Fluid heterogeneity detection based on the asymptotic distribution of the time-averaged mean squared displacement in single particle tracking experiments

Kui Zhang, Katelyn P R Crizer, Mark H Schoenfisch, David B Hill and Gustavo Didier

1 Department of Mathematics, Tulane University, New Orleans, LA 70118, United States of America
2 The Marsico Lung Institute and Department of Physics and Astronomy, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, United States of America
3 Department of Chemistry, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, United States of America

E-mail: gidier@tulane.edu

Received 11 May 2018, revised 27 August 2018
Accepted for publication 12 September 2018
Published 3 October 2018

Abstract

A tracer particle is called anomalously diffusive if its mean squared displacement grows approximately as $\sigma^2 t^\alpha$ as a function of time $t$ for some constant $\sigma^2$, where the diffusion exponent satisfies $\alpha \neq 1$. In this article, we use recent results on the asymptotic distribution of the time-averaged mean squared displacement (Didier and Zhang 2017 J. Time Ser. Anal. 38 395–416) to construct statistical tests for detecting physical heterogeneity in viscoelastic fluid samples starting from one or multiple observed anomalously diffusive paths. The methods are asymptotically valid for the range $0 < \alpha < 3/2$ and involve a mathematical characterization of time-averaged mean squared displacement bias and the effect of correlated disturbance errors. The assumptions on particle motion cover a broad family of fractional Gaussian processes, including fractional Brownian motion and many fractional instances of the generalized Langevin equation framework. We apply the proposed methods in experimental data from treated Pseudomonas aeruginosa biofilms generated by the collaboration of the Hill and Schoenfisch Labs at UNC-Chapel Hill.

Keywords: anomalous diffusion, mean squared displacement, asymptotic distribution, viscoelastic fluid, fluid heterogeneity

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

In this paper, we start from the asymptotic distribution of the time-averaged mean squared displacement of nanometric tracer particles [20] to construct statistical protocols for detecting physical fluid heterogeneity. The assumptions on particle motion cover a broad family of fractional Gaussian processes, including fractional Brownian motion and many instances of the generalized Langevin equation framework. The testing protocols allowed providing more accurate quantitative analysis of experimental data from the Hill and Schoenfisch Labs (UNC-Chapel Hill), and the results reported in [73] were generally confirmed.

Improvements in light microscopy, fluorescence techniques, nanoparticle synthesis and high-speed video have ushered in a flurry of experimental activity [81]. Single particle tracking has become a common tool in many scientific areas, such as colloid physics [34], the microrheology of complex fluids [35, 45, 54, 83] and the study of nanobiophysical systems, both in vivo and in vitro [2, 12]. This includes the diffusion of single molecules, e.g. proteins, on biopolymers such as DNA or microtubules, on surfaces or in lipid membranes, inside in vivo cells and in actin solutions [4, 26, 30, 32, 33, 46, 61, 67, 80, 82, 84, 89, 95], as well as two-dimensional biological membranes [58] and heterogeneous tracer diffusion and first passage characteristics in two-dimensional crowded environments [23].

Of primary concern in the analysis of particle path data is the ensemble mean squared displacement (MSD), where $X$ is the tracer particle’s position. A basic dynamic characterization of the latter is given by the relation

$$\langle X^2(t) \rangle = \sigma^2 t^\alpha, \quad \sigma^2, \alpha > 0, \quad t \geq 0, \quad \xi := (\log \sigma^2, \alpha).$$  \hspace{1cm} (1)

In (1), $\alpha$ is the diffusion exponent and $\sigma^2 = 2D$, where $D$ is the diffusivity constant. The parameter value $\alpha = 1$ corresponds to classical diffusion. If $\alpha \neq 1$, the stochastic process $X$ is said to be anomalously diffusive, more specifically sub- or superdiffusive depending on whether $\alpha < 1$ or $\alpha > 1$, respectively. Anomalous diffusion may emerge, for example, as a consequence of binding-unbinding events, of geometrical constraints on the particle’s movement, or of fluid viscoelasticity [49, 57, 76, 77].

The dominant statistical technique in the biophysical literature for estimating the parameters $\sigma^2$ and $\alpha$ is based on the so-named time-averaged mean squared displacement (TAMSD). Suppose that a single particle experiment generates a tracer bead sample path with observations $X(j)$, $j = 1, \ldots, N$. The pathwise statistic

$$M_N(\tau) := \frac{1}{N-\tau} \sum_{j=1}^{N-\tau} \{X(j+\tau) - X(j)\}^2$$

is the TAMSD at lag value $\tau$, i.e. the statistical counterpart of the MSD $\langle X^2(\tau) \rangle$. One generates an estimator of $\xi = (\log \sigma^2, \alpha)$ by means of the linear regression

$$\log M_N(\tau_k) = \log \sigma^2 + \alpha \log \tau_k + \varepsilon_k, \quad k = 1, \ldots, m,$$

possibly over several independent particle paths, where $m$ is the number of lag values used and $\{\varepsilon_k\}_{k=1}^m$ is a random vector with an unspecified distribution and correlated entries (see (7) and (22)). Plots of TAMSD curves as a function of the lag value $\tau$, often on a log–log scale, are widely reported as part of anomalous diffusion data analysis (e.g. [50, 90]). The choice of lag values $\tau_1, \ldots, \tau_m$ reflects the analyst’s visual perception of the range where the slope of the TAMSD curves stabilize and thus indicate the true diffusive regime and power law.

The potential heterogeneity of fluid samples in fields such as microrheology implies that estimating $\xi$ from single trajectories is of great interest [9, 11, 51, 93]. However, the
experimental and statistical difficulties involved in estimating $\xi$ based on the regression system (3) have been pointed out by many authors. A non-exhaustive list of issues includes limited fluorophore lifetimes, proteins diffusing out of the field of view, finite-resolution imaging and motion blurring due to camera integration times, measurement errors, the presence of drifts and intra-path correlation [3, 8, 10, 36, 40, 56, 60, 71, 93]. Such difficulties call for a deeper understanding of the stochastic behavior of the TAMSD and, accordingly, a wealth of literature on the subject has developed. Starting from an underlying fractional stochastic process, several properties of the TAMSD such as ergodicity were established [6, 13, 17, 38, 39, 59, 75]. In particular, finite sample exact characterizations and mathematically convenient approximations to the distribution of the TAMSD under Gaussianity are provided in [1, 6, 7, 22, 27–29, 65, 71, 79]. In [79], assuming an observed fractional Brownian motion (see example 2.2), it is shown that the standard TAMSD-based estimator is consistent, with vanishing bias and variance.

We say that a cumulative distribution function (c.d.f.) $F$ gives the asymptotic distribution of a sequence of random variables $\{W_N\}_{N \in \mathbb{N}}$ if the c.d.f. $F_N(x)$ of $W_N$ converges to $F(x)$ at every $x \in \mathbb{R}$ where $F$ is continuous. Results on convergence in distribution such as the classical central limit theorem (CLT; see example 2.1) have a number of interesting statistical consequences. Typically, statements are robust, i.e. they hold for a multitude of models. Moreover, they naturally lead to useful data analysis protocols such as confidence intervals and hypothesis tests with error margins that are quantifiable and whose accuracy provably increases at an explicit rate (e.g. $\sqrt{N}$ for the CLT). In the probability literature, the study of the asymptotic distribution of sums of functions of Gaussian random variables has been carried out over many decades now (see [21, 24, 25, 31, 52, 69, 74, 85, 86] for just a few references). In the context of anomalous diffusion modeling, in turn, the related asymptotic distribution of the TAMSD was established in [20] for a broad class of Gaussian fractional stochastic processes. It was shown that the convergence in distribution of the TAMSD occurs at different rates, and that the limiting distribution may be Gaussian or non-Gaussian, all depending on the value of the diffusion exponent $\alpha$. This made it possible, for example, to construct asymptotically valid confidence intervals for the anomalous diffusion parameters starting from a single observed particle path.

In this paper, we propose particle path-based statistical protocols for detecting fluid heterogeneity that builds upon the TAMSD’s asymptotic distribution. The protocols test fluid heterogeneity in two different experimental situations, namely,

(i) assuming local physical homogeneity, whether different regions of the fluid are heterogeneous;
(ii) assuming global physical homogeneity of each fluid sample, whether two samples from each fluid are heterogeneous.

Hereinafter, these two senses are referred to as intra- and interfluid heterogeneity, respectively. The testing methodology is based on an improved single-path TAMSD-based estimation technique. To construct the latter, we tackle two of the main issues involved in TAMSD-based estimation, namely: (a) the presence of bias in log-TAMSD-based methods; and (b) the effect of correlated disturbances $\{\varepsilon_k\}_{k=1,...,m}$ in (3). Starting from a concentration inequality [5], we address these issues by providing mathematical characterizations of the bias and finite sample estimation variance which are by themselves of interest, as well as by introducing procedures for bias-correction and nearly optimal estimation under intra-path correlation. Motivated by applications in viscoelastic diffusion, the single-path estimation and heterogeneity testing protocols are mathematically established for $0 < \alpha < 3/2$, which covers all the subdiffusive
range and part of the superdiffusive regime, and are asymptotically valid. For the sake of completeness, we also discuss and provide computational studies on the strong superdiffusivity range $3/2 \leq \alpha < 2$ (see remark 3.2 on the difficulties involved in dealing with the possibly non-Gaussian asymptotic distribution of the TAMSD). To guide experimental practice under common technical constraints such as limited camera recording time, we also apply the proposed tools in investigating the difference between observing longer particle paths and using a larger number of particle paths of given length. To illustrate the use of the protocols in physical practice, we make inferences on fluid viscoelasticity with data from the Hill and Schoenfisch Labs (UNC-Chapel Hill) on biofilm eradication, as first reported and described in [73].

The paper is organized as follows. In section 2, we summarize the key mathematical results on the asymptotic distribution of the TAMSD. In section 3, assuming a single observed path of realistic length, we characterize the bias and the variance in TAMSD-based estimation to construct the improved single-path estimator and compare it with the standard TAMSD-based estimator in terms of statistical performance. In section 4, assuming multiple observed paths, we use the estimator developed in section 3 to construct statistical testing protocols for intra- and interfluid heterogeneity detection. In section 5, we model and test fluid heterogeneity through experimental data. For the reader’s convenience, appendix A contains mathematically accurate statements of the results in section 2 and [20]. Appendix B–D contain all new mathematical results and their proofs. Newly designed Matlab routines containing the estimation and testing protocols will be made available on the authors’ websites at the time of publication.

2. Background

Before we revisit the results in [20] on the asymptotic behavior of the TAMSD, for the sake of exposition we consider some classical results from probability theory.

Example 2.1. Consider independent and identically distributed random variables $X_1, \ldots, X_N$, each with mean $\langle X_1 \rangle = \mu$ and finite variance $\text{Var}X_1 := \langle X_1^2 \rangle - \langle X_1 \rangle^2 = \varphi^2 > 0$. If $\bar{X}_N = N^{-1} \sum_{i=1}^N X_i$ denotes the sample mean, then the celebrated central limit theorem states that, for large $N$, the distribution of the standardized sample mean approaches that of a standard normal, i.e.

$$\sqrt{N} \left( \frac{\bar{X}_N - \mu}{\varphi} \right) \xrightarrow{d} \mathcal{N}(0, 1), \quad N \to \infty. \quad (4)$$

Apart from naturally leading to confidence intervals and hypothesis tests, the convergence (4) also implies that $\bar{X}_N$ is a consistent estimator of $\mu$, namely, it converges in probability to $\mu$. This is so because

$$\bar{X}_N - \mu = \left( \frac{\varphi}{\sqrt{N}} \right) \sqrt{N} \left( \frac{\bar{X}_N - \mu}{\varphi} \right) \xrightarrow{p} 0, \quad N \to \infty. \quad (5)$$

The zero limit in probability in (5) stems from the fact that the vanishing term $\varphi/\sqrt{N} \to 0$ multiplies a standardized sample mean that converges in distribution (4) (see [78]).

Apart from distinct assumptions on the observations, the claims in [20] on the asymptotic behavior of the TAMSD are reminiscent of the classical statements (4) and (5), with two
The rate of convergence is not typically \( \sqrt{N} \) in biophysical modeling; (ii) the asymptotic distribution of the TAMSD is not necessarily Gaussian.

So, consider the random vector

\[
\left( M_N(\tau_1), \ldots, M_N(\tau_m) \right),
\]

namely, a vector of TAMSD terms (2) at \( m \) different lag values, obtained from one path of a Gaussian, stationary increment process. Fitting (3) and (6) by means of ordinary least squares (OLS) regression is the most intuitive way of constructing an estimator of the diffusion parameter vector \( \xi = (\log \sigma^2, \alpha) \). This corresponds to the common practice in the biophysical literature, both in experimental and methodological work (e.g. [11, 50, 51, 90] among many references). Throughout this paper, 

\[
E_{\text{stand}} = (L_{\text{stand}}, A_{\text{stand}})
\]

denotes this standard estimator (see (22) for a precise expression). In this framework, we need to make the lag sizes \( \tau_1, \ldots, \tau_m \) themselves go to infinity, though no faster than the sample size \( N \). This mathematically expresses the practical analysis of anomalous diffusion data: the lag size has to be

(\( L1 \)) large enough for the TAMSD regime to become log-linear;

(\( L2 \)) but, at the same time, not too large because of the increased variance of the TAMSD.

For a generic lag value \( \tau \), we can model this idea by writing

\[
\infty \leftarrow \tau \ll N.
\]

The limit and inequality in (8) express (\( L1 \)) and (\( L2 \)), respectively (the accurate mathematical statements are given by condition (A.3); see also figure 1).

The asymptotic distribution of the TAMSD random vector (6) after centering is briefly described in table 1. This leads to the asymptotic behavior of the standard estimator (7), which is summarized in table 2 in terms of convergence rate, asymptotic distribution and consistency. In both cases, the value of \( \alpha \) determines the convergence rate and the nature of the asymptotic distribution. In particular, over almost the whole strong superdiffusivity range (i.e. over \( 3/2 < \alpha < 2 \)), the asymptotic distribution is non-Gaussian (Rosenblatt-type; see theorem A.1 and [74, 85, 88, 91]). For any instance, by an argument analogous to (5), the standard estimator is consistent, i.e.

\[
E_{\text{stand}} \overset{P}{\rightarrow} \xi.
\]

The family of stochastic processes for which the limits in distribution in tables 1 and 2 hold is broad and contains a number of popular models. Three examples are fractional Brownian motion (fBm), fractional instances of the generalized Langevin equation (GLE) and the (integrated) fractional Ornstein-Uhlenbeck process (fiOU).

