Maximum likelihood estimates of diffusion coefficients from single-particle tracking experiments

Jakob Tómas Bullerjahn and Gerhard Hummer

ARTICLES YOU MAY BE INTERESTED IN

Maximum likelihood estimates of diffusion coefficients from single-particle tracking experiments

View Online   Export Citation   CrossMark

Jakob Tómas Bullerjahn[1] (ID) and Gerhard Hummer[1,2,a] (ID)

AFFILIATIONS

[1] Department of Theoretical Biophysics, Max Planck Institute of Biophysics, 60438 Frankfurt am Main, Germany
[2] Institute of Biophysics, Goethe University Frankfurt, 60438 Frankfurt am Main, Germany

[a] Author to whom correspondence should be addressed: gerhard.hummer@biophys.mpg.de

ABSTRACT

Single-molecule localization microscopy allows practitioners to locate and track labeled molecules in biological systems. When extracting diffusion coefficients from the resulting trajectories, it is common practice to perform a linear fit on mean-squared-displacement curves. However, this strategy is suboptimal and prone to errors. Recently, it was shown that the increments between the observed positions provide a good estimate for the diffusion coefficient, and their statistics are well-suited for likelihood-based analysis methods. Here, we revisit the problem of extracting diffusion coefficients from single-particle tracking experiments subject to static noise and dynamic motion blur using the principle of maximum likelihood. Taking advantage of an efficient real-space formulation, we extend the model to mixtures of subpopulations differing in their diffusion coefficients, which we estimate with the help of the expectation–maximization algorithm. This formulation naturally leads to a probabilistic assignment of trajectories to subpopulations. We employ the theory to analyze experimental tracking data that cannot be explained with a single diffusion coefficient. We test how well a dataset conforms to the assumptions of a diffusion model and determine the optimal number of subpopulations with the help of a quality factor of known analytical distribution. To facilitate use by practitioners, we provide a fast open-source implementation of the theory for the efficient analysis of multiple trajectories in arbitrary dimensions simultaneously.

I. INTRODUCTION

Single-particle tracking methods are routinely used to monitor the erratic motion of labeled macromolecules in their native environment, such as molecular motors moving along the cytoskeletal network,[1,2] transcription factors binding to DNA,[3] or receptor proteins diffusing in cell membranes.[4] The mode of transport and associated transport coefficients are inferred from the measured trajectories using microscopic models mimicking the globally observed behavior, which can range from ordinary to anomalous and confined diffusion, as well as active transport.[5]

The simplest mode of transport in a dense fluid medium is free diffusion, which is fully characterized by a diffusion coefficient D. However, single-particle tracking experiments are plagued by static localization errors and dynamic motion blur, which have to be properly accounted for when estimating diffusion coefficients. Static errors originate from various noise sources in the experimental setup, such as spatial resolution of the instrument and noise in the detection and processing electronics,[6] and are commonly modeled via additive Gaussian noise.[7–9] Motion blur arises from the camera's finite frame integration time, during which the particle position is smeared out. This effect depends on the illumination profile of the shutter, which in most cases is uniform,[8] but can, in general, be represented by any non-negative function $s(t)$ that integrates to unity over the frame integration time.[9]

Traditionally, the estimation of diffusion coefficients has relied on the fact that the slope of the mean-squared displacement (MSD) is directly proportional to D.[10] This procedure, however, has some serious drawbacks. For instance, the most common estimator for the MSD of finite time series introduces correlations between the

MSD values at different time lags, which, in turn, cause the estimate to suffer if too many MSD values are used for the fit.[11,12] To remedy these shortcomings, one can either consider the optimal number of MSD values for the fit[12] or explicitly account for the above-mentioned correlations in the fit procedure.[13] Departing from the idea that the MSD is needed to determine D, one arrives at the highly efficient covariance-based estimator (CVE),[14] which is based on the much simpler statistics of the position increments between two observations. Even before, Berglund[9] used the sparse covariance matrix of the increments to construct an approximate maximum likelihood estimator (MLE) operating in discrete Fourier space. This estimator asymptotically approaches the exact MLE in the limit of infinitely long trajectories. It was later shown[15] that replacing the Fourier transform with a discrete sine transform results in an orthogonal basis in which the MLE of Ref. 9 is exact.

