Quantifying the degree of persistence in random amoeboid motion based on the Hurst exponent of fractional Brownian motion

Natallia Makarava,[1] Stephan Menz,[2] Matthias Theves,[3] Wilhelm Huisinga,[2] Carsten Beta,[3] and Matthias Holschneider[1,2]
[1]*Interdisciplinary Center for Dynamics of Complex Systems, University of Potsdam, Karl-Liebknecht Str. 24/25, D-14476 Potsdam, Germany*
[2]*Institute of Mathematics, University of Potsdam, Karl-Liebknecht Str. 24/25, D-14476 Potsdam, Germany*
[3]*Institute of Physics and Astronomy, University of Potsdam, Karl-Liebknecht Str. 24/25, D-14476 Potsdam, Germany*
(Received 5 June 2014; published 6 October 2014)

Amoebae explore their environment in a random way, unless external cues like, e.g., nutrients, bias their motion. Even in the absence of cues, however, experimental cell tracks show some degree of persistence. In this paper, we analyzed individual cell tracks in the framework of a linear mixed effects model, where each track is modeled by a fractional Brownian motion, i.e., a Gaussian process exhibiting a long-term correlation structure superposed on a linear trend. The degree of persistence was quantified by the Hurst exponent of fractional Brownian motion. Our analysis of experimental cell tracks of the amoeba *Dictyostelium discoideum* showed a persistent movement for the majority of tracks. Employing a sliding window approach, we estimated the variations of the Hurst exponent over time, which allowed us to identify points in time, where the correlation structure was distorted ("outliers"). Coarse graining of track data via down-sampling allowed us to identify the dependence of persistence on the spatial scale. While one would expect the (mode of the) Hurst exponent to be constant on different temporal scales due to the self-similarity property of fractional Brownian motion, we observed a trend towards stronger persistence for the down-sampled cell tracks indicating stronger persistence on larger time scales.

DOI: 10.1103/PhysRevE.90.042703 PACS number(s): 87.17.Jj, 05.10.Gg, 05.40.Fb

I. INTRODUCTION

Many motile eukaryotic cells move by extending pseudopodia, i.e., temporary protrusions of the cell membrane that are driven by the actin cytoskeleton [1]. Pseudopod formation at the cell front is typically associated with cytoskeletal contractions at the back, resulting in a displacement of the cell's center of mass. This concerted action of frontal membrane protrusion, substrate adhesion, and back retraction is generally known as amoeboid motion [2,3]. In the case of many biological functions, like the healing of a wound or the spreading of metastatic cancer cells, amoeboid motility is guided by external cues [4–6]. Chemical gradients as well as thermal, electrical, or mechanical stimuli may bias the position and direction of pseudopod extension, such that cells move, on average, in the direction of the external cues [7]. A well-established and widely used model organism for studies on eukaryotic motility is the soil amoeba *Dictyostelium discoideum* [8]. For this single-celled microorganism the receptor-mediated signaling pathways as well as the force generating cytoskeletal machinery have been investigated in great detail (see, e.g., [4] and references therein).

Besides the intracellular dynamics, the analysis of cellular motion patterns has been a focus of motility research over the past years. For example, data driven Langevin-based models were introduced for the random motion of human keratinocytes and fibroblasts [9] and have later been applied to describe *Dictyostelium* motility [10–13]. Besides random locomotion in isotropic environments, this method has also been extended to chemotactic motion in the presence of a chemical gradient [14].

The analysis of experimental cell tracks showed that even in the absence of directional cues, cells most likely extend new pseudopodia in a direction similar to the prior direction of movement, a phenomenon called persistence. An example of a *Dictyostelium* trajectory can be seen in Fig. 2. Mathematically such a behavior can be modeled by a correlated random walk [15–19]. Persistence respectively antipersistence is a characteristic property of Gaussian self-similar processes. Self-similarity means that with some exponent $H \in (0,1)$, known as the Hurst exponent, and any rescaling factor $a > 0$ [20] we have

$$\{X(at), t \in T\} =_d \{a^H X(t), t \in T\}, \tag{1}$$

where $=_d$ refers to identity in distribution (i.e., all the finite dimensional distributions agree). In this paper we focus on the only Gaussian self-similar processes, which is the fractional Brownian motion. The Hurst exponent may then be used to distinguish between antipersistent ($H < 0.5$) and persistent ($H > 0.5$) behavior.

