HAL
open science

# Ergodic least-squares estimators of the generalize diffusion coefficient for fractional Brownian motion

Denis Boyer, David S. Dean, Carlos Mejia-Monasterio, Gleb Oshanin

## To cite this version:

Denis Boyer, David S. Dean, Carlos Mejia-Monasterio, Gleb Oshanin. Ergodic least-squares estimators of the generalize diffusion coefficient for fractional Brownian motion. Physical Review E : Statistical, Nonlinear, and Soft Matter Physics, 2013, 87 (3), pp.030103. 10.1103/PhysRevE.87.030103 . hal-00820816

**HAL Id: hal-00820816**

**https://hal.science/hal-00820816v1**

Submitted on 1 Mar 2018

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

PHYSICAL REVIEW E **87**, 030103(R) (2013)

# Ergodic least-squares estimators of the generalized diffusion coefficient for fractional Brownian motion

Denis Boyer,[1,2,*] David S. Dean,[3,†] Carlos Mejía-Monasterio,[4,5,‡] and Gleb Oshanin[6,§]

[1]*Instituto de Física, Universidad Nacional Autónoma de México, Distrito Federal 04510, Mexico*

[2]*Centro de Ciencias de la Complejidad, Universidad Nacional Autónoma de México, Distrito Federal 04510, Mexico*

[3]*Université de Bordeaux and CNRS, Laboratoire Ondes et Matière d'Aquitaine, UMR No. 5798, F-33400 Talence, France*

[4]*Laboratory of Physical Properties, Technical University of Madrid, Avenida Complutense s/n 28040, Madrid, Spain*

[5]*Department of Mathematics and Statistics, University of Helsinki, P.O. Box 68, FIN-00014 Helsinki, Finland*

[6]*Laboratoire de Physique Théorique de la Matière Condensée, UMR No. 7600, CNRS, Université Pierre et Marie Curie, 4 place Jussieu, 75252 Paris, Cedex 5, France*

(Received 4 February 2013; published 29 March 2013)

We analyze a class of estimators of the generalized diffusion coefficient for fractional Brownian motion $B_t$ of known Hurst index $H$, based on weighted functionals of the single-time square displacement. We show that for a certain choice of the weight function these functionals possess an ergodic property and thus provide the true ensemble-average generalized diffusion coefficient to any necessary precision from single-trajectory data, but at the expense of a progressively higher experimental resolution. Convergence is fastest around $H \simeq 0.30$, a value in the subdiffusive regime.

DOI: 10.1103/PhysRevE.87.030103 PACS number(s): 02.50.−r, 05.10.Gg, 82.37.−j, 87.80.Nj

Single-molecule spectroscopy techniques allow the tracking of single particles over a wide range of time scales [1–3]. In complex media such as living cells, a number of recent studies have reported evidence for subdiffusive transport of particles such as proteins [4], viruses [5], chromosome monomers [6], mRNA [7], and lipid granules [8]. Subdiffusion is typically characterized by a sublinear growth with time of the mean square displacement (MSD) $\mathbb{E}(\mathbf{B}_t^2) = Kt^\nu$ with $\nu < 1$, where $\mathbf{B}_t$ is the particle position at time $t$, $\mathbb{E}$ denotes the ensemble average, and $K$ is a generalized diffusivity.

A growing body of single-trajectory studies suggest that fractional Brownian motion (FBM), among the variety of stochastic processes that produce subdiffusion, may be a model particularly relevant to subcellular transport. Fractional Brownian motion is a Gaussian continuous-time random process with stationary increments and is characterized by a so-called Hurst index $H = \nu/2$. If $H < 1/2$, trajectories are subdiffusive with increments that are negatively and long-range correlated [9]. Such correlations were observed in subdiffusing mRNA molecules [10], RNA proteins, and chromosomal loci [4] within *E. coli* cells. Similarly, FBM can be used to describe the dispersion of apoferritin proteins in crowded dextran solutions [11] and of lipid molecules in lipid bilayers [12].

