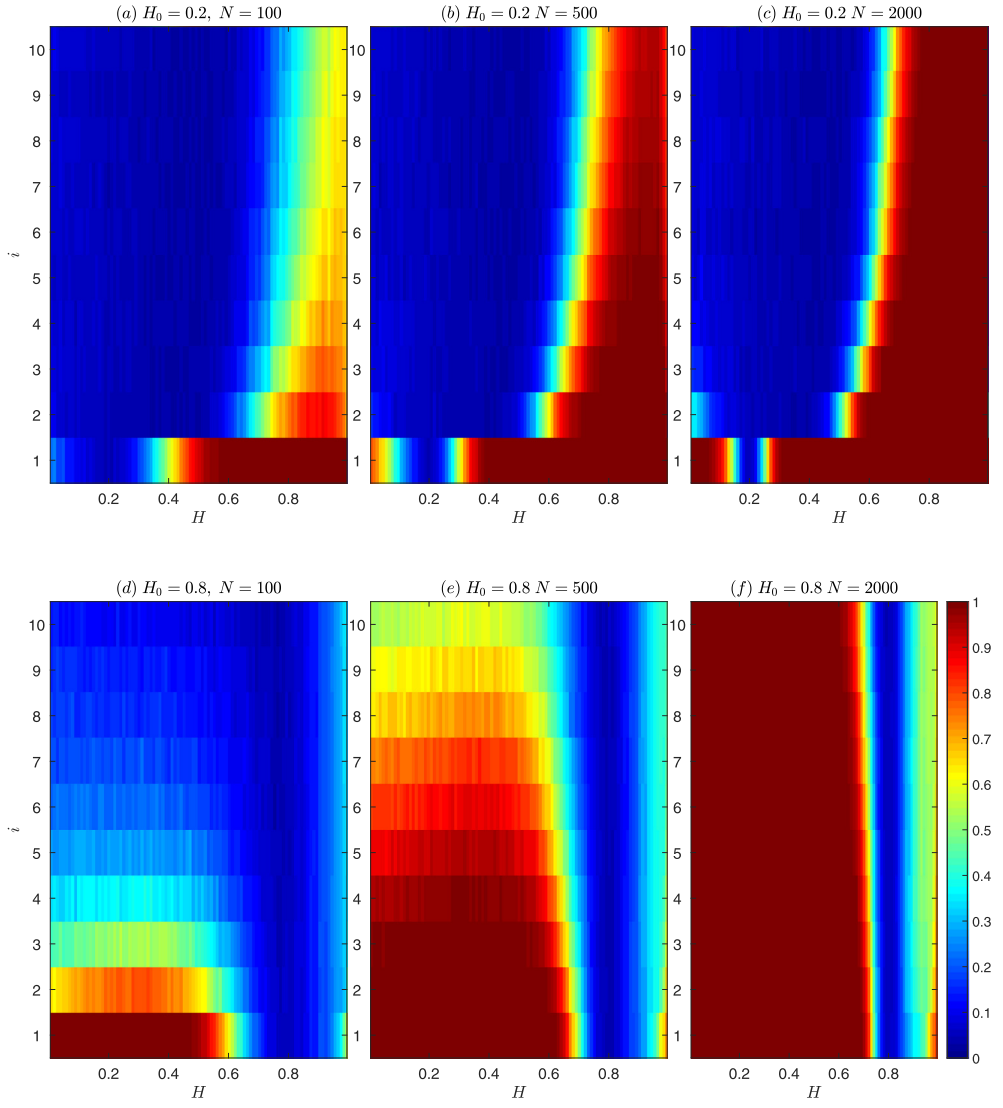


**Fig. 5.** The power of the test for two null hypotheses, namely FBM with $H_0 = 0.2$ (panels $(a)$–$(c)$) and FBM with $H_0 = 0.8$ (panels $(d)$–$(f)$) based on the $\hat{\gamma}_Y(i)$ statistic for $H \in \{0.01, 0.02, \cdots, 0.99\}$ for small values of $i$ parameters ($i = \{1, 2, \ldots, 10\}$) and three different trajectory lengths $N \in \{100, 500, 2000\}$. In order to calculate the power of the test for each trajectory length we simulated 1000 trajectories of FBM with the corresponding $H$ parameter. (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

interval $[0, 1]$). However, we need to take into consideration the influence of the $\tau$ parameter in the proposed approach. In the simulation study presented in the previous sections, we demonstrated the effectiveness of the statistical test based on EAM for selected values of the $\tau$ parameters. As it was shown, depending on the value of $\tau$ we could obtain slightly different effectiveness of the test. This is related to the fact, that although the $\widehat{AM}_X(\tau)$ statistic for a given zero-mean Gaussian process with stationary increments has generalized $\chi^2$ distribution, however, the weights in Eq. (10) are different for different values of $\tau$ values.