Despite the fact that the CVE is orders of magnitude faster than likelihood-based methods,[14] the latter can still be beneficial, e.g., to incorporate prior knowledge of the parameters in a Bayesian manner or to globally analyze trajectories of different lengths. The latter issue is common in single-particle tracking because individual molecules stochastically disappear as a result of photobleaching. Importantly, MLEs are minimally affected by blinking events and other interruptions of the trajectory recording, in the sense that the estimates and their uncertainty depend almost exclusively on the number of observed particle positions, and not the number of trajectories. We note, though, that the CVE can also overcome missing positions in the trajectory by lumping together increments of the resulting trajectory segments.[16] However, correlations in longer trajectories may reveal deviations from regular diffusion.

Here, we revisit the problem of minimizing the negative log-likelihood, but instead of transforming the data, we work in real space and exploit the symmetries of the increment covariance matrix to develop a fast and reliable numerical solution scheme. We develop MLEs of the mean-squared positional uncertainty a^2 in single-particle tracking and of the diffusional spread σ^2 during a single time step Δt,

$$\sigma^2 = 2D\Delta t, \tag{1}$$

from which one obtains the MLE of the diffusion coefficient as $D = \sigma^2/(2\Delta t)$, irrespective of the dimension d. The MLEs can also be used to analyze molecular dynamics simulation trajectories by setting the motion blur coefficient to $B = 0$. The coefficient a^2 then accounts for fast non-diffusional spread in the particle position.

We extend the likelihood formulation to a mixture model, which assumes that the trajectories to be analyzed originate from different subpopulations, each characterized by a distinct diffusion coefficient. In comparison to some established models, such as those that account for diffusion in inhomogeneous media[17–24] or those allowing for multiple diffusive regimes within a single trajectory,[25–29] our approach may seem somewhat restrictive. However, what it lacks in generality it makes up with speed, efficiency, and compactness. Furthermore, we provide rigorous statistical tests to assess whether the data comply with the theory assumptions or if more demanding models are needed to explain the data.

This paper is structured as follows. In Sec. II A, we review the minimal stochastic model of Ref. 9 for diffusive trajectories smeared out by static localization noise and dynamic motion blur.

Section II B is dedicated to the negative log-likelihood of the trajectory increments and the efficient $\mathcal{O}(N)$-algorithm used to numerically minimize it. To test the assumptions of the underlying diffusion model, we introduce a quality factor in Sec. II C that can be inspected either visually or via a non-parametric test, such as the Kolmogorov–Smirnov (KS) test or Kuiper's test. In Sec. II D, we generalize the likelihood function to a mixture of subpopulations, each with a distinct diffusion coefficient. We minimize this joint likelihood with an expectation–maximization (EM) algorithm. A novel selection criterion for the optimal number of subpopulations, based on quality factor statistics, is introduced and tested on synthetic data in Sec. II E. Section III A explores ways to determine whether a small sample of short trajectories is governed by diffusive dynamics or not. In Sec. III B, we use the mixture MLEs to analyze single-molecule tracking experiments[30] reporting on the diffusive dynamics of human mission elapsed time (MET) receptor tyrosine kinase in live cells. An outlook on possibilities to incorporate elements of Bayesian analysis into the theory is presented in Sec. IV. Our findings are summarized in Sec. V and implemented in a data analysis package written in Julia.[31] Detailed derivations and background information on the theory can be found in Appendixes A–D.

II. THEORY

A. Microscopic model

A detailed stochastic model capturing the effective dynamics observed in single-particle tracking experiments was introduced by Savin and Doyle[8] for a uniform illumination profile and later generalized to arbitrary profiles by Berglund.[9] In what follows, we review the model by Berglund, on which we then build the MLE in Sec. II B.

Consider a freely diffusing particle in one dimension with diffusion coefficient D whose dynamics is described with a time-continuous Wiener process Y, satisfying

$$\dot{Y}(t) = \frac{\sigma}{\sqrt{\Delta t}}\xi(t). \tag{2}$$

Here, a dot indicates the time derivative and $\xi(t)$ denotes Gaussian white noise fully characterized by $\langle\xi(t)\rangle = 0$ and $\langle\xi(t)\xi(t')\rangle = \delta(t - t')$, with $\delta(t)$ being the Dirac δ-function. The particle motion is captured by using a camera with frame integration time Δt, during which the camera shutter may be fully or partially open. This gives rise to the shutter function $s(t) \geq 0$ that smears out the particle position over the integration time Δt, resulting in the process

$$Z_i = \int_0^{\Delta t} d\tau\, s(\tau)Y(\tau + i\Delta t),$$
$$\int_0^{\Delta t} d\tau\, s(\tau) = 1 \tag{3}$$

for $i = 1, 2, \ldots$. Furthermore, we assume that each frame is affected by additional measurement noise, which we model as Gaussian with variance $a^2/2$. The observed motion of the particle is therefore described by the following stochastic process:

$$X_i = Z_i + \frac{a}{\sqrt{2}}R_i, \tag{4}$$

where R is a normal distributed random variable with $\langle R_i \rangle = 0$ and $\langle R_i R_j \rangle = \delta_{i,j}$, and $\delta_{i,j}$ is the Kronecker delta that evaluates to one if $i = j$ and zero otherwise. Note that due to linearity, the Gaussian nature of ξ and R is inherited by processes X, Y, and Z.