II. METHODS

A. Cell culture and experimental setup

Motility experiments were performed with the AX2 wild-type strain of the soil amoebae *Dictyostelium discoideum*. Cells were grown at 22° in an overnight shaking culture in 50-mL flasks containing HL5 medium (14 g/L peptone, 7 g/L yeast extract, 13.5 g/L glucose, 0.5 g/L KH_2PO_4, and 0.5 g/L Na_2HPO_4, ForMedium Ltd., Norfolk, UK). At the beginning of every experiment, we diluted cells from the shaking culture by 1:200 to a density of approximately 10^7 cells/mL using HL5. A 1-mL droplet from the diluted suspension was then dispersed onto the well of a culture dish (fluorodish cell culture dish 35 mm, World Precision Instruments, Sarasota, Florida, USA) and cells were given 10 min to attach to the glass bottom. The dish was then mounted on the stage of an inverted microscope (Olympus IX-71, Tokyo, Japan). To follow cell motility on the glass surface, images were recorded every 2 s for a duration of 180 min using an EoSens MC 1362/63 B/W camera (Mikrotron, Munich, Germany) and an

Olympus UPFLN60XOI objective with settings for differential interference contrast (DIC).

To extract the centroid position of each cell from the image sequences, a customized segmentation algorithm was written in MATLAB 8.0 R2012b (Mathworks, Natick, Massachusetts, USA). First, a standard Sobel edge detection together with an Isodata thresholding was applied on the original images to detect the cell boundaries [21,22]. We then dilated the resulting image with a 3×3 mask and used an internal MATLAB function to fill regions inside the cell [23]. In a last step, the cell shape in the image was corrected by four subsequent convolutions with a 3×3 mask. After a second thresholding based on the Isodata data algorithm (see above), the binary images accurately captured the morphology of the cells, and we could determine the centroid position of each cell from the geometric center of the corresponding pixel ensemble.

The cell positions in each frame were then linked together to form trajectories in time using the particle tracking algorithm of Crocker and Grier [24]. Similar to a next neighbor algorithm, the tracking process consisted of calculating and minimizing the sum over the squared displacements of all possible links between the cell positions in two subsequent frames. Note that during an experimental recording, cells may enter or leave the field of view. Also, when two cells come into contact with each other, the image segmentation program can no longer separate them. In this case, one of the two tracks will end and a new one will start, once the cells separate again. Finally, tracks may end when the segmentation program loses a cell due to image quality problems. Once the cell is detected again, a new track will start. Taken together, these different scenarios result in a distribution of tracks of different length with most of them shorter than the total measurement time.

B. Fractional Brownian motion

Fractional Brownian motion (fBm) is a self-similar, continuous, zero-mean Gaussian process which starts at zero. Its covariance at time points t and u is [25,26]

$$\mathrm{Cov}\big(B_t^H, B_u^H\big) = \tfrac{1}{2}(|t|^{2H} + |u|^{2H} - |t-u|^{2H}). \qquad (2)$$

The parameter $H \in (0,1)$ is the Hurst exponent. Fractional Brownian motion $\{B_t^H, t \in \mathbb{R}\}$ with Hurst exponent H is the only self-similar process with self-similarity exponent H in the class of Gaussian processes [20]. Recall that Gaussian process means that all finite-dimensional distributions are multivariate Gaussian distributions. Larger values of the Hurst exponent imply a smoother, less volatile, and less rough behavior. For $0.5 < H < 1$ the behavior of the process is persistent, whereas for $0 < H < 0.5$, it is antipersistent. Antipersistence means that the increments of the process are negatively correlated with correlations quickly decaying to zero, and the process becomes a so-called short-memory process:

$$\sum_{n=1}^{\infty} \mathbb{E}\big[\big|\big(B_1^H - B_0^H\big)\big(B_{n+1}^H - B_n^H\big)\big|\big] < \infty. \qquad (3)$$