Because the domain and probability distribution of the EAM statistic depends (among model parameters) on the $\tau$ parameter, a key issue in the second step of the testing procedure is the selection of this parameter. Optimizing the choice of $\tau$ based on the CDF form (14) seems to be technically difficult because we do not know the exact formulas for eigenvalues in the representation (10) of the EAM statistic (in the case of FBM there are no such known formulas). Moreover, the choice of the optimal $\tau$ should depend on the parameters of the analyzed process (in the case of FBM from $H$ and $D$), which we do not know a priori. Therefore, keeping the application aspect and simplicity, we propose a testing algorithm with a holistic approach. It consists of testing the model for all possible parameters with $\widehat{AM}_X(\tau)$ for the selected $\tau$ values and ultimately accepting the hypothesis $H_0$ (about the model with specific parameters) for which the majority of the test statistics fell into the acceptance region. The holistic algorithm for FBM can be summarized as follows.

Step 1: Compute the set of all possible EAM values for trajectory at any possible $\tau$, so the set $T = \{\widehat{AM}_X(\tau) : \tau = 2, \dots, N\}$, where $N+1$ is the length of the trajectory. Set the testing confidence level $\alpha$.
Step 2: Set the grid of parameters of the considered model, e.g. for FBM it could be $H_j \in \{0.01, 0.02, \cdots, 0.99\}$.
Step 3: For each $j$ (equivalently for each $H_j$).
Step 3a: For each $\tau = 2, \dots, K$ compute the quantile region

$$\left[Q^j_{\alpha/2}(\tau), Q^j_{1-\alpha/2}(\tau)\right], \tag{18}$$

where $Q^j_p(\tau)$ is the quantile of order $p$ of the generalized $\chi^2$ distribution given in Eq. (10) with parameters depending on $H_j$ and $K$ is selected such that $2 \le K \le N$. It should be highlighted that in this approach $K-1$ values of $\tau$ are taken in the testing procedure. In the extreme case $K = N$, the testing proceeds for all possible $\tau$ values. The quantiles $Q^j_p(\tau)$ are calculated by using the exact formula of CDF for the test statistic given in Eq. (14). Equivalently Step 3a is the second step of the testing procedure described in Section 3.2 for any $\tau = 2, \dots, K$ and specific $H_j$.
Step 3b: Compute the average number of acceptance

$$T_j = \frac{1}{K-1}\sum_{\tau=2}^{K} \mathbb{1}\left(\widehat{AM}_X(\tau) \in \left[Q^j_{\alpha/2}(\tau), Q^j_{1-\alpha/2}(\tau)\right]\right). \tag{19}$$

Step 4: Accept the hypothesis of the model for trajectory with $H_j$ and

$$j = \arg_j \max\left\{T_j : T_j \ge 1-\epsilon\right\}, \tag{20}$$

where $\epsilon$ is a small number (chosen according to the domain knowledge about a specific data or experiment). If the set $\{T_j : T_j \ge 1-\epsilon\}$ is empty then reject the hypothesis about the testing model for the analyzed data. In case there is no evidence to reject the hypothesis about the testing model, the $H_j$ for which the $T_j$ takes maximum value we denote as $\widehat{H}$.

The complexity of the presented algorithm depends on the chosen grid of parameters of the model, for example for FBM it could be $H_j \in \{0.01, 0.02, \cdots, 0.99\}$, and on the choice of $K$, so the number of time lags for computing EAM statistic $\widehat{AM}_X(\tau)$. The statistical likelihood of the presented testing procedure depends on the choice of $\epsilon$ parameter, which can model the sharpness of the decision boundary about the rejecting/accepting of the conjectured hypothesis. The choice of that parameter should match the expert knowledge, reflect experimental realities and specific properties of the measured data.

### *6.1. The holistic testing procedure - simulation study*

In this part we present the simulation study demonstrating the effectiveness of the holistic procedure described above. To this end, we simulate $M = 1000$ trajectories of length $N = 100$ of the FBM. In our study we consider two $H_0$ values corresponding to the sub-diffusive and super-diffusive case, namely $H_0 = 0.2$ and $H_0 = 0.8$. Thus, finally we simulate $M = 1000$ trajectories from FBM with $H_0 = 0.2$ and $H_0 = 0.8$. For each simulated trajectory we use the holistic procedure described in Steps 1-3 for $K = N = 100$. In Fig. 6 we demonstrate the mean values of $T_j$ (along all trajectories) defined in Eq. (19), where $H_j \in \{0.01, 0.02, \cdots, 0.99\}$.

As one can see, for the sub-diffusive case, the maximum mean of $T_j$ values is reached for $H_j = 0.23$ while for the super-diffusive case - for $H_j = 0.82$. Thus, the estimated values are close to the theoretical ones. Moreover, the mean of $T_j$ values are close to 1 in the neighborhood of the tested $H_0$ values, which indicates for almost all $\tau$s the value of the EAM statistic falls into the constructed acceptance regions (18) when $H_j$ is close to the theoretical values 0.2 and 0.8, respectively. The simple simulation study indicates that the holistic procedure gives acceptable results and can be useful for real analysis.