Whereas the determination of an anomalous exponent from data has been extensively studied, as it demonstrates deviation from standard Brownian motion (BM), the problem of estimating the generalized diffusion constant $K$ has received much less attention. It appears that $K$ is much more sensitive than $\nu$ to many biological factors and its precise determination can potentially yield valuable information about the kinetics of transcription, translation, and other physicobiological processes. The generalized diffusivity of RNA molecules in bacteria is greatly affected (either positively or negatively) by perturbations, for instance, treatment with antibiotic drugs, which have, however, a negligible effect on $\nu$ [4]. Likewise, the coefficient $K$ of lipids in membranes is strongly reduced by small cholesterol concentrations, whereas $\nu$ remains unchanged [12]. In the context of search problems, a particle following a subdiffusive FBM actually explores the three-dimensional space more compactly than a BM and can have a higher probability of eventually encountering a nearby target [13]. The larger the value of $K$, the faster this local exploration.

In this paper, generalizing our previous results for standard BM [14], we present a method to estimate the ensemble-average diffusivity $K$ from the analysis of single FBM trajectories of *a priori* known anomalous exponent. Estimating diffusion constants from data is not an easy task when trajectories are few and ensemble averages cannot be performed. Brownian motion and FBM are ergodic processes and time averages tend to ensemble averages, but convergence can be slow [15]. For finite trajectories of finite resolution, variations by orders of magnitude have been observed for estimators of the normal diffusion coefficient obtained from single particles moving along DNA [16], in the plasma membrane [2] or in the cytoplasm of mammalian cells [17]. Large fluctuations are also manifest in subdiffusive cases [4,12].

A broad dispersion in the measures of the diffusion coefficient raises important questions about optimal fitting methodologies. A reliable estimator must possess an ergodic property so that its most probable value should converge to the true ensemble average independently of the trajectory considered and its variance should vanish as the observation time increases. Recently, much effort has been invested in the analysis of this challenging problem and several different estimators have been analyzed, based, e.g., on the sliding time-average square displacement [18,19], the mean

*boyer@fisica.unam.mx
†david.dean@u-bordeaux1.fr
‡carlos.mejia@upm.es
§oshanin@lptmc.jussieu.fr

 ©2013 American Physical Society

length of a maximal excursion [20], the maximum likelihood approximation [21–25], and optimal weighted least-squares functionals [14].

Our aim here is to determine an ergodic least-squares estimator for the generalized diffusion coefficient when the underlying stochastic motion is given by a FBM. The estimators considered here are single-time quantities, unlike others based on fits of two-time quantities such as the time-average MSD.

Let us consider a fractional Brownian motion $B_t$ in one dimension with $B_0 = 0$ and zero expectation value for all $t \in [0,T]$, where $T$ is the total observation time. The covariance function of the process is given by [9]

$$\begin{aligned}\mathrm{Cov}(B_t,B_s) &= \mathbb{E}\{(B_t - \mathbb{E}\{B_t\})(B_s - \mathbb{E}\{B_s\})\}\\ &= \frac{K}{2}(t^{2H} + s^{2H} - |t-s|^{2H}),\end{aligned} \tag{1}$$

where $D$ $(=K/2)$ is the generalized diffusion coefficient and the Hurst exponent $H \in (0,1)$. The Hurst index describes the raggedness of the resulting motion, with a higher value leading to a smoother motion. Standard Brownian motion is a particular case of the FBM corresponding to $H = 1/2$. As already mentioned, for $H < 1/2$ the increments of the process are negatively correlated, so the FBM is subdiffusive. In contrast, for $H > 1/2$ the increments of the process are positively correlated and superdiffusive behavior is observed.

We consider a single trajectory $B_t$, that is, a particular realization of a FBM process with a known $H$, and write the following weighted least-squares functional:

$$F = \frac{1}{2}\int_0^T dt\, W(t)\big(B_t^2 - K_f t^{2H}\big)^2, \tag{2}$$

where $W(t)$ is some weighting function to be determined afterward and $K_f$ is a trial parameter. We call $K_f$ an estimate of the generalized diffusion coefficient from the single trajectory $B_t$ if it minimizes $F$. Calculating the partial derivative $\partial F/\partial K_f$, setting it to zero, and solving the resulting equation for $u = K_f/K$, we find the following least-squares estimator of the generalized diffusion coefficient $K$:

$$u \equiv \frac{K_f}{K} = \frac{1}{K}\frac{\int_0^T dt\,\omega(t)B_t^2}{\int_0^T dt\, t^{2H}\omega(t)}, \tag{3}$$

where we have introduced the notation

$$\omega(t) = t^{2H}W(t). \tag{4}$$

Note that the estimator $u$ measures the ratio of the observed generalized diffusion coefficient for a single given trajectory relative to the ensemble-average value. Moreover, $\mathbb{E}\{u\} \equiv 1$ holds for any arbitrary $\omega(t)$, making it possible to compare the effectiveness of different choices of $\omega(t)$. It is worthwhile remarking that $u$ is given by a single-time integration (a local functional) and thus differs from other estimates used in the literature that involve two-time integrals (see, e.g., [15]).