In contrast to this, persistent behavior is characterized by positive correlation of two consecutive increments and a slow decay of the correlation function [20,27]: The dependence between $B_1^H - B_0^H$ and $B_{n+1}^H - B_n^H$ decays slowly as n tends to infinity and

$$\sum_{n=1}^{\infty} \mathbb{E}\big[\big(B_1^H - B_0^H\big)\big(B_{n+1}^H - B_n^H\big)\big] = \infty. \qquad (4)$$

For the time-series analysis, we consider the following model for each of the coordinates of the trajectory,

$$\mathcal{Y}_t = F_t\beta + \lambda B_t^H = \beta_1 + \beta_2 t + \lambda B_t^H, \qquad (5)$$

with a 1×2 system matrix $F_t = [1,t]$ at time t, linear trend coefficients $\beta = [\beta_1,\beta_2]^T \in \mathbb{R}^2$ with offset β_1 and slope β_2, amplitude $\lambda \in \mathbb{R}_+$, and a fBm with Hurst exponent H. The linear trend coefficients allow one to account for an overall trend in the experimental track data. In the absence of any cues, we would expect $\beta_2 = 0$, i.e., there is no preferred direction. Note that this kind of model has the structure of a linear mixed effect model, where the fixed effects that are common to all time points are the offset and slope, whereas the random effects are correlated fluctuations around this common behavior.

This linear mixed effect structure becomes more transparent through the following matrix-vector notation. Observing the processes at times $t_1, \ldots, t_N$ yields the vector-valued version,

$$\mathcal{Y} = F\beta + \lambda u, \quad u \sim N(0, \mathbf{\Sigma}_H), \qquad (6)$$

with $\mathcal{Y} = [\mathcal{Y}_{t_1}, \ldots, \mathcal{Y}_{t_N}]^T$ and a $N \times 2$ system matrix F whose kth row is $F_{t_k} = [1,t_k]$. The covariance matrix $\mathbf{\Sigma}_H = (\mathbf{\Sigma}_{H;i,j})$ is obtained from Eq. (2) as

$$\mathbf{\Sigma}_{H;i,j} = \tfrac{1}{2}(|t_i|^{2H} + |t_j|^{2H} - |t_i - t_j|^{2H}). \qquad (7)$$

We consider prior distributions for β, for λ, and for H and apply the Bayesian formula to invert for these parameters from observed data. See [28] for general background on mixed effects models and [29,30] for details on the Bayesian estimation of H from the observation of $\mathcal{Y}$. As in this reference, for the analysis of the experimental *Dictyostelium discoideum* cell track data, we took a scaling invariant Jeffreys-like prior for λ [31] with density $\pi(\lambda) \sim \lambda^{-1}$, a flat, (or diffuse) prior for β with density $\pi(\beta) \simeq 1$, as well as a flat prior for H with density $\pi(H) = \chi_{[0,1]}$ [32]. This results in the posterior distribution of H given $\mathcal{Y}$ with density,

$$\mathbb{P}(H|\mathcal{Y}) = C|\mathbf{\Sigma}_H|^{-1/2}\big|F^{\mathrm{T}}\mathbf{\Sigma}_H^{-1}F\big|^{-1/2}R^{2-N}(H), \qquad (8)$$

with some normalization coefficient C and expressions,

$$R^2(H) = (\mathcal{Y} - F\beta^*(H))^{\mathrm{T}}\mathbf{\Sigma}_H^{-1}(y\mathcal{Y} - F\beta^*(H)), \qquad (9)$$

$$\beta^*(H) = \big(F^{\mathrm{T}}\mathbf{\Sigma}_H^{-1}F\big)^{-1}F^{\mathrm{T}}\mathbf{\Sigma}_H^{-1}\mathcal{Y}. \qquad (10)$$

Note that Σ_H is positive definite so that the square root of $R^2(H)$ is a well-defined positive real number. The quantity $\beta^*(H)$ is identically to the best linear unbiased predictor (BLUP) for the fixed effects β at given H and $R(H)$ is the associated misfit of the slope measured in terms of the covariance Σ_H. As a point estimator of H we chose the mode of the posterior distribution density,

$$\hat{H} = \mathrm{argmax}_H\,\mathbb{P}(H|\mathcal{Y}).$$

Moreover, the profiled (at $\hat{H}$) posterior distribution of the offset and slope parameters β are Gaussian with mode $\beta^*(\hat{H})$ and