## 7. Real data

In this part, we present the analysis of real data using the presented methodology. Namely, we investigate the dynamics of polystyrene microspheres embedded in agarose hydrogels. We recorded and tracked the motion of individual 50-nm beads in a 1.5% agarose gel, (Krapf et al., 2019). This single-particle tracking technique (Metzler et al., 2014a), is often referred to as passive microrheology, where the thermal motion of embedded probe particles is analyzed to extract the viscoelastic properties of the surrounding complex medium, (Levine and Lubensky, 2000). A 1.5% agarose gel was prepared from agarose powder (Genesee Scientific, San Diego, CA) by dissolving it in phosphate-buffered saline. Carboxylate-modified polystyrene microspheres with 50-nm nominal diameter (Bangs Laboratories, Fishers, IN) were first heated to 60 °C in 0.5% Tween 20 and introduced into the agarose solution. The solution was allowed to mix at 60 °C for 15 min, transferred to a hot glass-bottom Petri dish, and left to slowly cool down to room temperature. The microspheres were imaged in an inverted microscope with a 40x objective (Olympus PlanApo, N.A. 0.95) and a scientific CMOS camera (Andor Zyla 4.2) operated at

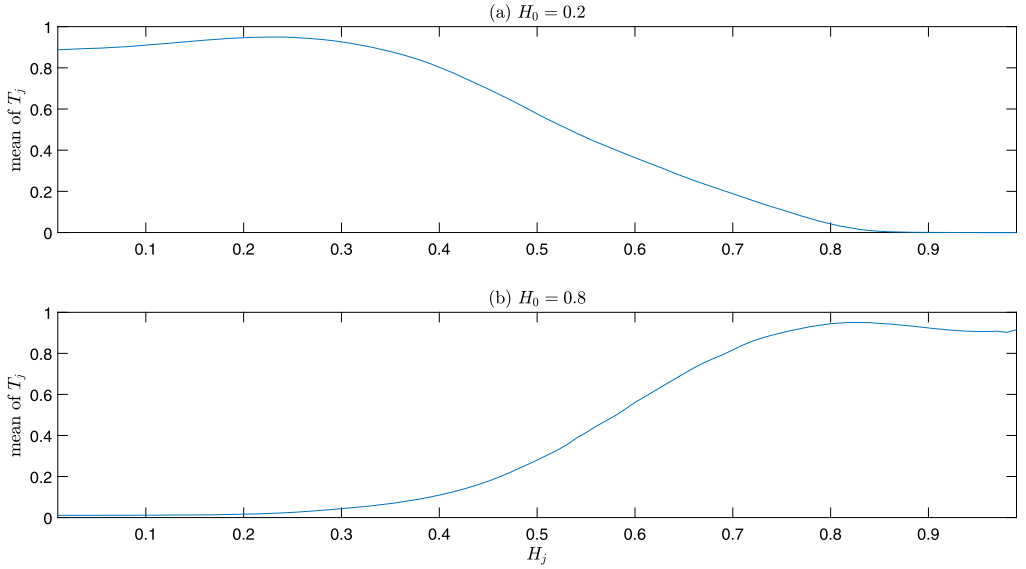


**Fig. 6.** The mean values of $T_j$ defined in Eq. (19) calculated for $M = 1000$ trajectories of FBM with $H_0 = 0.2$ (top panel) and $H = 0.8$ (bottom panel).

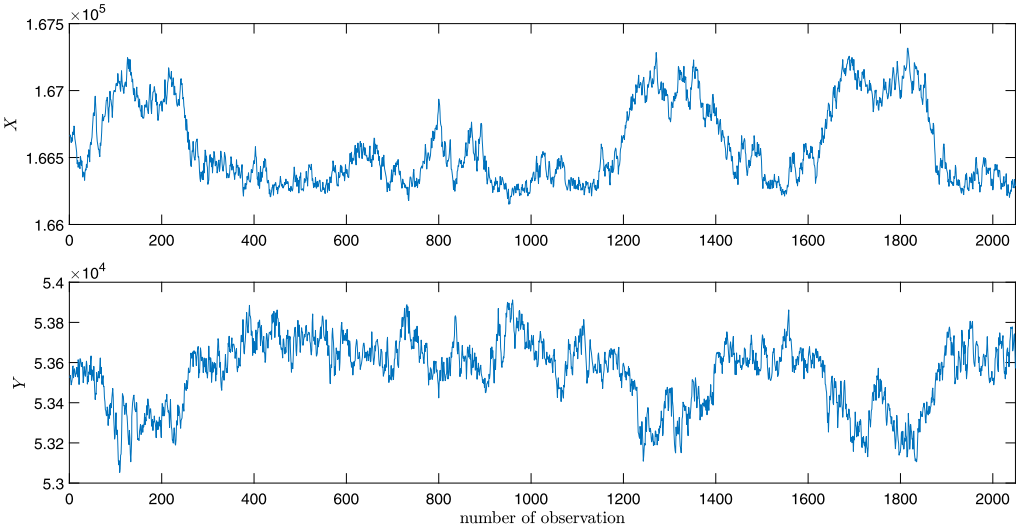


**Fig. 7.** Sample trajectories of $X$ and $Y$ coordinates.

71 frames/s, (Krapf et al., 2019, 2018). A total of 2048 images were recorded. The motion of 20 microspheres was tracked in two dimensions in LABVIEW using a crosscorrelation-based tracking algorithm (Gosse and Croquette, 2002).