According to Eqs. (2)–(4), the observed mean-squared displacement (MSD) is given by

$$\langle (X_i - X_0)^2 \rangle = a^2 + \sigma^2(i - 2B), \tag{5}$$

which is, in comparison to $\langle [Y(i\Delta t) - Y(0)]^2 \rangle = i\sigma^2$, corrupted by the static error a^2 and the dynamic error $2\sigma^2 B$. The latter is characterized by the motion blur coefficient

$$B = \frac{1}{\Delta t} \int_0^{\Delta t} d\tau\, S(\tau)[1 - S(\tau)], \tag{6}$$

where $S(t)$ gives the relative amount of exposure up to the time t, namely,

$$S(t) = \int_0^t d\tau\, s(\tau) \sim \begin{cases} 0, & t = 0 \\ 1, & t = \Delta t. \end{cases}$$

For uniform illumination, we have $s(t) = \Delta t^{-1}$, which results in a motion blur coefficient of $B = 1/6$. Generally, $0 \le B \le 1/4$ must hold. In the case of (near-)perfect time resolution, e.g., when analyzing molecular dynamics trajectories, one has $B = 0$.

B. Maximum likelihood estimation

The probability to observe a one-dimensional time series $\vec{X} = (X_0, X_1, X_2, \ldots, X_N)^T$ is given by the Gaussian joint probability distribution function $p(\vec{X}) \propto \exp\left(-[\vec{X} - \langle\vec{X}\rangle]^T \Sigma_X^{-1} [\vec{X} - \langle\vec{X}\rangle]/2\right)$, which can be reinterpreted as a likelihood function for the parameters a^2 and σ^2. These enter the dense covariance matrix Σ_X whose inversion requires $\mathcal{O}(N^3)$ operations and therefore make the variation of a^2 and σ^2, when maximizing the likelihood, computationally expensive.

The likelihood function can be expressed more economically in terms of the position increments $\Delta_i = X_i - X_{i-1}$, with $i = 1, 2, \ldots, N$. Like X_i, they are also Gaussian distributed, their expectation values read $\langle \Delta_i \rangle = 0$, and the elements of the corresponding covariance matrix Σ are given by

$$\Sigma_{i,j} = a^2 \Sigma'_{i,j} + \sigma^2 \Sigma''_{i,j}, \tag{7}$$

$$\Sigma'_{i,j} = \delta_{i,j} - \frac{\delta_{i-1,j}}{2} - \frac{\delta_{i,j-1}}{2}, \tag{8}$$

$$\Sigma''_{i,j} = (1 - 2B)\delta_{i,j} + B(\delta_{i-1,j} + \delta_{i,j-1}), \tag{9}$$

i.e., Σ is a linear combination of the constant tridiagonal matrices Σ' and Σ''. The likelihood of observing the increments $\vec{\Delta} = (\Delta_1, \Delta_2, \ldots, \Delta_N)^T$ in one dimension thus has the form

$$\ell(\vec{\Delta} \mid a^2, \sigma^2) = \sqrt{\frac{1}{(2\pi)^N |\Sigma|}} \exp\left(-\frac{1}{2} \vec{\Delta}^T \Sigma^{-1} \vec{\Delta}\right), \tag{10}$$

which results in the following negative log-likelihood (up to a negligible constant):

$$\mathcal{L}(\vec{\Delta} \mid a^2, \sigma^2) = \frac{1}{2} \vec{\Delta}^T \Sigma^{-1} \vec{\Delta} + \frac{1}{2} \ln(|\Sigma|). \tag{11}$$

Because Σ is tridiagonal, Eq. (11) can be evaluated in an efficient manner. For example, the Thomas algorithm[32] and a three-term recurrence relation[33] can be used to calculate $\Sigma^{-1}\vec{\Delta}$ and $|\Sigma|$, respectively, in $\mathcal{O}(N)$ operations. Yet, due to the fact that the covariance matrix here is also a symmetric Toeplitz matrix, the recurrence relation can be solved analytically (see Appendix A), giving

$$\ln(|\Sigma|) = N \ln(\alpha) + (N+1)\ln\left(\frac{1+q}{2}\right)$$
$$+ \ln\left(1 - \left[\frac{1-q}{1+q}\right]^{N+1}\right) - \ln(q). \tag{12}$$