Further on, from a straightforward calculation the variance of the estimator $u$ is, for arbitrary weight function $\omega(t)$,

$$\mathrm{Var}(u) = \frac{1}{K^2}\frac{\int_0^T\int_0^T dt\,ds\,\omega(t)\omega(s)\mathrm{Cov}\big(B_t^2,B_s^2\big)}{\big(\int_0^T dt\, t^{2H}\omega(t)\big)^2}, \tag{5}$$

where $\mathrm{Cov}(B_t^2,B_s^2)$ is the covariance function of a squared FBM trajectory

$$\mathrm{Cov}\big(B_t^2,B_s^2\big) = \mathbb{E}\big\{\big(B_t^2 - \mathbb{E}\{B_t^2\}\big)\big(B_s^2 - \mathbb{E}\{B_s^2\}\big)\big\}. \tag{6}$$

This function can be calculated exactly using Eq. (1) to give

$$\begin{aligned}\mathrm{Cov}\big(B_t^2,B_s^2\big) &= 2\,\mathrm{Cov}^2(B_t,B_s)\\ &= \frac{K^2}{2}(t^{2H} + s^{2H} - |t-s|^{2H})^2.\end{aligned} \tag{7}$$

Inserting the latter expression into Eq. (5) and noticing that the kernel is a symmetric function of $t$ and $s$, we have

$$\mathrm{Var}(u) = \frac{\int_0^T\int_0^t dt\,ds\,\omega(t)\omega(s)[t^{2H} + s^{2H} - (t-s)^{2H}]^2}{\big(\int_0^T dt\, t^{2H}\omega(t)\big)^2}. \tag{8}$$

Following Ref. [14], we choose

$$\omega(t) = (t_0 + t)^{-\alpha}, \tag{9}$$

where $t_0$ is a lag time and $\alpha$ a tunable exponent. In a discrete-time description, $t_0$ can be set equal to the interval between successive measurements [14]. We thus identify $t_0$ as a resolution parameter in the present continuous description. We also note that in [14], it was proven that a power-law weight function of the type in Eq. (9) was optimal among all weight functions. Fixing $t_0$ and scanning over different values of $\alpha$, we seek the value for which the variance of $u$ is smallest. Hopefully, for a such value, the variance should vanish in the limit of infinite resolution or infinite data size, i.e., when the parameter $\epsilon = t_0/T$ tends to zero. To check the latter point, we consider first the limit of an infinitely long observation time $\epsilon = 0$. For $\alpha < \gamma_H = 1 + 2H$ the integrals in Eq. (8) can be performed exactly yielding

$$\begin{aligned}\mathrm{Var}(u) = \frac{\gamma_H - \alpha}{2}\bigg(&\frac{1}{1-\alpha} + \frac{2}{\gamma_H - \alpha}\\ &+ \frac{1}{2\gamma_H - 1 - \alpha} - 2\frac{\Gamma(1-\alpha)\Gamma(\gamma_H)}{\Gamma(1+\gamma_H-\alpha)}\\ &+ \frac{\Gamma(1-\alpha)\Gamma(2\gamma_H - 1) - 2\Gamma(\gamma_H)\Gamma(\gamma_H - \alpha)}{\Gamma(2\gamma_H - \alpha)}\bigg),\end{aligned} \tag{10}$$

where $\Gamma(\cdot)$ is the Gamma function. In contrast, for $\alpha > \gamma_H = 1 + 2H$ and $\epsilon = 0$, the result in Eq. (8) can be conveniently represented as a single integral

$$\begin{aligned}\mathrm{Var}(u) = &\frac{\Gamma(2\gamma_H)\Gamma(2\alpha - 2\gamma_H)\Gamma^2(\alpha)}{\Gamma^2(\alpha - \gamma_H)\Gamma^2(\gamma_H)}\\ &\times \int_0^1 [1 + (1-x)^{2H} - x^{2H}]^2\, {}_2F_1(\alpha, 2\gamma_H, 2\alpha; x),\end{aligned} \tag{11}$$