In our study, we consider separately the $X$ and $Y$ coordinates. For each coordinate, there are 20 trajectories available. In each case, the trajectory length $N + 1 = 2049$. In Fig. 7 we show one sample trajectory corresponding to $X$ (top panel) and $Y$ (bottom panel) coordinates, respectively. Before the analysis, we subtract from the data the first observed value (in order to start the data from zero) and divide them by the square root of the sample second moment (standard deviation with sample mean equal to 0 as we presumed that the underlying process has mean 0) of the increment series (to consider the case $D = 1$). To all 20 trajectories corresponding to the $X$ and $Y$ coordinates, we have applied the first step of the testing procedure described in Section 3.1. More precisely, for each $\tau \in \{2, \cdots, N\}$ we calculated the $\widehat{AM}(\tau)$ statistic for each trajectory corresponding to $X$ and $Y$ coordinates. We denote them $\widehat{AM}_X(\tau)$ and $\widehat{AM}_Y(\tau)$, respectively. Finally, we present the mean (calculated for all trajectories) of the obtained values along the $\tau$ parameter. In Fig. 8 we demonstrate the mean of EAM statistic values for both coordinates.

The results presented in Fig. 8 indicate that the trajectories correspond to the sub-diffusive process, the mean of the EAM statistics are decreasing functions of $\tau$ and take negative values. In the second step of the testing procedure described in Section 6, we will check if the analyzed trajectories can be described by the FBM. Because the true value of the Hurst

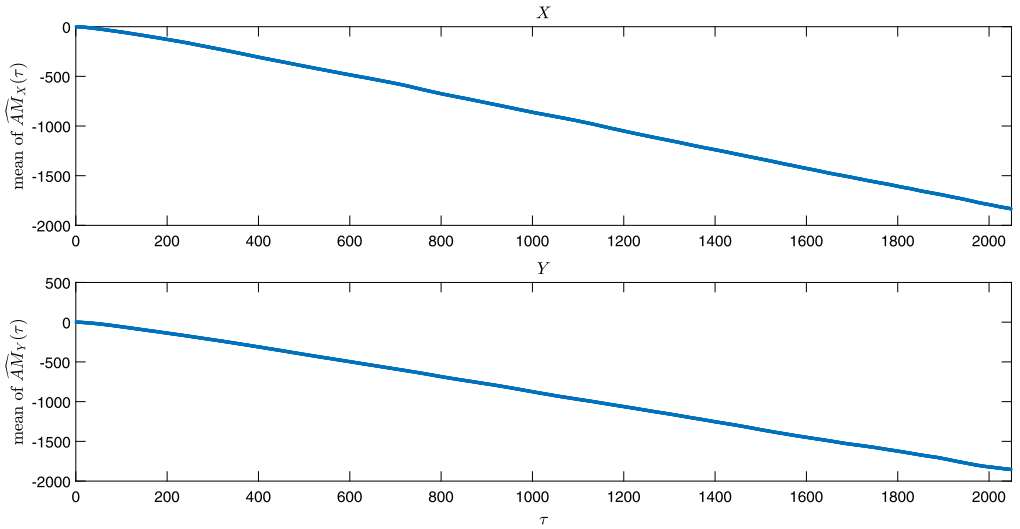


**Fig. 8.** The mean of $\widehat{AM}_X(\tau)$ and $\widehat{AM}_Y(\tau)$ for real trajectories corresponding to $X$ (top panel) and $Y$ (bottom panel) coordinates.

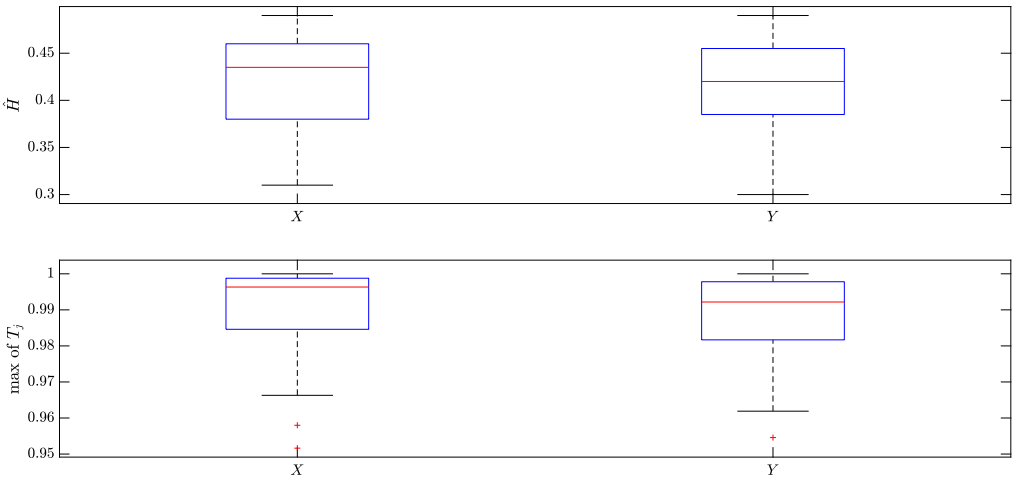


**Fig. 9.** Top panel: the boxplots of $H_j$ values (for all 20 trajectories) corresponding to $j$ satisfying Eq. (20) for X and Y coordinates. Bottom panel: the boxplots of corresponding maximum $T_j$ values for X and Y coordinates.

