Showcasing research from the group of Prof. Ralf Metzler, University of Potsdam, Germany

Bayesian analysis of single-particle tracking data using the nested-sampling algorithm: maximum-likelihood model selection applied to stochastic-diffusivity data

This study sets up a Bayesian framework to rank possible stochastic processes governing single-particle trajectories, including Brownian motion, models with non-Gaussian displacements, and anomalous diffusion, also with measurement noise. We compute relative model probabilities and respective model parameters. Our emphasis is on the nested-sampling protocol for the stochastic-diffusivity model. Applications to computer-generated and single-particle-tracking data are presented. Image sources: The Royal Society (Phil. Trans. 1763, 53, 370 418) and pixabay.com.

As featured in:

See Ralf Metzler et al., Phys. Chem. Chem. Phys., 2018, 20, 29018.

ROYAL SOCIETY OF CHEMISTRY

rsc.li/pccp

Registered charity number: 207890

PCCP

PAPER

Cite this: *Phys. Chem. Chem. Phys.*, 2018, **20**, 29018

Bayesian analysis of single-particle tracking data using the nested-sampling algorithm: maximum-likelihood model selection applied to stochastic-diffusivity data

Samudrajit Thapa,[a] Michael A. Lomholt,[b] Jens Krog,[b] Andrey G. Cherstvy[a] and Ralf Metzler *[b]

Received 26th June 2018,
Accepted 10th September 2018

DOI: 10.1039/c8cp04043e

rsc.li/pccp

We employ Bayesian statistics using the nested-sampling algorithm to compare and rank multiple models of ergodic diffusion (including anomalous diffusion) as well as to assess their optimal parameters for *in silico*-generated and real time-series. We focus on the recently-introduced model of Brownian motion with "diffusing diffusivity"—giving rise to widely-observed non-Gaussian displacement statistics—and its comparison to Brownian and fractional Brownian motion, also for the time-series with some measurement noise. We conduct this model-assessment analysis using Bayesian statistics and the nested-sampling algorithm on the level of individual particle trajectories. We evaluate relative model probabilities and compute best-parameter sets for each diffusion model, comparing the estimated parameters to the true ones. We test the performance of the nested-sampling algorithm and its predictive power both for computer-generated (idealised) trajectories as well as for real single-particle-tracking trajectories. Our approach delivers new important insight into the objective selection of the most suitable stochastic model for a given time-series. We also present first model-ranking results in application to experimental data of tracer diffusion in polymer-based hydrogels.

I. Introduction

Following the single-particle tracking (SPT) studies of Perrin,[1] the seminal work of Nordlund[2]—using moving film plates to record long-time individual mercury droplet trajectories—paved the way for standardised SPT setups. The latter are now routinely used to track fluorescently-labelled molecules and submicron tracers.[3–9] Over the last decades, significant insight was achieved on the behaviour of both thermally-driven and active tracers for a broad class of biological, soft-matter, and other complex systems.[9–14] Many of these systems exhibit standard Fickian diffusion or Brownian motion (BM):[15–19] the mean squared displacement (MSD) of a tracer grows linearly with time,

$$\langle \mathbf{r}^2(t)\rangle = \int_{-\infty}^{\infty} \mathbf{r}^2 P(\mathbf{r},t)\mathrm{d}\mathbf{r} = 2dD_1 t. \quad (1)$$

Here D_1 is the diffusion coefficient, d is the spatial dimension, and the angular brackets denote ensemble averaging. The probability density function $P(\mathbf{r},t)$ describes the diffusive spreading of particles starting at $\mathbf{r}(0) = 0$ with time t as the Gaussian propagator

$$P(\mathbf{r},t) = \frac{\exp\left(-\frac{\mathbf{r}^2}{4D_1 t}\right)}{(4\pi D_1 t)^{d/2}}. \quad (2)$$

For anomalous diffusion processes the MSD acquires the power-law scaling,[9,13,20–27]

$$\langle \mathbf{r}^2(t)\rangle = 2dD_\alpha t^\alpha, \quad (3)$$

where the anomalous diffusion exponent distinguishes sub-diffusion ($0 < \alpha < 1$) or superdiffusion ($1 < \alpha < 2$). The special cases include Brownian motion for $\alpha = 1$ and ballistic motion for $\alpha = 2$. Here, D_α denotes the generalised diffusion coefficient.[23,25]

Various physical reasons may cause the emergence of anomalous diffusion (both transient and persistent),[23,25] see Fig. 1. This fact is reflected in a multitude of mathematical models employed to describe the experimental MSDs of the form (3).[9,13,23–25] The list of relevant anomalous-diffusion models includes fractional Brownian motion (FBM)[28–30] and fractional Langevin equation motion (both based on long-ranged increment correlations),[23,31–35] continuous-time random walks with

[a] *Institute for Physics & Astronomy, University of Potsdam, 14476 Potsdam-Golm, Germany*
[b] *MEMPHYS, Department of Physics, Chemistry & Pharmacy, University of Southern Denmark, 5230 Odense M, Denmark. E-mail: rmetzler@uni-potsdam.de*

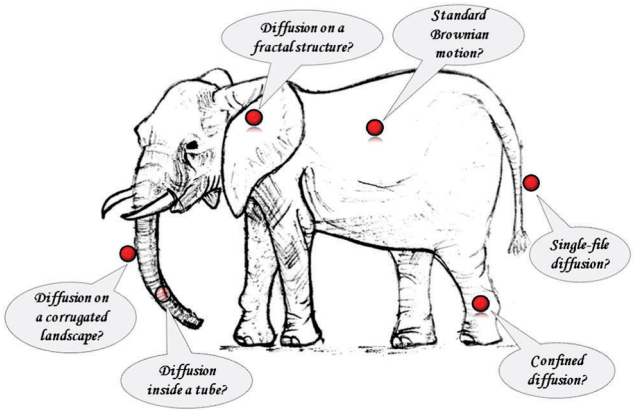

Fig. 1 Pictorial representation of the relevance of various diffusion models. Different body-parts of the elephant reflect distinct physical environments for tracer diffusion. Image courtesy Prerna Shaha.

scale-free waiting time distributions,[23,27,36–38] random walks on fractal structures,[20,39] heterogeneous diffusion processes with space-dependent diffusion coefficients,[40–46] and scaled BM with time-varying diffusivities $D_1(t) \simeq t^{\alpha-1}$ (see ref. 25 and 47–52).

A surging amount of experimental and *in silico* studies have been reported in which the MSD follows the law (1) of BM, while the associated displacement distribution $P(x,t)$ shows strong deviations from the Gaussian (2). Specifically, the Granick group demonstrated that colloids diffusing along phospholipid tubes display a displacement distribution of exponential shape[53] (in one dimension),

$$P(x,t) = \frac{\mathrm{e}^{-|x|/\lambda(t)}}{2\lambda(t)}, \tag{4}$$

where $\lambda(t) = \sqrt{D_1 t}$ is the spreading length.[54] Similar Fickian yet non-Gaussian diffusion was observed in a range of soft-matter and biological systems.[54–62]

Fickian diffusion (1) in conjunction with a non-Gaussian displacement distribution may be obtained from superstatistics as a superposition of individual Gaussian processes with given diffusivity,

$$P(x,t) = \int_0^\infty \pi(D_1) P(x,t|D_1)\mathrm{d}D_1, \tag{5}$$

weighted by the distribution $\pi(D_1)$.[63,64] The exponential law (4) then emerges from the exponential form[54,55,65–69]

$$\pi(D_1) = \exp(-D_1/\langle D_1\rangle)/\langle D_1\rangle. \tag{6}$$

Such superstatistics may be relevant for particles diffusing in spatial patches characterised by different D_1 values (for multi-state trajectories[70] in heterogeneous media), or for particles with spread diffusivities in a homogeneous environment.[168]

In some cases, the system exhibits a non-Gaussian displacement distribution over the entire observation,[53–55,58,59,69] while other systems show a crossover to a Gaussian distribution characterised by an effective diffusivity.[53,54,61] To explain a possible crossover mathematical models based on concepts of fluctuating diffusivities[72] were recently developed.[65–68,73–75] One central idea is the "diffusing diffusivity" (DD) model introduced by Slater and Chubynsky.[65] According to it, the particle motion is governed by a Langevin equation driven by white Gaussian noise whose associated diffusivity, however, is a stochastic process itself. In other words, the diffusivity fluctuates in time. Indeed, it was shown by Slater and Chubynsky[65] as well as Jain and Sebastian[66,74] that for confined diffusivity dynamics the system is initially characterised by an exponential displacement distribution, which eventually shows a crossover to a Gaussian. Chechkin and coworkers presented a minimal DD model[67] based on the mathematical concept of subordination.[76] This formulation allows for an explicit analytical solution helping one to study the non-Gaussian to Gaussian crossover. The latter is associated with the built-in time scale of the confined diffusivity dynamics following a squared Ornstein–Uhlenbeck process.[67] This DD model was recently explored for generalised Gamma distribution $\pi(D_1)$ as well as for non-equilibrium initial conditions of the diffusivity distribution.[68,169]

Here we ask the question how we can faithfully analyse experimental time-series and extract the stochastic mechanism underlying the measured dynamics. Our method is based on Bayesian statistics and nested-sampling algorithm. We specifically develop this method to include ergodic stochastic processes of the BM, FBM, and DD families. Based on synthesised data we demonstrate that this objective analysis indeed provides a reliable determination of the underlying stochastic process. Specifically, Section II outlines the basic concepts of Bayesian inference and nested sampling. The implementation of the different stochastic models is discussed in Section III. Our main results are presented in Section IV, first for a number of *in silico*-generated trajectories, followed by the analysis of real experimental time-series.[60] A general discussion is given in Section V. In the Appendices we collect the mathematical and conceptual details of the model-ranking analysis, to support the statements in the main text.

II. General concepts of Bayesian inference and nested-sampling

Here, we introduce the basic ideas of the objective analysis method combining Bayesian inference with nested sampling.

A number of mathematical concepts—such as the time-averaged MSD,[27,83,84] mean maximal-excursion method,[85] detrended fluctuation analysis,[86] weighted least-squares method,[87] *p*-variation test,[88,89] *etc.*—were employed to analyse various features of time-series recorded in SPT experiments, see also the recent reviews.[9,23,25,26] Typically, a single quantity is insufficient to unambiguously identify the underlying mathematical model of diffusion. For instance, the time-averaged MSD—ubiquitously used in data analysis—is a linear function of lag time both for BM and continuous-time random walks with power-law waiting-time distributions.[23,25,27] To determine the mathematical nature of a given physical diffusion process complementary statistical indicators can be used to discriminate possible models.[59,90–95,170]

A. Bayesian model selection

How can one approach such complex stochastic motion, determine the physical mechanisms giving rise to the observed dynamics, and extract physical parameters? We here resolve some of these issues using the Bayesian statistical approach[101–110] that provides a systematic strategy for a *statistical* model comparison and ranking.

Bayesian inference methods are being employed ubiquitously and successfully across multiple sciences: cosmology and astrophysics,[111–114] general[115–117] and high-energy physics,[118] ecology,[119,120] evolution,[121] environmental science,[122] systems biology,[123] genetics,[124] financial risk analysis,[125] and other complex dynamical systems,[126] including SPT-data analysis.[127] Recent superstatistical and time-dependent Bayesian approaches for heterogeneous random walks should also be mentioned here.[128] The Bayesian method—invented at first by Bayes,[101] extended and refined later by Laplace[102] and, *i.e.*, by Jeffreys[103]—assigns relative probabilities based on likelihood functions evaluated for each model and involves a single time-series as its input. This fact makes this method especially useful for the SPT-data analysis, where often few but long traces are recorded.[8,9,25,129]

We focus on the model-comparison and parameter-estimation approach for the DD model[65,67,68] as an example of increased complexity due to its built-in crossover dynamics. We apply the nested-sampling method developed mainly by Skilling[105,130–132] (see ref. 115 for other sampling methods) to estimate the model evidence. This method was employed previously in cosmological analyses, see, *e.g.*, ref. 111 and 113. Our current computer code generalises and extends the Bayesian-FBM code developed recently,[133] see also ref. 141. This generalised approach enables us to compare relative probabilities of multiple diffusion models: BM, BM with a superimposed measurement noise (noisy BM), FBM, noisy FBM, and BM with DD (or simply DD, the main focus of this study). The computer code in Matlab is publicly available on GitHub.[142]

B. Details of the Bayesian analysis

The inference principle and model-comparison strategy is based on Bayes' theorem[103–106]

$$P(M_i|\mathrm{Data}) = \frac{P(\mathrm{Data}|M_i)P(M_i)}{P(\mathrm{Data})}. \quad (7)$$

In our language, $P(M_i|\mathrm{Data})$ is the probability that model M_i is realised given the data, $P(\mathrm{Data}|M_i)$ is the probability of data given model M_i, $P(M_i)$ is the prior probability of M_i, and $P(\mathrm{Data})$ is the probability of data. The rank of model M_i is given by the ratio of its evidence

$$Z_i = P(\mathrm{Data}|M_i) \quad (8)$$

to the sum of evidences for all other N_{models} possible models taken into consideration, where $i = 1...N_{\mathrm{models}}$. The respective statistical model probability—assuming one of the models is true—is

$$P(M_k) = \frac{P(M_k|\mathrm{Data})}{\sum_{i=1}^{N_{\mathrm{models}}} P(M_i|\mathrm{Data})} = \frac{Z_k}{\sum_{i=1}^{N_{\mathrm{models}}} Z_i}. \quad (9)$$

The models are set to be equally-probable before analysing the data, that is $P(M_i) = P(M_j)$. Each model M_i has a vector of unknown parameters,[111] $\boldsymbol{\theta}_i = \{\theta_1,\theta_2,\ldots,\theta_N\}_i$ and—based on general physical intuition or knowledge about the range of these parameters—some prior probability functions $\pi(\boldsymbol{\theta}_i) = P(\boldsymbol{\theta}_i|M_i)$ are being assigned to them. Here, N is the total number of model parameters. Hereafter, we use the function $\pi(\cdot)$ of respective arguments to denote prior probabilities, as in the classical literature[105] (not to be mixed with the constant π).

The evidence Z_i can be expressed *via* the likelihood function of data for a given model, $\mathcal{L}_i(\boldsymbol{\theta}_i) = P(\mathrm{Data}|\boldsymbol{\theta}_i, M_i)$, the main concept of Bayesian statistics, as

$$Z_i = \int \mathcal{L}_i(\boldsymbol{\theta}_i)\pi(\boldsymbol{\theta}_i)\mathrm{d}\boldsymbol{\theta}_i. \quad (10)$$

We are also interested in the *posterior* probability which can be expressed using relation (7) as

$$P(\boldsymbol{\theta}_i|M_i, \mathrm{Data}) = \frac{P(\mathrm{Data}|\boldsymbol{\theta}_i, M_i)P(\boldsymbol{\theta}_i|M_i)}{P(\mathrm{Data}|M_i)} = \frac{\mathcal{L}_i(\boldsymbol{\theta}_i)\pi(\boldsymbol{\theta}_i)}{Z_i}. \quad (11)$$

The computation of respective evidence values is central for model ranking and parameter estimation, as performed below. Note that expression (10) becomes progressively difficult to compute analytically as the number of model parameters increases and the dimension of $\boldsymbol{\theta}_i$ grows.