**Example 2.2.** Together with the continuous time random walk, fBm is one of the most popular models of anomalous diffusion [2, 87]. For some value of the so-named Hurst parameter \( H \in (0, 1) \) and a variance parameter \( D > 0 \), a fBm \( B_H(t) \) is the only Gaussian, stationary increment process with covariance function

\[
\langle B_H(s)B_H(t) \rangle = D \left( |t|^{2H} + |s|^{2H} - |t-s|^{2H} \right), \quad s, t \in \mathbb{R}.
\]

The particular parameter value \( H = 1/2 \) corresponds to the ordinary Brownian motion (Wiener process). In view of (10), which implies exact self-similarity, for fBm the MSD scaling relation (1) holds as an equality, i.e.
\[
\langle B^2_H(t) \rangle = \sigma^2 t^\alpha, \quad t \in \mathbb{R},
\]
where
\[
\sigma^2 = 2D, \quad \alpha = 2H. \tag{12}
\]

**Example 2.3.** The GLE has been used as a universal model of anomalous diffusion in the biophysical field of microrheology [53, 66, 68, 96]. A subclass of interest of the GLE framework is the fractional GLE family [18, 43, 44], which is obtained almost surely as the solution of the stochastic differential equation
\[
m \frac{dV(t)}{dt} = -\lambda \int_{-\infty}^t \Gamma(t-s)V(s)ds dt + dB_H(t), \quad 1/2 < H < 1. \tag{13}
\]
In (13), \(m, \lambda > 0\) and the memory kernel has the form \(\Gamma(t) = 2H(2H-1)|t|^{2H-2}, t \neq 0\), which is a consequence of invoking the fluctuation–dissipation relation [19, 51]. The integrated fractional generalized Langevin process (ifGL) is given by \(X(t) = \int_0^t V(s)ds, t > 0\), where \(\{V(t)\}_{t \geq 0}\) is a solution of the fractional GLE. For the ifGL, relation (1) holds with \(\alpha = 2(1-H)\) (subdiffusive) as \(t \to \infty\).

**Example 2.4.** The ifOU is given by \(X(t) = \int_0^t V(s)ds, t > 0\), where the so-called fractional Ornstein–Uhlenbeck process \(\{V(t)\}_{t \geq 0}\) is the almost surely continuous solution to the fBm-driven Langevin equation
\[
dV(t) = -\lambda V(t)dt + \varphi dB_H(t), \quad t \geq 0, \quad \lambda, \quad \varphi > 0, \quad 0 < H < 1 \tag{14}
\]
(see [14, 70]). The ifOU process is a mathematically convenient model of anomalous diffusion. In the subdiffusive range, it displays a similar correlation structure to that of the ifGL process. For the ifOU, relation (1) holds with (12) as \(t \to \infty\).

**Remark 2.1.** The results in [20] do not cover some important anomalous diffusion models such as continuous time random walks. For this family of models, limit theorems typically involve distinct nonstandard asymptotic distributions depending on the assumptions (see, for instance, [41, 55] and references therein; for general guidelines on the use of the TAMSD, see [42]).

### 3. Improved TAMSD-based estimation

The standard estimator \(E_{\text{stand}} = (L_{\text{stand}}, A_{\text{stand}})\) in (7) has at least two significant shortcomings: finite sample bias and suboptimal performance in the presence of correlation among the regression disturbance terms \(\{\varepsilon_k\}_{k=1, \ldots, N}\). We propose a single-path improved estimation protocol that addresses these issues. Accordingly, it involves two components, which we describe next. These two components involve asymptotically valid mathematical expressions for finite-sample bias and variance. Hereinafter, different lag values are expressed as
\[
\tau_k = w_k \tau, \quad w_1 < \ldots < w_k, \tag{15}
\]
for fixed constants \(w\), where \(\tau = \tau(N)\) grows as function of \(N\).
Bias correction. In TAMSD-based scaling analysis, there at least two sources of bias. First, bias appears if the particle movement is not exactly self-similar (not a fBm), i.e. \( \langle X^2(t) \rangle \neq \sigma^2 t^\alpha \) over a range of \( t \).

In fact, the deviation of the MSD from exact self-similarity or power scaling is generally controlled by the relation

\[
\frac{\langle X^2(t) \rangle}{\sigma^2 t^\alpha} - 1 \leq \frac{C}{p^\delta} \quad \text{for large } t, \tag{16}
\]

for some constant \( \sigma^2 > 0 \), where the deviation parameter \( \delta > 0 \) mostly depends on the high frequency behavior of the particle motion (see proposition A.1). Second, even under self-similarity, bias stems from the elementary fact that the logarithm of the ensemble...
average and the ensemble average of the logarithm are distinct (e.g. [62–64, 92]). In the context of (3), this means that
\[ \langle \log M_N(\tau) \rangle \neq \log \langle M_N(\tau) \rangle = \alpha \log \tau + \log \sigma^2, \quad \tau \in \mathbb{N}. \]
So, by reinterpreting \( \log M_N(\tau) \) itself as an estimator of \( \alpha \log \tau + \log \sigma^2 \), we can express the bias involved in TAMSD-based estimation as
\[ \langle \log M_N(\tau) \rangle - \left( \alpha \log \tau + \log \sigma^2 \right) = -\frac{\tau}{N} \beta_N(\alpha, \tau) + O\left( \frac{1}{\tau^0} \right) + O\left( \frac{1}{N} \right) \] (17)
for the same \( \delta > 0 \) as in (16) (for \( 0 < \alpha < 3/2 \)—see theorem C.1; see also remark 3.2 on the range \( 3/2 \leq \alpha < 2 \)). The term of order \( O(\tau^{-\delta}) \), then, is mostly determined by the high frequency behavior of the anomalously diffusive particle (see figure 1 and expressions (A.2) and (A.4)). In (17), the main bias factor is given by the function
\[ \beta_N(\alpha, \tau) = \frac{1}{4\tau} \sum_{i=-N+1}^{N-1} \left( \frac{1}{N} - \frac{|i|}{N} \right) \left\{ \left| \frac{i}{\tau} + 1 \right|^\alpha - 2 \left| \frac{i}{\tau} \right|^\alpha + \left| \frac{i}{\tau} - 1 \right|^\alpha \right\}^2. \] (18)
Note that (18) depends on the unknown parameter \( \alpha \). So, we use \( A_{\text{stand}} \) and (18) to define an estimator of the bias vector by
\[ \left( \beta_N(A_{\text{stand}}, \tau_k) \right)_{k=1, \ldots, m}. \] (19)

(b) Accounting for disturbance correlation. In linear estimation theory, the method for dealing with correlated random errors is called generalized least squares (GLS). In fact, the resulting GLS estimator is the best linear unbiased estimator, since it outperforms its OLS counterpart in terms of mean squared error (MSE) (see [15]).

In the context of TAMSD-based estimation, to better understand the difference between the standard, OLS-based estimator and the related GLS-based estimator, recast the vector system (3) as the regression model
\[ z = X \xi + \varepsilon. \] (20)
In (20), the term \( \xi \) is as in (1), and the dependent variable and the regressor are given by, respectively,
\[ z = \left( \log M_N(\tau_k) \right)_{k=1, \ldots, m}, \quad X = \begin{pmatrix} 1 & \log \tau_1 \\ \vdots & \vdots \\ 1 & \log \tau_m \end{pmatrix}. \] (21)
It is well known that the expression

\[ E_{\text{stand}} := (X^T X)^{-1} X^T z = (L_{\text{stand}}, A_{\text{stand}}) \]

(22)
gives the standard estimator (7) generated by the OLS solution to the system (20). By contrast, let

\[ \Upsilon(\xi) = \left(v_{k_1 k_2}(\xi)\right)_{k_1, k_2 = 1, \ldots, m}, \]

(23)
be the finite sample covariance matrix of the vector \( z \) as in (21). The GLS solution is given by

\[ (X^T \Upsilon(\xi)^{-1} X)^{-1} X^T \Upsilon(\xi)^{-1} z, \]

(24)
which involves the unknown matrix (23). In practice, then, one needs to estimate such a matrix. For this purpose, we first establish the entrywise expansion

\[ v_{k_1 k_2}(\xi) = \tau N^{\zeta(\alpha, \tau_{k_1}, \tau_{k_2})} + O\left(\frac{\tau N^{1-\delta}}{N}\right) + o\left(\frac{\tau N}{N}\right), \quad k_1, k_2 = 1, \ldots, m \]

(25)
(for 0 < \( \alpha < 3/2 \)—see theorem C.2; see also remark 3.2 on the range \( 3/2 \leq \alpha < 2 \)). In (25), the main variance factor is given by

\[ \zeta(\alpha, \tau_{k_1}, \tau_{k_2}) = \frac{1}{2\tau} \sum_{j=-N+1}^{N-1} \left( 1 - \frac{|i|}{N} \right) \left( \frac{i}{\sqrt{\tau_{k_1} \tau_{k_2}}} + \sqrt{\frac{\tau_{k_1}}{\tau_{k_2}}} \alpha - \frac{i}{\sqrt{\tau_{k_1} \tau_{k_2}}} + \sqrt{\frac{\tau_{k_2}}{\tau_{k_1}}} \right)^2. \]

(26)
Note that expression (26) does not involve the constant \( \sigma^2 \), but it is still a function of the unknown parameter \( \alpha \). Second, and in view of this, we can use \( A_{\text{stand}} \) and (26) to define an estimator of the covariance matrix by

\[ \Upsilon(A_{\text{stand}}) := \left( \frac{\tau}{N} \zeta(\alpha, \tau_{k_1}, \tau_{k_2}) \right)_{k_1, k_2 = 1, \ldots, m}. \]

(27)
Drawing upon (a) and (b), we can further construct an improved estimator of \( \xi \) by a quasi-GLS procedure based on the estimator \( \Upsilon(A_{\text{stand}}) \) and by replacing (3) with the bias-corrected regression system

\[ \log M_N(\tau_k) + \frac{\tau_k}{N} \beta_N(A_{\text{stand}}, \tau_k) = \log \sigma^2 + \alpha \log \tau_k + \varepsilon_k, \quad k = 1, \ldots, m. \]

(28)
The resulting estimator can be expressed as

\[ E = (L, A) = (X^T \Upsilon^{-1}(A_{\text{stand}}) X)^{-1} X^T \Upsilon^{-1}(A_{\text{stand}}) \]

(29)
where \( X \) is again as in (21) and

\[ y = \begin{pmatrix} \log M_N(\tau_1) + \frac{\tau_1}{N} \beta_N(A_{\text{stand}}, \tau_1) \\ \vdots \\ \log M_N(\tau_m) + \frac{\tau_m}{N} \beta_N(A_{\text{stand}}, \tau_m) \end{pmatrix}. \]

(30)
For the reader’s convenience, the construction of the estimator $E$ is summarized in the form of pseudocode in appendix D.

To compare the performances of $E$ and $E_{\text{stand}}$, we generated 1000 independent paths of length $2^{10}$ and estimated the diffusion exponent based on the two methods. Figure 2 displays the results in terms of Monte Carlo bias, standard deviation and square root MSE. The improved estimator $E$ outperforms the usual estimator $E_{\text{stand}}$ by any of the three criteria for different values of $\alpha$.

**Remark 3.1.** Note that the main bias and variance factors $\beta_N(\alpha, \tau)$ and $\varsigma_N(\alpha, \tau_1, \tau_2)$ in (18) and (26), respectively, converge as $N \to \infty$ (see lemma C.1). Moreover, after standardization, the estimator (29) is provably asymptotically normal and consistent for $0 < \alpha < 3/2$ (see (35) in section 4 and proposition C.1). See also remark 3.2 on the range $3/2 \leq \alpha < 2$.

**Remark 3.2.** Although we do not provide proofs in this paper, the methods developed in this section and also in section 4 can be extended to the strongly superdiffusive range $3/2 \leq \alpha < 2$. For example, due to nonstandard convergence rates, expressions (17) and (25) hold after replacing $O(\tau)$ with $O(2^{\log N}/N)$ (for $\alpha = 3/2$) or $O((\tau)^{3-2\alpha})$ (for $3/2 < \alpha < 2$). Likewise, the asymptotic non-Gaussian distribution of the estimator (35) in section 4, with nonstandard convergence rates, can be established. However, inference involving the non-standard limiting distribution can be cumbersome, and the computational studies in this section and in section 4 show that the methods in the proposed format work reasonably well for realistic path lengths. See also [20], remark 2, on how to construct asymptotically valid confidence intervals for $\xi$ based on the standard estimator $E_{\text{stand}}$ assuming prior knowledge that $3/2 < \alpha < 2$.

### 4. Testing heterogeneity

Single particle tracking experiments with viscoelastic diffusion often generate data in the form of multiple particle paths. As discussed in the Introduction, fluid heterogeneity can be tested in the *intra-* and *interfluid* senses. The pathwise framework constructed in section 3 can be used in new testing protocols with good finite sample and asymptotic properties. We remind the reader that the mathematical statements cover the diffusion exponent range $0 < \alpha < 3/2$ (subdiffusive and mildly superdiffusive range), though we also include computational experiments for the strongly superdiffusive range $3/2 \leq \alpha < 2$.

Tables 3 and 4 display the proposed framework. For each type of fluid heterogeneity, they show the appropriate hypotheses and testing procedures, respectively. In the remainder of this section, we provide a detailed description of the protocols. To set the notation, we recall that, for a given a hypothesis test, the conditional probability

$$ P(H_0 \text{ is rejected} \mid \xi \text{ satisfies } H_0) := \epsilon \in [0, 1] $$

is called the size (or significance level) of the test, whereas the function

$$ \xi \mapsto P(H_0 \text{ is rejected} \mid \xi \text{ satisfies } H_1) \in [0, 1] $$

is called the power of the test.
4.1. Intrafluid heterogeneity

Suppose $\nu \in \mathbb{N}$ bead diffusion paths of length $N$ from a single fluid sample are available. If the fluid is physically homogeneous, it is expected to generate particle paths with nearly identical parameter values $\xi$. The alternative is that $\xi_i \neq \xi_j$ for some pair $i, j$, namely, their anomalous diffusion parameters differ. These two possibilities, labeled $H_0$ and $H_a$, respectively, are listed in the row ‘intrafluid’ in table 3.

Starting from the $\nu$ particle paths, let $E_i, \quad i = 1, \ldots, \nu,$

be vector-valued estimators as in (29). For the purpose of constructing a test statistic, we need a normalized (standardized) estimator. Note that the variance of the GLS-type solution (24) is given by

$$(X^T Y^{-1}(\xi) X)^{-1}$$

(34)

(see [15]). So, define a standardized estimator by

$$Z_i = \begin{pmatrix} Z_{i,1} \\ Z_{i,2} \end{pmatrix} = \Lambda^{-1/2}(A_{\text{stand},i}) E_i, \quad i = 1, \ldots, \nu,$$

(35)

where

$$\Lambda(A_{\text{stand},i}) := (X^T Y^{-1}(A_{\text{stand},i}) X)^{-1}$$

(36)

and the variance estimator $\Upsilon(A_{\text{stand},i})$ is given by (27). Then, (35) converges in distribution to $\nu$ independent and identically distributed normal random vectors with uncorrelated entries (see proposition C.1). So, for $Z_j = \nu^{-1} \sum_{i=1}^{\nu} Z_{ij},$ let

$$S_j^2 = \frac{1}{\nu - 1} \sum_{i=1}^{\nu} (Z_{ij} - \bar{Z}_j)^2, \quad j = 1, 2,$$

(37)

be the normalized and decorrelated sample variances of $\{Z_{ij}\}_{i=1,\ldots,\nu}, \quad j = 1, 2,$ as in (35). Then, under $H_0$,
as $N \to \infty$, where $X_1$ and $X_2$ are independent random variables. To test heterogeneity at significance level $\epsilon$, we can use Bonferroni-type correction (e.g. [15], section 5.3) and reject the null hypothesis $H_0$ if

$$R_{\text{intra}} : (\nu - 1)S_1^2 > \chi^2_{\nu - 1, \epsilon/2} \quad \text{or} \quad (\nu - 1)S_2^2 > \chi^2_{\nu - 1, \epsilon/2}, \quad (38)$$

where $\chi^2_{\nu - 1, \epsilon/2}$ is a chi-square quantile (see table 4, ‘intrafluid’ rows).