exponent is unknown, thus we have applied the holistic approach presented in Section 6. However, according to the first step we restrict ourselves only to the sub-diffusive case, thus finally we take $H_j \in \{0.01, 0.02, \cdots, 0.5\}$ and take $K = N$.

In Fig. 9 (top panel) we present the boxplots of $H_j$ values (for all 20 trajectories) corresponding to $j$ satisfying Eq. (20), i.e. the Hurst exponent for which the $T_j$ takes the maximum value that is close to 1. The "estimated" values of $H_j$ are denoted as $\widehat{H}$. Separately we demonstrate the results for $X$ and $Y$ coordinates. On the bottom panel of Fig. 9, we demonstrate the boxplots of corresponding maximum $T_j$ values. One can see, the maximum $T_j$s are close to 1 that indicates for almost all $\tau$s the value of the EAM statistic falls into the constructed acceptance regions. Thus, we can not reject the preliminary assumption of FBM. The median of the $\widehat{H}$ for X coordinate is equal to 0.435 while for $Y$ coordinate - 0.420. This confirms the sub-diffusive regime for both coordinates.

The final conclusion is that we can not reject the hypothesis that the analyzed real data correspond to the FBM with a sub-diffusive regime. Using only the first step of the testing procedure, we could recognize the anomalous type without the model specification. The second step gives us more detailed information about the theoretical model behind the data. Moreover, we could also identify the corresponding estimated Hurst exponent.

Our results are consistent with agarose hydrogels being a material with viscoelastic properties, (Ahearne et al., 2005), and, thus, the particles embedded in them are expected to exhibit sub-diffusion. A previous study of 50 nm microspheres in 1.5% agarose gel showed that the power spectral density of these particles matches that of FBM with a mean Hurst exponent $H = 0.36$, (Krapf et al., 2019). The results presented here agree with these previous results and provide unequivocal evidence

to assess anomalous diffusion, even in the challenging scenario where the Hurst exponent is close to 0.5, i.e., close to the normal diffusion case.

## 8. Conclusions

In this paper, we have studied anomalous diffusion dynamics for Gaussian processes differing from the classical BM normal diffusion scenario. In this analysis, we have applied the statistic called the empirical anomaly measure, recently introduced in Maraj et al. (2021). In this paper, we have analyzed the probabilistic properties of the EAM using its quadratic form representation. The mathematical foundations of the EAM properties were the basis for the introduction of the statistical test for anomalous diffusion behavior detection. The proper identification of anomalous regime is crucial in various real applications. Thus, the considered problem seems to be of great value.

The provided simulation study indicates the effectiveness of the proposed approach. We demonstrate the results for the FBM, the classical Gaussian process with anomalous diffusive behavior. However, the proposed methodology is universal and the theoretical probabilistic properties are proved for any Gaussian process with stationary increments. Thus, the testing procedure can be easily extended to any process with such properties.

As it was mentioned, the EAM arises as a linear combination of the sample ACVF of the increment process for any value of its arguments and the appropriate lags, thus in some sense, it can be considered as the generalization of the sample ACVF. Thus, in our simulation study, we compare the results of the testing procedure (the second step of the methodology) for the EAM and the test presented in Balcerek and Burnecki (2020) based on the sample ACVF.

Finally, we demonstrate a holistic approach that can be applied by practitioners dealing with anomalous diffusive data. In this approach, all values of EAM are considered, thus the whole information about the process is taken into account in the testing procedure. We illustrate the methodology for real trajectories describing the motion of polystyrene beads in agarose hydrogels. We found that these particles exhibit sub-diffusion in a manner consistent with FBM, which points to the viscoelastic properties of the medium.

In a future study, we plan to extend the methodology for the more general case, namely, for general Gaussian processes without the assumption of stationary increments. One of the examples is the process, called scaled Brownian motion, (Bodrova et al., 2019). However, in this case, the ACVF (and thus sample ACVF) depends on two arguments, thus the methodology needs to be carefully modified.