C. Nested-sampling analysis: computation protocol and uncertainties

The nested-sampling method enables us to reduce the multi-dimensional integral (10) to a one-dimensional one, reducing the computational costs. As suggested by Skilling,[105,130–132] the evidence functions (10) can be evaluated explicitly as

$$Z_i = \int_0^1 \mathcal{L}_i(X_i)\mathrm{d}X_i, \quad (12)$$

i.e., the area under the curve $\mathcal{L}_i(X_i)$. The function $\mathcal{L}_i(X_i)$ is defined as the inverse of $X_i(\lambda)$ called below the "prior mass"[130,133]—the "fraction" of the probability—in the regions of parameter space where the likelihood exceeds a threshold value λ, namely[133]

$$X_i(\lambda) = \int_{\mathcal{L}_i(\theta_i) > \lambda} \pi(\boldsymbol{\theta}_i)\mathrm{d}\boldsymbol{\theta}_i. \quad (13)$$

The function $\mathcal{L}_i(X_i)$ is related to $X_i(\lambda)$ by $\mathcal{L}_i(X_i(\lambda)) = \lambda$. By definition, $\mathcal{L}_i(X_i)$ is a monotonically decreasing function, so the integral (12) can be replaced by a sum and estimated as

$$Z_i^{j_{\max}} \approx \sum_{k=1}^{j_{\max}} \mathcal{L}_{i,k} \times w_k. \quad (14)$$

Here, $w_j = (\Delta X)_j$ are the weights stemming from splitting the integration interval over X in eqn (12) and $j_{\max}$ is the index at which the iteration procedure is terminated, see below for details.[171,172]

The procedure for estimating the evidence (14) as used here goes as follows (see also the flowchart in Fig. 2 and ref. 133):

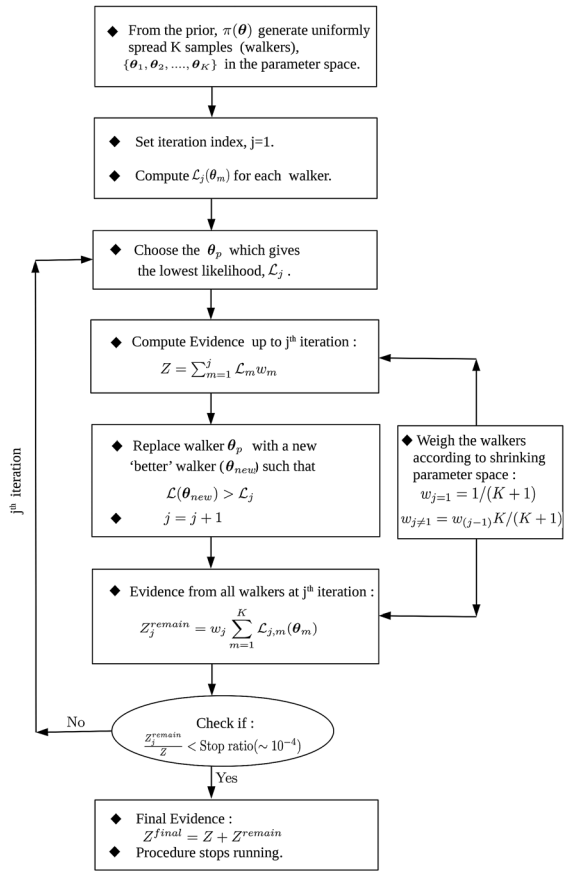

Fig. 2 Flowchart of the nested-sampling algorithm, see text for details on individual items of the procedure.

(1) For each model M_i we define an appropriate likelihood function $\mathcal{L}_i$ and prior distributions $\pi(\boldsymbol{\theta}_i)$ for each parameter $\boldsymbol{\theta}_i$ (the first lower index here indicates the model number). We generate K independent walkers $\boldsymbol{\theta}_{i,j=1,m}$ with $m = 1, 2, \ldots, K$—as points in the parameter space—randomly from $\pi(\boldsymbol{\theta}_i)$.

(2) At the first iteration step $j = 1$ we select the walker with the lowest likelihood, $\boldsymbol{\theta}_{i,\min}^{(1)}$. The corresponding likelihood $\mathcal{L}_i\left(\boldsymbol{\theta}_{i,\min}^{(1)}\right) \equiv \mathcal{L}_{i,1}$ contributes the 1st term in eqn (14), with the weight $w_1 = 1/(K+1)$.[130–132] We remove the walker $\boldsymbol{\theta}_{i,\min}^{(1)}$ from the set of K walkers (the upper index here is the iteration number).

(3) At iteration step $j > 1$ we generate a new walker by choosing one walker randomly, making a copy of it, and prescribing the copy to perform a random walk. This walker is set independent of the remaining $K-1$ walkers. The random walk follows the Metropolis–Hastings algorithm,[160,161] see also ref. 133. This walker is constrained such that its likelihood is greater than $\mathcal{L}_{i,(j-1)}$ (the second lower index indicates the iteration number).

(4) We select the walker $\boldsymbol{\theta}_{i,\min}^{(j)}$ with the lowest likelihood $\mathcal{L}_{i,j}$. Its contribution is then weighted with the factor $w_j = w_{j-1}K/(K+1)$, as K walkers are uniformly distributed from $X_i = 0$ to $X_i = X_i\left(\mathcal{L}_{i,(j-1)}\right)$[130–132] and thus the "prior mass" shrinks on average by $K/(K+1)$ on each iteration. This nested-like update procedure for the walkers gives the name for the method itself. As the prior-mass-range shrinks after each iteration, for a monotonically decreasing $\mathcal{L}_i(X_i)$ we move along the abscissa from $X_i = 1$ towards $X_i = 0$ and, consequently, from lower to higher values of $\mathcal{L}_i(X_i)$.[131,132]

(5) After the jth iteration step, the evidence for all remaining walkers is computed as

$$Z_{i,j}^{\text{remain}} = w_j \sum_{m=1}^{K} \mathcal{L}_{i,j,m}\left(\boldsymbol{\theta}_{i,j,m}\right). \tag{16}$$

The indices in the likelihood function denote the model number i, the iteration number j, and the walker number m, in order of their appearance. The quantity (16) is related to the total evidence accumulated up to the $(j-1)$th iteration, computed as the ratio

$$\text{Ratio}_{i,j} = \frac{Z_{i,j}^{\text{remain}}}{\sum_{k=1}^{j-1} \mathcal{L}_{i,k} w_k}. \tag{17}$$

(6) The iteration loop is stopped once $\text{Ratio}_{i,j} < 10^{-4}$. We define the last index as $j = j_{\max}$ and compute the final evidence (including the last contribution) for the model M_i using eqn (14) as

$$Z_i^{\text{final}} = Z_i^{(j_{\max}-1)} + Z_{i,j_{\max}}^{\text{remain}}. \tag{18}$$

For the nested-sampling procedure, following the classical studies[105,130] the uncertainty in computing $\log(Z_i)$ is estimated as the ratio $\sqrt{\mathcal{H}_i/K}$, where the information is[104,105]

$$\mathcal{H}_i = \int_0^1 \frac{\mathcal{L}_i(X_i)}{Z_i} \log\left(\frac{\mathcal{L}_i(X_i)}{Z_i}\right) \mathrm{d}X_i. \tag{19}$$

Replacing this integral by a discrete sum, following the procedure (18) this quantity is estimated as

$$\begin{aligned} \mathcal{H}_i^{j_{\max}} \approx \mathcal{H}_i^{(j_{\max}-1)} + \mathcal{H}_{i,j_{\max}}^{\text{remain}} = \sum_{k=1}^{j_{\max}-1} w_k \frac{\mathcal{L}_{i,k}}{Z_i} \log\left(\frac{\mathcal{L}_{i,k}}{Z_i}\right) \\ + w_{j_{\max}} \sum_{m=1}^{K} \frac{\mathcal{L}_{i,j_{\max},m}}{Z_i} \log\left(\frac{\mathcal{L}_{i,j_{\max},m}}{Z_i}\right). \end{aligned} \tag{20}$$

Uncertainties of computing $\log_{10} Z_i$, as provided for different models in Table 1, are then estimated as

$$\Delta(\log_{10} Z_i) = \pm\sqrt{\mathcal{H}_i/K} \times \log_{10} e \approx \pm\frac{1}{2.3}\sqrt{\mathcal{H}_i/K}. \tag{21}$$

Finally, for the estimated model parameters the moment of order $\mathcal{M}$ is evaluated as the weighted sum

$$\begin{aligned} \left\langle \boldsymbol{\theta}_i^{\mathcal{M}} \right\rangle \approx \sum_{j=1}^{(j_{\max}-1)} \left(\boldsymbol{\theta}_{i,\min}^{(j)}\right)^{\mathcal{M}} \frac{\mathcal{L}_{i,j} w_j}{Z_i} \\ + w_{j_{\max}} \sum_{m=1}^{K} \left(\boldsymbol{\theta}_{i,\min,m}^{(j_{\max})}\right)^{\mathcal{M}} \frac{\mathcal{L}_{i,j_{\max},m}}{Z_i}. \end{aligned} \tag{22}$$

These expressions enable us to compute mean values and error bars[105] for the model parameters in the nested-sampling algorithm, see below.

Table 1 Bayesian model-comparison nested-sampling results for (18) and (9) obtained for a single computer-generated BM trajectory with the parameters $D_1 = 0.5$, $\Delta t = 1$, $N_{\text{points}} = 300$. The estimated BM diffusivity is $D_1 = 0.496 \pm 0.028$, close to the real value. The FBM exponent for this tracer was found to be ≈ 0.5. The error values for $\log_{10} Z$ are nested-sampling-intrinsic errors, see eqn (21): we refer to eqn (23) for the standard model-comparison ranges to which these uncertainties can be compared

Model	$\log_{10} Z$	Model probability
BM	-370.891 ± 0.068	0.765
Noisy BM	-371.685 ± 0.076	0.123
FBM	-371.780 ± 0.080	0.099
Noisy FBM	-372.631 ± 0.089	0.014
DD	-408.593 ± 0.317	0.000

Table 2 Model-ranking results for a single FBM trajectory with $\sigma_H = 1$, $H = 0.25$, $\Delta t = 1$, $N_{\text{points}} = 300$. The estimated step deviation is $\sigma_H = 0.987 \pm 0.031$ and the Hurst exponent is $H = 0.273 \pm 0.023$

Model	$\log_{10} Z$	Model probability
BM	-366.836 ± 0.064	0.000
Noisy BM	-356.657 ± 0.074	0.002
FBM	-353.973 ± 0.082	0.782
Noisy FBM	-354.529 ± 0.087	0.217
DD	-402.563 ± 0.318	0.000

Following the classical Bayesian literature,[103,135,139] a difference in probability-realisation for two models in the Bayesian framework can be quantified in terms of their evidence values. Specifically, one model is called significantly favoured with respect to another one if for their Bayes factor $B_{i,j} = Z_i/Z_j$ one has

$$1/2 < \log_{10}[Z_i/Z_j] < 1. \tag{23}$$

When the condition $1 < \log_{10}[Z_i/Z_j] < 2$ is satisfied, the ith model is strongly favoured over the jth model, while if comparison of log-evidence yields $\log_{10}[Z_i/Z_j] < 2$ the ith model is decisively preferred in this model-ranking analysis. We refer the reader to Tables 1 and 2 to compare the differences in values of computed $\log_{10} Z_i$ and $\Delta(\log_{10} Z_i)$ for respective diffusion models used in the Bayesian analysis. The uncertainties on $\log_{10} Z_i$ we obtain are always smaller than the ranges of $\log_{10}[Z_i]$-differences listed above and, thus, they do not affect our model-comparison statements. Following the standard procedure in the Bayesian model-selection approach, the errors are only computed for $\log_{10}[Z_i]$, and not for the probability itself. The latter is evaluated from respective evidences *via* eqn (9). The interested reader can consult Sections 3.6.1 and 9.2.3 of ref. 105 for the general error-propagation analysis and for some subtle features and uncertainties involved upon transforming $\log_{10} Z_i$ to Z_i and, finally, to P_i. We do not discuss the issue of ΔP_i in the current study.

III. Application of model-prediction algorithm to stochastic models

Here we discuss several models of ergodic diffusion used in the Bayesian approach with nested-sampling algorithm below. For each model we specify its likelihood function, relevant model parameters, and their prior distributions. The new part in this section is the implementation of the DD model consisting of the coupled stochastic eqn (35). We use one-dimensional notation ($d = 1$); in higher dimensions the results below would represent a single component.

A. Brownian motion

The long-time diffusion of a Brownian particle obeys the overdamped Langevin equation[13,19,23,25,143]

$$\mathrm{d}x(t)/\mathrm{d}t = \sqrt{2D_1} \times \xi(t) \tag{24}$$

driven by the white Gaussian noise $\xi(t)$ with zero mean and autocorrelation $\langle \xi(t_1)\xi(t_2)\rangle = \delta(t_1 - t_2)$. The probability density function of displacements from the initial position $x(0) = 0$ after some time is then the Gaussian (2). BM is ergodic such that ensemble- and time-averaged physical observables coincide at long times.[23,25,144]

As particle displacements $\Delta x_j = x_j - x_{(j-1)}$ at the jth step are independent and identically distributed random variables, for the nested-sampling method we use a Gaussian likelihood function,[141]

$$\mathcal{L}_{\text{BM}}(\boldsymbol{\theta}_{\text{BM}}) = \prod_{j=2}^{(N_{\text{points}})} \frac{\exp\left(-\dfrac{(\Delta x_j)^2}{4(D_1)_j \Delta t}\right)}{\sqrt{4\pi (D_1)_j \Delta t}}. \tag{25}$$

Particle displacements are sampled equidistantly in time in the analysis below; here j is the index along the trajectory. The only BM parameter is D_1 determining the magnitude of the MSD (1). The index j for $(D_1)_j$ is relevant for the DD model where the diffusion constant changes between time-steps. For the step-deviation parameter σ we choose Jeffrey's prior,[103,111,113,145] suitable when the scale of a parameter is large or unknown. The relation of σ on each step to the respective diffusion coefficient is

$$(D_1)_j = \sigma_j^2/[2(\Delta t)]. \tag{26}$$

For the parameter range $\theta \in [\theta_{\min}, \theta_{\max}]$ Jeffrey's prior is[104,173]

$$\pi(\theta) = \begin{cases} 1/[\theta \log(\theta_{\max}/\theta_{\min})], & \theta_{\min} < \theta < \theta_{\max} \\ 0, & \text{otherwise} \end{cases}. \tag{27}$$

Motivated by previous studies[146,147] we consider noisy BM: the noise in SPT experiments is typically due to particle-localisation errors and motion blur.[8,70,129] Given the actual particle position x_j^{act} and the experimentally observed positions x_j^{obs} at time-step j, we assume that x_j^{obs} includes a Gaussian measurement noise η_j with zero mean and variance

$$\langle \eta_j^2 \rangle = \sigma_{\text{Noise}}^2. \tag{28}$$

The observed particle positions are then given by

$$x_j^{\text{obs}} = x_j^{\text{act}} + \eta_j \tag{29}$$

and the corresponding likelihood function (using eqn (26)) is[129,141]

$$\mathcal{L}_{\mathrm{NoisyBM}}\left(\boldsymbol{\theta}_{\mathrm{NoisyBM}}\right)=\prod_{j=2}^{N_{\mathrm{points}}}\frac{\exp\left(-\frac{\left(x_j^{\mathrm{obs}}-\tilde{x}_j\right)^2}{2\tilde{\sigma}_j^2}\right)}{\sqrt{2\pi\tilde{\sigma}_j^2}}. \tag{30}$$

Here, two sets of new variables are iteratively defined for $2 < j \leq N_{\mathrm{points}}$ as (see ref. 141 and note a different index for the initial point)

$$\tilde{x}_j = x_{j-1}-\frac{{\sigma_{\mathrm{Noise}}}^2}{\tilde{\sigma}_{j-1}}\left(x_{j-1}-\tilde{x}_{j-1}\right), \tilde{\sigma}_j^2=\sigma^2+{\sigma_{\mathrm{Noise}}}^2\left(2-\frac{{\sigma_{\mathrm{Noise}}}^2}{\tilde{\sigma}_{j-1}{}^2}\right). \tag{31}$$

For the initial step $j = 2$ in eqn (30) we set $\tilde{x}_2 = x_2$ and $\tilde{\sigma}_2{}^2 = \sigma^2 + 2{\sigma_{\mathrm{Noise}}}^2$. In addition to D_1, in the model of noisy BM we implement a uniform prior for σ_{Noise},

$$\pi(\theta)=\begin{cases}1/[\theta_{\max}-\theta_{\min}], & \theta_{\min}<\theta<\theta_{\max}\\ 0, & \text{otherwise}\end{cases}. \tag{32}$$

We refer the reader here to ref. 129 for another treatment of localisation errors in SPT time-series as well as for the maximum-likelihood estimation method aimed at assessing the values of diffusion coefficients of the tracers.