To check the size of the test (38) over finite samples, we conducted a Monte Carlo study with 50 simulated paths of length $2^{12}$ and recorded whether or not the null hypothesis $H_0$ is rejected at $\epsilon = 0.05$ significance level. This procedure was repeated 2000 times. Since each outcome is a Bernoulli trial (reject or not $H_0$), the simulation rejection rate follows a binomial distribution with $n = 2000$ and $p = 0.05$. As shown in figure 3, left plot, the observed simulation rejection rate was around 0.05 and within the 95% confidence interval (for $0 < \alpha < 3/2$), as expected. Unreported computational experiments for different significance levels lead to analogous conclusions.

Figure 4 displays Monte Carlo power curves for the intrafluid test. The study was conducted with a total of $\nu = \nu_1 + \nu_2$ paths, where $\nu_1$ and $\nu_2$ have diffusion exponents $\alpha_1 = 1$ and $\alpha_2$, respectively, and $\alpha_2 = 0.8$ (left plot) or $\alpha_2 = 0.7$ (right plot). In each plot, the x-axis represents the proportion of paths

$$\frac{\nu_2}{\nu_1 + \nu_2} \quad (39)$$

with diffusion exponent $\alpha = \alpha_2$. The power curves quickly converge to 1 as a function of the ratio (39), especially for the more distinguishable value $\alpha_2 = 0.7 < 1 = \alpha_1$.

When $\alpha > 3/2$, under $H_0$ the estimators (33) converge in distribution to $\nu$ independent and identically distributed non-Gaussian random vectors. Hence, so do the estimators (35). In this case, the marginal distributions of the decorrelated vector $((\nu - 1)S_1^2, (\nu - 1)S_2^2)$ do not approach chi-squared distributions. In computational experiments, the size of the intrafluid test (38) did not significantly deviate from the 0.05 target for $\alpha = 1.6$, indicating that the non-standard asymptotic behavior is not a concern for paths of length $2^{12}$. Deviation was significant for the extreme value $\alpha = 1.8$, suggesting that approximating the test size by simulation may be generally recommendable for greater accuracy (see figure 3, left plot).
4.2. Interfluid heterogeneity

Now suppose \( \nu_I \) and \( \nu_{II} \) paths, \( \nu_I, \nu_{II} \in \mathbb{N} \),

\[ (40) \]
are obtained from two physically homogeneous fluid samples I and II, respectively. We are interested in testing whether the samples I and II are homogeneous, namely, whether or not particle diffusion in the fluid samples displays the same underlying parameter value \( \xi \). These two possibilities, labeled \( H_0 \) and \( H_{a} \), respectively, are described in the row ‘interfluid’ in table 3.

Figure 3. Intra- and interfluid heterogeneity: test sizes. Monte Carlo sizes of the intrafluid test (left plot; see (38)) and interfluid test (right plot; see (46)) as a function of the diffusion exponent \( \alpha \) (x-axis). For every value of \( \alpha \), each of 2000 Monte Carlo runs consisted of generating 50 independent paths of length \( 2^{12} \) and conducting a test at \( \epsilon = 0.05 \) (see (31)) or, equivalently, 95% confidence level. The Monte Carlo rejection rate is very close to the theoretical value of \( \epsilon = 0.05 \) for almost all values of \( \alpha \).

Figure 4. Intrafluid heterogeneity: test power. Simulations were conducted with a total of \( \nu = \nu_1 + \nu_2 \) paths, where \( \nu_1 \) and \( \nu_2 \) of them displayed diffusion exponents \( \alpha = \alpha_1 = 1 \) and \( \alpha = \alpha_2 \), respectively. In each plot, the y-axis represents the observed test power, or rejection rates (see (32)), and the x-axis is the proportion of paths with \( \alpha = \alpha_2 \) (see (39)), starting at 0.02. The total number of Monte Carlo runs is 2000. Left plot: \( \alpha_2 = 0.8 \). Right plot: \( \alpha_2 = 0.7 \).
Since multiple (independent) particle paths are assumed available for each fluid sample, we can construct an estimator involving all available TAMSD terms. In fact, first define the overall average mean squared displacement over \( \nu \) TAMSD terms (AMSD) by

\[
M^*_N(\tau) = \frac{1}{\nu} \sum_{\ell=1}^{\nu} M_N(\tau)_\ell. \tag{41}
\]

By independence,

\[
\langle M^*_N(\tau) \rangle = \langle M_N(\tau) \rangle = \langle X^2(\tau) \rangle, \quad \text{Var} M^*_N(\tau) = \frac{1}{\nu} \text{Var} M_N(\tau). \tag{42}
\]

Then, AMSD-type estimators

\[
E^* = (L^*, A^*) \tag{43}
\]

can be obtained by applying the pseudocode in appendix D after replacing TAMSD terms \( M_N(\tau_k) \) with their AMSD counterparts \( M^*_N(\tau_k) \), \( k = 1, \ldots, m \). Given two fluid samples I and II, let \( E^*_I \) and \( E^*_II \) be their respective AMSD-type estimators. Their finite sample covariance matrices are given by \( \nu^{-1}_I \Lambda(\xi_I) \) and \( \nu^{-1}_II \Lambda(\xi_{II}) \), respectively (see (34)). By analogy with (25) and (36), we can define their AMSD-type estimators

\[
\frac{1}{\nu}_I \Lambda(A^*_I), \frac{1}{\nu}_II \Lambda(A^*_II). \tag{44}
\]

Figure 5 displays a study of the accuracy of \( \Lambda(A^*_\bullet) \) as an estimator. It plots Monte Carlo variances of the estimator \( E \) as well as their estimates \( \Lambda(A^*_\bullet) \) for several values of the parameter \( \alpha \). The latter nearly perfectly match the former in the subdiffusive range. A slight deviation appears in the strongly superdiffusive range, but still within an acceptable margin.

Hence, we define the standardized estimators

\[
Z^*_j = \begin{pmatrix} Z^*_I,1 \\ Z^*_I,2 \end{pmatrix} = \sqrt{\nu} \Lambda^{-1/2}(A^*_j) E^*_j, \quad j = I, II. \tag{45}
\]

In view of proposition C.1, these estimators are also asymptotically normal for \( 0 < \alpha < 3/2 \). Hence, let

\[
T_1 = \frac{Z^*_{I,1} - Z^*_{II,1}}{\sqrt{2}}, \quad T_2 = \frac{Z^*_{I,2} - Z^*_{II,2}}{\sqrt{2}} \tag{45}
\]

be the associated test statistics. The rejection region is given by

\[
R_{\text{inter}} : |T_1| > z_{\epsilon/4} \quad \text{or} \quad |T_2| > z_{\epsilon/4}, \tag{46}
\]

where \( z_{\epsilon/4} \) is a standard Normal quantile (see table 4, row ‘interfluid’). In (46), the probability \( \epsilon/2 \) stems, first, from applying a Bonferroni-type correction to a double testing region (hence yielding \( \epsilon/2 \) significance level in each), and second, from the fact that in each region the test statistic distribution is two-sided.

To check the test’s size over finite samples, we produced a 2000-run Monte Carlo study based on two sets of 50 paths with the same diffusion exponent, where tests were conducted at significance level \( \epsilon = 0.05 \). As shown in figure 3, right plot, the rejection rate was close to 0.05, as expected.

In figure 6, we investigate the interfluid test power as a function of the path lengths and number of paths. The x-axis represents the difference between the diffusion exponents from two fluids, namely,
\[ \alpha = |\alpha_I - \alpha_{II}|, \]

whereas the \( y \)-axis is the test power at \( \epsilon = 0.05 \). From top to bottom, the three plots correspond to \( \alpha_{\text{min}} = \min\{\alpha_I, \alpha_{II}\} = 0.2, 1.0, 1.8 \), respectively, for various combinations of realistic values of \( N \) and \( \nu = \nu_I = \nu_{II} \). In all cases, the power curves start at around 0.05, as expected, and quickly approach 1 as a function of \( \delta_\alpha \) as defined in (47). Larger path lengths, larger number of particle paths as well as not very large values of \( \alpha_{\text{min}} \) are associated with faster convergence of power curves to 1.

4.3. Interfluid heterogeneity: more or longer paths under technical–experimental constraints?

Figure 6 also illustrates the following phenomenon. For the subdiffusive and diffusive cases, there is no visible difference between doubling the path lengths or the number of paths. However, in the strongly superdiffusive range, doubling the number of paths increases the test power more than doubling the path lengths.

In real world lab conditions, conducting single particle experiments involves coping with technical restrictions. For example, there may be limited camera recording time, tracer particles may slip out of the field of view or there may be a limit on the total number of tracer particles per fluid sample while still assuming that particles diffuse independently. So, assuming technical–experimental restrictions are in place, it is relevant to ask: what is the difference between

**Method I:** recording the movement of a larger number of particles \( (\nu) \) over a fixed period of time (hence, keeping constant the average sample path length \( N \)); and

**Method II:** recording the same number of particles \( \nu \) over a longer period of time (hence, yielding a larger average \( N \))? 

We answer this question in the framework of interfluid heterogeneity testing.

In the regimes of Methods I and II, we investigate the performance of the AMSD-type estimator (43) in terms of bias, standard deviation and square root MSE. Bearing in mind
expression (42), by a similar reasoning to the one leading to expression (17) for a single observed path, the bias of $\log_2 M_N^*(\tau)$ as an estimator of $\alpha \log \tau + \log \sigma^2$ is given by

$$O(\tau^{-\delta}) - \left\langle \frac{(M_N^*(\tau) - \langle X^2(\tau) \rangle)^2}{2\langle X^2(\tau) \rangle^2} \right\rangle = O(\tau^{-\delta}) - \frac{1}{\nu} \left\langle \frac{(M_N(\tau) - \langle X^2(\tau) \rangle)^2}{2\langle X^2(\tau) \rangle^2} \right\rangle.$$ (48)

Moreover, by the independence of particle paths, we can approximate the variance of $\log_2 M_N^*(\tau)$ by

$$\text{Var}\left( \log M_N^*(\tau) \right) \approx \frac{\text{Var} M_N(\tau)}{\nu \langle X^2(\tau) \rangle^2}.$$ (49)

The performance of the estimator (43) in the two regimes depends on the interplay between the bias and variance components (48) and (49), respectively. In a computational experiment, we applied the following procedure.

1. Start out in the same setting: $2^4$ paths of length $2^8$ for each method, run 500 Monte Carlo simulations to get the bias, standard deviation and square root MSE of $A$ for Method I and II;
2. for Method I, fix the path length and at each step generate $2^4$ times the previous number of paths and redo the Monte Carlos experiments;
3. for Method II, fix the number of paths and at each step generate paths of length $2^4$ times the previous length, multiply all lags by 2 and redo the Monte Carlos experiments;
4. repeat 2. and 3. three times.

For ease of comparison, table 5 displays the multiple instances generated. Note that, at each step, the total number of points recorded

$$\nu \times N$$ (50)

is identical for the two methods.

We compare the results in figure 7, top and middle plots, where the diffusion exponent is set to $\alpha = 0.6$ and 1.0, respectively. Method II has smaller bias and square root MSE. The reason is that, when $\nu$ is large enough, the term $O(\tau^{-\delta})$ dominates the bias. Thus, increasing the number of paths $\nu$ does not reduce the bias. However, increasing the path length $N$ means that the TAMSD terms $M_N(\tau)$ with larger lag values $\tau$ can be used in the regression procedure. This implies a reduction in magnitude of the term $O(\tau^{-\delta})$, and hence, smaller bias. Method I displays smaller standard deviation because a 16-fold increase in $\nu$ reduces the standard error by a factor of 1/4. Meanwhile, noting that $\langle X^2(\tau) \rangle \sim \sigma^2 \tau^\alpha$, expression (A.5) implies that for Method II the standard deviation is proportional to $\sqrt{\tau/N}$. By multiplying $N$ by 16 and $\tau$ by 2, the standard error is reduced by a factor of $1/2 \sqrt{2}$.

In figure 7, bottom plot, we set $\alpha = 1.8$. For this parameter value, the convergence rate of the TAMSD-based estimators is slower than $\sqrt{\tau/N}$. Method II still shows a smaller bias by comparison to Method I, as expected. However, since $\delta$ increases as a function of $\alpha$ (see expression (A.4)), then $O(\tau^{-\delta})$ shrinks with $\alpha$. Therefore, the component $O(\tau^{-\delta})$ carries less weight in the estimator’s bias for the superdiffusive case than for the subdiffusive case. Since $\alpha > 3/2$, again by expression (A.5) (see also remark 3.2) the standard deviation for Method II is proportional to $(\tau/N)^{3-\alpha} = (\tau/N)^{0.2}$. Thus, again assuming a 16-fold increase in $N$ and a 2-fold increase in $\tau$, the standard error is reduced by a factor of $1/8^{0.2}$, which is much slower than the standard error reduction factor of $1/2 \sqrt{2}$ for Method I. These are the two reasons why Method I displays smaller square root MSE than Method II.
Figure 6. Interfluid heterogeneity: test power. For various values of $N$ (see (2)) and $\nu = \nu_1 = \nu_\Pi$ (see (40)), we plot test power (y-axis; see (32)) as a function of $\delta_\alpha$ (x-axis; see (47)). The parameter values are given by $\alpha_{\min} := \min\{\alpha_1, \alpha_\Pi\}$ in the range $\alpha_{\min} = 0.2, 1.0, 1.8$ (top, middle and bottom, respectively).
5. Analysis of experimental data: heterogeneity of treated \textit{P. aeruginosa} biofilms

The Hill and Schoenfisch Labs at UNC-Chapel Hill produced data from experiments on disruption and eradication of \textit{P. aeruginosa} biofilms using nitric oxide-releasing chitosan oligosaccharides \cite{73}. For the reader’s convenience, we provide a brief description of the experiments.

Cystic fibrosis (CF) lung disease is caused by defective chloride transport, resulting in thickened, dehydrated mucus. This restricts bacterial motility and promotes \textit{P. aeruginosa} biofilm formation. Inhaled tobramycin is currently the only antibiotic recommended for the treatment of both initial and chronic \textit{P. aeruginosa} infections in patients with CF. While inhaled tobramycin is effective at eradicating bacteria within biofilms, it fails to physically remove the structural remnants of the biofilm from the airways. This may lead to biofilm regrowth and the development of antibiotic-resistant infections. Therefore, an ideal anti-biofilm therapeutic for CF would both eradicate bacteria and physically degrade the biofilm, facilitating clearance from the airways.

On the other hand, nitric oxide (NO) is an endogenously produced diatomic free radical with significant antibacterial activity against \textit{P. aeruginosa} biofilms. Atomic force microscopy revealed that NO exposure causes structural damage to the membranes of planktonic Gram-negative bacteria, including \textit{P. aeruginosa}. The interest lies in the utility of NO-releasing chitosan oligosaccharides to both eradicate and physically alter \textit{P. aeruginosa} biofilms and in comparing its effect with tobramycin. In order to measure the physical changes to bacterial biofilms caused by NO-releasing chitosan oligosaccharides, movements of fluorescent tracer particles embedded in \textit{P. aeruginosa} biofilms were imaged at 60 frames per second for 30 seconds on an inverted microscope at 40× magnification. The tracer particle displacement as a function of time was quantified using Video Spot Tracker software (Center for Computer Integrated Systems for Microscopy and Manipulation, University of North Carolina at Chapel Hill); see \cite{73} for more details.