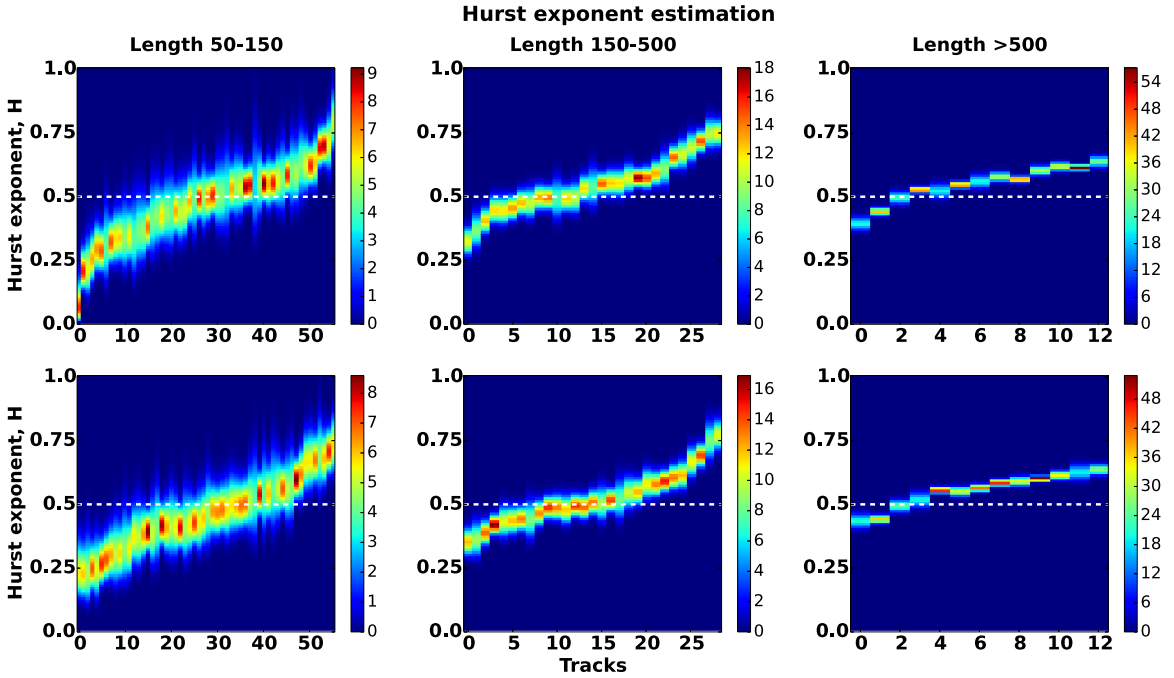

FIG. 1. (Color online) Estimated Hurst exponents along the X coordinate *(top)* and along the Y coordinate *(bottom)*. The white dashed line corresponds to a Hurst exponent of $H = 0.5$, separating the persistent area with $H > 0.5$ from the antipersistent area with $H < 0.5$. The posterior density of H is color coded. Dark colors for low density values to increasingly brighter colors for higher values. The tracks are sorted with increasing most probable Hurst exponent $\hat{H}$. Therefore it is clearly visible that for larger tracks most of the Hurst exponents are in the persistent region $H > 0$.

covariance matrix $F^{\mathrm{T}}\boldsymbol{\Sigma}_{\hat{H}}^{-1}F$. The marginalized and profiled (at $\hat{H}$) posterior density,

$$p(\lambda) = C\lambda^{1-N}\exp\left(-\frac{R^2(\hat{H})}{2\lambda^2}\right), \tag{11}$$

of the amplitude λ is obtained by integration over β.