In the theoretical considerations of the introduced methodology, there is a strong assumption of the Gaussianity of the underlying process. Unfortunately, if anomalous diffusion is studied, this assumption is an important limitation. There are examples of real phenomena, where the "pure" FBM is not a proper model and rather its various extensions are more suitable. One of these extensions is the superstatistical FBM, see e.g. in Molina-García et al. (2016); Maćkała and Magdziarz (2019); Itto and Beck (2021), considered as an anomalous diffusion model appropriate for living systems. Further, the superstatistical FBM is the FBM with the random diffusion coefficient. In the literature, various distributions of the $D$ parameter (see Eq. (1)) are considered, see e.g. Maćkała and Magdziarz (2019). The superstatistical FBM is not a Gaussian process and, thus, the theoretical results presented in this paper cannot be directly applied. However, the methodology based on the EAM statistic can still be used. In Maraj et al. (2021) the methodology based on EAM was demonstrated for general second-order anomalous diffusion processes. Thus, the only assumption that was crucial there was the existence of the second moment of the underlying process. This assumption is satisfied when the diffusion coefficient is a finite-variance random variable. In Maraj et al. (2021) it was demonstrated that the EAM statistic for any second-order sub-diffusive process is a decreasing function and takes negative values, while for the super-diffusive case, it is an increasing and positive-valued function. To prove this special property of EAM the assumption of Gaussian distribution of the process was not needed. Thus, in order to determine the anomalous diffusion behavior and the type of anomaly, the EAM-based methodology can be used. However, to introduce the statistical test for superstatistical FBM based on EAM, the corresponding acceptance region (or equivalently $p_{value}$) should correspond to the distribution of the EAM statistic for the considered process. In this case, the theory of quadratic forms for Gaussian processes cannot be applied. As it was demonstrated in Maraj et al. (2021), the acceptance region can also be calculated using simulated trajectories corresponding to the tested process (with given values of the parameters). Thus, the presented methodology can also be extended also for processes that are based on FBM. In future studies, we plan to analyze the theoretical properties of EAM for processes based on FBM, such a superstatistical FBM.

The next possible extension is the adaptation of the methodology to the anomalous diffusion processes with infinite variance, such as non-Gaussian $\alpha-$stable-based processes. However, in this case, the theoretical ACVF is not defined, thus it needs to be replaced by alternative dependency measures adequate for infinite-variance processes, see for instance Samorodnitsky and Taqqu (1994). The processes with infinite variance are widely discussed in the context of anomalous diffusion, see e.g. Campagnola et al. (2015); Janczura et al. (2011), thus it seems to be reasonable to consider also this approach.

The last possible extension is related to the multidimensional processes analysis and the proper modification of the EAM in such a case. The ACVF in Eq. (3) needs to be replaced by cross-dependence measures, like cross-covariance for finite-variance processes or alternative cross-dependency measures for infinite-variance processes, see e.g. Grzesiek et al. (2019, 2020). This direction seems to be very promising, as the experimental data in many cases are two- or three-dimensional. In the future research, we also plan to include the intelligent methods into the classical statistical methodology and use the novel approach for the anomalous diffusion processes analysis, see e.g. Szarek et al. (2020); Szarek (2021). This approach

seems to be especially important when the process under consideration is disturbed by the additive noise that affects the testing results.

## Acknowledgements

The work of A.W. was supported by National Center of Science under Opus Grant No. 2020/37/B/HS4/00120 “Market risk model identification and validation using novel statistical, probabilistic, and machine learning tools”.