Fluid heterogeneity has been correlated with increased viscoelasticity for complex biological materials such as sputum \cite{16}. In the experiments we describe, the effect of antibacterial treatment on biofilm heterogeneity was thus evaluated at different concentrations based on tracer particle displacement data. In table 6, we use the data to test the intrafluid heterogeneity of \textit{P. aeruginosa} biofilms after tobramycin treatments at concentrations levels 25, 50, 100, 200, and 400 \(\mu\)g ml\(^{-1}\). From each of these fluid samples, we randomly select 100 paths of length \(N = 1800\). An application of the intrafluid test \cite{38} produces strong evidence (negligible \(p\)-values) of intrafluid heterogeneity in every sample. This conclusion matches those reported in \cite{73}. Since no homogeneous fluid samples are detected from any of these five samples, we do not perform the interfluid heterogeneity test \cite{46}.

In table 7, we apply \cite{38} in the testing of intrafluid heterogeneity of \textit{P. aeruginosa} biofilms after COS2-NO treatment at concentration levels 1, 2, 4, 8, and 16 mg ml\(^{-1}\). COS2 releases
Figure 7. Bias, standard deviation, and square root MSE for Methods I (blue) and II (red). The x-axis denotes the total number of recorded data points (see (50)). Top: $\alpha = 0.6$. Middle: $\alpha = 1.0$. Bottom: $\alpha = 1.8$. 
NO, which cause the physical disruption and eradication of biofilms [73], and also reduces the viscoelastic properties of mucus [72]. As before, 100 paths of length 1800 were randomly selected for each concentration level. At concentrations 1 or 2 mg ml\(^{-1}\), the \(p\)-values are still less than machine error, which indicates strongly significant heterogeneity. As the concentration level increases to 4 and 8 mg ml\(^{-1}\), the \(p\)-values also increase. At concentration level 16 mg ml\(^{-1}\), the \(p\)-value reaches 0.18. Hence, we fail to reject the null hypothesis of intrafluid homogeneity. This provides evidence that the COS2-NO treatment is effective at eradicating \(P.\ aeruginosa\) biofilms. Once again, this analysis confirms the conclusions reported in [73].

### Table 6. Intrafluid biofilm heterogeneity testing after treatment with tobramycin at concentration levels 25, 50, 100, 200, 400 \(\mu\)g ml\(^{-1}\). 100 independent paths of length 1800 were randomly selected for each concentration level.

| Tobramycin (\(\mu\)g ml\(^{-1}\)) | \(p\)-value |
|---------------------------------|-------------|
| 25                              | <10\(^{-16}\) |
| 50                              | <10\(^{-16}\) |
| 100                             | <10\(^{-16}\) |
| 200                             | <10\(^{-16}\) |
| 400                             | <10\(^{-16}\) |

### Table 7. Intrafluid biofilm heterogeneity testing after treatment with COS2-NO at concentration levels 1, 2, 4, 8, 16 mg ml\(^{-1}\). 100 independent paths of length 1800 were randomly selected for each concentration level.

| COS2-NO (mg ml\(^{-1}\)) | \(p\)-value |
|---------------------------|-------------|
| 1                         | <10\(^{-16}\) |
| 2                         | <10\(^{-16}\) |
| 4                         | 2 \times 10\(^{-13}\) |
| 8                         | 3 \times 10\(^{-9}\) |
| 16                        | 0.18         |

### Table 8. Interfluid biofilm heterogeneity testing after treatment with COS2-NO at concentration level 16 mg ml\(^{-1}\). 100 independent paths of length 1800 were randomly selected for each group.

| COS2-NO 16 mg ml\(^{-1}\) | \(p\)-value |
|---------------------------|-------------|
| Group A versus Group B    | 0.9996      |
| Group A versus Group C    | 0.9998      |
| Group B versus Group C    | 0.9999      |

### 6. Conclusion

Motivated by applications in viscoelastic diffusion, in this paper we start from the TAMSD’s asymptotic distribution for a broad class of Gaussian fractional stochastic processes [20] to
propose statistical protocols that make use of single-particle tracking data in the detection of fluid heterogeneity.

The testing methodology is based on an improved TAMSD-type estimator. To construct this estimator, we tackle two of the main issues involved in TAMSD-based estimation, namely, we mathematically characterize: (a) the finite-sample bias in log-TAMSD-based methods; and (b) the effect of disturbance correlation. The theoretical results on (a) and (b) allow us to propose a nearly optimal estimator by combining a bias-correction procedure and a generalized least squares-type regression solution. The improved TAMSD-based estimator (29) is asymptotically normal for $0 < \alpha < 3/2$, and computational experiments show that the new estimator outperforms the standard TAMSD-based estimator both in terms of bias and square root MSE for values of $\alpha$ over the whole parameter range $(0, 2)$.

The estimator (29) is used in the construction of protocols for fluid heterogeneity detection in two different experimental situations, namely, when testing: (i) whether different regions of the same fluid are heterogeneous (intrafluid heterogeneity); or (ii) whether two samples from each homogenous fluid are heterogeneous (interfluid heterogeneity). Reflecting the asymptotic behavior of (29), for $0 < \alpha < 3/2$ the test statistics (37) and (45) for intra- and interfluid heterogeneity are asymptotically chi-square and asymptotically normally distributed, respectively. This ensures the tests and associated quantiles are asymptotically valid. Computational experiments confirm that the tests’ significance levels are accurate over finite samples, and that the tests display high power even for relatively small deviations from the null hypotheses.

In all cases, for the sake of completeness, we discuss and provide computational studies on the strong superdiffusivity range $3/2 \leq \alpha < 2$. Although this may not affect physical areas of application where subdiffusion is prevalent, this research points to one difficulty involved in TAMSD-based modeling, namely, a potentially non-Gaussian (Rosenblatt-type) limiting distribution with an intricate cumulant structure (see expression (A.10)).

The constructed framework helps to shed light on the effect of common technical experimental constraints such as limited camera recording time: we characterize the difference between observing longer particle paths and using a larger number of particle paths of given length.

We apply the protocols in physical practice by making inferences on fluid viscoelasticity with data from the Hill and Schoenfisch Labs (UNC-Chapel Hill), as first reported and described in [73]. The testing protocols reveal that COS2-NO treatment is effective in eradicating *P. aeruginosa* biofilms, since greater concentration levels of the treatment clearly lead to greater fluid homogeneity as detected by tracer particle displacement data.

The research contained in this paper points to a number of interesting questions. From a modeling standpoint, it would be convenient to construct a heterogeneity testing framework for single particle experiments that, unlike TAMSD-based methods, mathematically covers the full anomalous diffusion parametric range $0 < \alpha < 2$ under the same limit parametric family of distributions. Moreover, ideally a heterogeneity testing framework should be robust with respect to nuisance trends and added experimental noise. In another research direction, the same questions can be asked for classes of anomalous diffusion models not covered by the results in [20] such as, for example, continuous time random walks and related stochastic processes.

**Acknowledgments**

DBH was partially supported by the awards DMS 1462992 (National Science Foundation), AI1 12029 and HL 108808 (National Institutes of Health), and Hill16XX0 (Cystic Fibrosis
Foundation). GD was partially supported by the prime award no. W911NF-14-1-0475 from the Biomathematics subdivision of the Army Research Office, USA. The authors would like to thank John Fricks and two anonymous reviewers for their suggestions and comments on this paper.

Appendix A. The asymptotic distribution of the TAMSD

In theorem A.1, we provide the asymptotic distribution of the TAMSD random vector after centering, which in corollary A.1 allows developing the asymptotic distribution of the standard estimator (7) or (22) of the diffusivity coefficient and diffusion exponent.

**Theorem A.1 (Didier and Zhang [20], theorem 1).** Suppose the particle motion is a Gaussian, stationary increment process whose covariance function admits the harmonizable (Fourier domain) representation

\[ \langle X(s)X(t) \rangle = C_\alpha^2 \int_R \frac{(e^{is} - 1)(e^{-is} - 1)}{x^2} f(x) \, dx, \quad \alpha \in (0, 2), \quad C_\alpha \neq 0. \]

(A.1)

In (A.1), the spectral density has the generic form

\[ f(x) = \frac{s(x)}{|x|^{\alpha/2 - 1/2}}, \]

for constants \( C_0, \delta_0, \varepsilon_0 > 0 \). Suppose the growth of the lag value term \( \tau = \tau(N) \in \mathbb{N} \) as in (15) with respect to the sample size \( N \) is given by

\[ \frac{\tau(N) \log^2(N)}{N} + \frac{N}{\tau(N)^{1+\delta/2}} \to 0, \quad N \to \infty, \]

(A.3)

where

\[ \delta = \min\{\alpha/2, \delta_0/2\}. \]

(A.4)

Also consider the rates of convergence

\[
\begin{align*}
0 < \alpha < 3/2 : & \quad \eta(N) = \sqrt{N}, \quad \zeta(\tau) = \tau^{\alpha + 1/2}, \\
\alpha = 3/2 : & \quad \eta(N) = \sqrt{N \log(N)}, \quad \zeta(\tau) = \tau^2; \\
3/2 < \alpha < 2 : & \quad \eta(N) = N^{\alpha - 1}, \quad \zeta(\tau) = \tau^2.
\end{align*}
\]

(A.5)

Then, as \( N \to \infty \),

\[ \left( \frac{N_k}{\eta(N_k)\zeta(\tau_k)} (M_N(\tau_k) - \langle X(\tau_k) \rangle_k) \right)_{k=1, \ldots, m} \overset{d}{\to} Z, \]

(A.6)

where, for \( k = 1, \ldots, m \), \( N_k = N - \tau_k \) is the number of available terms in each TAMSD sum (2) and \( \tau_k \) is given by (15). In (A.6), the distribution of the random vector \( Z \) can be described as follows.

1. If \( 0 < \alpha < 3/2, \) then \( Z \sim N(0, \Sigma) \) (a m-variate Gaussian distribution), where the entry \( k_1, k_2 \) of the matrix \( \Sigma = \Sigma(\alpha) \) is given by
\[ \Sigma_{k_1,k_2} = 2w_{k_1}^{-\alpha-1/2}w_{k_2}^{-\alpha-1/2} \left( \frac{C_\alpha}{C_H} \right)^4 \left\| \hat{G}(y;w_{k_1},w_{k_2}) \right\|_{L^2(\mathbb{R})}^2, \]  \quad (A.7)

\( k_1, k_2 = 1, \ldots, m \). In (A.7), we define
\[ C_H = \sqrt{\pi^{-1} H(2H) \sin(H\pi)} \]
and
\[ \hat{G}(y;w_{k_1},w_{k_2}) = C_H^2 \frac{(e^{i\omega_{k_1}y} - 1)(e^{-i\omega_{k_2}y} - 1)}{|y|^{\alpha+1}}; \quad (A.8) \]

(ii) if \( \alpha = 3/2 \), then \( Z \sim \mathcal{N}(0, \Sigma) \), where the entry \( k_1, k_2 \) of the matrix \( \Sigma = \Sigma(\alpha) \) is given by
\[ \Sigma_{k_1,k_2} = 4\vartheta^2, \quad k_1, k_2 = 1, \ldots, m, \quad (A.9) \]
and
\[ \vartheta = \left( \frac{C_\alpha}{C_H} \right)^2 \frac{\alpha(\alpha-1)}{2}; \]

(iii) if \( 3/2 < \alpha < 2 \), \( Z \) follows a multivariate Rosenblatt-type distribution whose characteristic function is given by
\[ \phi_Z(t) = \exp \left\{ \sum_{k=2}^{\infty} \frac{2i\vartheta \sum_{k=1}^{m} t_k}{s} c_s(\alpha) \right\} \quad (A.10) \]
around the origin. In (A.10), for \( s \geq 2 \), \( c_s(\alpha) \) is given by
\[ \int_{[0,1]^s} |x_1 - x_2|^{\alpha-2}|x_2 - x_3|^{\alpha-2} \cdots |x_s - x_1|^{\alpha-2} dx_1 \cdots dx_s. \quad (A.11) \]

Consider again the regression system (20) and recall that expression (22) gives the standard estimator generated by the OLS solution to the system (20). The following corollary describes the asymptotic distribution of the standard estimator (22).

**Corollary A.1 (Didier and Zhang [20], corollary 1).** Suppose the assumptions of theorem A.1 hold. Then, as \( N \to \infty \),
\[ \frac{N^\tau}{\eta(N)\zeta(\tau)} \left( \begin{array}{c} \log \tau - \log \sigma^2 \\ A - \alpha \end{array} \right) \xrightarrow{d} \left( \begin{array}{c} U^T \\ -U^T \end{array} \right) \Lambda Z, \]
\quad (A.12)

In (A.12),
\[ \Lambda = A(\sigma^2, \alpha) = \text{diag}(\zeta(w_1)/(\sigma^2 w_1^\alpha), \ldots, \zeta(w_m)/(\sigma^2 w_m^\alpha)); \]
\quad (A.13)
\[ \eta(\cdot), \zeta(\cdot) \text{ and } Z \text{ are as in theorem A.1, and} \]
\[ U^T = \frac{1}{c_w} \left( \sum_{k=1}^{m} \log(w_k/w_1), \ldots, \sum_{k=1}^{m} \log(w_k/w_m) \right) \]
\quad (A.14)
with constant
\[ c_w = m \sum_{k=1}^{m} \log^2(w_k) - \left( \sum_{k=1}^{m} \log(w_k) \right)^2. \]  
(A.15)

In particular, the standard estimator (22) is consistent, namely, relation (9) holds.

Proposition A.1 (Didier and Zhang [20], proposition 1, (i)). Under the assumptions
of theorem A.1, there is a constant \( \sigma^2 > 0 \) such that (16) holds for some \( C > 0 \), where \( \delta > 0 \) is given by (A.4).

Appendix B. Some lemmas

In this section, we present some lemmas that are used to prove theorems C.1 and C.2 in
appendix C. Throughout this section, we assume \( 0 < \alpha < 3/2 \) and the conditions of theorem A.1. In
proofs, whenever convenient \( C \) denotes a constant that may change from one line to the next.

In light of (A.6), define the standardized statistic
\[ \varpi(\tau) = \frac{M_N(\tau)}{\langle X^2(\tau) \rangle}. \]  
(B.1)

In particular,
\[ \varpi(\tau) \xrightarrow{P} 1, \quad N \to \infty, \]  
(B.2)
so that a Taylor expansion can be applied to \( \log \varpi(\tau) \) around 1. Meanwhile, we define the standardized increment
\[ W_j(\tau) = \frac{X(j + \tau) - X(j)}{\sqrt{\langle X^2(\tau) \rangle}}. \]  
(B.3)

We will use the following results in our proofs. The first one is the classical Isserlis theorem,
which reduces the higher moments of a multivariate normal vector to its second moments. The
second one is a concentration inequality that will allow us to establish sharp bounds on the
tails of centered quadratic forms.

Theorem B.1 (Isserlis, [37]). Let \( (Z_1, Z_2, \ldots, Z_{2N}) \) be a zero mean, multivariate normal
random vector. Then,
\[ \langle Z_1 Z_2 \ldots Z_{2N} \rangle = \sum \prod \langle ZZ \rangle, \]
where the notation \( \sum \prod \) stands for summing over all distinct ways of partitioning \( Z_1, \ldots, Z_{2N} \)
into pairs \( Z_i, Z_j \) and each summand is a product of these \( N \) pairs.

Theorem B.2 ([5, 47]). Let \( Z_1, \ldots, Z_N \overset{i.i.d.}{\sim} \mathcal{N}(0, 1) \) and consider constants
\( \eta_1, \ldots, \eta_N \geq 0 \), not all zero. Let \( \| \eta \|_2 \) and \( \| \eta \|_\infty \) be the Euclidean square and sup norms of
the vector \( \eta = (\eta_1, \ldots, \eta_N)^T \). Also, define the random variable \( X = \sum_{i=1}^{N} \eta_i \langle Z_i^2 - 1 \rangle \). Then, for every \( x > 0 \),
\[ \mathbb{P}(X \geq 2 \| \eta \|_2 \sqrt{x} + 2 \| \eta \|_\infty x) \leq \exp(-x), \]
\[ \mathbb{P}(X \leq -2 \| \eta \|_2 \sqrt{x}) \leq \exp(-x). \]

The following lemma describes some basic properties of the central moments of (B.1).