B. Fractional brownian motion

The mathematical concept of FBM was developed by Mandelbrot and van Ness,[29,30] see also a similar earlier process considered by Kolmogorov.[28] This anomalous diffusion process obeys the overdamped Langevin equation $\mathrm{d}x(t)/\mathrm{d}t = \xi_{\mathrm{fGn}}(t)$ driven by the fractional Gaussian noise $\xi_{\mathrm{fGn}}(t)$ with zero mean and long-ranged correlation,[23,25,29,30,32]

$$\langle\xi_{\mathrm{fGn}}(t_1)\xi_{\mathrm{fGn}}(t_2)\rangle = 2H(2H-1)D_H\times|t_1-t_2|^{2(H-1)}, \tag{33}$$

where $t_1 \neq t_2$. Note the change of sign in eqn (33) from negative (anti-persistent FBM) to positive (persistent FBM) when H crosses the value 1/2 from below. In eqn (33), D_H is the generalised diffusion coefficient and the Hurst index H relates to α in eqn (3) as $H = \alpha/2$. The displacement distribution for free FBM is Gaussian[25]

$$P(x,t)=\frac{\exp\left(-\frac{x^2}{4D_Ht^{2H}}\right)}{\sqrt{4\pi D_Ht^{2H}}}, \tag{34}$$

and the position autocorrelation is $\langle x(t_1)x(t_2)\rangle = D_H(t_1^{2H}+t_2^{2H}-|t_1-t_2|^{2H})$.[29,30] At $t_1 = t_2 = t$ this reduces to the MSD (3). Similar to BM, FBM describes ergodic diffusion.[25,32,34] Details on priors and likelihood functions for FBM are given in Appendix A1, see also ref. 133 and 148. The FBM process was simulated with the help of computer code from ref. 133.

C. Brownian motion with "diffusing diffusivity"

The minimal DD model is governed by the system of stochastic differential equations[67]

$$\frac{\mathrm{d}x(t)}{\mathrm{d}t}=\sqrt{2D_1(t)}\times\xi(t), \tag{35a}$$

$$D_1(t)=Y^2(t), \tag{35b}$$

$$\frac{\mathrm{d}Y(t)}{\mathrm{d}t}=-\frac{Y(t)}{\tau}+\varepsilon\times\bar{\eta}(t), \tag{35c}$$

where $x(t)$ is the particle position, $D_1(t)$ the time-dependent diffusion coefficient defined in terms of squared auxiliary variable $Y(t)$. The latter is given by the Ornstein–Uhlenbeck process, eqn (35c), with the amplitude of white Gaussian noise ε. Here τ is the relaxation time.[143,149]

The integration of eqn (35b) and (35c) provides the solution[67,143]

$$Y(t)=Y(0)\mathrm{e}^{-t/\tau}+\int_0^t\varepsilon\bar{\eta}(t')\mathrm{e}^{-(t-t')/\tau}\mathrm{d}t'. \tag{36}$$

The relaxation time τ—specifying for how long the Ornstein–Uhlenbeck process maintains its temporal correlations—is imperative for Bayesian analysis and nested-sampling algorithm, see Appendix A3. From eqn (35) for a given initial condition $Y_1 = Y(0)$ we obtain the mean and variance as

$$\langle Y(t)\rangle = Y_1\exp(-t/\tau) \tag{37}$$

and

$$\langle Y^2(t)\rangle-\langle Y(t)\rangle^2 = D_\bigstar[1-\exp(-2t/\tau)]/2, \tag{38}$$

respectively, where the characteristic diffusion coefficient is $D_\bigstar = \tau\varepsilon^2$.

To illustrate the DD process Fig. 3 compares results for the time-averaged MSD[7,25]

$$\overline{\delta^2(\Delta)}=\frac{1}{T-\Delta}\int_0^{T-\Delta}[x(t+\Delta)-x(t)]^2\mathrm{d}t \tag{39}$$

and the corresponding averages over an ensemble of N trajectories,

$$\left\langle\overline{\delta^2(\Delta)}\right\rangle = N^{-1}\sum_{i=1}^{N}\overline{\delta_i^2(\Delta)}, \tag{40}$$

for both BM and DD models. We see that, as expected, for the DD model the time-averaged MSD in the limit of short lag times shows considerably more spread. The latter is reflected in larger values of the ergodicity breaking parameter (see inset in Fig. 3) defined by[25,27]

$$\mathrm{EB}(\Delta)=\left\langle\left(\overline{\delta^2(\Delta)}\right)^2\right\rangle\Big/\left\langle\overline{\delta^2(\Delta)}\right\rangle^2-1. \tag{41}$$

The BM asymptote for this parameter is[25,32,144]

$$\mathrm{EB}_{\mathrm{BM}}(\Delta)=4\Delta/(3T). \tag{42}$$

Fig. 3 also confirms that for the DD model the EB at short and long lag times exceeds that for pure BM, as expected. As the total length T of trajectories is varied, in Fig. 9 (Appendix A4) we observe the reciprocal dependence $\mathrm{EB}(\Delta = 1, T) \simeq 1/T$, similar to the

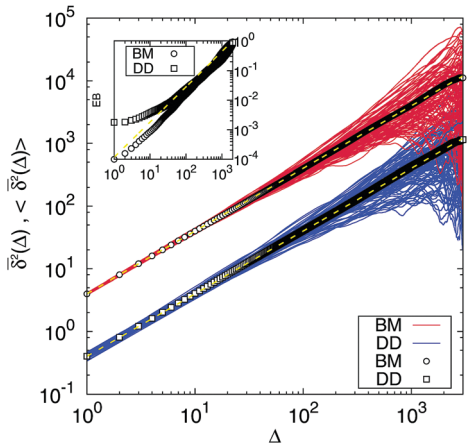

Fig. 3 Simulations results for $N = 10^2$ time-averaged MSDs (39) for BM (red curves) and DD (blue curves) models ($N_{\text{points}} \times \delta t = T = 3000$). The respective ensemble averaged asymptotes $\langle x^2(t)\rangle_{\text{BM}} = 4D_1 t$ and $\langle x^2(t)\rangle_{\text{DD}} = 2D_\star t$ are the dashed lines in the main plot. The ensemble averages are shown by the symbols. The inset illustrates the evolution of the ergodicity breaking parameter *versus* lag time, with the BM asymptote (42) being the dashed line. Parameters: $D_1 = 1$, $\Delta t = 1$ for the BM model and $\tau = 5$, $D_\star = 0.2$, $\Delta t = 1$ for the DD model.

variation for BM.[25,174] For the DD model, details of priors and likelihood functions as well as nested-sampling update procedure are given in Appendices A2 and A3.

IV. Main results

A. Computer-simulated data sets

We first focus on the results of nested-sampling runs obtained for two-dimensional surrogate trajectories for different diffusion models of Section III. We take particle motion along x and y directions as independent, allowing us to extend the likelihood functions to two spatial dimensions through the product of the one-dimensional functions. We compare the statistical probability (9) for the models of BM, noisy BM, FBM, noisy FBM, and DD and estimate the relevant model parameters. Unless specified otherwise, we use the prior ranges of Table 3. We run the nested-sampling code on the computer-generated trajectories with $N_{\text{points}} = 300$ for the BM, FBM, and DD models. This number of points corresponds to that in the experimental SPT data set considered in Section IVB, see Fig. 8.

For a single BM trajectory $x(t)$ we show in Table 1 that the Bayesian analysis using nested-sampling algorithm indeed assigns the highest evidence to BM with the probability $P(\text{BM}) \approx 0.765$. For a single FBM trajectory of the same length the method gives $P(\text{FBM}) \approx 0.782$ (see Table 2). Tables 1 and 2 present the validation of our model-prediction results for a single trajectory generated *in silico*. The total probability of the respective models (BM and FBM) in these tables is composed from the noise-free and noise-containing contributions. Summing the respective probabilities we reach a high level of model-assessment confidence, namely $\gtrsim 0.9$. Intrinsic nested-sampling-related errors in determining $\log_{10} Z$ values for a single trace, see eqn (21), are listed in Tables 1 and 2. Regarding the parameter estimation procedure, the likelihood function for the FBM model was shown to have a symmetric, bell-like variation as function of D_H and H.[141] Roughly speaking, in the Bayesian approach we consider probabilities $P \gtrsim 0.95$ as a strong evidence for a model, while the minimal significance level is often set at $P \approx 0.05$.

Finally, for the DD model we test the Bayesian analysis with the nested-sampling algorithm in more detail. The distribution of the model parameters is presented in Fig. 4 and 11. We observe an asymmetric, longer-tailed distribution of $D_\star$ and τ around the point of maximal likelihood, for all entries of the iteration procedure up to $j = j_{\max}$. This results in 10^3 selected entries in Fig. 11. A three-dimensional projection showing the approach of the likelihood function (normalised to the maximal value in the set to its maximum) is shown in Fig. 4. For this data set, we analyse $j_{\max} \approx 37\,000$ points and show 50% of them with the highest likelihood values (to improve the visibility).

For DD trajectories with a varying number N_{points} of points the nested-sampling predictions are shown in Fig. 5. We observe that the estimates of τ and $D_\star$ improve as N_{points} increases, as expected. The probability of a correct model prediction—that is, the fraction of DD traces predicted by nested-sampling runs for computer-generated DD traces from eqn (3)—also increases for longer trajectories, see Fig. 5 and 7. This probability is quantified by respective numbers in the brackets Fig. 5a and b for $N = 10^2$ trajectories analysed in total in this plot. We note that for the DD model as N_{points} increases, the computational expenses to run

Table 3 Summary of the likelihood functions, relevant parameters, and prior distributions for the parameters of diffusion models used for Bayesian statistics with nested-sampling algorithm. The range of priors are also provided, unless specified in the text otherwise

Model	Likelihood, $\mathcal{L}(\boldsymbol{\theta})$	Model parameters, $\boldsymbol{\theta}$	Priors, $\pi(\boldsymbol{\theta})$	Prior range
BM	Eqn (25)	Step deviation, σ	Eqn (27)	$\sigma \in [10^{-3}, 10^3]$
Noisy BM	Eqn (30)	Step deviation, σ	Eqn (27)	$\sigma \in [10^{-3}, 10^3]$
		Noise magnitude, σ_{Noise}	Eqn (32)	$\sigma_{\text{Noise}} \in [0, 1]$
FBM	Eqn (S2)	Hurst index, H	Eqn (32)	$H \in [0, 1]$
		Step deviation, σ_H	Eqn (27)	$\sigma_H \in [10^{-3}, 10^3]$
Noisy FBM	Eqn (S5)	Hurst index, H	Eqn (32)	$H \in [0, 1]$
		Step deviation, σ_H	Eqn (27)	$\sigma_H \in [10^{-3}, 10^3]$
		Noise magnitude, σ_{Noise}	Eqn (32)	$\sigma_{\text{Noise}} \in [0, 1]$
DD	Eqn (S9)	Relaxation time, τ	Eqn (27)	$\tau \in [\Delta t, \Delta t \times N_{\text{points}}]$
		Scaling parameter, $D_\star$	Eqn (27)	$D_\star \in [10^{-2}, 10^2]$
		Ornstein–Uhlenbeck Process, Y_1	Eqn (S13)	
		Ornstein–Uhlenbeck Process, $[Y_2, Y_3, \ldots, Y_{N_{\text{points}}-1}]$	Eqn (S16)	

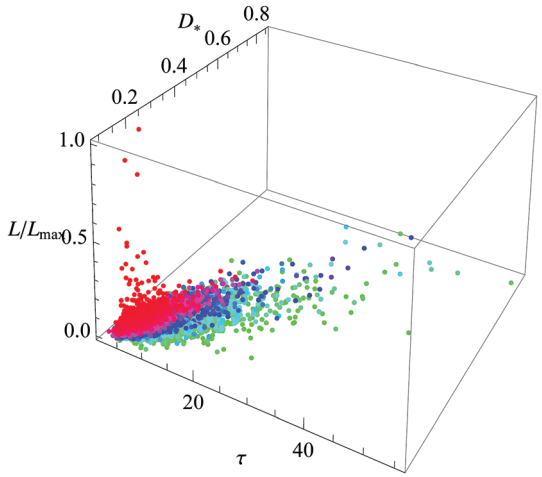

Fig. 4 Likelihood function from nested-sampling runs, presented in three dimensions *versus* the parameters of the DD model. Colours change from green to red as likelihood values increase. We show only the half of the data points with the highest likelihood values for better visibility (see text for details).

the nested-sampling algorithm to a prescribed precision grow rapidly (and take considerable time on a standard workstation). For example, for $N = 100$ DD traces consisting each of $N_{\text{points}} = 300$ points our Bayesian-nested-sampling computation to the standard accuracy takes ~ 15 hours on a standard workstation in order to rank BM, FBM, and DD diffusion models. The main reason for a long time requires to assess the DD model is the proportionality of the number of local diffusion coefficients (DD model parameters) to be estimated to the number of points in the trajectory, N_{points}.

As further support for our algorithm, Fig. 10 shows the distributions of diffusion coefficients: the actual values used for generating the DD trajectories and the nested-sampling-estimated values are shown for $N = 10^2$ traces. The agreement with the exponential distribution (S14) expected for the Ornstein–Uhlenbeck process, shown as the line in Fig. 10, is fairly good over an extended region of the distribution.

To study the effect of the sampling time-step (the respective time resolution in SPT experiments) on the model-prediction results, we generate DD trajectories with the same time-step $\Delta t = 1$, but of varying total length, see Fig. 6. To do so, a long trajectory is generated and then only 300 points are taken from it, separated by a new sampling time-step, $\Delta T > \Delta t$. One may think of this as if an experimental setup has a time resolution ΔT, while the essential particle motion actually occurs at time intervals $\sim \Delta t$. This also enables different data-sampling strategies for the available experimental time-series, see also ref. 148. Note here that the parameters ΔT and τ are not related to each other. The ratio $\Delta T/\Delta t$ defines the sampling "frequency" of a given time-series in the case when not its every point is taken for the analysis. The parameter τ is the relaxation time of the Ornstein–Uhlenbeck process, see eqn (35c).