III. RESULTS

We analyzed an ensemble of time-lapse microscopy data of *Dictyostelium* movement in the absence of external cues

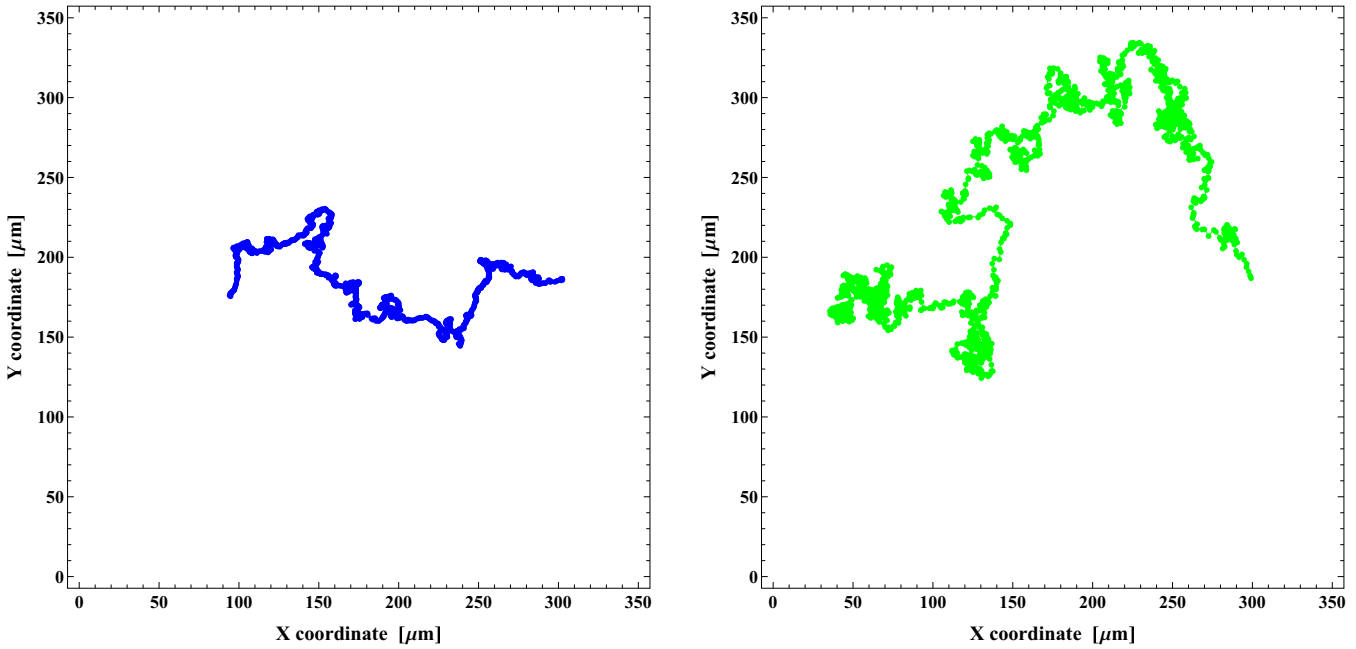

FIG. 2. (Color online) (left) Illustrative original track (No. 683) and (right) reconstructed track data of the movement of a *Dictyostelium* cell in the absence of external cues.

TABLE I. The point estimation of the Hurst exponent H, trend coefficients (offset, slope) β, and amplitude λ, estimated separately based on the X and Y coordinates of a single track (No. 683, shown in Fig. 2).

	X coordinate	Y coordinate
$\hat{H}$	0.565	0.605
$\hat{\beta}$	$[301.28, -0.12]^{\mathrm{T}}$	$[186.17, -0.039]^{\mathrm{T}}$
$\hat{\lambda}$	1.45	1.65

by estimating the Hurst exponents (separately for the X and Y coordinates) as well as the offset β_1, slope β_2, and the amplitude λ. The posterior densities of the Hurst exponents for all tracks with a length larger than 50 data points along X and Y coordinates are shown in Fig. 1. For each track, the posterior density distribution of H is color coded along the vertical axis. The tracks are ordered by increasing $\hat{H}$. The width of the posterior density clearly depends on the length of the signal, with more concentrated distributions resulting from longer tracks. Thus, for the tracks with length more than 500 data points, the estimator gave the sharpest results (compare different panels in Fig. 1). Moreover, we see that with increasing track length, the Hurst exponents displayed most often a persistent behavior ($H > 0.5$).