**Lemma B.1.** As \( N \to \infty \),
\[ \langle (\varpi(\tau) - 1)^2 \rangle = O\left( \frac{\tau}{N} \right). \]  

**Proof.** Expression (B.5) (for \( \kappa \geq 3 \)) can be proved by adapting the argument for establishing expression (C.24) in [94], while making use of the bound (B.11) and lemma B.2 (expressions (B.8) and (B.9)). So, for the reader’s convenience, we establish (B.4) (for \( \kappa = 2 \)). The left-hand side of (B.4) can be rewritten as
\[ \frac{1}{N^2} \left\langle \sum_{k_1=1}^{N} \sum_{k_2=1}^{N} \left( \frac{(X + k_1) - X(k_1)}{X^2(\tau)} - 1 \right) \left( \frac{(X + k_2) - X(k_2)}{X^2(\tau)} - 1 \right) \right\rangle = \frac{1}{N^2} \left\langle \sum_{k_1=1}^{N} \sum_{k_2=1}^{N} (W_{k_1}^2(\tau) - 1)(W_{k_2}^2(\tau) - 1) \right\rangle. \]

By applying the Isserlis theorem (theorem B.1),
\[ \langle (W_{k_1}^2(\tau) - 1)(W_{k_2}^2(\tau) - 1) \rangle = \langle W_{k_1}^2(\tau)W_{k_2}^2(\tau) \rangle - \langle W_{k_1}^2(\tau) \rangle - \langle W_{k_2}^2(\tau) \rangle + 1 \]
\[ = 2\langle W_{k_1}(\tau)W_{k_2}(\tau) \rangle = \frac{2}{(X^2(\tau))^2} \gamma^{2}(k_1 - k_2). \]

where
\[ \gamma^{2}(k_1 - k_2) = \langle (X(k_1 + \tau) - X(k_1))(X(k_2 + \tau) - X(k_2)) \rangle. \]

Thus, (B.6) can be recast as
\[ \frac{2}{N^2(X^2(\tau))^2} \sum_{k_1,k_2=1}^{N} \gamma^{2}(k_1 - k_2). \]

Note that
\[ \frac{2}{N^2(X^2(\tau))^2} = O\left( \frac{1}{N^{2-2x}} \right) = O\left( \frac{\tau}{\zeta(\tau)N^{-2}(N)} \right), \]

where \( \zeta(\tau), \eta(N) \) are defined by (A.5). Then, by (B.7), lemmas C.1–C.4 in [20], expression (B.6) is of the order \( O\left( \frac{\tau}{N^{2-2x}} \right) \), as claimed. □
The next lemma draws upon theorem B.2 and lemma B.1 to construct a concentration inequality for (B.1) (see also [22]).

**Lemma B.2.** Fix $-\infty < r < 1/2 < 3/2 < r'$. Then, for any $0 < \xi < 1/2$ and some $C > 0$,

$$
P(\varpi(\tau) \leq r) \leq \exp \left\{ - C \left( \frac{N}{\tau} \right)^{\xi} \right\} \quad (B.8)
$$

and

$$
P(\varpi(\tau) \geq r') \geq \exp \left\{ - C \left( \frac{N}{\tau} \right)^{\xi} \right\}. \quad (B.9)
$$

**Proof.** Let $W_j(\tau), j = 1, \ldots, N$ be as in (B.3). Then, for $\varpi(\tau)$ as in (B.2), we can write

$$
\varpi(\tau) = \frac{1}{N} \sum_{j=1}^{N} W_j^2(\tau) = \frac{1}{N} \mathbf{W}_N^T \mathbf{W}_N,
$$

where $\mathbf{W}_N = (W_1(\tau), \ldots, W_N(\tau))^T$ is a multivariate normal vector with covariance matrix $\Gamma$. Consider the spectral decomposition $\mathbf{Q} \Lambda \mathbf{Q}^T = \Gamma$, where $\mathbf{Q}$ is a $N \times N$ orthogonal matrix and $\Lambda = \text{diag}\{\lambda_1, \ldots, \lambda_N\}$ is a $N \times N$ diagonal matrix. Then, $\mathbf{W}_N \overset{d}{=} \mathbf{Q} \Lambda^{1/2} \mathbf{Z}_N$, where $\mathbf{Z}_N = (Z_1, \ldots, Z_N)^T \sim \mathcal{N}(0, I_N)$. $I_N$ is the $N \times N$ identity matrix, and $\overset{d}{=}$ denotes equality in distribution. Therefore,

$$
\varpi(\tau) \overset{d}{=} \frac{1}{N} (\mathbf{Q} \Lambda^{1/2} \mathbf{Z}_N)^T \mathbf{Q} \Lambda^{1/2} \mathbf{Z}_N = \frac{1}{N} \mathbf{Z}_N^T \Lambda \mathbf{Z}_N = \sum_{j=1}^{N} \eta_{j,N} Z_j^2, \quad (B.10)
$$

where $\eta_{j,N} = \frac{\lambda_j}{N}$. Let $\mathbf{\eta}_N = (\eta_{j,N})_{j=1}^{N}$ be the vector of coefficients $\eta_{j,N}$. By expression (B.4) in lemma B.1,

$$
\|\mathbf{\eta}_N\|_\infty^2 \leq \|\mathbf{\eta}_N\|_2^2 = \text{Var} \, \varpi(\tau) = \langle (\varpi(\tau) - 1)^2 \rangle = O \left( \frac{\xi}{N} \right). \quad (B.11)
$$

By theorem B.2, by using the same argument as in the proof of lemma C.3 in [94], and applying the bound (B.11),

$$
P(\varpi(\tau) \leq r) = P \left( \sum_{j=1}^{N} \eta_{j,N} (Z_j^2 - 1) \leq r - 1 \right) \leq \exp \left\{ - \frac{C}{\|\mathbf{\eta}_N\|_2^2} \right\} \leq \exp \left\{ - C \left( \frac{N}{\tau} \right)^{\xi} \right\}
$$

for some $C > 0$. Thus, (B.8) follows. To show (B.9), it suffices to adapt the proof of expression (C.34) in [94]. In fact, by (B.11), $0 < \xi < 1/2$ and theorem B.2,

$$
P(\varpi(\tau) \geq r') = P \left( \sum_{j=1}^{N} \eta_{j,N} (Z_j^2 - 1) \geq r' - 1 \right)
$$

$$
\leq P \left( \sum_{j=1}^{N} \eta_{j,N} (Z_j^2 - 1) \geq 2 \|\mathbf{\eta}_N\|_2 \left( \frac{N}{\tau} \right)^{\xi/2} + 2 \|\mathbf{\eta}_N\|_\infty \left( \frac{N}{\tau} \right)^{\xi} \right) \leq \exp \left\{ - C \left( \frac{N}{\tau} \right)^{\xi} \right\}.
$$

\hfill \Box

The following lemma is used in the proof of lemma B.6.
Lemma B.3. Let $p \geq 1$, there is a constant $K_p$ only depending on $p$ such that
\[
\langle|\log \varpi(\tau)|^p\rangle \leq K_p.
\]
\[\text{(B.12)}\]

Proof. By (B.10), $\varpi(\tau)$ is a nonnegative weighted sum of independent chi-squared random variables, where not all weights are zero. Then, relation (B.12) is a consequence of expression (96) in [63], p 184.

The following lemma can be shown based on lemma B.2 and an adaptation of the proof of expressions (C.38) and (C.39) in [94], which pertains to higher order (cross)moments of wavelet variance terms.

Lemma B.4. Let $\kappa_1, \kappa_2 \in \mathbb{N} \cup \{0\}$, $\kappa_1 + \kappa_2 \geq 3$, and fix $0 < r < 1/2$. Then, as $N \to \infty$,
\[
\langle(\varpi(\tau_1) - 1)^{\kappa_1}(\varpi(\tau_2) - 1)^{\kappa_2}\rangle = O\left(\left(\frac{\tau}{N}\right)^2\right).
\]
\[\text{(B.13)}\]

Proof. By (B.10), $\varpi(\tau)$ is a nonnegative weighted sum of independent chi-squared random variables, where not all weights are zero. Then, relation (B.12) is a consequence of expression (96) in [63], p 184.

The following lemma can be shown based on lemma B.2 and an adaptation of the proof of expressions (C.38) and (C.39) in [94], which pertains to higher order (cross)moments of wavelet variance terms.

Lemma B.4. Let $\kappa_1, \kappa_2 \in \mathbb{N} \cup \{0\}$, $\kappa_1 + \kappa_2 \geq 3$, and fix $0 < r < 1/2$. Then, as $N \to \infty$,
\[
\langle(\varpi(\tau_1) - 1)^{\kappa_1}(\varpi(\tau_2) - 1)^{\kappa_2}\rangle = O\left(\left(\frac{\tau}{N}\right)^2\right).
\]
\[\text{(B.13)}\]

Proof. By (B.10), $\varpi(\tau)$ is a nonnegative weighted sum of independent chi-squared random variables, where not all weights are zero. Then, relation (B.12) is a consequence of expression (96) in [63], p 184.

The following lemma can be shown based on lemma B.2 and an adaptation of the proof of expressions (C.38) and (C.39) in [94], which pertains to higher order (cross)moments of wavelet variance terms.

Lemma B.4. Let $\kappa_1, \kappa_2 \in \mathbb{N} \cup \{0\}$, $\kappa_1 + \kappa_2 \geq 3$, and fix $0 < r < 1/2$. Then, as $N \to \infty$,
\[
\langle(\varpi(\tau_1) - 1)^{\kappa_1}(\varpi(\tau_2) - 1)^{\kappa_2}\rangle = O\left(\left(\frac{\tau}{N}\right)^2\right).
\]
\[\text{(B.13)}\]

Proof. By (B.10), $\varpi(\tau)$ is a nonnegative weighted sum of independent chi-squared random variables, where not all weights are zero. Then, relation (B.12) is a consequence of expression (96) in [63], p 184.

The following lemma can be shown based on lemma B.2 and an adaptation of the proof of expressions (C.38) and (C.39) in [94], which pertains to higher order (cross)moments of wavelet variance terms.

Lemma B.4. Let $\kappa_1, \kappa_2 \in \mathbb{N} \cup \{0\}$, $\kappa_1 + \kappa_2 \geq 3$, and fix $0 < r < 1/2$. Then, as $N \to \infty$,
\[
\langle(\varpi(\tau_1) - 1)^{\kappa_1}(\varpi(\tau_2) - 1)^{\kappa_2}\rangle = O\left(\left(\frac{\tau}{N}\right)^2\right).
\]
\[\text{(B.13)}\]

Proof. By (B.10), $\varpi(\tau)$ is a nonnegative weighted sum of independent chi-squared random variables, where not all weights are zero. Then, relation (B.12) is a consequence of expression (96) in [63], p 184.

The following lemma can be shown based on lemma B.2 and an adaptation of the proof of expressions (C.38) and (C.39) in [94], which pertains to higher order (cross)moments of wavelet variance terms.

Lemma B.4. Let $\kappa_1, \kappa_2 \in \mathbb{N} \cup \{0\}$, $\kappa_1 + \kappa_2 \geq 3$, and fix $0 < r < 1/2$. Then, as $N \to \infty$,
\[
\langle(\varpi(\tau_1) - 1)^{\kappa_1}(\varpi(\tau_2) - 1)^{\kappa_2}\rangle = O\left(\left(\frac{\tau}{N}\right)^2\right).
\]
\[\text{(B.13)}\]

Proof. By (B.10), $\varpi(\tau)$ is a nonnegative weighted sum of independent chi-squared random variables, where not all weights are zero. Then, relation (B.12) is a consequence of expression (96) in [63], p 184.

The following lemma can be shown based on lemma B.2 and an adaptation of the proof of expressions (C.38) and (C.39) in [94], which pertains to higher order (cross)moments of wavelet variance terms.

Lemma B.4. Let $\kappa_1, \kappa_2 \in \mathbb{N} \cup \{0\}$, $\kappa_1 + \kappa_2 \geq 3$, and fix $0 < r < 1/2$. Then, as $N \to \infty$,
\[
\langle(\varpi(\tau_1) - 1)^{\kappa_1}(\varpi(\tau_2) - 1)^{\kappa_2}\rangle = O\left(\left(\frac{\tau}{N}\right)^2\right).
\]
\[\text{(B.13)}\]

Proof. By (B.10), $\varpi(\tau)$ is a nonnegative weighted sum of independent chi-squared random variables, where not all weights are zero. Then, relation (B.12) is a consequence of expression (96) in [63], p 184.
\[
\left\langle \frac{(X(j_1 + \tau_1) - X(j_1))(X(j_2 + \tau_2) - X(j_2))}{\sqrt{\langle X^2(\tau_1) \rangle} \sqrt{\langle X^2(\tau_2) \rangle}} \right\rangle
\]

\[
= \left\langle \frac{(B_H(j_1 + \tau_1) - B_H(j_1))(B_H(j_2 + \tau_2) - B_H(j_2))}{\sqrt{\langle B^2_H(\tau_1) \rangle} \sqrt{\langle B^2_H(\tau_2) \rangle}} \right\rangle (1 + O(\tau^{-\delta})), \tag{B.17}
\]

where \( B_H \) is a standard fBm with Hurst parameter given by relation (12). By (B.3), (B.17) and expression (10) for the covariance function of fBm,

\[
\langle W_{j_1}(\tau_1)W_{j_2}(\tau_2) \rangle = \frac{1}{2} \left\{ \left| \frac{j_1 - j_2}{\sqrt{\tau_1\tau_2}} + \frac{\tau_1}{\sqrt{\tau_1\tau_2}} \right|^\alpha - \left| \frac{j_1 - j_2}{\sqrt{\tau_1\tau_2}} + \frac{\tau_1}{\sqrt{\tau_1\tau_2}} - \frac{\tau_2}{\sqrt{\tau_1\tau_2}} \right|^\alpha \right\} (1 + O(\tau^{-\delta})). \tag{B.18}
\]

Since \( \langle W_{j_1}(\tau_1)W_{j_2}(\tau_2) \rangle = \langle W_{j_1+k}(\tau_1)W_{j_2+k}(\tau_2) \rangle \), then by expression (B.16) we can rewrite (B.15) as

\[
\frac{1}{2N} \sum \frac{1}{N} \sum_{j_1,j_2 = 1}^{N-1} (2\langle W_{j_1}(\tau_1)W_{j_2}(\tau_2) \rangle)^2. \tag{B.19}
\]

Relation (B.14) is now a consequence of (B.18) and (B.19).

\[\square\]

Lemma B.6.