Here, $q = \sqrt{1 - 4\beta^2/\alpha^2} \ge 0$, with $\alpha = a^2 + \sigma^2(1 - 2B)$ and $\beta = -a^2/2 + \sigma^2 B$. If $q \to 0$, which implies either $\sigma \to 0$ or $a \to 0$ with $B = 1/4$, the expression above reduces to

$$\lim_{q \to 0} \ln(|\Sigma|) = \ln(N+1) + N \ln(\alpha/2). \tag{13}$$

Equations (12) and (13) can be modified to account for non-symmetric tridiagonal Toeplitz matrices, as demonstrated in Ref. 34.

A global analysis of M d-dimensional trajectories of different lengths $N_m + 1$ is realized by minimizing

$$\mathcal{L}\left(\{\vec{\Delta}_{m,n}\}_{m=1,\ldots,M}^{n=1,\ldots,d} \mid a^2, \sigma^2\right) = \frac{1}{2} \sum_{m=1}^{M} \sum_{n=1}^{d} \left[\vec{\Delta}_{m,n}^T \Sigma_m^{-1} \vec{\Delta}_{m,n} + \ln(|\Sigma_m|)\right] \tag{14}$$

with respect to the same one-dimensional parameters a^2 and σ^2 as before due to the assumption of isotropic motion. Here, the quantities $\vec{\Delta}_{m,n}$ and Σ_m are defined as before, except that they vary with the trajectory ($m = 1, \ldots, M$) and dimension ($n = 1, \ldots, d$), respectively. In general, the minimization has to be conducted numerically, but thanks to the linear dependence of Σ on the parameters, it becomes analytically tractable on the boundaries, where either a^2 or σ^2 becomes zero. The remaining parameter can then be estimated as follows:

$$a^2\big|_{\sigma^2=0} = \frac{1}{d\mathcal{N}_M} \sum_{m=1}^{M} \sum_{n=1}^{d} \vec{\Delta}_{m,n}^T \Sigma_m'^{-1} \vec{\Delta}_{m,n}, \tag{15}$$

$$\sigma^2\big|_{a^2=0} = \frac{1}{d\mathcal{N}_M} \sum_{m=1}^{M} \sum_{n=1}^{d} \vec{\Delta}_{m,n}^T \Sigma_m''^{-1} \vec{\Delta}_{m,n}, \tag{16}$$

where $\mathcal{N}_M = \sum_{m=1}^{M} N_m$ denotes the total number of considered increments and N_m is the number of entries in the vector $\vec{\Delta}_{m,n}$, independent of n. Note that the vector-matrix products are all strictly positive because matrices Σ', Σ'', and Σ, as well as their inverses, are all positive definite (see Appendix B). If both a^2 and σ^2 are greater than zero, Eq. (14) can effectively be reduced to a one-dimensional optimization problem in the vein of Ref. 14, where the sine-transformed counterpart of Eq. (11) was considered. We thereby introduce the

new parameters a^2 and $\phi = \sigma^2/a^2$, which make the covariance matrix and the log-likelihood function separable, i.e., $\boldsymbol{\Sigma}_m = a^2 \tilde{\boldsymbol{\Sigma}}_m(\phi) \ \forall m$ and

$$\mathcal{L}\left(\{\vec{\Delta}_{m,n}\} \mid a^2, \phi\right) = \frac{1}{2} \sum_{m=1}^{M} \sum_{n=1}^{d} \left[\frac{1}{a^2} \vec{\Delta}_{m,n}^T \tilde{\boldsymbol{\Sigma}}_m^{-1} \vec{\Delta}_{m,n} \right.$$
$$\left. + N_m \ln(a^2) + \ln(|\tilde{\boldsymbol{\Sigma}}_m|) \right]. \qquad (17)$$