In Fig. 2 (left), the coordinates of a single illustrative track (No. 683) are depicted. The point estimates of the offset, slope, and amplitude for this particular track are given in Table I. As expected, the slope β_2 is close to zero for both the X and the Y coordinates. The Bayesian estimation of the Hurst exponent is depicted in Fig. 4 (left). To investigate whether the fBm model is capable of generating *in silico* tracks that resemble features of the experimental tracks, we simulated the fBm process based on the estimated parameters. A realization of the simulated (reconstructed) track data jointly with the corresponding experimental track is shown in Fig. 2. The estimated parameter values on which this simulation is based

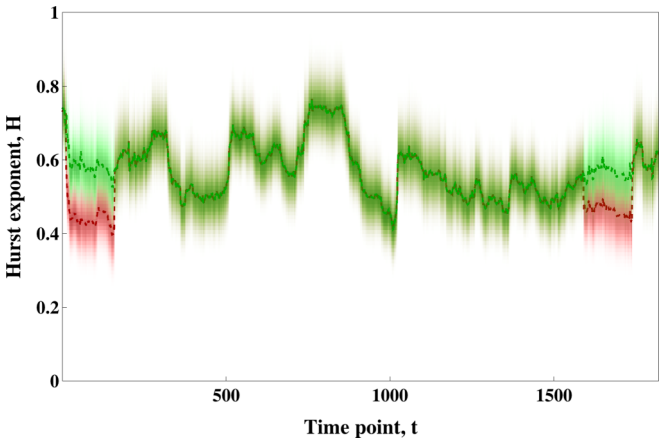

FIG. 3. (Color online) Projection of Bayesian estimation of the Hurst exponent for the X coordinate of some *Dictyostelium* cell as a function of time for some illustrative track (No. 683). In red color we depicted posterior distributions before deletion of the outliers, whereas in green they are shown after deletion. Shift to persistent area is clearly observed.

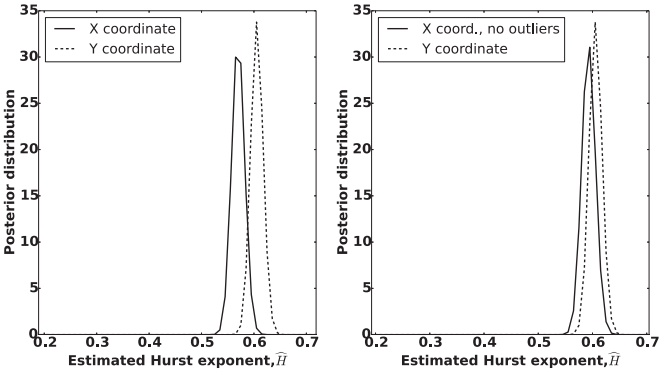

FIG. 4. Posterior density of X (solid line) and Y coordinate (dashed line) of an illustrative track (No. 683); based on original track data (left) and after outliers were detected and deleted (right).

are given in Table I. A first visual inspection showed overall similarities—a more detailed analysis requiring an in-depth characterization of the features of the experimental tracks was beyond the scope of the present paper.

The length of the chosen track (No. 683 with 1969 time points) allowed us also to analyze a potential time dependence of the Hurst exponent. We used a sliding window in which we calculated the posterior distribution of the Hurst exponent. The width of the window was 150 data points and was chosen based on the validation test [29]. The corresponding time evolution of the posterior probability for the Hurst exponent of the X coordinate time series is shown in Fig. 3. As can be seen in Fig. 3, the posterior densities have some significant jumps from one point in time to the next. This occurred whenever the correlation structure of the analyzed signal gets destroyed by outliers.

To identify possible outliers, i.e., points where the correlation structure is not compatible with the correlations of a fBm, we consider the differences between neighboring maximal posterior values of the Hurst exponent in the shifting window estimates. A difference exceeding 0.1 was used as the definition of an outlier. With this approach, we detected three outliers in the illustrative track. For comparison, we deleted the outliers from the track and re-estimated the individual parameters. The difference before and after outlier deletion is illustrated in Fig. 3. Outlier deletion clearly shifted the posterior distribution of the Hurst exponent to the more persistent regime (see Fig. 4 right). Furthermore, it can be seen that now the posterior value of the Hurst exponent of the X coordinate time series was almost equal to the one of the Y coordinate time series of the original track. Note that as far as only outliers along the X coordinate are deleted, this will only influence the X coordinate. Whenever the track has outliers along the X and Y coordinates simultaneously, the

TABLE II. The point estimation of the Hurst exponent H before and after deleting outliers.