\[
\langle \log \varpi(\tau_1) \log \varpi(\tau_2) \rangle = \langle (\varpi(\tau_1) - 1)(\varpi(\tau_2) - 1) \rangle + o\left(\frac{\tau}{N}\right). \tag{B.20}
\]

\[\textbf{Proof.}\] For notational simplicity, assume \( k_1 = 1 \) and \( k_2 = 2 \). Let

\[
S_1 = \langle \log \varpi(\tau_1) \log \varpi(\tau_2) \rangle - \langle \log \varpi(\tau_1) \log \varpi(\tau_2) \rangle \mathbf{1}_{\{\min\{\varpi(\tau_1), \varpi(\tau_2)\} > 1\}}.
\]

\[
S_2 = \langle \log \varpi(\tau_1) \log \varpi(\tau_2) \rangle \mathbf{1}_{\{\min\{\varpi(\tau_1), \varpi(\tau_2)\} > 1\}} - \langle (\varpi(\tau_1) - 1)(\varpi(\tau_2) - 1) \rangle \mathbf{1}_{\{\min\{\varpi(\tau_1), \varpi(\tau_2)\} > 1\}}.
\]

\[
S_3 = \langle (\varpi(\tau_1) - 1)(\varpi(\tau_2) - 1) \rangle \mathbf{1}_{\{\min\{\varpi(\tau_1), \varpi(\tau_2)\} > 1\}} - \langle (\varpi(\tau_1) - 1)(\varpi(\tau_2) - 1) \rangle.
\]

Note that

\[
\langle \log \varpi(\tau_1) \log \varpi(\tau_2) \rangle = S_1 + S_2 + S_3.
\]

Therefore, establishing (B.20) is equivalent to showing that \( S_1 + S_2 + S_3 = o\left(\frac{\tau}{N}\right) \). It suffices to show that

\[
\max\{|S_1|, |S_2|, |S_3|\} = o\left(\frac{\tau}{N}\right). \tag{B.21}
\]

Let \( 0 < r < 1/2 \). We start off with \( S_2 \) by writing out the almost sure Taylor expansion...
\[ \log \varpi(\tau)1_{\varpi(\tau) > r} = \left\{ (\varpi(\tau) - 1) - \frac{1}{2} \left( \frac{\varpi(\tau) - 1}{\sigma_+^2(\varpi(\tau))} \right)^2 \right\} 1_{\varpi(\tau) > r}, \]  
(B.22)

where \( \sigma_+^2(\varpi(\tau)) \in \min\{\varpi(\tau), 1\}, \max\{\varpi(\tau), 1\} \). Then,

\[ \langle \log \varpi(\tau_1) \log \varpi(\tau_2) 1_{\min\{\varpi(\tau_1), \varpi(\tau_2)\} > r} \rangle \]
\[ = \langle (\varpi(\tau_1) - 1)(\varpi(\tau_2) - 1) 1_{\min\{\varpi(\tau_1), \varpi(\tau_2)\} > r} \rangle \]
\[ - \frac{1}{2} \langle (\varpi(\tau_1) - 1) \left( \frac{\varpi(\tau_2) - 1}{\sigma_+^2(\varpi(\tau_2))} \right)^2 1_{\min\{\varpi(\tau_1), \varpi(\tau_2)\} > r} \rangle \]
\[ - \frac{1}{2} \langle \left( \frac{\varpi(\tau_1) - 1}{\sigma_+^2(\varpi(\tau_1))} \right)^2 (\varpi(\tau_2) - 1) 1_{\min\{\varpi(\tau_1), \varpi(\tau_2)\} > r} \rangle \]
\[ + \frac{1}{4} \langle \left( \frac{\varpi(\tau_1) - 1}{\sigma_+^2(\varpi(\tau_1))} \right)^2 \left( \frac{\varpi(\tau_2) - 1}{\sigma_+^2(\varpi(\tau_2))} \right)^2 1_{\min\{\varpi(\tau_1), \varpi(\tau_2)\} > r} \rangle \].  
(B.23)

The second, third and fourth terms can be bounded by a similar argument, so we only develop the latter. Recast

\[ \left( \frac{\varpi(\tau) - 1}{\sigma_+^2(\varpi(\tau))} \right)^2 1_{\varpi(\tau) > r} = \left( \frac{\varpi(\tau) - 1}{\sigma_+^2(\varpi(\tau))} \right)^2 \left( 1_{\varpi > 1/2} 1_{\varpi(\tau) > r} + 1_{\varpi(\tau) > 1/2} \right) \]
\[ \leq \frac{2}{r^2} \left( (\varpi(\tau_1) - 1)^2 1_{1/2 > \varpi(\tau_1)} (\varpi(\tau_2) - 1)^2 1_{1/2 > \varpi(\tau_2)} \right) \]
\[ + \frac{1}{(r/2)^2} \left( (\varpi(\tau_1) - 1)^2 1_{\varpi(\tau_1) > 1/2} (\varpi(\tau_2) - 1)^2 1_{\varpi(\tau_2) > 1/2} \right) \]
\[ + \frac{1}{(r/2)^2} \left( (\varpi(\tau_1) - 1)^2 1_{\varpi(\tau_1) > 1/2} (\varpi(\tau_2) - 1)^2 1_{\varpi(\tau_2) > 1/2} \right) \]
\[ + \frac{1}{(1/2)^4} \left( (\varpi(\tau_1) - 1)^2 1_{\varpi(\tau_1) > 1/2} (\varpi(\tau_2) - 1)^2 1_{\varpi(\tau_2) > 1/2} \right). \]  
(B.25)

By (B.13), the fourth term in (B.25) is bounded by

\[ O\left( \left( \frac{\tau}{N} \right)^2 \right). \]  
(B.26)
By the Cauchy–Schwarz inequality, (B.8) and (B.13), the first term in the sum (B.25) is bounded by

\[
\frac{1}{r^2} \sqrt{\langle (\varpi(\tau_1) - 1)^4(\varpi(\tau_1) - 1)^4 \rangle} \sqrt{\langle 1_{\{\varpi(\tau_1) > r\}} 1_{\{\varpi(\tau_2) > r\}} \rangle} 
\leq \frac{1}{r^2} O\left(\frac{\tau}{N}\right) \sqrt{P(1/2 > \varpi(\tau_1) > r) P(1/2 > \varpi(\tau_2) > r)} 
\leq \frac{1}{r^2} O\left(\frac{\tau}{N}\right) \exp \left\{- C \left(\frac{N}{\tau}\right)^{\xi} \right\} = o\left(\frac{\tau}{N}\right). \quad (B.27)
\]

Again by the Cauchy–Schwarz inequality, (B.8) and (B.13), the second term in the sum (B.25) is bounded by

\[
\frac{4}{r^2} \sqrt{\langle (\varpi(\tau_1) - 1)^4(\varpi(\tau_1) - 1)^4 \rangle} \sqrt{\langle 1_{\{\varpi(\tau_1) > 1/2\}} 1_{\{\varpi(\tau_2) > r\}} \rangle} 
\leq \frac{4}{r^2} O\left(\frac{\tau}{N}\right) \sqrt{P(\varpi(\tau_1) > 1/2) P(1/2 > \varpi(\tau_2) > r)} 
\leq \frac{4}{r^2} O\left(\frac{\tau}{N}\right) \exp \left\{- C \left(\frac{N}{\tau}\right)^{\xi} \right\} = o\left(\frac{\tau}{N}\right). \quad (B.28)
\]

An analogous bound holds for the third term in the sum (B.25). Therefore, \|S_j\| is bounded by the right-hand side of (B.21). To tackle \(S_5\), rewrite it as

\[
\langle (\varpi(\tau_1) - 1)(\varpi(\tau_2) - 1) 1_{\{\varpi(\tau_1) > r\}} 1_{\{\varpi(\tau_2) \leq r\}} \rangle 
+ 1_{\{\varpi(\tau_1) \leq r\}} 1_{\{\varpi(\tau_2) > r\}} + 1_{\{\varpi(\tau_1) \leq r\}} 1_{\{\varpi(\tau_2) \leq r\}} \rangle. \quad (B.29)
\]

By the Cauchy–Schwarz inequality, (B.8) and (B.13), the first term on the right-hand side of (B.29) is bounded by

\[
\sqrt{\langle (\varpi(\tau_1) - 1)^2(\varpi(\tau_2) - 1)^2 \rangle} \sqrt{P(\varpi(\tau_2) \leq r)} 
\leq O\left(\frac{\tau}{N}\right) \exp \left\{- C \left(\frac{N}{\tau}\right)^{\xi} \right\} = o\left(\frac{\tau}{N}\right).
\]

Similar bounds hold for the remaining terms on the right-hand side of (B.29). Therefore, \|S_j\| is also bounded by the right-hand side of (B.21). As for \(S_1\), it can be reexpressed as

\[
\langle \log \varpi(\tau_1) \log \varpi(\tau_2) \rangle \left( 1_{\{\varpi(\tau_1) > r\}} 1_{\{\varpi(\tau_2) \leq r\}} + 1_{\{\varpi(\tau_1) \leq r\}} 1_{\{\varpi(\tau_2) > r\}} + 1_{\{\varpi(\tau_1) \leq r\}} 1_{\{\varpi(\tau_2) \leq r\}} \right). \quad (B.30)
\]

Note that, by lemma B.3, \(\langle \log^4 \varpi(\tau) \rangle\) is bounded. Then, by applying the Cauchy–Schwarz inequality twice, the first term on the right-hand side of (B.30) is bounded by

\[
\sqrt{\langle \log^2 \varpi(\tau_1) \log^2 \varpi(\tau_2) \rangle} \sqrt{P(\varpi(\tau_2) \leq r)} 
\leq \left( \langle \log^4 \varpi(\tau_1) \rangle \langle \log^4 \varpi(\tau_2) \rangle \right)^{1/4} \exp \left\{- C \left(\frac{N}{\tau}\right)^{\xi} \right\} = o\left(\frac{\tau}{N}\right).
\]


Similar bounds hold for the remaining terms on the right-hand side of (B.30). Therefore, $|S_1|$ is bounded by the right-hand side of (B.21). This shows (B.20).

**Lemma B.7.**

$$\langle \log \varpi(\tau) \rangle + \frac{1}{2}((\varpi(\tau) - 1)^2) = O\left(\frac{T}{N}\right). \quad (B.31)$$

**Proof.** Fix $0 < r < 1/2$. Let

$$T_1 = \langle \log \varpi(\tau) \rangle - \langle \log \varpi(\tau) 1_{\{\varpi(\tau) > r\}} \rangle,$$

$$T_2 = \langle \log \varpi(\tau) 1_{\{\varpi(\tau) > r\}} \rangle + \frac{1}{2}((\varpi(\tau) - 1)^2 1_{\{\varpi(\tau) > r\}}),$$

$$T_3 = \frac{1}{2}((\varpi(\tau) - 1)^2) - \frac{1}{2}((\varpi(\tau) - 1)^2 1_{\{\varpi(\tau) > r\}}).$$

Recall that, by lemma B.3, $\langle \log^2 \varpi(\tau) \rangle$ is bounded. Thus, by the Cauchy–Schwarz inequality and lemma B.2,

$$T_1 = \langle \log \varpi(\tau) 1_{\{\varpi(\tau) \leq r\}} \rangle \leq \sqrt{\langle \log^2 \varpi(\tau) \rangle} \sqrt{\mathbb{P}(\varpi(\tau) \leq r)} \leq \sqrt{\langle \log^2 \varpi(\tau) \rangle} \exp \left\{ - C \left(\frac{N}{T}\right)^{\frac{1}{2}} \right\} = o\left(\frac{T}{N}\right). \quad (B.32)$$

By a similar reasoning, we can further prove that

$$T_3 = o\left(\frac{T}{N}\right). \quad (B.33)$$

Now, we turn to $T_2$. By an almost sure Taylor expansion,

$$\log \varpi(\tau) 1_{\{\varpi(\tau) > r\}} = \left\{ (\varpi(\tau) - 1) - \frac{1}{2}((\varpi(\tau) - 1)^2 + \frac{1}{3} \left(\frac{\varpi - 1}{\sigma^2(\varpi)}\right)^3 \right\} 1_{\{\varpi(\tau) > r\}},$$

where $\sigma^2(\varpi(\tau)) \in [\min\{\varpi(\tau), 1\}, \max\{\varpi(\tau), 1\}]$. Then, $T_2$ is bounded by

$$\left| \langle (\varpi(\tau) - 1) 1_{\{\varpi(\tau) > r\}} \rangle \right| + \frac{1}{3} \left\| \left(\frac{\varpi - 1}{\sigma^2(\varpi)}\right)^3 1_{\{\varpi(\tau) > r\}} \right\|. \quad (B.34)$$

Since $\langle \varpi(\tau) - 1 \rangle = 0$, by the Cauchy–Schwarz inequality and lemmas B.1 and B.2, the first term in (B.34) can be bounded by

$$\left| \langle (\varpi(\tau) - 1) 1_{\{\varpi(\tau) > r\}} \rangle \right| = \left| \langle (\varpi(\tau) - 1) 1_{\{\varpi(\tau) > r\}} \rangle - \langle \varpi(\tau) - 1 \rangle \right|$$

$$= \left| \langle (\varpi(\tau) - 1) 1_{\{\varpi(\tau) \leq r\}} \rangle \right| \leq \sqrt{\langle (\varpi(\tau) - 1)^2 \rangle} \sqrt{\mathbb{P}(\varpi(\tau) \leq r)} = o\left(\frac{T}{N}\right).$$

Meanwhile, by the Cauchy–Schwarz inequality and lemmas B.2 and B.4, the second term in (B.34) is bounded by
\[
\frac{1}{3\tau^3} |((\varpi - 1)1_{\{1/2 > \varpi(\tau) > 1\}}| + \frac{1}{3(1/2)^3} |((\varpi - 1)1_{\{\varpi(\tau) > 1/2\}}|) \leq O\left(\frac{\tau}{N}\right). \quad (B.35)
\]

Thus,

\[
T_2 = O\left(\frac{\tau}{N}\right). \quad (B.36)
\]

Relations (B.32), (B.33) and (B.36) imply (B.31).

\[\square\]

**Appendix C. Bias and variance of \(E_{\text{stand}}\) and the asymptotic distribution of \(E\)**

We are now in a position to prove theorems C.1 and C.2 and proposition C.1, which give, respectively, asymptotically valid characterizations of the bias and variance involved in TAMSD-based estimation, and the asymptotic distribution of the standardized estimator (35).

The proof of theorem C.1 is a consequence of a Taylor expansion, followed by using estimates of the decay of TAMSD moments. Constructing such estimates requires using a concentration inequality (e.g. [5, 48]), which was done in appendix B.

**Theorem C.1.** For \(0 < \alpha < 3/2\), under the assumptions of theorem A.1, (17) holds.

**Proof.** The left-hand side of (17) can be rewritten as

\[
\left\langle \log \frac{M_N(\tau)}{X^2(\tau)} \right\rangle + \log \langle X^2(\tau) \rangle = \langle \log \varpi(\tau) \rangle + \frac{\langle X^2(\tau) \rangle}{\sigma^2 \tau^\alpha}. \quad (C.1)
\]

By proposition 1 in [20], we can rewrite the second sum term on the right-hand side of (C.1) as

\[
\log(1 + O(\tau^{-\delta})) = O(\tau^{-\delta}), \quad N \to \infty.
\]

By lemmas B.5 and B.7, we can recast the first sum term on the right-hand side of (C.1) as

\[
-\frac{1}{2} \left((\varpi(\tau) - 1)^2 + O\left(\frac{\tau}{N}\right) \right)
\]

\[
= -\frac{1}{4N} \sum_{i=N+1}^{N-1} \left( 1 - \frac{|i|}{N} \right) \left| \frac{i}{\tau} + 1 \right| \left| \frac{i}{\tau} - 1 \right|^\alpha \left| \frac{i}{\tau} \right|^\alpha + O(\tau^{-\delta}) + O\left(\frac{\tau}{N}\right).
\]

Thus, (17) follows.

\[\square\]

Next, the proof of theorem C.2 relies on Taylor expansions of the moments of the logarithm of the TAMSD.