In Fig. 6 we present the results of the model-ranking analysis for two values of τ. For trajectories generated with $\tau = 5$ (panel (a)) for $\Delta T \gtrsim 5\Delta t$ we detect progressively higher probabilities for the pure BM model, which replaces the DD model realised for the sampling time $\Delta T = \Delta t$. This fact can be explained

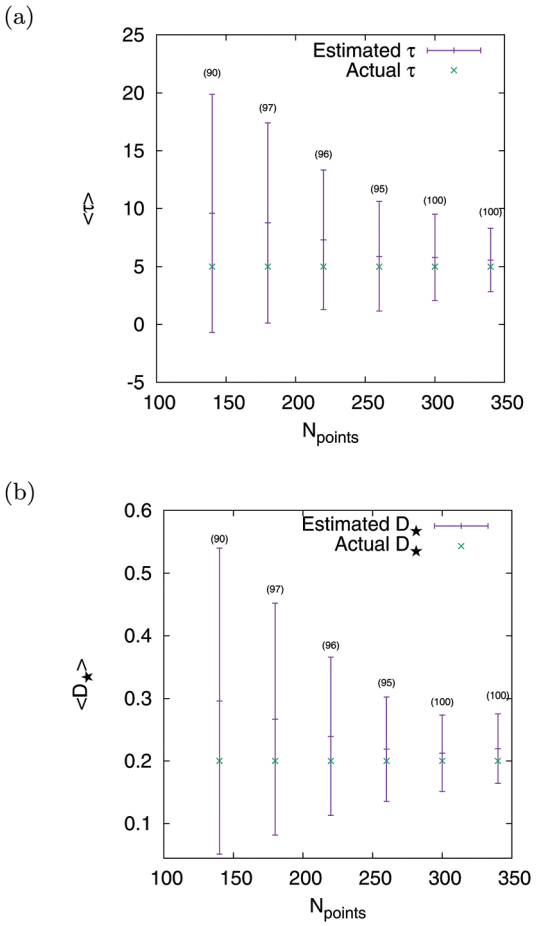

Fig. 5 Ensemble-averaged estimates of τ and $D_\star$ from nested-sampling runs on DD trajectories with varying N_{points} generated with the parameters $\tau = 5$, $D_\star = 0.2$, $\Delta t = 1$. In the brackets we list the numbers of traces (out of $N = 100$ traces in total) which give the highest DD probabilities. Only these high-probability trajectories were used to obtain the ensemble-averaged values $\langle\tau\rangle$ and $\langle D_\star\rangle$ shown in panels (a) and (b), respectively. Error bars were computed using eqn (22).

as follows. For larger intermission times between recording events of particle positions the exponentially decaying correlations of the diffusivities in the DD model cause the effective BM-Gaussian behaviour encoded by the crossover of the DD model. We observe a similar trend for DD trajectories generated with $\tau = 10$, but the onset of the crossover from the DD to BM model naturally takes place at longer sampling times, at $\Delta T \gtrsim 10\Delta t$. The sampling time-step thus has a crucial role for the data analysis, as expected, see also ref. 148. Finally, based on $N = 20$ DD trajectories in Fig. 12 we show how the BM model becomes progressively more favoured by the nested-sampling algorithm as the sampling time-step for the traces increases from $\Delta T/\Delta t = 1$ to 3, and finally to 7, see panels (a), (b), and (c) of Fig. 12, respectively.

Moreover, we examined how the total length of *in silico*-generated DD trajectories affects the confidence level of the model-prediction nested-sampling results. In Fig. 7 we present the results of the analysis of relative likelihood functions and DD-model probabilities for $N = 100$ trajectories with varying N_{points}. This figure illustrates the evolution of fractions

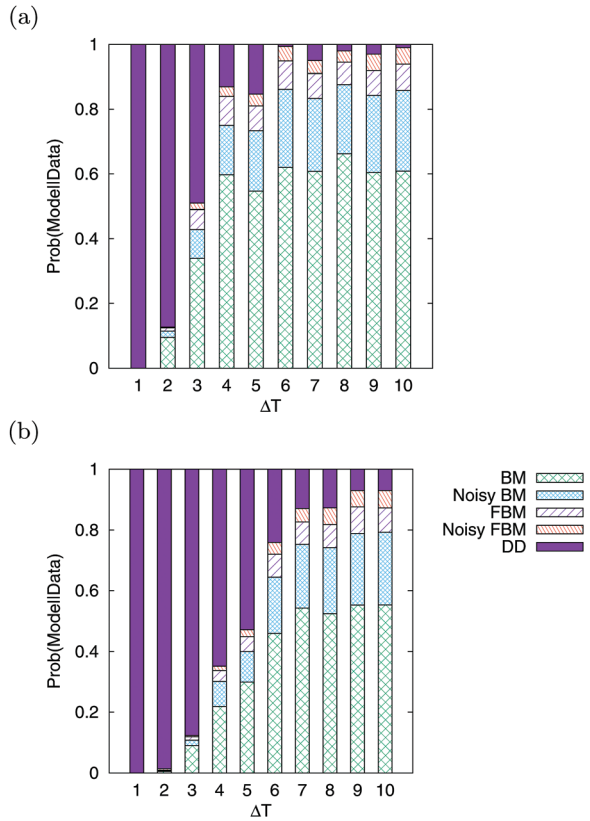

Fig. 6 (a) Bayesian model-ranking results with nested-sampling algorithm for computer-generated DD trajectories of varying length, with the parameters $\tau = 5$, $D_\star = 0.2$, $\Delta t = 1$. Averaging over $N = 10^2$ DD time-series was performed; model-prediction results for individual traces are shown in Fig. 12. The number of analysed points from each trajectory is kept fixed at $N_{\text{points}} \approx 300$ *via* choosing a suitable ΔT so that the total trace length is $T = 300 \times \Delta T$. The value of the sampling time $\Delta T/\Delta t$ is shown in the plots in terms of the elementary time-step, Δt. The value of ΔT controls the sampling "frequency" of the data sets. (b) The same as in panel (a) but for $\tau = 10$.

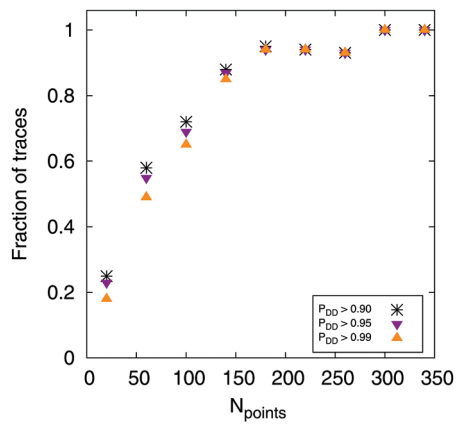

Fig. 7 Model-prediction nested-sampling results for the DD model, as computed after averaging over $N = 10^2$ *in silico*-generated trajectories of varying length. We plot the results using the level of confidence of our algorithm for detecting respective model probabilities above 0.9, 0.95, and 0.99, as indicated in the legend. The graph shows the fraction of trajectories for which this level of confidence for probability prediction is reached.

of trajectories predicted by the nested-sampling method with probabilities of the DD model in the range $P_{\text{DD}} > 90\%$, 95%, and 99%. We find that, as intuitively expected, for larger number of points in the DD time-series we detect progressively higher percentage of traces which satisfy a respective level of probability confidence.

In contrast, when the number of points reaches $N_{\text{points}} \gtrsim 300$ the Bayesian method with nested-sampling algorithm predict the DD model as strongly dominant, with respective probability values approaching unity. We refer the reader also to Fig. 13 where statistical results for probability distributions of the BM, FBM, and DD models for respective pure *in silico*-generated trajectories are presented, in panels (a), (b), and (c), correspondingly. From Fig. 13 we observe that the distributions of nested-sampling-predicted model probabilities can be skewed and non-symmetric (with respect to the maximum), in particular at the conditions when the values are not very close to unity. The latter are realised, *i.e.*, for relatively short tracer trajectories and, thus, for insufficient statistics for the nested-sampling algorithm to predict the respective models precisely.

Finally, we also propose how to analyse traces considerably longer than $\sim$300 points, a rather common situation in modern SPT experiments. One can segment a time-series into multiple sets of $N_{\text{points}} \sim 300$ points. This number of points enables the nested-sampling approach to delivers very confident model-prediction results (see Fig. 7 and 13c). Moreover, for segmented traces the nested-sampling algorithm produces results at much smaller computational costs, as compared to non-fragmented trajectories, the fact particular important for the DD model with its large parameter space. This procedure uses the fact that the DD trajectories are stationary. Conversely, for a DD process with non-equilibrium initial conditions[68] or continuous-time random walks with scale-free waiting-time distributions,[23,25,38] this type of analysis will reveal the inherently non-stationary character of the dynamics.

In addition to inherent errors of the nested-sampling method, as listed in Tables 1 and 2 for single-trace results, in Fig. 13 we present the results for the statistical uncertainties coupled with nested-sampling-determination of model probabilities for multiple trajectories. Namely, in Fig. 13 we present the distributions of nested-sampling model probabilities for the models of BM, FBM, and DD diffusion, computed for three ensembles of $N = 100$ traces of different lengths, as generated *in silico* for each model (idealised trajectories). We observe, typically, that for longer trajectories the values of model probabilities become less spread and the means of the distributions approaches unity. This means that the nested-sampling analysis delivers more confident predictions for longer trajectories of all three diffusion models considered here, as expected.

B. Experimental data sets

As an application of our model-ranking algorithm to real SPT data, we show in Fig. 8a Bayesian results with nested-sampling algorithm for data of anomalously-diffusing micron-sized tracer beads in mucin hydrogels at pH = 7[60,150] and in the absence of added salt (data courtesy C. E. Wagner, MIT). We refer the reader here to the thorough investigation of ref. 151 for the details on

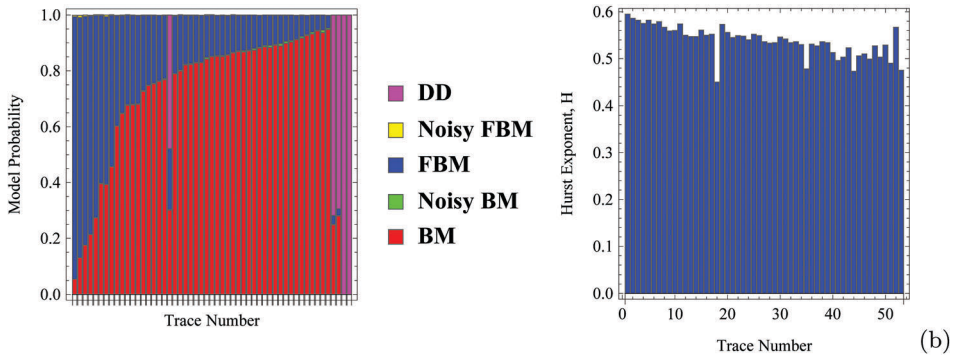

Fig. 8 Bayesian-nested-sampling model-ranking results (a) and Hurst exponent results (b), as obtained for tracer diffusion in mucin gels measured in ref. 60 (at pH = 7 and with N_{points} = 300 points). The results are plotted *versus* the respective trace number in the set. The trajectories in both panels are ordered according to a decreasing nested-sampling-predicted probability of the FBM model, see ref. 151 for further details.

Bayesian model-ranking results for different experimental conditions in mucin gels; here we present only some preliminary results. The diffusive properties of tracers in mucin gels are further described, *i.e.*, in ref. 152 and 153. We refer the reader also to ref. 150 where extensive and comparative data analyses of tracer diffusion in mucin hydrogels at different conditions are presented, based on standard statistical quantifiers.

Our analysis predicts that under pH = 7 conditions the BM model dominates the data of tracer diffusion in mucin gels, with FBM being the second most-frequently identified model. Finally, only 4 or 5 trajectories out of N = 53 in this set are identified to obey the DD model with high probability. Note that in Fig. 8a and b the traces are ordered according to a decreasing probability of the FBM model (the blue bars), from left to right. For most traces the model probabilities appear to be splitted mainly between BM and FBM. For a single trajectory, namely trace #18, however, all three models (BM, FBM, and DD) contributed significantly to the results. This makes this trajectory to look like an outlier in the presented data set, Fig. 8a.

We confirmed (results not shown) that anomalous scaling exponents and generalised diffusion coefficients predicted by the nested-sampling approach yield values close to those obtained from power-law fit of individual time-averaged MSDs.[151] We also found that when tracer trajectories are ordered according to decreasing probability of the FBM model, as in Fig. 8a, the respective Hurst exponents become progressively more normal (from left to right in Fig. 8b). In other words, for those experimental trajectories for which BM and FBM models coexist, the Hurst exponent becomes progressively closer to H = 1/2 as BM starts dominating over FBM (regarding model probability values). This indicates nearly normal diffusion, as one would expect, see ref. 151 for further analysis.

V. Discussion

Imperfect or insufficient experimental data in the presence of measurement uncertainties[8,70,105,154] challenge the development of maximum-likelihood Bayesian model-prediction and parameter-estimation approaches. This is the main focus of the current study. Often, based on standard statistical classifiers, a diffusion model cannot be identified unambiguously from SPT data. The recent, extremely fast growth of the amount of experimental SPT data sets and results from supercomputing studies as well as the expansion of available computer resources enable researchers to test a broader spectrum of possible diffusion models against a given ensemble of time-series of the particles.

We here present results of the model-ranking Bayesian analysis using the nested-sampling algorithm for BM, FBM, and DD diffusion models. BM and FBM *in silico*-generated data were considered, also in the presence of "measurement" noise.[133,146,147] The nested-sampling algorithm developed for the DD model is the main novelty of the current analysis, see Appendix A3. We successfully tested this approach on computer-generated and experimentally garnered data sets for tracer diffusion. For the latter case we concluded that—similar to results of ref. 60—individual tracer trajectories can have roughly Gaussian displacement distributions. However, depending on external conditions (for instance, different pH values, salt concentration, *etc.*[60]) the distribution of tracer displacements for the entire population can reveal pronounced non-Gaussian features, see ref. 150 and 151 for details. In general, for the FBM model with results shown in Fig. 8 each tracer particle has a specific Hurst exponent and generalised diffusion coefficient. In contrast, for an ensemble of tracers with a medium-imposed distribution of exponents and diffusivities (due to heterogeneities of local microenvironments[71,128]) non-Gaussian features, absent for diffusion of individual particles, may emerge. Diffusion is therefore Gaussian for individual particles but non-Gaussian for their population.

Other models of diffusion—such as hidden Markov models[70,111,117,155–157] with multiple diffusive states and transitions between them as well as diffusion in heterogeneous energy landscapes[158,159]—can be integrated into the current code too. This would provide a broader range of possible models to "filter" the experimental data through. In the future, also non-ergodic anomalous-diffusion processes are planned to be implemented into this Bayesian analysis using the nested-sampling algorithm. Some algorithms for microscopic force-reconstruction based on

maximum-likelihood estimator—aimed at restoring the force-fields acting on a Brownian particle from its displacements, see the recent study[167]—can be another perspective for future research.

Conflicts of interest

There are no conflicts of interest to declare.

Appendix A: details of diffusion models and computation analysis

1. Priors and likelihood functions for the FBM model

The nested-sampling method was applied recently by some of the authors to the FBM model[133] (see also ref. 148 and a complementary recent study[87]), also in the presence of a measurement noise and drift. We refer the reader also to ref. 162–164 for a Bayesian-based framework of parameter estimation for FBM. The two FBM parameters to be determined are the generalised diffusion coefficient D_H and the Hurst exponent H. Similar to eqn (26), we define the step deviation σ_H *via* $D_H = \frac{{\sigma_H}^2}{2(\Delta t)^{2H}}$.[133] Keeping in mind a wide range of possible σ_H values, Jeffrey's prior is used for this parameter. In contrast, for a limited range of the Hurst exponent, with $H \in [0,1]$, a uniform prior is used.[111]

To obtain a quadratic form structurally-similar to that of BM in eqn (30), for FBM the Bayesian procedure is as follows. For the data set of N_{points} positions at evenly-spaced time intervals (the time-step is Δt), particle displacements after n diffusion steps, $\Delta x_n = x_n - x_{n-1}$ have the auto-covariance function[133]

$$\gamma(k) = \langle \Delta x_n \Delta x_{n+k} \rangle = D_H (\Delta t)^{2H} \left[|k+1|^{2H} + |k-1|^{2H} - 2|k|^{2H} \right]. \tag{S1}$$

This function only depends on the time-step difference k due to the stationarity of increments of FBM.[25,148] From particle displacements recorded along the trajectory we create the column-vector $\Delta \boldsymbol{x}_{(N_{\text{points}}-1)}$ with its transpose $\Delta \boldsymbol{x}^T_{(N_{\text{points}}-1)}$. The likelihood function is then[133]

$$\mathcal{L}_{\text{FBM}}(\boldsymbol{\theta}_{\text{FBM}}) = \frac{\exp\left(-\frac{1}{2} \Delta \boldsymbol{x}^T_{(N_{\text{points}}-1)} {\boldsymbol{\Gamma}_{(N_{\text{points}}-1)}}^{-1} \Delta \boldsymbol{x}_{(N_{\text{points}}-1)} \right)}{(2\pi)^{N/2} \left| \boldsymbol{\Gamma}_{(N_{\text{points}}-1)} \right|^{1/2}}, \tag{S2}$$

where ${\boldsymbol{\Gamma}_{(N_{\text{points}}-1)}}^{-1}$ is the inverse of the $(N_{\text{points}}-1) \times (N_{\text{points}}-1)$ covariance matrix with the elements

$$(\boldsymbol{\Gamma}_{(N_{\text{points}}-1)})_{m,n} = \langle \Delta x_m \times \Delta x_n \rangle = \gamma(m-n) \tag{S3}$$

and determinant $|\boldsymbol{\Gamma}_{(N_{\text{points}}-1)}|$. Here the angular brackets denote averaging over fractional Gaussian noise, eqn (33). For two- and three-dimensional motion the likelihood function is the product of respective components (S2).[133]