	X coordinate	Y coordinate	X coordinate, deleted outliers
$\hat{H}$	0.565	0.605	0.595

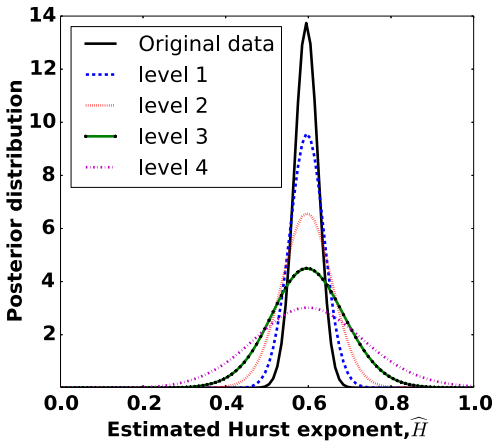

FIG. 5. (Color online) Posterior density of the Hurst exponent obtained by the Bayesian approach applied to fBm with Hurst exponent $H = 0.595$ (based on 1000 realizations).

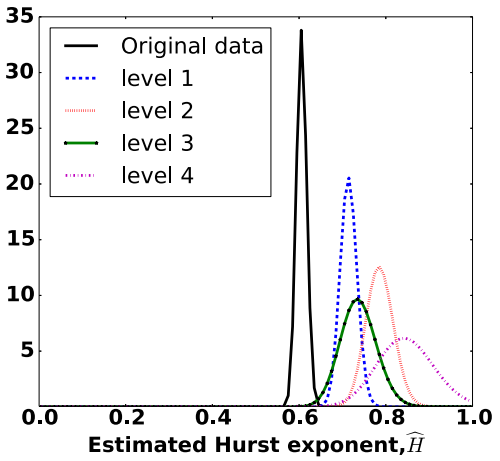

FIG. 6. (Color online) Posterior density of the Hurst exponent obtained by the Bayesian approach applied to the illustrative track (No. 683); see Fig. 2 (left).

deletion of them will also shift the posterior densities for X and Y coordinates, respectively. Here, the same analysis for the Y coordinate did not show additional outliers. Thus, the value of the estimated Hurst exponent of the Y coordinate time series was virtually unaffected by deleting the outliers (data not shown).

Next, we analyzed the scale dependency of the Hurst exponent to quantify to what extent the experimental cell track data are scaling invariant. To this end, we down-sampled the time series and analyzed the resulting coarse-grained time series. A down-sampling by 2^ℓ is then called the signal at level ℓ.

If the time series where indeed self-similar, the same Hurst exponent should be observed on all levels. This is indeed the case for synthetic data: Based on a Cholesky decomposition of $\mathbf{\Sigma}_H = L^T L$ with $H = 0.595$ (chosen as in Table II), we generated the random noise using $L^T[e_1, \ldots, e_N]^T$ with iid. standard Gaussian random variables $e_1, \ldots, e_N$. The resulting posterior distributions of the Hurst exponent corresponding to the scales $a = 2^\ell$ with levels $\ell = 1, \ldots, 4$ are depicted in Fig. 5. As can be seen, all posteriors are centered at the chosen Hurst exponent of $H = 0.595$. For the experimental *Dictyostelium* track data, however, we observed some variation in posterior distributions for different levels of down-sampling—with a shift towards more persistent behavior at the larger scales, as can be seen in Fig. 6. We therefore speculate, that random *Dictyostelium* movement does not fully resemble fBm. While the Hurst exponent of fBm provides a useful way to quantify the degree of persistence in our experimental cell tracks, the property of self-similarity is not fully preserved in our data, displaying a stronger persistence on larger time scales. However, this approach provides a useful way to readily explore the correlation structure of the cell tracks in order to identify outliers in our data. Future work will concentrate on improving our modeling approach in order to capture the properties of our data in more detail.

IV. CONCLUSIONS

In this paper, we illustrated the application of a Bayesian estimation of a self-similarity exponent (Hurst exponent) to the analysis of experimental cell tracks of the amoeba *Dictyostelium discoideum*. We used the statistical framework of a linear mixed effects model based on fractional Brownian motion. Our analysis allowed one to quantify the degree of persistence in terms of the Hurst exponent, as well as its dependence of different spatial and temporal scales. The latter enable also the identification of possible outliers in the data.