**Theorem C.2.** For \(0 < \alpha < 3/2\), under the assumptions of theorem A.1, expression (25) holds.

**Proof.** For \(k_1, k_2 = 1, \ldots, m\), rewrite

\[
u_{k_1, k_2} = \text{Cov}(\log M_N(\tau_{k_1}), \log M_N(\tau_{k_2}))
\]

\[
= \langle \log M_N(\tau_{k_1}) - \langle \log M_N(\tau_{k_1}) \rangle \rangle \langle \log M_N(\tau_{k_2}) - \langle \log M_N(\tau_{k_2}) \rangle \rangle
\]
\[
\mathbb{E} \left( \log \varpi(\tau_{k_1}) - \langle \log \varpi(\tau_{k_1}) \rangle \right) \mathbb{E} \left( \log \varpi(\tau_{k_2}) - \langle \log \varpi(\tau_{k_2}) \rangle \right) \\
= \langle \log \varpi(\tau_{k_1}) \log \varpi(\tau_{k_2}) \rangle - \langle \log \varpi(\tau_{k_1}) \rangle \langle \log \varpi(\tau_{k_2}) \rangle.
\] (C.2)

By lemmas B.1 and B.7,\[
\langle \log \varpi(\tau_{k_1}) \rangle = O \left( \frac{\tau}{N} \right).
\]
Therefore, (C.2) can be reexpressed as\[
\langle \log \varpi(\tau_{k_1}) \log \varpi(\tau_{k_2}) \rangle + o \left( \frac{\tau}{N} \right).
\] (C.3)

By lemmas B.5, B.6 and C.1 (expression (C.20)), expression (25) holds. \(\Box\)

The proof of proposition C.1 builds upon Taylor expansions and characterizing the asymptotic behavior of the standardization term in the definition of the estimator (35).

**Proposition C.1.** Under the assumptions of theorem A.1, suppose \(0 < \alpha < 3/2\). Then, the estimator (35) satisfies\[
\Lambda^{-1/2}(A_{\text{stand}})(E - \xi) \xrightarrow{d} \mathcal{N}(0, I), \quad N \rightarrow \infty,
\] (C.4)
where the vector \(\xi\) is given by (1). In particular, the estimator is consistent, i.e.\[
E \xrightarrow{p} \xi.
\]

**Proof.** Recast the estimator (35) as\[
Z = (X^T Y^{-1}(A_{\text{stand}})X)^{1/2}(X^T Y^{-1}(A_{\text{stand}})X)^{-1} X^T Y^{-1}(A_{\text{stand}})y.
\] (C.5)

Rewrite \(Y(\alpha) = Y(\alpha, N)\) as to express the dependence of the latter matrix on \(N\). Define\[
\left( \frac{N}{\tau} Y(A_{\text{stand}}, N) \right)^{-1} := S(A_{\text{stand}}, N) = \left( s_{k_1, k_2}(A_{\text{stand}}, N) \right)_{k_1, k_2 = 1, \ldots, m},
\] (C.6)
\[
s_N(A_{\text{stand}}) = \sum_{k_1 = 1}^m \sum_{k_2 = 1}^m s_{k_1, k_2}(A_{\text{stand}}, N) \in \mathbb{R}.
\]

By (23), (25) and (C.20), we can write\[
S(A_{\text{stand}}, N) \xrightarrow{p} S(\alpha) = \left( s_{k_1, k_2}(\alpha) \right)_{k_1, k_2 = 1, \ldots, m}, \quad N \rightarrow \infty,
\] (C.7)
and\[
s(\alpha) := \sum_{k_1 = 1}^m \sum_{k_2 = 1}^m s_{k_1, k_2}(\alpha) \in \mathbb{R}
\] (C.8)
for some constant matrix \(S(\alpha)\). For notational simplicity, write \(s_N(A_{\text{stand}}, N)\) and \(s_{k_1, k_2}(N) = s_{k_1, k_2}(A_{\text{stand}}, N)\). Then,
\[ X^T S(\xi) X = \left( \sum_{k_1=1}^{m} \sum_{k_2=1}^{m} s_{N} \log \tau_{k_1, k_2} s_{N}(N) - \sum_{k_1=1}^{m} \sum_{k_2=1}^{m} (\log \tau_{k_1, k_2}) s_{N}(N) \right). \]

By a simple calculation and relation (C.7),

\[ c_w(N) := \det(X^T S(A_{\text{stand}}, N) X) \]

\[ = s_{N} \sum_{k_1=1}^{m} \sum_{k_2=1}^{m} \log w_{k_1, k_2} s_{N}(N) - \left( \sum_{k_1=1}^{m} \sum_{k_2=1}^{m} \log w_{k_1, k_2} s_{N}(N) \right)^2 \]

\[ \overset{p}{\rightarrow} s(\alpha) \sum_{k_1=1}^{m} \sum_{k_2=1}^{m} \log w_{k_1, k_2}(\alpha) - \left( \sum_{k_1=1}^{m} \sum_{k_2=1}^{m} \log w_{k_1, k_2}(\alpha) \right)^2 = c_w(\alpha). \]

Moreover, by (C.6),

\[ (X^T Y^{-1}(A_{\text{stand}}, N)^{-1} X^T Y^{-1}(A_{\text{stand}}, N) = (X^T S(A_{\text{stand}}, N) X)^{-1} X^T S(A_{\text{stand}}, N) \]

\[ = \frac{1}{e_w(N)} \log \tau \left\{ \sum_{k_1=1}^{m} \sum_{k_2=1}^{m} \log w_{k_1, k_2} s_{N}(N) - \sum_{k_1=1}^{m} \sum_{k_2=1}^{m} \log w_{k_1, k_2} s_{N}(N) \right\} \]

\[ =: \left( \log \tau \left( a_{n,j} + b_{n,j} \right) \right)_{j=1, \ldots, m}, \]

where the sequences of constants \( \{a_{n,j}\}_{N \in \mathbb{N}} \) and \( \{b_{n,j}\}_{N \in \mathbb{N}} \) converge to constants \( a_j \) and \( b_j \), respectively, for \( j = 1, \ldots, m \).

Recall that, for a symmetric positive definite matrix

\[ M = \begin{pmatrix} m_{11} & m_{12} \\ m_{12} & m_{22} \end{pmatrix}, \]

we can write its square root in closed form as

\[ M^{1/2} = \frac{1}{\sqrt{\text{tr}(M) + 2\sqrt{\det(M)}}} \begin{pmatrix} m_{11} + \sqrt{\det(M)} \\ m_{12} \\ m_{22} + \sqrt{\det(M)} \end{pmatrix}. \]

Therefore,

\[ (X^T S(A_{\text{stand}}) X)^{1/2} = \frac{1}{\sqrt{\text{tr}(X^T S(A_{\text{stand}}) X) + 2\sqrt{c_w(N)}}}. \]
\[
\left( s_N + \sqrt{c_w(N)} \right) \sum_{k_1=1}^{m} \sum_{k_2=1}^{m} \left( \log \tau_{k_1} \right) s_{k_1,k_2}(N) = \sum_{k_1=1}^{m} \sum_{k_2=1}^{m} \left( \log \tau_{k_1} \log \tau_{k_2} \right) s_{k_1,k_2}(N) + \sqrt{c_w(N)}. \tag{C.11}
\]

Note that
\[
tr(X^T S(A_{\text{stand}})X) + 2 \sqrt{c_w(N)} \sim \log^2 \tau s_N. \tag{C.12}
\]

By expressions (C.7), (C.8), (C.10)–(C.12),
\[
\begin{align*}
&\left( X^T Y^{-1}(A_{\text{stand}}, N)X \right)^{1/2} \left( X^T Y^{-1}(A_{\text{stand}}, N)X \right)^{-1} X^T Y^{-1}(A_{\text{stand}}, N) \\
&= \sqrt{\frac{N}{\tau}} \left( X^T S(A_{\text{stand}}, N)X \right)^{1/2} \left( X^T S(A_{\text{stand}}, N)X \right)^{-1} X^T S(A_{\text{stand}}, N) \\
&= \sqrt{\frac{N}{\tau}} \frac{1}{\sqrt{tr(X^T S(A_{\text{stand}}, N)) + 2 \sqrt{c_w(N)} c_w(N)}} \\
&\left( a_{n_j} \left[ \log \tau \right] (s_N + \sqrt{c_w(N)}) - \sum_{k_1=1}^{m} \sum_{k_2=1}^{m} \left( \log \tau_{k_1} \right) s_{k_1,k_2}(N) \right) + b_{n_j} (s_N + \sqrt{c_w(N)}) \\
&\left( a_{n_j} \left[ \log \tau \right] \sum_{k_1=1}^{m} \sum_{k_2=1}^{m} \left( \log \tau_{k_1} \right) s_{k_1,k_2}(N) \\
&- \sum_{k_1=1}^{m} \sum_{k_2=1}^{m} \left( \log \tau_{k_1} \log \tau_{k_2} \right) s_{k_1,k_2}(N) - \sqrt{c_w(N)} \right) \\
&+ b_{n_j} \left[ \sum_{k_1=1}^{m} \sum_{k_2=1}^{m} \left( \log \tau_{k_1} \right) s_{k_1,k_2}(N) \right] \\
&= \sqrt{\frac{N}{\tau}} \frac{1}{\sqrt{tr(X^T S(A_{\text{stand}}, N)) + 2 \sqrt{c_w(N)} c_w(N)}} \\
&\left( a_{n_j} \left[ \log \tau \right] \sqrt{c_w(N)} - \sum_{k_1=1}^{m} \sum_{k_2=1}^{m} \left( \log \omega_{k_1} \right) s_{k_1,k_2}(N) \right) + b_{n_j} (s_N + \sqrt{c_w(N)}) \\
&\left( a_{n_j} \left[ \log \tau \right] \sum_{k_1=1}^{m} \sum_{k_2=1}^{m} \left( \log \omega_{k_1} \right) s_{k_1,k_2}(N) \\
&- \sum_{k_1=1}^{m} \sum_{k_2=1}^{m} \left( \log \omega_{k_1} \log \omega_{k_2} \right) s_{k_1,k_2}(N) \\
&- \sqrt{c_w(N)} \right) + b_{n_j} \left[ \log \tau s_N + \sum_{k_1=1}^{m} \sum_{k_2=1}^{m} \left( \log \omega_{k_1} \right) s_{k_1,k_2}(N) \right] \\
&= \sqrt{\frac{N}{\tau}} \frac{1}{\sqrt{s(\alpha) c_w(\alpha)}} \\
&\left( a_{j} \sqrt{c_w(\alpha)} \right) \left( b_{j} s(\alpha) - a_{j} \sum_{k_1=1}^{m} \sum_{k_2=1}^{m} \left( \log \omega_{k_1} \right) s_{k_1,k_2}(\alpha) \right) \\
&= \sqrt{\frac{N}{\tau}} \Psi \in \mathbb{R}^{2 \times m}, \tag{C.13}
\end{align*}
\]

as \( N \to \infty \). For \( y \) and \( X \) as in (30), rewrite the left-hand side of expression (C.4) as
\[
\left( X^T Y^{-1}(A_{\text{stand}}, N)X \right)^{1/2} (E - \xi)
\]
\[(X^T Y^{-1}(A_{\text{stand}, N})X)^{1/2}(X^T Y^{-1}(A_{\text{stand}, N})X)^{-1}X^T Y^{-1}(A_{\text{stand}, N})(y - X\xi). \]  

(C.14)

Recast

\[ y = \left( \log[M_N(\tau_k) e^{\frac{\beta_N(A_{\text{stand}, \tau_k})}{\sigma^2 \tau_k^\alpha}}] \right)_{k=1, \ldots, m}. \]

By entrywise first order Taylor expansions,

\[ y - X\xi = \left( \log \left( \frac{M_N(\tau_k) e^{\frac{\beta_N(A_{\text{stand}, \tau_k})}{\sigma^2 \tau_k^\alpha}}}{\sigma^2 \tau_k^\alpha} \right) - 1 \right)_{k=1, \ldots, m} + \left( o\left( \frac{M_N(\tau_k) e^{\frac{\beta_N(A_{\text{stand}, \tau_k})}{\sigma^2 \tau_k^\alpha}}}{\sigma^2 \tau_k^\alpha} - 1 \right) \right)_{k=1, \ldots, m}. \]  

(C.15)

However, for \( k = 1, \ldots, m \), the first term on the right-hand side of (C.15) can be reexpressed as

\[ \frac{M_N(\tau_k)}{\sigma^2 \tau_k^\alpha} \left( e^{\frac{\beta_N(A_{\text{stand}, \tau_k})}{\sigma^2 \tau_k^\alpha}} - 1 \right)_{k=1, \ldots, m} + \frac{M_N(\tau_k)}{\sigma^2 \tau_k^\alpha} - 1. \]  

(C.16)

Again by a first order Taylor expansion,

\[ e^{\frac{\beta_N(A_{\text{stand}, \tau_k})}{\sigma^2 \tau_k^\alpha}} - 1 = \frac{\beta_N(A_{\text{stand}, \tau_k})}{N} + o\left( \frac{\tau_k}{N} \right). \]  

(C.17)

Therefore, by (C.13), (C.16) and (C.17), we can rewrite relation (C.14) as

\[ (\Psi + o_P(1)) \sqrt{\frac{N}{\tau}} \left( \frac{M_N(\tau_k)}{\sigma^2 \tau_k^\alpha} - 1 \right)_{k=1, \ldots, m} + o_P\left( \sqrt{\frac{\tau}{N}} \right). \]

(C.18)

Expression (C.4) is a consequence of (C.18) and theorem A.1, where the estimator (C.5) is asymptotically standardized. \( \Box \)

The following lemma establishes the convergence of the main bias and variance factors and is used in the proofs of proposition C.1 and theorem C.2.