As discussed in ref. 133 we also consider FBM with a measurement noise (noisy FBM), defined by eqn (29). In this case, the particle displacements have the auto-covariance function[133]

$$\gamma^{\text{obs}}(k) = \langle \Delta x_n^{\text{obs}} \times \Delta x_{n+k}^{\text{obs}} \rangle = \begin{cases} \gamma^{\text{act}}(0) + 2{\sigma_{\text{Noise}}}^2, & \text{for } k = 0 \\ \gamma^{\text{act}}(1) - {\sigma_{\text{Noise}}}^2, & \text{for } k = 1 \\ \gamma^{\text{act}}(0), & \text{for } 1 < k < (N_{\text{points}} - n - 1) \end{cases}, \tag{S4}$$

where $\gamma^{\text{act}}(k)$ is defined by eqn (S1). Here, the index k denotes the step number. The likelihood function for noisy FBM is then given by the quadratic Gaussian-like form[133]

$$\mathcal{L}_{\text{NoisyFBM}}(\boldsymbol{\theta}_{\text{NoisyFBM}}) = \frac{\exp\left(-\frac{1}{2} \left(\Delta \boldsymbol{x}^{\text{obs}}_{(N_{\text{points}}-1)} \right)^T \left(\boldsymbol{\Gamma}^{\text{obs}}_{(N_{\text{points}}-1)} \right)^{-1} \left(\Delta \boldsymbol{x}^{\text{obs}}_{(N_{\text{points}}-1)} \right) \right)}{(2\pi)^{N/2} \left| \boldsymbol{\Gamma}^{\text{obs}}_{(N_{\text{points}}-1)} \right|^{1/2}}, \tag{S5}$$

where, analogously to eqn (S3), we have

$$\left(\boldsymbol{\Gamma}^{\text{obs}}_{(N_{\text{points}}-1)} \right)_{m,n} = \langle \Delta x_m^{\text{obs}} \times \Delta x_n^{\text{obs}} \rangle = \gamma^{\text{obs}}(m-n). \tag{S6}$$

The noisy FBM model has three parameters: step deviation σ_H, Hurst index H, and measurement noise strength σ_{Noise}, see eqn (28). For the parameters σ_H and H we use the same prior distributions (27) and (32), whereas for σ_{Noise} a uniform prior (32) is used.

2. Priors and likelihood functions for the DD model

We now address how to estimate the likelihood function and prior distributions of the parameters for the DD model. To generate the "random walkers" from prior distributions prescribed for each of the parameters in the set $\boldsymbol{\theta}_{\text{DD}}$ we compute the cumulative distribution function,[165] CDF(θ). We construct the "inverse prior function" $\text{CDF}^{-1}(u)$ that converts a random number u drawn from a uniform distribution (on an interval[165]) into the corresponding coordinate $\boldsymbol{\theta}_{\text{DD}}$ in the parameter space of $(N_{\text{points}} + 1)$ dimensions[133] for the specific rules of this conversion. The model parameters are then

$$\boldsymbol{\theta}_{\text{DD}} = \{\tau, D_\star, (D_1)_1, (D_1)_2, \ldots, (D_1)_{(N_{\text{points}}-1)}\}, \tag{S7}$$

where $\{(D_1)_1, (D_1)_2, \ldots, (D_1)_{N_{\text{points}}-1}\}$ are instantaneous values of the diffusion coefficient computed at intermediate times

$$\left\{ \frac{t_1 + t_2}{2}, \frac{t_2 + t_3}{2}, \ldots, \frac{t_{(N_{\text{points}}-1)} + t_{N_{\text{points}}}}{2} \right\} \tag{S8}$$

for each trajectory. The number of parameters for the DD model is $N_{\text{par}} = (N_{\text{points}} + 1)$. Particle diffusivities at time instances (S8) are then given by the respective squares of the Ornstein–Uhlenbeck process, eqn (35b), namely $\{{Y_1}^2, {Y_2}^2, \ldots, {Y_{(N_{\text{points}}-1)}}^2\}$.

For a trajectory $x(t) = \{x(t_1), x(t_2), \ldots, x(t_{N\text{points}})\}$—given as solution of the set of eqn (35) or as acquired from an SPT-experiment—we then find the best-possible set of parameters $\boldsymbol{\theta}_{\text{DD}}$ using the nested-sampling algorithm, as detailed in

Appendix A3. Now, the likelihood function can be formulated similar to the BM framework (25): as the position increments Δx_j of the particle are independent random variables, given the set of diffusivities $\{(D_1)_1,(D_1)_2,\ldots,(D_1)_{N_{\text{points}}-1}\}$ they are the product of Gaussian likelihoods for each increment,

$$\begin{aligned}\mathcal{L}_{\text{DD}}&\left(\left\{\Delta x(t_1),\Delta x(t_2),\ldots\Delta x\left(t_{N_{\text{points}}-1}\right)\right\}\right.\\&\left.\left|\left\{(D_1)_1,(D_1)_2,\ldots(D_1)_{N_{\text{points}}-1}\right\}\right.\right)\\&=\prod_{j=1}^{N_{\text{points}}-1}\frac{\exp\left(-\frac{(\Delta x_j)^2}{4(D_1)_j\Delta t}\right)}{\sqrt{4\pi(D_1)_j\Delta t}}.\end{aligned}\tag{S9}$$

For the parameters τ and $D_\star$ Jeffrey's prior (27) is used. The log function in this prior reflects the fact that when the magnitude of a parameter is not known, we have to sample a broad range of possible values. We construct the inverse prior function

$$\text{CDF}_{\text{Jeff}}{}^{-1}(u_{\text{DD}})=\theta_{\text{DD}}\tag{S10}$$

of a random number u from a uniform distribution in the range (0,1) by inverting (S10) to get[133]

$$u(\theta)=\int_{\theta_{\min}}^{\theta}\pi(\theta')\mathrm{d}\theta'=\frac{1}{\log(\theta_{\max}/\theta_{\min})}\int_{\theta_{\min}}^{\theta}\frac{\mathrm{d}\theta'}{\theta'}.\tag{S11}$$

After integration, from eqn (S10) we find for the inverse prior function of τ and $D_\star$

$$\text{CDF}_{\text{Jeff}}{}^{-1}(u)=\theta_{\min}\times\exp[u\log(\theta_{\max}/\theta_{\min})].\tag{S12}$$

We need to ensure that the process $Y(t)$ is stationary starting from $t=0$. To this end, for the first step at $j=1$ we generate Y_1 values from the equilibrium distribution of the Ornstein–Uhlenbeck process,[67]

$$\pi(Y_1|D_\star)=(\pi D_\star)^{-1/2}\exp(-{Y_1}^2/D_\star).\tag{S13}$$

In the long time, stationary limit integration of (S13) shows that the Ornstein–Uhlenbeck diffusion coefficients are distributed roughly exponentially (see ref. 67 and also Fig. 10),

$$\pi((D_1)_1)=(\pi D_\star(D_1)_1)^{-1/2}\exp[-(D_1)_1/D_\star].\tag{S14}$$

For all other steps $j>1$, using the Ornstein–Uhlenbeck Gaussian displacement distribution,[67,143,149] we employ a Gaussian prior. Thus, for an arbitrary k value we define

$$\{Y_{]k[}\}\equiv\{Y_1,Y_2,\ldots,Y_{k-1},Y_{k+1},\ldots,Y_{(N_{\text{points}}-1)}\}\tag{S15}$$

defined as the set of all Y values except Y_k. For $\{Y_{]1[}\}$ the prior is chosen as (using the results of ref. 143 and 149, see also ref. 166)

$$\pi\left(\{Y_{]1[}\}|\{\tau,D_\star,Y_1\}\right)=\frac{\exp\left[-\frac{\sum_{l=2}^{(N_{\text{points}}-1)}\left(Y_l-Y_{l-1}\mathrm{e}^{-\Delta t/\tau}\right)^2}{D_\star(1-\mathrm{e}^{-2\Delta t/\tau})}\right]}{\left[\pi D_\star\{1-\mathrm{e}^{-2\Delta t/\tau}\}\right]^{\frac{(N_{\text{points}}-1)-1}{2}}}.\tag{S16}$$

This expression utilises the mean and variance (37) of the Ornstein–Uhlenbeck process.[175]

The inverse prior functions (S13) and (S16) for the first ($l=1$) and all other steps ($l=2,3,\ldots,N_{\text{points}}-1$) is obtained *via* the inversion of eqn (S11), respectively, as

$$\text{CDF}_{Y_1}{}^{-1}(u_1)=Y_1=\sqrt{D_\star}\,\text{erf}^{-1}(2u_1-1)\tag{S17}$$

and

$$\begin{aligned}&\text{CDF}_{Y_l}{}^{-1}(u_l)=Y_l\\&=Y_{l-1}\mathrm{e}^{-\Delta t/\tau}+\sqrt{D_\star(1-\mathrm{e}^{-2\Delta t/\tau})}\times\text{erf}^{-1}(2u_l-1).\end{aligned}\tag{S18}$$

Here, the error function is $\text{erf}(x)=2\pi^{-1/2}\int_0^x\exp(-t^2)\mathrm{d}t$ and its inverse is $\text{erf}^{-1}(x)$. Inverting expression (S18) we get the cumulative representation

$$u_l(Y_l)=\int_{-\infty}^{Y_l}\frac{\exp\left[-\frac{\left(Y_l'-Y_{l-1}\mathrm{e}^{-\Delta t/\tau}\right)^2}{D_\star(1-\mathrm{e}^{-2\Delta t/\tau})}\right]\mathrm{d}Y_l'}{\sqrt{\pi D_\star(1-\mathrm{e}^{-2\Delta t/\tau})}}.\tag{S19}$$

3. Nested-sampling update procedure for the DD model

In the DD model the parameters τ and $D_\star$ at each iteration step are chosen from a log-uniform Jeffrey's prior. However, for $(N_{\text{points}}-1)$ nested-sampling parameters stemming from the Ornstein–Uhlenbeck process $Y_l(t)$ we ensure that they are sampled from the correct distribution,[104] as prescribed by their priors (S13) and (S16). As discussed in Section II, after the jth nested-sampling iteration we have $(K-1)$ walkers with a set of likelihoods $\{\mathcal{L}_{\text{DD},j,m}\}$, with $m=\{1,2,\ldots,K-1\}$, which are all greater than a certain value. The update procedure of the walkers employs the Metropolis–Hastings method,[104] as outlined in Section IIC. We ensure that the new positions of newly-copied walkers at each step come from the correct distribution. The detailed-balance condition implies that displacement distributions of the random walk converge to eqn (27), (S13), and (S16).

The joint distribution for the whole set $\{Y_1,Y_2,\ldots,Y_{(N_{\text{points}}-1)}\}$ is the product of the respective distributions (S13) and (S16),

$$\pi(\{Y_j\}|\{\tau,D_\star\})=\pi(Y_1|D_\star)\times\pi(\{Y_{]1[}\}|\{\tau,D_\star,Y_1\}).\tag{S20}$$

To update the τ and $D_\star$ values, we select new values τ_{new} and $D_{\star,\text{new}}$ from the distribution (27) and accept them with the Metropolis-criterion-based acceptance ratio,[160]

$$\text{Ratio}_j=\frac{\pi\left(\{Y_j\}|\{\tau_{\text{new}},D_{\star,\text{new}}\}\right)}{\pi\left(\{Y_j\}|\{\tau_{\text{old}},D_{\star,\text{old}}\}\right)}.\tag{S21}$$

Similarly, to update a single Y_k value we rewrite the left-hand side in eqn (S20) using the "chain rule" for conditional probabilities,

$$P(\{A,B\}|C)=P(A|\{B,C\})\times P(B|C),\tag{S22}$$

where $A\equiv Y_k$, $B\equiv\{Y_{]k[}\}$, and $C=\{\tau,D_\star\}$, that gives

$$\pi\left(Y_k|\{\{Y_{]k[}\},\tau,D_\star\}\right)=\frac{\pi\left(\{Y_j\}|\{\tau,D_\star\}\right)}{\pi\left(\{Y_{]k[}\}|\{\tau,D_\star\}\right)}.\tag{S23}$$

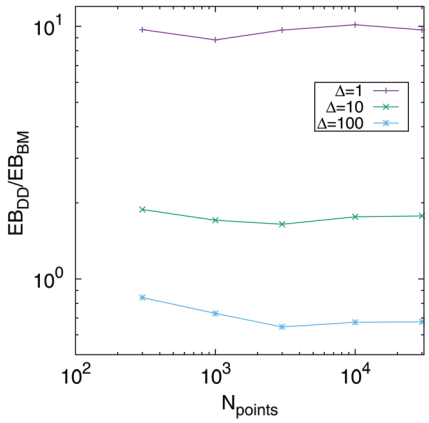

Fig. 9 Ergodicity breaking parameter (41) for the DD model (35) normalised to the respective value for BM, EB_{BM}, plotted for $\Delta/\Delta t = \{1,10,100\}$ and varying trajectory length $T = N_{points} \times \Delta t$. Both lag times Δ and trace lengths T have units of time, but are listed here as numbers *via* setting the elementary time $\Delta t = 1$. The parameters of Fig. 3 were used, with averaging performed over $N = 10^3$ traces.

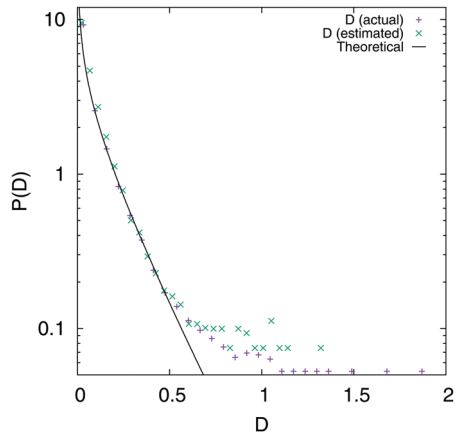

Fig. 10 Normalised distributions of estimated and actual diffusion coefficients, used to generate $N = 100$ trajectories with $N_{points} = 300$ from the DD model. Other parameters are the same as in Fig. 11. The prior (S16) along the time-series was used to compute the actual distribution. The solid line is the theoretical prediction (S14) for the chosen τ and $D_\star$.

Putting expressions (S13) and (S16) into this relation, marginalising the denominator of eqn (S23) over the Y_k value,

$$\pi(\{Y_{]k[}\}|\{\tau, D_\star\}) = \int \pi(\{Y_j\}|\{\tau, D_\star\})\,\mathrm{d}Y_k, \tag{S24}$$

we find for eqn (S23) that

$$\pi(Y_k|\{\{Y_{]k[}\}, \tau, D_\star\}) = \frac{\exp\left(-\frac{(Y_k - \Gamma_k)^2}{2\Sigma_k^2}\right)}{(2\pi\Sigma_k^2)^{1/2}}. \tag{S25}$$

Here, the additional model parameters $\{\Gamma_k, \Sigma_k^2\}$ are expressed as $\left\{Y_2 e^{-\Delta t/\tau}, \frac{D_\star}{2}(1 - e^{-2\Delta t/\tau})\right\}$, $\left\{\frac{Y_{k+1} + Y_{k-1}}{2\cosh(\Delta t/\tau)}, \frac{D_\star}{2}\tanh(\Delta t/\tau)\right\}$, and $\left\{Y_{(N_{points}-2)} e^{-\Delta t/\tau}, \frac{D_\star}{2}\tanh(\Delta t/\tau)\right\}$ for $k = \{1,\ 1 < k < (N_{points} - 1),\ (N_{points} - 1)\}$, respectively.