## References

Ahearne, M., Yang, Y., El Haj, A.J., Then, K.Y., Liu, K.K., 2005. Characterizing the viscoelastic properties of thin hydrogel-based constructs for tissue engineering applications. J. R. Soc. Interface 2, 455–463.
Arias Velásquez, R.M., Mejía Lara, J.V., 2020. Gaussian approach for probability and correlation between the number of COVID-19 cases and the air pollution in Lima. Urban Clim. 33, 100664.
Bachoc, F., Bevilacqua, M., Velandia, D., 2019. Composite likelihood estimation for a Gaussian process under fixed domain asymptotics. J. Multivar. Anal. 174, 104534.
Balcerek, M., Burnecki, K., 2020. Testing of fractional Brownian motion in a noisy environment. Chaos Solitons Fractals 140, 110097.
Beck, C., Cohen, E.G.D., 2003. Superstatistics. Physica A 322, 267–275.
Beran, J., 1994. Statistics for Long-Memory Processes. Chapman & Hall, New York.
Beran, J., Möhrle, S., Ghosh, S., 2016. Testing for Hermite rank in Gaussian subordination processes. J. Comput. Graph. Stat. 25, 917–934.
Bodrova, A.S., Chechkin, A.V., Sokolov, I.M., 2019. Scaled Brownian motion with renewal resetting. Phys. Rev. E 100, 012120.
Bretscher, O., 2005. Linear Algebra with Applications, 5th edition. Prentice Hall.
Campagnola, G., Nepal, K., Schroder, B., Peersen, O.B., Krapf, D., 2015. Superdiffusive motion of membrane-targeting C2 domains. Sci. Rep. 5, 17721.
Chechkin, A.V., Seno, F., Metzler, R., Sokolov, I.M., 2017. Brownian yet non-Gaussian diffusion: from superstatistics to subordination of diffusing diffusivities. Phys. Rev. X 7, 021002.
Coeurjolly, J.F., Porcu, E., 2018. Fast and exact simulation of complex-valued stationary Gaussian processes through embedding circulant matrix. J. Comput. Graph. Stat. 27, 278–290.
Davis, C.B., Hans, C.M., Santner, T.J., 2021. Prediction of non-stationary response functions using a Bayesian composite Gaussian process. Comput. Stat. Data Anal. 154, 107083.
Ernst, D., 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, 3429–3432.
Ernst, P.A., Brown, L.D., Shepp, L., Wolpert, R.L., 2017. Stationary Gaussian Markov processes as limits of stationary autoregressive time series. J. Multivar. Anal. 155, 180–186.
Fuliński, A., 2011. Anomalous diffusion and weak nonergodicity. Phys. Rev. E 83, 061140.
Gal, N., Lechtman-Goldstein, D., Weihs, D., 2013. Particle tracking in living cells: a review of the mean square displacement method and beyond. Rheol. Acta 52, 425–443.
Gerber, F., Nychka, D.W., 2021. Parallel cross-validation: a scalable fitting method for Gaussian process models. Comput. Stat. Data Anal. 155, 107113.
Gosse, C., Croquette, V., 2002. Magnetic tweezers: micromanipulation and force measurement at the molecular level. Biophys. J. 82, 3314–3329.
Grzesiek, A., Sikora, G., Teuerle, M., Wyłomańska, A., 2020. Spatio-temporal dependence measures for alpha-stable bivariate AR(1) models. J. Time Ser. Anal. 41 (3), 454–475.
Grzesiek, A., Teuerle, M., Wyłomańska, A., 2019. Cross-codifference for bidimensional VAR(1) time series with infinite variance. Commun. Stat., Simul. Comput.
Höfling, F., Franosch, T., 2013. Anomalous transport in the crowded world of biological cells. Rep. Prog. Phys. 76 (4), 046602.
Horn, R.A., Johnson, C.R., 2013. Matrix Analysis, 2nd ed. Cambridge University Press.
Itto, Y., Beck, C., 2021. Superstatistical modelling of protein diffusion dynamics in bacteria. J. R. Soc. Interface 18, 20200927.
Janczura, J., Orzeł, S., Wyłomańska, A., 2011. Subordinated alpha-stable Ornstein-Uhlenbeck process as a tool of financial data description. Physica A 390, 4379–4387.
Kepten, E., Weron, A., Sikora, G., Burnecki, K., Garini, Y., 2015. Guidelines for the fitting of anomalous diffusion mean square displacement graphs from single particle tracking experiments. PLoS ONE 10, e0117722.
Klages, R., Radons, G., Sokolov, I.M. (Eds.), 2008. Anomalous Transport: Foundations and Applications. Wiley.
Kolmogorov, A., 1940. Wienersche Spiralen und einige andere interessante Kurven im HilbertschenRaum. Proc. Acad. Sci. USSR 26, 115–118.
Krapf, D., 2015. Mechanisms underlying anomalous diffusion in the plasma membrane. Curr. Top. Membr. 75, 167–207.
Krapf, D., Campagnola, G., Nepal, K., Peersen, O.B., 2016. Strange kinetics of bulk-mediated diffusion on lipid bilayers. Phys. Chem. Chem. Phys. 18, 12633.
Krapf, D., Lukat, N., Marinari, E., Metzler, R., Oshanin, G., Selhuber-Unkel, C., Squarcini, A., Stadler, L., Weiss, M., Xu, X., 2019. Spectral content of a single non-Brownian trajectory. Phys. Rev. X 9, 011019.
Krapf, D., Marinari, E., Metzler, R., Oshanin, G., Xu, X., Squarcini, A., 2018. Power spectral density of a single Brownian trajectory: what one can and cannot learn from it. New J. Phys. 20, 023029.
Krapf, D., Metzler, R., 2019. Strange interfacial molecular dynamics. Phys. Today 72, 48–54.
Kusmierz, L., Majumdar, S.N., Sabhapandit, S., Schehr, G., 2014. First order transition for the optimal search time of Lévy flights with resetting. Phys. Rev. Lett. 113, 220602.
Lee, T., Bilionis, I., Tepole, A.B., 2020. Propagation of uncertainty in the mechanical and biological response of growing tissues using multi-fidelity Gaussian process regression. Comput. Methods Appl. Mech. Eng. 359, 112724.
Levine, A.J., Lubensky, T.C., 2000. One- and two-particle microrheology. Phys. Rev. Lett. 85, 1774–1777.
Li, X., Yuan, C., Li, X., Wang, Z., 2020. State of health estimation for Li-ion battery using incremental capacity analysis and Gaussian process regression. Energy 190, 116467.
Maćkała, A., Magdziarz, M., 2019. Statistical analysis of superstatistical fractional Brownian motion and applications. Phys. Rev. E 99, 012143.
Magdziarz, M., Weron, A., 2007. Fractional Fokker-Planck dynamics: stochastic representation and computer simulation. Phys. Rev. E 75, 056702.
Maraj, K., Szarek, D., Sikora, G., Wyłomańska, A., 2021. Empirical anomaly measure for finite-variance processes. J. Phys. A, Math. Theor. 54, 024001.
Mariñas-Collado, I., Bowman, A., Macaulay, V., 2019. A phylogenetic Gaussian process model for the evolution of curves embedded in d-dimensions. Comput. Stat. Data Anal. 137, 285–298.
Mathai, A.M., Provost, S.B., 1992. Quadratic Forms in Random Variables: Theory and Applications. Marcel Dekker, New York.
Metzler, R., Jeon, J.H., Cherstvy, A.G., Barkai, E., 2014a. Anomalous diffusion models and their properties: non-stationarity, non-ergodicity, and ageing at the centenary of single particle tracking. Phys. Chem. Chem. Phys. 16, 24128–24164.