**Lemma C.1.** For \( 0 < \alpha < 3/2 \), consider the main bias and variance factors (18) and (26), respectively, under the assumptions of theorem A.1. Then, there are functions \( \beta(\alpha, \cdot) > 0 \) and \( \varsigma(\alpha, \cdot, \cdot) > 0 \) such that

\[ \beta_N(A_{\text{stand}, \tau_k}) \xrightarrow{P} \left( \beta(\alpha, \tau_k) \right)_{k=1, \ldots, m}, \]

(C.19)

\[ \varsigma_N(A_{\text{stand}, \tau_k, \tau_{k1}})_{k_1, k_2=1, \ldots, m} \xrightarrow{P} \left( \varsigma(\alpha, k_1, k_2) \right)_{k_1, k_2=1, \ldots, m}, \]

(C.20)

as \( N \to \infty \).
Proof. From expression (25) for the variance term $v_{k_1k_2}(\xi)$, recast
\[
\varsigma_N(\alpha, \tau_{k_1}, \tau_{k_2}) = \frac{1}{2} \left\{ \sum_{|i| \leq \tau} - \sum_{\tau + 1 \leq |i| \leq N-1} \right\} \left( 1 - \frac{|i|}{N} \right) \left\{ \left| \frac{i}{\tau \sqrt{w_{k_1} w_{k_2}}} + \sqrt{\frac{w_{k_2}}{w_{k_1}}} \right|^2 \right\}.
\]

The second sum term on the right-hand side of (C.21) is bounded by
\[
\frac{1}{2} \left( 1 - \frac{\tau}{N} \right) \sum_{\tau + 1 \leq |i| \leq N-1} \left\{ \left| \frac{i}{\tau \sqrt{w_{k_1} w_{k_2}}} + \sqrt{\frac{w_{k_2}}{w_{k_1}}} \right|^2 \right\}.
\]

as $N \to \infty$. On the other hand, by condition (A.3) and the dominated convergence theorem, the first sum term on the right-hand side of (C.21) converges to
\[
\varsigma(\alpha, k_1, k_2) := \frac{1}{2} \int_{-1}^{1} \left\{ \left| \frac{y}{\sqrt{w_{k_1} w_{k_2}}} + \sqrt{\frac{w_{k_2}}{w_{k_1}}} \right|^2 \right\} dy > 0.
\]

Moreover, by corollary A.1, $A_{\text{stand}} \Rightarrow \alpha$. So, pick a small enough $\varepsilon_0$ such that $\alpha \in (\varepsilon_0, 3/2 - \varepsilon_0)$. Let $A = \{ \omega : A_{\text{stand}}(\omega) \in (\alpha - \frac{\varepsilon_0}{2}, \alpha + \frac{\varepsilon_0}{2}) \}$. In the set $A$, by a simple adaptation of the argument leading to (C.22) and the convergence to (C.23),
\[
\varsigma_N(A_{\text{stand}}, \tau_{k_1}, \tau_{k_2}) \to \varsigma(\alpha, k_1, k_2), \quad N \to \infty,
\]
where $\mathbb{P}(A) \to 1$. This shows (C.20). A similar argument can be used to show (C.19).
Appendix D. Pseudocode for generating the improved TAMSD-based estimator $E$

Generating the improved pathwise estimator $E$ (see (29))

Input:
- one observed particle path $\{X_1, X_2, \ldots, X_N\}$ of length $N$;
- regression lag values $\tau_k$, $k = 1, \ldots, m$ (typically, $\tau_k = \tau_{w_k}$, $w_1 < \ldots < w_m$, $\tau \ll N$);
- the expression for the asymptotic covariance matrix $\Upsilon(\alpha)$ as a function of $\alpha$;

Step 1: obtain the standard estimator $A_{\text{stand}}$ over the chosen lag values;
Step 2: estimate the asymptotic covariance matrix $\Upsilon(\xi)$ by means of $\Upsilon(A_{\text{stand}})$ (see (27));
Step 3: use $A_{\text{stand}}$ and the estimator (19) of the bias vector to produce the bias-corrected regression system (28);
Step 4: obtain the estimator $E$ by means of $\Upsilon(A_{\text{stand}})$-based GLS on the bias-corrected regression system (28).

ORCID iDs

Gustavo Didier ⚖ https://orcid.org/0000-0003-1363-5868

References

[1] Andreanov A and Grebenkov D 2012 Time-averaged MSD of Brownian motion J. Stat. Mech. P07001
[2] Barkai E, Garini Y and Metzler R 2012 Strange kinetics of single molecules in living cells Phys. Today 65 29–35
[3] Berglund A J 2010 Statistics of camera-based single-particle tracking Phys. Rev. E 82 011917
[4] Bertseva E, Grebenkov D S, Schmidhauser P, Gribkova S, Jeney S and Forró I 2012 Optical trapping microrheology in cultured human cells Eur. Phys. J. E 35 63
[5] Boucheron S, Lugosi G and Massart P 2013 Concentration Inequalities: a Nonasymptotic Theory of Independence (Oxford: Oxford University Press)
[6] Boyer D, Dean D S, Mejía-Monasterio C and Oshanin G 2012 Optimal estimates of the diffusion coefficient of a single Brownian trajectory Phys. Rev. E 85 031136
[7] Boyer D, Dean D S, Mejía-Monasterio C and Oshanin G 2013 Distribution of the least-squares estimators of a single Brownian trajectory diffusion coefficient J. Stat. Mech. P04017
[8] Briane V, Kervrann C and Vимond M 2018 Statistical analysis of particle trajectories in living cells Phys. Rev. E 97 062121
[9] Burnecki K 2012 FARIMA processes with application to biophysical data J. Stat. Mech. P05015
[10] 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 1–11
[11] Burnecki K, Kepten E, Janczura J, Bronshtein I, Garini Y and Weron A 2012 Universal algorithm for identification of fractional Brownian motion. A case of telomere subdiffusion Biophys. J. 103 1839–47
[12] Burnecki K, Muszkiewicz M, Sikora G and Weron A 2012 Statistical modelling of subdiffusive dynamics in the cytoplasm of living cells: a FARIMA approach Europhys. Lett. 98 10004
[13] Burov S, Jeon J H, Metzler R and Barkai E 2011 Single particle tracking in systems showing anomalous diffusion: the role of weak ergodicity breaking Phys. Chem. Chem. Phys. 13 1800–12
[14] Cheridito P, Kawaguchi H and Maejima M 2003 Fractional Ornstein–Uhlenbeck processes Electron. J. Probab. 8 1–14
[15] Christensen R 2011 Plane Answers to Complex Questions: the Theory of Linear Models (New York: Springer)
[16] Dawson M, Wirtz D and Hanes J 2003 Enhanced viscoelasticity of human cystic fibrotic sputum correlates with increasing microheterogeneity in particle transport J. Biol. Chem. 278 50393–401
[17] Deng W and Barkai E 2009 Ergodic properties of fractional Brownian–Langevin motion Phys. Rev. E 79 011112
[18] Didier G and Fricks J 2014 On the wavelet-based simulation of anomalous diffusion J. Stat. Comput. Simul. 84 697–723
[19] Didier G, McKinley S A, Hill D B and Fricks J 2012 Statistical challenges in micro rheology J. Time Ser. Anal. 33 724–43
[20] Didier G and Zhang K 2017 The asymptotic distribution of the pathwise mean squared displacement in single particle tracking experiments J. Time Ser. Anal. 38 395–416
[21] Dobrushin R and Major P 1979 Non-central limit theorems for non-linear functional of Gaussian fields Probab. Theory Relat. Fields 50 27–52
[22] Gajda J, Wyłomańska A, Kantz H, Chechkin A and Sikora G 2018 Large deviations of time-averaged statistics for Gaussian processes Stat. Probab. Lett. 143 47–55
[23] Ghosh S K, Cherstvy A G, Grebenkov D S and Metzler R 2016 Anomalous, non-Gaussian tracer diffusion in crowded two-dimensional environments New J. Phys. 18 013027
[24] Giraitis L and Surgailis D 1985 Clt and other limit theorems for functionals of Gaussian processes Z. Wahrscheinlichkeitstheor. Verwandle Geb. 70 191
[25] Giraitis L and Surgailis D 1990 A central limit theorem for quadratic forms in strongly dependent linear variables and its application to asymptotical normality of Whittle’s estimate Probab. Theory Relat. Fields 86 87–104
[26] Gorman J and Greene E C 2008 Visualizing one-dimensional diffusion of proteins along DNA Nat. Struct. Mol. Biol. 15 768–74
[27] Grebenkov D 2011 Probability distribution of the time-averaged mean-square displacement of a Gaussian process Phys. Rev. E 84 031124
[28] Grebenkov D 2011 Time-averaged quadratic functionals of a Gaussian process Phys. Rev. E 83 061117
[29] Grebenkov D 2013 Optimal and suboptimal quadratic forms for noncentered Gaussian processes Phys. Rev. E 88 032140
[30] Grebenkov D S, Vahabi M, Bertseva E, Forró L and Jeney S 2013 Hydrodynamic and subdiffusive motion of tracers in a viscoelastic medium Phys. Rev. E 88 040701
[31] Guyon X and León J 1989 Convergence en loi des H-variations d’un processus Gaussien stationnaire sur R Ann. IHP: Probab. Stat. 25 265–82
[32] Halford S E and Marko J F 2004 How do site-specific DNA-binding proteins find their targets? Nucl. Acids Res. 32 3040–52
[33] Helenius J, Brouhard G, Kalaidzidis Y, Diez S and Howard J 2006 The depolymerizing kinesin Helenius J, Brouhard G, Kalaidzidis Y, Diez S and Howard J 2006 The depolymerizing kinesin
[34] Hess S T, Girirajan T and Mason M D 2006 Ultra-high resolution imaging by fluorescence photoactivation localization microscopy Biophys. J. 91 4258–72
[35] Hess S T, Girirajan T and Mason M D 2006 Ultra-high resolution imaging by fluorescence photoactivation localization microscopy Biophys. J. 91 4258–72
[36] Hozé N and Hochman D 2017 Statistical methods for large ensembles of super-resolution stochastic single particle trajectories in cell biology Annu. Rev. Stat. Appl. 4 189–223
[37] Isserlis L 1916 On certain probable errors and correlation coefficients of multiple frequency distributions with skew regression Statist. Probab. Lett. 11 185–90
[38] Jeon J H, 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
[39] 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
[43] Kou S C 2008 Stochastic modeling in nanoscale biophysics: subdiffusion within proteins Ann. Appl. Stat. 2 501–35
[44] Kou S C and Xie X S 2004 Generalized Langevin equation with fractional Gaussian noise: subdiffusion within a single protein molecule Phys. Rev. Lett. 93 180603
[45] Lai S, Wang Y, Cone R, Wirtz D and Hanes J 2009 Altering mucus rheology to solidify human mucus at the nanoscale PLoS One 4 e4294
[46] Lasne D, Blab G A, Bercaud S, Heine M, Groc L, Choquet D, Cognet L and Lounis B 2006 Single nanoparticle photothermal tracking (SNaPT) of 5 nm gold beads in live cells Biophys. J. 91 4598–604
[47] Laurent B and Massart P 2000 Adaptive estimation of a quadratic functional by model selection Ann. Stat. 28 1302–38
[48] Ledoux M 2005 The Concentration of Measure Phenomenon vol 89 (Providence, RI: American Mathematical Society)
[49] Levine A and Lubensky T 2000 One-and two-particle microrheology Phys. Rev. Lett. 85 1774–7
[50] Lieleo G, Vladescu I and Ribbeck K 2010 Characterization of particle translocation through mucin hydrogels Biophys. J. 98 1782
[51] Lysy M, Pillai N, Hill D B, Forest M G, Mellnik J, Vasquez P and McKinley S A 2016 Model comparison for single particle tracking in biological fluids J. Am. Stat. Assoc. 111 1413–26
[52] Major P 1981 Limit theorems for non-linear functionals of Gaussian sequences Z. Wahrscheinlichkeitstheor. Verwandte Geb. 57 129–8
[53] Mason E, Roueff F and Taqqu M S 2007 Central limit theorem for the log-regression wavelet analysis of single-particle tracking in biological fluids Phys. Rev. E 74 1250–3
[54] Metzler R, Jeon J H and Cherstvy A 2016 Non-Brownian diffusion in lipid membranes: experiments and simulations Biochimica Biophys. Acta 1858 2451–67
[55] Metzler R, Tejedor V, Jeon J, He Y, Deng W, Burov S and Barkai E 2009 Analysis of single particle trajectories: from normal to anomalous diffusion Acta Phys. Pol. B 40 1315–31
[56] Michalet X and Berglund A J 2012 Optimal diffusion coefficient estimation in single-particle tracking Phys. Rev. E 85 061916
[57] Minoura I, Katayama E, Sekimoto K and Muto E 2010 One-dimensional Brownian motion of charged nanoparticles along microtubules: a model system for weak binding interactions Biophys. J. 98 1589–97
[58] Moulines E, Roueff F and Taqqu M S 2007 Central limit theorem for the log-regression wavelet estimation of the memory parameter in the Gaussian semi-parametric context Fractals 15 301–13
[59] Moulines E, Roueff F and Taqqu M 2007 On the spectral density of the wavelet coefficients of long-memory time series with application to the log-regression estimation of the memory parameter J. Time Ser. Anal. 28 155–87
[60] Moulines E, Roueff F and Taqqu M 2008 A wavelet Whittle estimator of the memory parameter of a nonstationary Gaussian time series Ann. Stat. 36 1925–56
[61] Nandi A, Heinrich D and Lindner B 2012 Distributions of diffusion measures from a local mean-square displacement analysis Phys. Rev. E 86 021926
[62] Nguyen H D and McKinley S A 2018 Anomalous diffusion and the generalized Langevin equation SIAM J. Math. Anal. 50 5119–60
[63] Nishimura S Y, Lord S J, Klein L O, Willets K A, He M, Lu Z, Twieg R J and Moerner W E 2006 Diffusion of lipid-like single-molecule fluorophores in the cell membrane J. Phys. Chem. B 110 8151–7
[64] Ottobre M and Pavliotis G 2011 Asymptotic analysis for the generalized Langevin equation Nonlinearity 24 1629
[65] Pipiras V and Taqqu M S 2017 Long-Range Dependence and Self-Similarity (Cambridge Series on Statistical and Probabilistic Mathematics) (Cambridge: Cambridge University Press)
[70] Prakasa Rao B L S 2010 Statistical Inference for Fractional Diffusion Processes (Wiley Series in Probability and Statistics) (New York: Wiley)

[71] Qian H, Sheetz M and Elson E 1991 Single particle tracking. Analysis of diffusion and flow in two-dimensional systems Biophys. J. 60 910–21

[72] Reighard K P, Ehre C, Rushton Z L, Ahonen M J R, Hill D B and Schoenfisch M H 2017 Role of nitric oxide-releasing chitosan oligosaccharides on mucus viscoelasticity ACS Biomater. Sci. Eng. 3 1017–26

[73] Reighard K P, Hill D B, Dixon G A, Worley B V and Schoenfisch M H 2015 Disruption and eradication of P. aeruginosa biofilms using nitric oxide-releasing chitosan oligosaccharides Biofouling 31 775–87

[74] Rosenblatt M 1961 Independence and dependence Proc. 4th Berkeley Symp. on Mathematical Statistics and Probability vol 2

[75] Sandev T and Metzler R and Tomovski Ž 2012 Velocity and displacement correlation functions for fractional generalized Langevin equations Fractional Calculus Appl. Anal. 15 426–50

[76] Saxton M 1994 Anomalous diffusion due to obstacles: a monte carlo study Biophys. J. 66 394–401

[77] Saxton M 1996 Anomalous diffusion due to binding: a monte carlo study Biophys. J. 70 1250–62

[78] Shiryaev A 2000 Probability Theory (New York: Springer)

[79] Sokolov I M 2008 Statistics and the single molecule

[80] Smith M B, Karatekin E, Gohlke A, Mizuno H, Watanabe N and Vavylonis D 2011 Interactive, computer-assisted tracking of speckle trajectories in fluorescence microscopy: application to actin polymerization and membrane fusion Biophys. J. 101 1794–804

[81] Sokolov I M 2012 Velocity and displacement correlation functions for fractional generalized Langevin equations Fractional Calculus Appl. Anal. 15 426–50

[82] Sonesson A W, Elofsson U M, Callisen T H and Brismar H 2007 Tracking single lipase molecules on a trimyristin substrate surface using quantum dots Langmuir 23 8352–6

[83] Tafvizi A, Mirny L A and van Oijen A M 2011 Dancing on DNA: kinetic aspects of search processes

[84] Taqqu M S 2003 Properties and numerical evaluation of the rosenblatt distribution Bernoulli 19 982–1005

[85] Veitch D and Abry P 1999 A wavelet-based joint estimator of the parameters of long-range dependence IEEE Trans. Inf. Theory 45 878–97

[86] Vestergaard C L, Blainey P C and Flyvbjerg H 2014 Optimal estimation of diffusion coefficients from single-particle trajectories Phys. Rev. E 89 022726

[87] Wieser S and Schütz G J 2008 Tracking single molecules in the live cell plasma membrane—do’s and do not’s Methods 46 131–40

[88] Zwanzig R 2001 Nonequilibrium Statistical Mechanics (Oxford: Oxford University Press)