To follow the prior distribution (S25) we update the Y_k values according to another Ornstein–Uhlenbeck process with a small increment Δs (note that Δs has the units of the diffusion coefficient here). Then, from eqn (35c) we find

$$\frac{\mathrm{d}Y_k(s)}{\mathrm{d}s} = -\frac{Y_k(s) - \Gamma_k}{\Sigma_k^2} + \eta(s), \tag{S26}$$

with $\langle \eta(s)\eta(s')\rangle = 2\delta(s - s')$. Thus, the update rule for Y_k is

$$\begin{aligned} Y_k(s + \Delta s) = \Gamma_k + (Y_k(s) - \Gamma_k)\exp\left(-\frac{\Delta s}{\Sigma_k^2}\right) \\ + \Sigma_k\sqrt{1 - \exp\left(-\frac{2 \times \Delta s}{\Sigma_k^2}\right)}\mathcal{N}(0,1), \end{aligned} \tag{S27}$$

where is the normal distribution $\mathcal{N}(0,1)$ on with zero mean and unit variance. Increasing the lag value Δs, the values of $Y_k(s + \Delta s)$ become independent of $Y_k(s)$.

4. Additional tables and figures

Here, we present additional figures supporting findings of the main text of the manuscript.

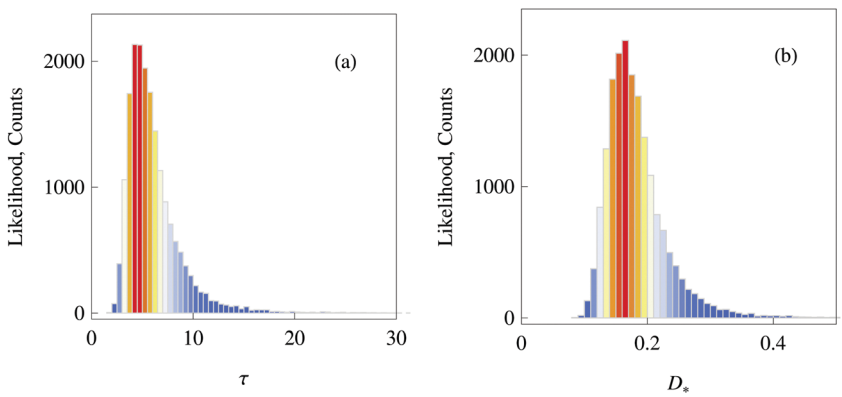

Fig. 11 Likelihood function distributions $p(\mathcal{L}_{DD})$ for a single DD trajectory from Fig. 3 in the vicinity of the maximum likelihood point, as function of $D_\star$ and τ. Parameters: $N_{points} = 300$, the actual parameters for generating the trace are $\tau = 5$, $D_\star = 0.2$, and $\Delta t = 1$.

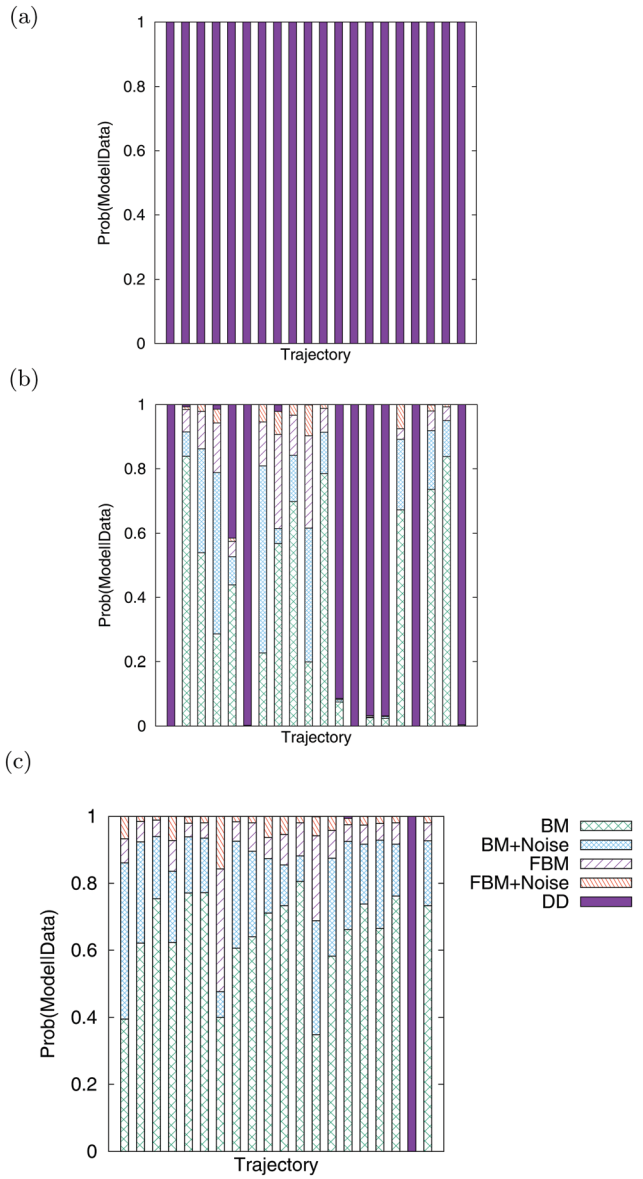

Fig. 12 Model-prediction single-trajectory nested-sampling results for N = 20 DD trajectories with τ = 5, Δt = 1, and $D_\star$ = 0.2, plotted for varying sampling time-step $\Delta T/\Delta t$ = 1, 3, and 7, in panels (a), (b), and (c), respectively. The colour and pattern scheme is the same as in Fig. 6.

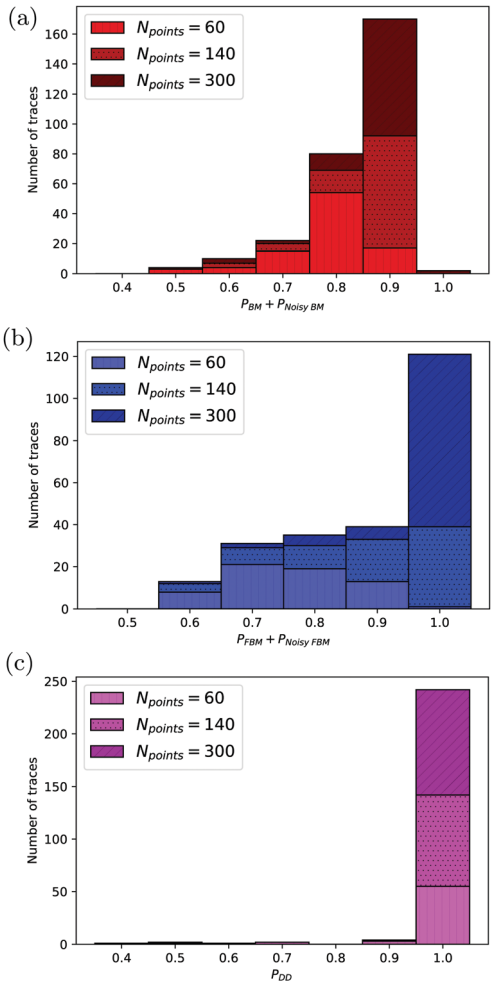

Fig. 13 Model-prediction nested-sampling results in terms of distribution of probabilities for the models of BM (panel (a)), FBM (b), and DD (c), as obtained for idealised computer-generated trajectories (ensemble of $N = 10^2$ traces is used). The colour scheme of Fig. 8 is employed here too, in order to distinguish the respective models of diffusion. In these graphs, the results for noise-free and noisy models are presented together for the models of BM and FBM, similar as model-comparison results in Tables 1 and 2. For all three diffusion models, only the traces resulting in a domination of the correct model are counted in the histograms.

Abbreviations

SPT	Single-particle tracking
MSD	Mean-squared displacement
DD	Diffusing diffusivity
BM	Brownian motion
FBM	Fractional Brownian motion

Acknowledgements

The authors thank K. Ribbeck and C. E. Wagner for scientific correspondence and providing/sharing the data set of ref. 60. We thank Prerna Shaha for the elephant drawing for Fig. 1 and to Alexey Cherstvy for further inspiration. J. K. and M. A. L. acknowledge funding from Danish Council for Independent Research Natural Sciences (FNU), Grant 4002-00428B. R. M. acknowledges financial support by the Deutsche Forschungsgemeinschaft (DFG Grants ME 1535/6-1, ME 1535/7-1). R. M. acknowledges support within a Humboldt Polish Honorary Research Scholarship (Foundation for Polish Science). S. T. acknowledges the Deutscher Akademischer Austauschdienst for his PhD Scholarship (Programme ID 57214224).

References

1 J. Perrin, Movement Brownien et réalité moleculaire (Bronwian motion and molecular reality), *Ann. Chim. Phys.*, 1909, **18**, 1.
2 I. Nordlund, A new determination of Avogadro's number from Brownian motion of small mercury spherules, *Z. Phys. Chem.*, 1914, **87**, 40.

3 E. W. Moerner and M. Orrit, Illuminating single molecules in condensed matter, *Science*, 1999, **238**, 1670.
4 J. M. Saxton and K. Jacobsen, Single-particle tracking: applications to membrane dynamics, *Annu. Rev. Biophys. Biomol. Struct.*, 1997, **26**, 373.
5 S. X. Xie, *et al.*, Single-molecule approach to molecular biology in living bacterial cells, *Annu. Rev. Biophys.*, 2008, **37**, 417.
6 C. Bräuchle, C. D. Lamb and J. Michaelis, *Single Particle Tracking and Single Molecule Energy Transfer*, Wiley-VCH Weinheim, Germany, 2012.
7 E. Barkai, Y. Garini and R. Metzler, Strange kinetics of single molecules in living cells, *Phys. Today*, 2012, **65**(8), 29.
8 C. Manzo and M. F. Garcia-Parajo, A review of progress in single particle tracking: from methods to biophysical insights, *Rep. Prog. Phys.*, 2015, **78**, 124601.
9 K. Nørregaard, R. Metzler, C. Ritter, K. Berg-Sørensen and L. Oddershede, Manipulation and motion of organelles and single molecules in living cells, *Chem. Rev.*, 2017, **117**, 4342.
10 T. Franosch, M. Grimm, M. Belushkin, F. M. Mor, G. Foffi, L. Forro and S. Jeney, Resonances arising from hydrodynamic memory in Brownian motion, *Nature*, 2011, **478**, 85.
11 T. Li, S. Kheifets, D. Medellin and M. G. Raizen, Measurement of the instantaneous velocity of a Brownian particle, *Science*, 2010, **328**, 1673.
12 R. X. Huang, I. Chavez, K. M. Taute, B. Lukic, S. Jeney, M. G. Raizen and E. L. Florin, Direct observation of the full transition from ballistic to diffusive Brownian motion in a liquid, *Nat. Phys.*, 2011, **7**, 576.
13 F. Höfling and T. Franosch, Anomalous transport in the crowded world of biological cells, *Rep. Prog. Phys.*, 2013, **76**, 046602.
14 K. Kroy and F. Cichos, Hot Brownian Motion, in *Diffusive Spreading in Nature, Technology and Society*, ed. A. Bunde, J. Caro, J. Kärger, and G. Vogl, Springer, 2018.
15 A. Fick, Über Diffusion, *Ann. Phys.*, 1855, **170**, 59.
16 A. Einstein, Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen, *Ann. Phys.*, 1905, **322**, 549.
17 W. Sutherland, A dynamical theory of diffusion for non-elecrolytes and the molecular mass of albumin, *Philos. Mag.*, 1905, **9**, 781.
18 M. von Smoluchowski, Zur kinetischen Theorie der Brownschen Molekularbewegung und der Suspensionen, *Ann. Phys.*, 1906, **21**, 756.
19 P. Langevin, Sur la théorie de movement brownien, *C. R. Hebd. Seances Acad. Sci., Ser. D*, 1908, **146**, 530.
20 J.-P. Bouchaud and A. Georges, Anomalous diffusion in disordered media: Statistical mechanisms, models and physical applications, *Phys. Rep.*, 1990, **195**, 127.
21 R. Metzler and J. Klafter, The random walk's guide to anomalous diffusion: a fractional dynamics approach, *Phys. Rep.*, 2000, **339**, 1.
22 R. Metzler and J. Klafter, The restaurant at the end of the random walk: recent developments in the description of anomalous transport by fractional dynamics, *J. Phys. A: Math. Gen.*, 2004, **37**, R161.
23 S. Burov, J.-H. Jeon, R. Metzler and E. Barkai, Single particle tracking in systems showing anomalous diffusion: the role of weak ergodicity breaking, *Phys. Chem. Chem. Phys.*, 2011, **13**, 1800.
24 I. M. Sokolov, Models of anomalous diffusion in crowded environments, *Soft Matter*, 2012, **8**, 9043.
25 R. Metzler, J.-H. Jeon, A. G. Cherstvy and E. Barkai, Anomalous diffusion models and their properties: non-stationarity, non-ergodicity, and ageing at the centenary of single particle tracking, *Phys. Chem. Chem. Phys.*, 2014, **16**, 24128.
26 Y. Meroz and I. M. Sokolov, A toolbox for determining subdiffusive mechanisms, *Phys. Rep.*, 2015, **573**, 1.
27 Y. He, S. Burov, R. Metzler and E. Barkai, Random time-scale invariant diffusion and transport coefficients, *Phys. Rev. Lett.*, 2008, **101**, 058101.
28 A. N. Kolmogorov, Wienersche Spiralen und einige andere interessante Kurven im Hilbertschen Raum, *C. R. (Dokl.) Acad. Sci. URSS*, 1940, **26**, 115.
29 B. B. Mandelbrot and W. J. van Ness, Fractional Brownian motions, fractional noises and applications, *SIAM Rev.*, 1968, **10**, 422.
30 B. B. Mandelbrot, *The Fractal Geometry of Nature*, W. H. Freeman, New York, 1982.
31 I. Goychuk, Viscoelastic subdiffusion: generalized Langevin equation approach, *Adv. Chem. Phys.*, 2012, **150**, 187.
32 W. Deng and E. Barkai, Ergodic properties of fractional Brownian-Langevin motion, *Phys. Rev. E*, 2009, **79**, 011112.
33 J.-H. Jeon and R. Metzler, Fractional Brownian motion and motion governed by the fractional Langevin equation in confined geometries, *Phys. Rev. E*, 2010, **81**, 021103.
34 J.-H. Jeon and R. Metzler, Inequivalence of time and ensemble averages in ergodic systems: exponential versus power-law relaxation in confinement, *Phys. Rev. E*, 2012, **85**, 021147.
35 M. Schwarzl, A. Godec and R. Metzler, Quantifying non-ergodicity of anomalous diffusion with higher order moments, *Sci. Rep.*, 2017, **7**, 3878.
36 H. Scher and E. W. Montroll, Anomalous transit-time dispersion in amorphous solids, *Phys. Rev. B*, 1975, **12**, 2455.
37 T. Neusius, I. M. Sokolov and J. C. Smith, Subdiffusion in time-averaged, confined random walks, *Phys. Rev. E*, 2009, **80**, 011109.
38 R. Hou, A. G. Cherstvy, R. Metzler and T. Akimoto, Biased continuous-time random walks for ordinary and equilibrium cases: facilitation of diffusion, ergodicity breaking and ageing, *Phys. Chem. Chem. Phys.*, 2018, **20**, 20827.
39 S. Havlin and D. Ben-Avraham, Diffusion in disordered media, *Adv. Phys.*, 2002, **51**, 187.
40 A. W. C. Lau and T. C. Lubensky, State-dependent diffusion: Thermodynamic consistency and its path integral formulation, *Phys. Rev. E*, 2007, **76**, 011123.
41 A. G. Cherstvy, A. V. Chechkin and R. Metzler, Anomalous diffusion and ergodicity breaking in heterogeneous diffusion processes, *New J. Phys.*, 2013, **15**, 083039.

42 A. G. Cherstvy and R. Metzler, Population splitting, trapping, and non-ergodicity in heterogeneous diffusion processes, *Phys. Chem. Chem. Phys.*, 2013, **15**, 20220.