[1] T. D. Pollard, Nature (London) **422**, 741 (2003).
[2] T. Lämmermann and M. Sixt, Current Opinion in Cell Biology **21**, 636 (2009).
[3] P. Friedl and K. Wolf, J. Cell Biol. **188**, 11 (2010).
[4] K. F. Swaney, C.-H. Huang, and P. N. Devreotes, Annual Review of Biophysics **39**, 265 (2010).
[5] A. Bahat and M. Eisenbach, Molecular and Cellular Endocrinology **252**, 115 (2006).
[6] M. Zhao, Seminars in Cell & Developmental Biology **20**, 674 (2009).
[7] P. J. M. Van Haastert and P. N. Devreotes, Nat. Rev. Mol. Cell Biol. **5**, 626 (2004).
[8] R. H. Kessin, *Dictyostelium: Evolution, Cell Biology, and the Development of Multicellularity* (Cambridge University Press, Philadelphia, 2001).
[9] D. Selmeczi, S. Mosler, P. H. Hagedorn, N. B. Larsen, and H. Flyvbjerg, Biophys. J. **89**, 912 (2005).
[10] H. Takagi, M. J. Sato, T. Yanagida, and M. Ueda, PLoS One **3**, e2648 (2008).
[11] H. U. Bödeker, C. Beta, T. D. Frank, and E. Bodenschatz, Europhys. Lett. **90**, 28005 (2010).
[12] L. Li, S. F. Nørrelykke, and E. Cox, PLoS One **3**, e2093435 (2008).
[13] L. Li, E. C. Cox, and H. Flyvbjerg, Phys. Biol. **8**, 046006 (2011).

[14] G. Amselem, M. Theves, A. Bae, E. Bodenschatz, and C. Beta, PLoS One **7**, e37213 (2012).
[15] C. S. Patlak, in *The Bulletin of Mathematical Biophysics*, Vol. 15 (Kluwer Academic Publishers, Dordrecht, 1953), pp. 311–338.
[16] R. L. Hall, J. Math. Biol. **4**, 327 (1977).
[17] E. A. Codling, M. J. Plank, and S. Benhamou, J. R. Soc. Interface **5**, 813 (2008).
[18] P. Dieterich, R. Klages, R. Preuss, and A. Schwab, Proc. Nat. Acad. Sci. U.S.A. **105**, 459 (2008).
[19] P. J. M. Van Haastert, PLoS Computational Biology **6**, e1000874 (2010).
[20] G. Samorodnitsky and M. S. Taqqu, *Stable Non-Gaussian Random Processes* (Chapman and Hall, New York, 1994).
[21] T. W. Ridler and S. Calvard, IEEE Transactions on Systems, Man and Cybernetics **8**, 630 (1978).
[22] R. C. Gonzales, R. E. Woods, and S. L. Eddins, *Digital Image Processing Using MATLAB*, 2nd ed. (Gatesmark Publishing, Knoxville, 2009).
[23] P. Soille, *Morphological Image Analysis: Principles and Applications* (Springer-Verlag, New York/Berlin, 1999).
[24] J. C. Crocker and D. G. Grier, J. Colloid Interface Sci. **179**, 298 (1996).
[25] A. N. Kolmogorov, Doklady Akad. Nauk SSSR **26**, 115 (1940).
[26] B. B. Mandelbrot and J. W. Van Ness, SIAM Rev. **10**, 422 (1968).
[27] G. Samorodnitsky, Foundations and Trends in Stochastic Systems **1**, 163 (2006).
[28] E. Demidenko, *Mixed Models: Theory and Applications* (Wiley Series in Probability and Statistics, Hoboken, 2004).
[29] N. Makarava, S. Benmehdi, and M. Holschneider, Phys. Rev. E **84**, 021109 (2011).
[30] N. Makarava and M. Holschneider, Eur. Phys. J. B **85**, 272 (2012).
[31] H. Jeffreys, Proc. R. Soc. London A **186**, 453 (1946).
[32] J. M. Van Noortwijk, H. J. Kalk, and E. H. Chbab, Heron **49**, 189 (2004).