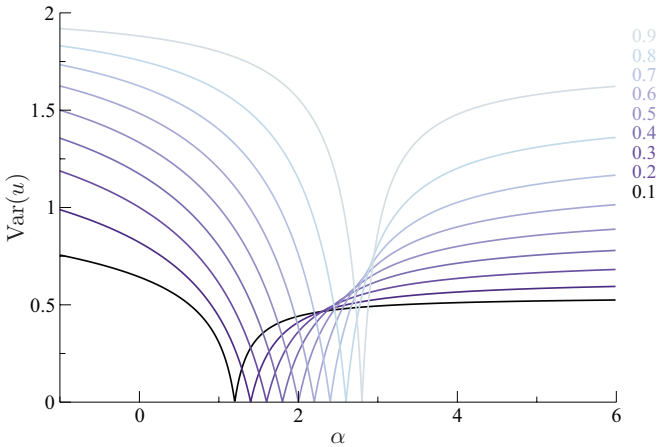


FIG. 1. (Color online) Variance in Eqs. (10) (for $\alpha < 1 + 2H$) and (11) (for $\alpha > 1 + 2H$) as a function of $\alpha$ for different values of the Hurst parameter $H$.

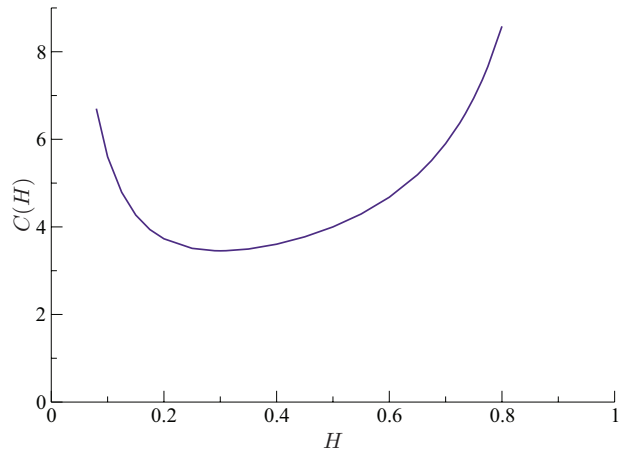


FIG. 2. (Color online) Prefactor in Eq. (12) as a function of the Hurst index.

where ${}_2F_1(\cdot)$ is the hypergeometric function. The integral in Eq. (11) can also be performed exactly by using the series representation of the hypergeometric function and then resumming the resulting series. However, the expression obtained is rather lengthy as it contains several hypergeometric functions ${}_3F_2(\cdot)$. In contrast, the result in the form of Eq. (11) can be tackled by MATHEMATICA; in addition, the asymptotic behavior can be easily extracted from it, so we prefer to work with the compact expression (11) rather than with an exact but cumbersome expression.

In Fig. 1 we show the dependence of the variance of the estimator $u$ on the exponent $\alpha$ for different values of the Hurst index $H$. We notice that for any fixed $H$, the variance vanishes as $\alpha$ approaches $\alpha = 1 + 2H$ and is nonzero for any other value. This means that for a fractional Brownian motion with Hurst index $H$ the estimators in Eq. (3) with power-law weight functions $\omega(t) = (t_0 + t)^{-\alpha}$ possess an ergodic property only when $\alpha = 1 + 2H$.

The last issue we discuss is that of the decay rate of the variance when $\epsilon$ is small but finite in the ergodic case $\alpha = 1 + 2H$. It is straightforward to show from Eq. (8) that in the limit $\epsilon \to 0$ the variance is given to leading order by

$$\mathrm{Var}(u) \sim \frac{C(H)}{\ln(1/\epsilon)}, \tag{12}$$

where $C(H)$ is a constant defined by

$$C(H) = \int_0^1 \frac{dx}{x^{1+2H}}[1 + x^{2H} - (1-x)^{2H}]^2, \tag{13}$$

which exists for any $H \in (0,1)$. This result generalizes that of Ref. [14] for ordinary Brownian motion. We conclude that the variance of the estimator vanishes logarithmically with the total observation time. In other words, the diffusion constant estimated from one trajectory by this method tends toward the correct value logarithmically slowly. The prefactor $C(H)$, which is displayed in Fig. 2, reaches a minimum at $H^* \simeq 0.30$. From Fig. 2 we notice that, keeping the resolution $\epsilon$ fixed, the variance of $u$ will typically be small for processes with $H \in [0.15, 0.6]$. This interval encompasses almost all the anomalous exponent values reported in single-particle studies. Conversely, the function $C(H)$ diverges as $H \to 0$ or 1. Therefore, we can expect that, even with the ergodic choice of $\alpha$, the estimates of the diffusion constant should become highly inaccurate for nearly localized or nearly ballistic FBM processes.