Metzler, R., Jeon, J.H., Cherstvy, A.G., Barkai, E., 2014b. Anomalous transport in the crowded world of biological cells. Phys. Chem. Chem. Phys. 16, 24128–24164.
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.
Michalet, X., Berglund, A.J., 2012. Optimal diffusion coefficient estimation in single-particle tracking. Phys. Rev. E 85, 061916.
Molina-García, D., Pham, T.M., Paradisi, P., Manzo, C., Pagnini, G., 2016. Fractional kinetics emerging from ergodicity breaking in random media. Phys. Rev. E 94, 052147.
Montroll, E.W., Weiss, G.H., 2017. Random walks on lattices. II. J. Math. Phys. 6, 167–181.
Oliveira, F.A., Ferreira, R.M., Lapas, L.C., Vainstein, M.H., 2019. Anomalous diffusion: a basic mechanism for the evolution of inhomogeneous systems. Front. Phys. 7, 7–18.
Sabri, A., Xu, X., Krapf, D., Weiss, M., 2020. Elucidating the origin of heterogeneous anomalous diffusion in the cytoplasm of mammalian cells. Phys. Rev. Lett. 125, 058101.
Samorodnitsky, G., Taqqu, M., 1994. Stable Non-Gaussian Random Processes. Chapman & Hall, New York.
Scher, H., 2017. Continuous time random walk (CTRW) put to work. Eur. Phys. J. B 90, 252.
Sikora, G., 2018. Statistical test for fractional Brownian motion based on detrending moving average algorithm. Chaos Solitons Fractals 114, 54–62.
Sikora, G., Burnecki, K., Wyłomańska, A., 2017a. Mean-squared displacement statistical test for fractional Brownian motion. Phys. Rev. E 95, 032110.
Sikora, G., Teuerle, M., Wyłomańska, A., Grebenkov, D., 2017b. Statistical properties of the anomalous scaling exponent estimator based on time averaged mean square displacement. Phys. Rev. E 96, 022132.
Sokolov, I.M., 2012. Models of anomalous diffusion in crowded environments. Soft Matter 8, 9043–9052.
Stroud, J.R., Stein, M.L., Lysen, S., 2017. Bayesian and maximum likelihood estimation for Gaussian processes on an incomplete lattice. J. Comput. Graph. Stat. 26, 108–120.
Szarek, D., 2021. Neural network-based anomalous diffusion parameter estimation approaches for Gaussian processes. Int. J. Adv. Eng. Sci. Appl. Math. 13, 257–269.
Szarek, D., Sikora, G., Balcerek, M., Jabłoński, I., Wyłomańska, A., 2020. Fractional dynamics identification via intelligent unpacking of the sample autocovariance function by neural networks. Entropy 22, 1322.
Thapa, S., Wyłomańska, A., Sikora, G., Wagner, C.E., Krapf, D., Kantz, H., Chechkin, A.V., Metzler, R., 2021. Leveraging large-deviation statistics to decipher the stochastic properties of measured trajectories. New J. Phys. 23, 013008.
Vilk, O., Orchan, Y., Charter, M., Ganot, N., Toledo, S., Nathan, R., Assaf, M., 2021. Ergodicity breaking and lack of a typical waiting time in area-restricted search of avian predators. arXiv preprint, arXiv:2101.11527.
Wang, B., Xu, A., 2019. Gaussian process methods for nonparametric functional regression with mixed predictors. Comput. Stat. Data Anal. 131, 80–90.
Wei, W., Balabdaoui, F., Held, L., 2017. Calibration tests for multivariate Gaussian forecasts. J. Multivar. Anal. 154, 216–233.
Weigel, A.V., Ragi, S., Reid, M.L., Chong, E.K.P., Tamkun, M.M., Krapf, D., 2012. Obstructed diffusion propagator analysis for single-particle tracking. Phys. Rev. E 85, 041924.
Weigel, A.V., Simon, B., Tamkun, M.M., Krapf, D., 2011. Ergodic and nonergodic processes coexist in the plasma membrane as observed by single-molecule tracking. Proc. Natl. Acad. Sci. USA 108, 6438–6443.
Zhou, J., 2011. Maximum likelihood ratio test for the stability of sequence of Gaussian random processes. Comput. Stat. Data Anal. 55, 2114–2127.
Zhou, Y., Xiao, Y., 2018. Joint asymptotics for estimating the fractal indices of bivariate Gaussian processes. J. Multivar. Anal. 165, 56–72.