43 A. G. Cherstvy and R. Metzler, Nonergodicity, fluctuations, and criticality in heterogeneous diffusion processes, *Phys. Rev. E*, 2014, **90**, 012134.

44 A. G. Cherstvy, A. V. Chechkin and R. Metzler, Ageing and confinement in non-ergodic heterogeneous diffusion processes, *J. Phys. A: Math. Theor.*, 2014, **47**, 485002.

45 M. Heidernätsch, *On the diffusion in inhomogeneous systems*, PhD thesis, TU Chemnitz, (2015).

46 A. G. Cherstvy and R. Metzler, Ergodicity breaking, ageing, and confinement in generalized diffusion processes with position and time dependent diffusivity, *J. Stat. Mech.: Theory Exp.*, 2015, P05010.

47 A. Fulinski, Communication: How to generate and measure anomalous diffusion in simple systems, *J. Chem. Phys.*, 2013, **138**, 021101.

48 A. Fulinski, Anomalous weakly nonergodic Brownian motions in nonuniform temperatures, *Acta Phys. Pol.*, 2013, **44**, 1137.

49 F. Thiel and I. M. Sokolov, Scaled Brownian motion as a mean-field model for continuous-time random walks, *Phys. Rev. E*, 2014, **89**, 012115.

50 J.-H. Jeon, A. V. Chechkin and R. Metzler, Scaled Brownian motion: a paradoxical process with a time dependent diffusivity for the description of anomalous diffusion, *Phys. Chem. Chem. Phys.*, 2014, **16**, 15811.

51 H. Safdari, A. G. Cherstvy, A. V. Chechkin, F. Thiel, I. M. Sokolov and R. Metzler, Quantifying the non-ergodicity of scaled Brownian motion, *J. Phys. A: Math. Theor.*, 2015, **48**, 375002.

52 A. Bodrova, A. V. Chechkin, A. G. Cherstvy, H. Safdari, I. M. Sokolov and R. Metzler, Underdamped scaled Brownian motion: (non-)existence of the overdamped limit in anomalous diffusion, *Sci. Rep.*, 2016, **6**, 30520.

53 B. Wang, S. M. Anthony, S. C. Bae and S. Granick, Anomalous yet Brownian, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**, 15160.

54 B. Wang, J. Kuo, S. C. Bae and S. Granick, When Brownian diffusion is not Gaussian, *Nat. Mater.*, 2012, **11**, 481.

55 S. Hapca, J. W. Crawford and I. M. Young, Anomalous diffusion of heterogeneous populations characterized by normal diffusion at the individual level, *J. R. Soc., Interface*, 2009, **6**, 111.

56 J. Guan, B. Wang and S. Granick, Even hard-sphere colloidal suspensions display Fickian yet non-Gaussian diffusion, *ACS Nano*, 2014, **8**, 3331.

57 D. Wang, C. He, M. P. Stoykovich and D. K. Schwartz, Nanoscale topography influences polymer surface diffusion, *ACS Nano*, 2015, **9**, 1656.

58 W. He, H. Song, Y. Su, L. Geng, B. J. Ackerson, H. B. Peng and P. Tong, Dynamic heterogeneity and non-Gaussian statistics for acetylcholine receptors on live cell membrane, *Nat. Commun.*, 2016, **7**, 11701.

59 J.-H. Jeon, M. Javanainen, H. Martinez-Seara, R. Metzler and I. Vattulainen, Protein crowding in lipid bilayers gives rise to non-Gaussian anomalous lateral diffusion of phospholipids and proteins, *Phys. Rev. X*, 2016, **6**, 021006.

60 C. E. Wagner, B. S. Turner, M. Rubinstein, G. H. McKinley and K. Ribbeck, A rheological study of the association and dynamics of MUC5AC gels, *Biomacromolecules*, 2017, **18**, 3654.

61 S. Gupta, J. U. De Mel, R. M. Perera, P. Zolnierczuk, M. Bleuel, A. Faraone and G. J. Schneider, Dynamics of phospholipid membranes beyond thermal undulations, *J. Phys. Chem. Lett.*, 2018, **9**, 2956.

62 A. G. Cherstvy, O. Nagel, C. Beta and R. Metzler, Non-Gaussianity, population heterogeneity, and transient superdiffusion in the spreading dynamics of amoeboid cells, *Phys. Chem. Chem. Phys.*, 2018, **20**, 23034.

63 C. Beck and E. G. D. Cohen, Superstatistics, *Phys. A*, 2003, **322**, 267.

64 C. Beck, Superstatistical Brownian motion, *Prog. Theor. Phys. Suppl.*, 2006, **162**, 29.

65 M. V. Chubynsky and G. W. Slater, Diffusing diffusivity: a model for anomalous, yet Brownian, diffusion, *Phys. Rev. Lett.*, 2014, **113**, 098302.

66 R. Jain and K. L. Sebastian, Diffusion in a crowded, rearranging environment, *J. Phys. Chem. B*, 2016, **120**, 3988.

67 A. V. Chechkin, F. Seno, R. Metzler and I. M. Sokolov, Brownian yet non-Gaussian diffusion: from superstatistics to subordination of diffusing diffusivities, *Phys. Rev. X*, 2017, **7**, 021002.

68 V. Sposini, A. Chechkin, F. Seno, G. Pagnini and R. Metzler, Random diffusivity from stochastic equations: comparison of two models for Brownian yet non-Gaussian diffusion, *New J. Phys.*, 2018, **20**, 043044.

69 T. J. Lampo, S. Stylianidou, M. P. Backlund, P. A. Wiggins and A. J. Spakowitz, Cytoplasmic RNA-protein particles exhibit non-Gaussian subdiffusive behavior, *Biophys. J.*, 2017, **112**, 1.

70 D. C. Young and J. Scrimgeour, Optimizing likelihood models for particle trajectory segmentation in multi-state systems, *Phys. Biol.*, 2018, **15**, 066003.

71 M. T. Valentine, P. D. Kaplan, D. Thota, J. C. Crocker, T. Gisler, R. K. Prud'homme, M. Beck and D. A. Weitz, Investigating the microenvironments of inhomogeneous soft materials with multiple particle tracking, *Phys. Rev. E*, 2001, **64**, 061506.

72 J. Luczka, P. Talkner and P. Hänggi, Diffusion of Brownian particles governed by fluctuating friction, *Phys. A*, 2000, **278**, 18.

73 A. G. Cherstvy and R. Metzler, Anomalous diffusion in time-fluctuating non-stationary diffusivity landscapes, *Phys. Chem. Chem. Phys.*, 2016, **18**, 23840.

74 R. Jain and K. L. Sebastian, Lévy flight with absorption: a model for diffusing diffusivity with long tails, *Phys. Rev. E*, 2017, **95**, 032135.

75 A. Cuetos, N. Morillo and A. Piatti, Is Fickian Yet Non-Gaussian Diffusion Ubiquitous?, E-print arXiv:1806.05116.

76 S. Bochner, *Harmonic Analysis and the Theory of Probability*, Berkeley University Press, Berkeley CA, 1960.

77 J. C. Cox and S. A. Ross, The valuation of options for alternative stochastic processes, *J. Financ. Econ.*, 1976, **3**, 145.

78 S. L. Heston, A closed-form solution for options with stochastic volatility with applications to bond and currency options, *Rev. Financ. Stud.*, 1993, **6**, 327.

79 J.-P. Fouque, G. Papanicolaou and K. R. Sircar, *Derivatives in Financial Markets with Stochastic Volatility*, Cambridge University Press, 2000.

80 O. E. Barndorff-Nielsen and N. Shephard, Econometric analysis of realized volatility and its use in estimating stochastic volatility models, *J. R. Stat. Soc. Ser. B*, 2002, **64**, 253.

81 P. Carr, H. Geman, D. B. Madan and M. Yor, Stochastic volatility for Lévy processes, *Math. Finance*, 2003, **13**, 345.

82 S. Paul, S. Ghosh and D. S. Ray, Reaction-diffusion systems with fluctuating diffusivity; spatio-temporal chaos and phase separation, *J. Stat. Mech.*, 2018, 033205.

83 I. Golding and E. C. Cox, Physical nature of bacterial cytoplasm, *Phys. Rev. Lett.*, 2006, **96**, 098102.

84 Y. M. Wang, R. H. Austin and E. C. Cox, Single molecule measurements of repressor protein 1D diffusion on DNA, *Phys. Rev. Lett.*, 2006, **97**, 048302.

85 V. Tejedor, O. Benichou, R. Voituriez, R. Jungmann, F. Simmel, C. Selhuber-Unkel, L. B. Oddershede and R. Metzler, Quantitative analysis of single particle trajectories: mean maximal excursion method, *Biophys. J.*, 2010, **98**, 1364.

86 K. Hu, P. C. Ivanov, Z. Chen, P. Carpena and H. E. Stanley, Effect of trends on detrended fluctuation analysis, *Phys. Rev. E*, 2001, **64**, 011114.

87 K. Fogelmark, M. A. Lomholt, A. Irbäck and T. Ambjörnsson, Fitting a function to time-dependent ensemble averaged data, *Sci. Rep.*, 2018, **8**, 6984.

88 M. Magdziarz, A. Weron, K. Burnecki and J. Klafter, Fractional Brownian motion versus the continuous-time random walk: a simple test for subdiffusive dynamics, *Phys. Rev. Lett.*, 2009, **103**, 180602.

89 K. Burnecki, E. Kepten, J. Janczura, I. Bronshtein, Y. Garini and A. Weron, Universal algorithm for identification of fractional Brownian motion. A case of telomere subdiffusion, *Biophys. J.*, 2012, **103**, 1839.

90 J.-H. Jeon, V. Tejedor, S. Burov, E. Barkai, C. Selhuber-Unkel, K. Berg-Sørensen, L. Oddershede and R. Metzler, In vivo anomalous diffusion and weak ergodicity breaking of lipid granules, *Phys. Rev. Lett.*, 2011, **106**, 048103.

91 M. S. Song, H. C. Moon, J.-H. Jeon and H. Y. Park, Neuronal messenger ribonucleoprotein transport follows an aging Lévy walk, *Nat. Commun.*, 2018, **9**, 344.

92 J.-H. Jeon, N. Leijnse, L. B. Oddershede and R. Metzler, Anomalous diffusion and power-law relaxation of the time averaged mean squared displacement in worm-like micellar solutions, *New J. Phys.*, 2013, **15**, 04501.

93 S. M. A. Tabei, S. Burov, H. Y. Kim, A. Kuznetsov, T. Huynh, J. Jureller, L. H. Philiosin, A. R. Dinner and N. F. Scherer, Intracelluar transport of insulin granules is a subordinated random walk, *Proc. Natl. Acad. Sci. U. S. A.*, 2013, **110**, 4911.

94 A. V. Weigel, B. Simon, M. M. Tamkun and D. Krapf, Ergodic and nonergodic processes coexist in the plasma membrane as observed by single-molecule tracking, *Proc. Natl. Acad. Sci. U. S. A.*, 2011, **108**, 6438.

95 X. Hu, L. Hong, M. D. Smith, T. Neusius, X. Cheng and J. C. Smith, The dynanmics of single protein molecules is non-equilibrium and self-similar over thirteen decades in time, *Nat. Phys.*, 2016, **12**, 171.

96 T. Kühn, T. O. Ihalainen, J. Hyväluoma, N. Dross, S. F. Willman, J. Langowski, M. Vihinen-Ranta and J. Timonen, Protein diffusion in mammalian cell cytoplasm, *PLoS One*, 2011, **6**, e22962.

97 S. K. Ghosh, A. G. Cherstvy, D. S. Grebenkov and R. Metzler, Anomalous, non-Gaussian tracer diffusion in crowded two-dimensional environments, *New J. Phys.*, 2016, **18**, 013027.

98 F. Thiel, F. Flegel and I. M. Sokolov, Disentangling sources of anomalous diffusion, *Phys. Rev. Lett.*, 2013, **111**, 010601.

99 E. Kepten, A. Weron, G. Sikora, R. Burnecki and Y. Garini, Guidelines for the fitting of anomalous diffusion mean square displacement graphs from single particle tracking experiments, *PLoS One*, 2015, **10**, e0117722.

100 F. Thiel and I. M. Sokolov, Weak ergodicity breaking in an anomalous diffusion process of mixed origins, *Phys. Rev. E*, 2014, **89**, 012136.

101 T. Bayes, An essay toward solving a problem in the doctrine of chances, *Philos. Trans. R. Soc. London*, 1763, **63**, 370.

102 P. S. de Laplace, Théorie analytique des probabilités, Courcier Imprimeur, Paris, 1812.

103 H. Jeffreys, *Theory of Probability*, Clarendon Press, Oxford, 1939; H. Jeffreys, *Scientific Inference*, Cambridge University Press, Cambridge, 1957.

104 C. J. D. MacKay, Information Theory, *Inference and Learning Algorithms''*, Cambridge University Press, 2003.

105 S. D. Sivia and J. Skilling, *Data Analysis: A Bayesian Tutorial*, Oxford University Press, 2006.

106 T. E. Jaynes and L. G. Bretthorst, *Probability Theory – The Logic of Science*, Cambridge University Press, 2003.

107 P. C. Gregory, *Bayesian Logical Data Analysis for the Physical Sciences*, Cambridge University Press, Cambridge, 2005.

108 K. Burnecki, E. Kepten, Y. Garini, G. Sikora and A. Weron, Estimating the anomalous diffusion exponent for single particle tracking data with measurement errors – An alternative approach, *Sci. Rep.*, 2015, **5**, 11306.

109 S. Bera, S. Paul, R. Singh, D. Ghosh, A. Kundu, A. Banerjee and R. Adhikari, Fast Bayesian inference of optical trap stiffness and particle diffusion, *Sci. Rep.*, 2017, **7**, 41638.

110 B. Efron, Why isn't everyone a Bayesian?, *Am. Stat.*, 1986, **40**, 1.

111 R. Trotta, Bayes in the sky: Bayesian inference and model selection in cosmology, *Contemp. Phys.*, 2008, **49**, 71.

112 P. Mukherjee, D. Parkinson and A. R. Liddle, A nested sampling algorithm for cosmological model selection, *Astroph. J. Lett.*, 2006, **638**, L51.

113 R. Trotta, G. Jöhannesson, I. V. Moskalenko, T. A. Porter, R. Ruiz de Austri and A. W. Strong, Constrains in cosmic-ray propagation models from global Bayesian analysis, *Astrophys. J.*, 2011, **729**, 106.

114 G. Jöhannesson, *et al.*, Bayesian analytics of cosmic ray propagation: evidence against homogeneous, *Astrophys. J.*, 2016, **824**, 16.

115 G. D'Agostini, Bayesian inference in processing experimental data: principles and basic applications, *Rep. Prog. Phys.*, 2003, **66**, 1383.

116 V. Dose, Bayesian inference in physics: case studies, *Rep. Prog. Phys.*, 2003, **66**, 1421.

117 U. von Toussaint, Bayesian inference in physics, *Rev. Mod. Phys.*, 2011, **83**, 943.

118 R. D. Cousins, The Jeffreys-Lindley paradox and discovery criteria in high energy physics, *Syntheses*, 2017, **194**, 395.

119 A. M. Ellison, Bayesian inference in ecology, *Ecol. Lett.*, 2004, **7**, 509.

120 R. King, *Bayesian Analysis for Population Ecology*, CRC Press, 2010.

121 M. A. Beaumont, Approximate Bayesian computation in evolution and ecology, *Annu. Rev. Ecol. & Evol. Syst.*, 2010, **41**, 379.

122 J. S. Clark, Why environmental scientists are becoming Bayesians, *Ecol. Lett.*, 2005, **8**, 2.

123 D. J. Wilkinson, Bayesian methods in bioinformatics and computational systems biology, *Briefings Bioinf.*, 2007, **8**, 109.