In conclusion, we have shown that the true ensemble-average generalized diffusion coefficient $K$ of a fractional Brownian motion of known Hurst index $H$ can be obtained from single-trajectory data using the weighted least-squares estimator in Eq. (3) with the weight function $\omega(t) = 1/(t_0 + t)^{1+2H}$. Such an estimator possesses an ergodic property so that $K$ can be evaluated with any necessary precision but at the expense of increasing the observation time $T$ (or decreasing $t_0$). A limitation of the present class of estimators, which are based on single-time functionals of $B_t^2$, is admittedly their slow convergence toward the ensemble average. Two-time functionals, based on the time-average MSD, for instance, exhibit faster convergence: For FBM with $H < 3/4$ the relative variance of the time-average MSD vanishes as $t_0/T$ [15]. Nevertheless, these other estimators might be more sensitive to measurement errors and may not be accurate when diffusion is no longer a pure process but a mixture of processes with different characteristic times. A quantitative comparison between estimators beyond the ideal cases considered here is a necessary future step.

G.O. acknowledges helpful discussions with M. Kleptsyna. D.S.D., C.M.-M., and G.O. were partially supported by the ESF Research Network "Exploring the Physics of Small Devices." C.M.-M. was supported by the European Research Council and the Academy of Finland.

[1] *Single Particle Tracking and Single Molecule Energy Transfer*, edited by C. Bräuchle, D. C. Lamb, and J. Michaelis (Wiley-VCH, Weinheim, 2010).

[2] M. J. Saxton and K. Jacobson, Annu. Rev. Biophys. Biomol. Struct. **26**, 373 (1997).

[3] T. G. Mason and D. A. Weitz, Phys. Rev. Lett. **74**, 1250 (1995).

[4] S. C. Weber, A. J. Spakowitz, and J. A. Theriot, Phys. Rev. Lett. **104**, 238102 (2010).
[5] G. Seisenberger *et al.*, Science **294**, 1929 (2001).
[6] I. Bronstein *et al.*, Phys. Rev. Lett. **103**, 018102 (2009).
[7] I. Golding and E. C. Cox, Phys. Rev. Lett. **96**, 098102 (2006).
[8] J.-H. Jeon *et al.*, Phys. Rev. Lett. **106**, 048103 (2011).
[9] B. Mandelbrot and J. W. van Ness, SIAM Rev. **10**, 422 (1968).
[10] M. Magdziarz, A. Weron, K. Burnecki, and J. Klafter, Phys. Rev. Lett. **103**, 180602 (2009).
[11] J. Szymanski and M. Weiss, Phys. Rev. Lett. **103**, 038102 (2009).
[12] J.-H. Jeon, H. Martinez-Seara Monne, M. Javanainen, and R. Metzler, Phys. Rev. Lett. **109**, 188103 (2012).
[13] G. Guigas and M. Weiss, Biophys. J. **94**, 90 (2008).
[14] D. Boyer, D. S. Dean, C. Mejía-Monasterio, and G. Oshanin, Phys. Rev. E **86**, 060101 (2012).
[15] W. Deng and E. Barkai, Phys. Rev. E **79**, 011112 (2009).
[16] Y. M. Wang, R. H. Austin, and E. C. Cox, Phys. Rev. Lett. **97**, 048302 (2006).
[17] M. Goulian and S. M. Simon, Biophys. J. **79**, 2188 (2000).
[18] D. S. Grebenkov, Phys. Rev. E **83**, 061117 (2011); **84**, 031124 (2011).
[19] A. Andreanov and D. S. Grebenkov, J. Stat. Mech. (2012) P07001.
[20] V. Tejedor *et al.*, Biophys. J. **98**, 1364 (2010).
[21] A. J. Berglund, Phys. Rev. E **82**, 011917 (2010).
[22] X. Michalet, Phys. Rev. E **82**, 041914 (2010); **83**, 059904(E) (2011).
[23] X. Michalet and A. J. Berglund, Phys. Rev. E **85**, 061916 (2012).
[24] D. Boyer and D. S. Dean, J. Phys. A **44**, 335003 (2011).
[25] D. Boyer, D. S. Dean, C. Mejía-Monasterio, and G. Oshanin, Phys. Rev. E **85**, 031136 (2012).