124 M. A. Beaumont and B. Rannala, The Bayesian revolution in genetics, *Nat. Rev. Genet.*, 2004, **5**, 251.

125 L. F. James, G. Müller and Z. Zhang, Stochastic volatility models based on OU-Gamma time change: theory and estimation, *J. Business & Econ. Stat.*, 2018, **36**, 75.

126 C. Mark, C. Metzner, L. Lautscham, P. L. Strissel, R. Strick and B. Fabry, Bayesian model selection for complex dynamic systems, *Nat. Commun.*, 2018, **9**, 1803.

127 M. Lysy, N. S. Pillai, D. B. Hill, M. G. Forest, J. Mellnik, P. Vasquez and S. A. McKinley, Model comparison and assessment for single particle tracking in biological fluids, *J. Am. Stat. Assoc.*, 2016, **111**, 1413.

128 C. Metzner, C. Mark, J. Steinwachs, L. Lautscham, F. Stadler and B. Fabry, Superstatistical analysis and modelling of heterogeneous random walks, *Nat. Commun.*, 2015, **6**, 7516.

129 D. Montiel, H. Cang and H. Yang, Quantitative characterization of changes in dynamical behavior for single-particle tracking studies, *J. Phys. Chem. B*, 2006, **110**, 19763.

130 J. Skilling, Nested sampling, *AIP Conf. Proc.*, 2004, **735**, 395.

131 J. Skilling, *Nested sampling for Bayesian computations*, Proc. Valencia, ISBA 8th World meeting on Bayesian Statistics, Benidorm (Alicante, Spain), June 1st–6th, (2006).

132 J. Skilling, Nested sampling for general Bayesian computation, *Bayesian Anal.*, 2006, **1**, 833.

133 J. Krog, L. H. Jacobsen, F. W. Lund, D. Wüstner and M. A. Lomholt, Bayesian model selection with fractional Brownian motion, arXiv:1804.01365.

134 A. C. Harvey, Time series models, Prentice Hall, 1993.

135 R. E. Krass and A. E. Raftery, Bayes factors, *J. Am. Stat. Assoc.*, 1995, **90**, 773.

136 K. P. Burnham and D. R. Anderson, Model selection and multimodel inference, Springer-Verlag, 2002.

137 M. Evans and T. Schwartz, Methods for approximating integrals in statistics with special emphasis on Bayesian integration problems, *Stat. Sci.*, 1995, **10**, 254.

138 J. M. Dickey, The weighted likelihood ratio, linear hypotheses on normal location parameters, *Annu. Math. Stat.*, 1971, **42**, 204.

139 R. Trotta, Applications of Bayesian model selection to cosmological parameters, *Mon. Not. R. Astron. Soc.*, 2007, **378**, 72.

140 A. R. Liddle, How many cosmological parameters, *Mon. Not. R. Astron. Soc.*, 2004, **351**, L49.

141 J. Krog and M. A. Lomholt, Bayesian inference with information content model check for Langevin equations, *Phys. Rev. E*, 2017, **96**, 062106.

142 FBM: https://github.com/mlomholt/fbm; DD: https://github.com/sthapa/bmdd.

143 G. E. Uhlenbeck and L. S. Ornstein, On the theory of the Brownian motion, *Phys. Rev.*, 1930, **36**, 823.

144 A. Andreanov and D. S. Grebenkov, Time-averaged MSD of Brownian Motion, *J. Stat. Mech.: Theory Exp.*, 2012, P07001.

145 H. Jeffreys, An invariant form for the prior probability in estimation problems, *Proc. R. Soc. London, Ser. A*, 1946, **186**, 453.

146 S. Eule and R. Friedrich, Describing the dynamics of processes consisting simultaneously of Poissonian and non-Poissonian kinetics, *Phys. Rev. E*, 2013, **87**, 032162.

147 J.-H. Jeon, E. Barkai and R. Metzler, Noisy continuous time random walks, *J. Chem. Phys.*, 2013, **139**, 121916.

148 K. Hinsen and G. R. Kneller, Communication: A multiscale Bayesian inference approach to analyzing subdiffusion in particle trajectories, *J. Chem. Phys.*, 2016, **145**, 151101.

149 S. Chandrasekhar, Stochastic problems in physics and astronomy, *Rev. Mod. Phys.*, 1943, **15**, 1.

150 C. E. Wagner, PhD thesis, *Micro- and macro-rheological studies of the structure and association dynamics of biopolymer gels*, Massachusetts Institute of Technology, 2018.

151 A. G. Cherstvy, S. Thapa, C. E. Wagner and R. Metzler, Non-Gaussian, non-ergodic, and non-Fickian diffusion of tracers in mucin hydrogels: Bayesian model-ranking analysis and behaviour of multiple statistical quantifiers, work in preparation, 2018.

152 O. Lieleg, I. Vladescu and K. Ribbeck, Characterization of particle translocation through mucin hydrogels, *Biophys. J.*, 2010, **98**, 1782.

153 O. Lieleg and K. Ribbeck, Biological hydrogels as selective diffusion barriers, *Trends Cell Biol.*, 2008, **21**, 543.

154 X. Michalet, Mean square displacement analysis of single-particle trajectories with localization error: Brownian motion in an isotropic medium, *Phys. Rev. E*, 2010, **82**, 041914.

155 R. Das, W. C. Cairo and D. Coombs, A hidden Markov model for single particle tracks quantifies dynamic interactions between LFA-1 and the actin cytoskeleton, *PLoS Comput. Biol.*, 2009, **5**, e1000556.

156 F. Perrson, M. Lindén, C. Unoson and J. Elf, Extracting intracellular diffusive states and transition rates from single-molecule tracking data, *Nat. Methods*, 2013, **10**, 265.

157 N. Monnier, Z. Barry, Y. H. Park, C. K. Su, Z. Katz, P. B. English, A. Dey, K. Pan, M. I. Cheeseman, H. R. Singer and M. Bathe, Inferring transient particle transport dynamics in live cells, *Nat. Methods*, 2015, **12**, 838.

158 B. J. Masson, D. Casanova, S. Türkcan, G. Voisinne, R. M. Popoff, M. Vergassola and A. Alexandrou, Inferring

maps of forces inside cell membrane microdomains, *Phys. Rev. Lett.*, 2009, **102**, 048103.
159 M. Slutsky, M. Kardar and L. A. Mirny, Diffusion in correlated random potentials, with applications to DNA, *Phys. Rev. E*, 2004, **69**, 061903.
160 N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth and A. H. Teller, Equations of state calculations by fast computing machines, *J. Chem. Phys.*, 1953, **21**, 1087.
161 W. K. Hastings, Monte Carlo sampling methods using Markov chains and their applications, *Biometrika*, 1970, **57**, 97.
162 Y. Kozachenko, A. Melnikov and Y. Mishura, On drift parameter estimation in models with fractional Brownian motion, *Statistics*, 2015, **49**, 35.
163 A. Beskos, J. Dureau and K. Kalogeropoulos, Bayesian inference for partially observed stochastic differential equations driven by fractional Brownian motion, *Biometrika*, 2015, **102**, 809.
164 B. L. S. Prakasa Rao, *Statistical Inference for Fractional Diffusion Processes*, John Wiley & Sons Ltd, 2010.
165 J. Klafter and I. M. Sokolov, First Steps in Random Walks: From Tools to Applications, Oxford University Press, 2011.
166 H. H. Strey, On the estimation of parameters from time traces originating from an Ornstein–Uhlenbeck process, arXiv:1805.05977, 2018.
167 L. P. García, J. D. Perez, G. Volpe, A. V. Arzola and G. Volpe, High-performance reconstruction of microscopic force fields from Brownian trajectories, arXiv:1808.05468, 2018.
168 We note that one of the first quantitative analyses of non-Gaussian (nearly exponential) displacement distributions for micron-sized tracers in complex heterogeneous media (agarose gels) with varying local microenvironments and particle-to-particle variations of diffusivities was presented in an SPT study by the Weitz group.[71]
169 We note that the concept of stochastic volatility[77,78] in financial economics and option-pricing models[79–81] based on geometric BM is analogous to the DD approach. Also, some extensions of the DD concept were developed recently for reaction-diffusion systems.[82]
170 Another complication arises due to the often highly non-trivial dynamics, especially in living biological cells combining spatial heterogeneity with crowding-induced viscoelasticity.[9,13,96,97] Also other complex systems often feature dynamics influenced by both annealed and quenched types of disorder, and thus combinations of ''elementary'' stochastic mechanisms or crossovers from one to the other mechanism are needed to provide a faithful description.[24,26,90,93,94,98–100] For instance, SPT measurements of potassium channels diffusion in plasma membranes exhibit a coexistence of ergodic ''fractal diffusion'' and non-ergodic continuous-time random walk-like motion,[94] or lipid granule motion in yeast cells show a crossover from continuous-time random walk dynamics to FBM-type motion.[90]
171 As mentioned before,[133] a particular statement on Bayesian model comparison is sensitive to both the amount of available data as well as type of priors and their ranges (see also the discussion in ref. 115). Typically, for a proper model the predictions turn out more confident when more data points are available, and when the respective priors are more restricted in range. The opposite is true for models which do not fit a given data set, where more data penalise statistical probabilities of a given model. Note also that the Occam's razor and Jeffreys-Lindley paradox favour simpler models for sparse data.[118,133]
172 How does the nested-sampling approach compare with the method of simulated annealing (thermodynamic integration)?[131] Similar to the nested-sampling procedure, the simulated-annealing method computes Z per eqn (12), being a thermal method however, it operates with a reduction of temperature to find maximum likelihood. For simulated annealing, the inverse temperature $\beta \propto 1/(k_B\mathcal{T})$ changes gradually from 0 to 1 and the walkers are sampled from the distribution $dP_{i,\beta} \propto \mathcal{L}_i^\beta dX_i$. Thus, the mean log-likelihood averaged over all walkers becomes

$$\langle \log \mathcal{L}_i \rangle_\beta = \frac{\int \mathcal{L}_i^\beta \log \mathcal{L}_i dX_i}{\int \mathcal{L}_i^\beta dX_i} = \frac{d}{d\beta}\left[\log\left(\int_0^1 \mathcal{L}_i^\beta dX_i\right)\right]. \quad (15)$$

To get the log-evidence, we integrate over β variations to find $\int_0^1 \langle \log \mathcal{L}_i \rangle_\beta d\beta = \log\left(\int \mathcal{L}_i dX_i\right) - \log\left(\int dX_i\right) = \log Z_i$. The annealing over $\beta = -\frac{d(\log X_i)}{d(\log \mathcal{L}_i)}$ therefore tracks the density of states of the system,[131] equivalent to -1/slope in $\log \mathcal{L}_i$ *versus* $\log X_i$ plot, while the nested-sampling method tracks likelihood variations varying $\log X_i$. This annealing procedure can fail when $\log \mathcal{L}_i$ is not a strictly-concave function of $\log X_i$, the situation realised for systems with several separated (steady) states.[131] The advantage of the nested-sampling method *versus* simulated-annealing approach is that the first one enables one to correctly assess the likelihood as long as $\mathcal{L}_i$ decreases monotonically with X_i. The nested-sampling method achieves the same accuracy for Z_i with a smaller number of likelihood evaluations, as compared to the simulated-annealing method.[112]
Alternatively, Z_i in eqn (12) can be evaluated directly using the Laplace method of integration for high-dimensional integrals. Note that one can reach an acceptable accuracy with this method only if the integrand of eqn (10) is close to a Gaussian.[137] When comparing two equiprobable models M_1 and M_2, the Bayes factor in favour of M_1 over M_2 is the ratio Z_1/Z_2.[135] Some model-selection methods are not based on computation of evidence *per se*, but rather on Bayes factors. The Savage–Dickey density ratio, see ref. 138, unlike the Laplace method, is applicable to any form of the likelihood function,[139] but it requires the models to be nested. For other model-selection approaches—which are not based on Bayes' theorem but implementing a ''information criterion''—the assumptions on a nested nature of the models and of the Gaussian nature of likelihood functions are not needed.[140] Two popular approaches of this kind are the Akaike- and Bayesian-information criteria. The first one is defined as $\mathrm{AIC} = -2\log \mathcal{L}_{\max} + 2N$, while the second one is $\mathrm{BIC} = -2\log \mathcal{L}_{\max} + N \log N_{\mathrm{points}}$. Here $\mathcal{L}_{\max}$

is the maximum likelihood. The better model is the one with lower AIC or BIC values. While the Akaike criterion is biased towards models with more parameters,[134,135] the choice of Bayesian-information-criterion is justified when the model complexity does not increase with the length of time-series, N_{points}.[136] This assumption does not hold, *e.g.*, for the DD model in the nested-sampling analysis because the number of model parameters N and the algorithm complexity grow with N_{points}.

The idea of Bayesian evidence Z_i in all the evidence-based model-comparison methods enables one not to use approximations of these information-criteria approaches. The nested-sampling method, as used in the main text, also enables one to use physical information and some experimental insights to construct and set a proper range for the priors. The nested-sampling method has a number of advantages. Specifically, it provides confidence intervals for the statistical analysis, ensures statistical robustness of the results, provides an algorithm to find best-fit model parameters, and it does so when several models are to be ranked.

173 We note the principle of maximum information entropy and entropic prior as other possible prior choices.[105,116]

174 We note that while the amplitude spread in Fig. 3 is due to the stochastic nature of the mathematical process at finite observation time T, in given measured data some additional amplitude spread will enter the time-averaged MSD due to measurement inaccuracy as well as heterogeneities on the level of individual trajectories. Diffusing tracer particles may experience patches with varying permeability, locally fluctuating environments, space-varying diffusive and binding-unbinding dynamics with the substrate. Confinement and "caging" effects may also be present. Another reason for the irreproducibility of the time-averaged MSD stems from heterogeneities on the level of the ensemble of tracers due to different starting conditions, possible dispersion of sizes, and different particle–substrate interactions.

175 We list here the physical dimensions of the main quantities and observables used in the text. In the results below some of them are presented as dimensionless, for the sake of simplicity. Starting with the BM and FBM models, MSDs are measured in m^2, generalised diffusion coefficients are in units $\text{m}^2\ \text{s}^{-\alpha}$, the standard D_1 and characteristic $D_\star$ diffusion coefficients are in $\text{m}^2\ \text{s}^{-1}$, anomalous exponents α and Hurst exponents $H = \alpha/2$ are unitless, times are in s with elementary time $\Delta t = 1$, step deviations σ and the quantity η are in m, the noise $\xi(t)$ for BM is in $1/\sqrt{\text{s}}$, the noise $\xi_{\text{fGn}}(t)$ in the FBM model is in $\text{m}\ \text{s}^{-1}$, the standard particle-displacement probability-distribution function $P(x,t)$ are in m^{-1}, and, finally, the displacements $\langle x^2(t)\rangle$ and $\left\langle \overline{\delta^2(\Delta)} \right\rangle$ are in m^2. For the Ornstein–Uhlenbeck process, the quantity $Y(t)$ is in $\text{m}/\sqrt{\text{s}}$, the noise magnitude ε is in $\text{m}\ \text{s}^{-1}$, the noise $\bar{\eta}(t)$ has the dimension of $1/\sqrt{\text{s}}$, variables $\gamma^{\text{act,obs}}$ are in m^2, and the variable s in eqn (S26) has the dimension of $\text{m}^2\ \text{s}^{-1}$. For the Bayesian model, the prior distributions $\pi(\theta)$ have dimensions of $1/\theta$ for a given parameter θ (this is true also for multidimensional priors so that, *e.g.*, the dimension of $\pi(\{Y_{]1[}\})$ is $1/[Y^{(N_{\text{points}}-2)}]$ or $1\Big/\left[\left(\text{m}\Big/\sqrt{\text{s}}\right)^{(N_{\text{points}}-2)}\right]$), the dimensions of the likelihood function and evidence Z depend on the number of points in the trace, namely it is $1/\text{m}^{(N_{\text{points}}-1)}$, the quantities X, weights w, and the information are unitless. Note that as we compare the likelihood functions or evidence for each model for trajectories with a given number of points, for relative model probabilities given by eqn (9) the units of functions Z_i cancel one another.