The determinants $\ln(|\tilde{\boldsymbol{\Sigma}}_m|)$ can be evaluated via Eqs. (12) and (13) by replacing α and β with $\tilde{\alpha} = 1 + \phi(1 - 2B_m)$ and $\tilde{\beta} = -1/2 + \phi B_m$, respectively. Equation (17) gets minimized with respect to a^2 for

$$a^2 = \frac{1}{d\mathcal{N}_M} \sum_{m=1}^{M} \sum_{n=1}^{d} \vec{\Delta}_{m,n}^T \tilde{\boldsymbol{\Sigma}}_m^{-1} \vec{\Delta}_{m,n} \qquad (18)$$

and therefore reduces to

$$\mathcal{L}\left(\{\vec{\Delta}_{m,n}\} \mid \phi\right) = \frac{d\mathcal{N}_M}{2} \ln\left(\sum_{m=1}^{M} \sum_{n=1}^{d} \vec{\Delta}_{m,n}^T \tilde{\boldsymbol{\Sigma}}_m^{-1} \vec{\Delta}_{m,n} \right) + \frac{d}{2} \sum_{m=1}^{M} \ln(|\tilde{\boldsymbol{\Sigma}}_m|), \qquad (19)$$

except for an additive constant $d\mathcal{N}_M/2 - d\mathcal{N}_M \ln(d\mathcal{N}_M)/2$, which can be neglected because it is independent of a^2 and σ^2. Minimizing Eq. (19) with respect to ϕ is a nonlinear one-dimensional optimization problem, which can be tackled conveniently using robust derivative-free algorithms, such as Brent's method[35] or golden-section search.[36] The Julia data analysis package[31] makes use of an implementation of Brent's method provided by the Optim package.[37]

To assess the uncertainty of the estimates, we turn to the standard errors $\delta\theta = \sqrt{\text{var}(\theta)}$ of the model parameters $\theta \in \{a^2, \sigma^2\}$. These are bounded from below by the Cramér–Rao bounds, which are computed from the Fisher information corresponding to the likelihood as

$$\delta a^2|_{\sigma^2=0} \geq \sqrt{\frac{2}{d\mathcal{N}_M}} a^2|_{\sigma^2=0},$$
$$\delta\sigma^2|_{a^2=0} \geq \sqrt{\frac{2}{d\mathcal{N}_M}} \sigma^2|_{a^2=0},$$
$$\delta a^2 \geq \sqrt{\frac{I_{2,2}}{I_{1,1}I_{2,2} - I_{1,2}^2}}, \qquad (20)$$
$$\delta\sigma^2 \geq \sqrt{\frac{(a^2)^2 I_{1,1} - 2\phi a^2 I_{1,2} + \phi^2 I_{2,2}}{I_{1,1}I_{2,2} - I_{1,2}^2}}.$$

The matrix elements $I_{i,j}$ of the Fisher information matrix $\mathbf{I}(a^2, \phi)$ are specified in Appendix C along with a detailed derivation of the above equations. The bounds in Eq. (20) become tight in the limit $\mathcal{N}_M \to \infty$ because the estimators are asymptotically unbiased (see Fig. 1). For extremely sparse datasets, where the lower bounds of Eq. (20) vastly underestimate the standard errors, other methods have to be employed to estimate the uncertainty, such as bootstrapping.
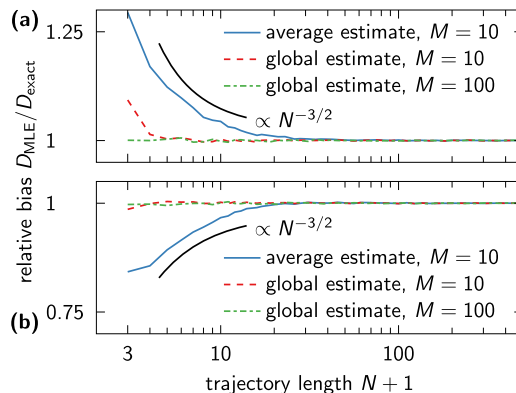

FIG. 1. Estimating relative bias in the diffusion coefficient MLE with respect to the trajectory length. We analyzed M simulated trajectories of equal lengths $N + 1$, either separately or collectively, using Eqs. (15), (16), (18), and (19). We considered two cases: a signal-to-noise ratio of (a) $\sigma^2/a^2 = 1/2$ and (b) $\sigma^2/a^2 = 2$. The resulting average diffusion coefficient (blue solid line), which was determined from the single-trajectory estimates and compared to the simulation input value D_{exact}, was either over- or under-estimated, respectively, for $N \lesssim 20$. As indicated by the guides to the eye (black solid lines), the bias is $\mathcal{O}(N^{-3/2})$, which becomes $\mathcal{O}(\mathcal{N}_M^{-3/2})$ for the global MLE estimates (red dashed line for $M = 10$ and green dashed-dotted line for $M = 100$). This is why the latter are virtually unbiased for almost all trajectory lengths. All results were averaged over multiple realizations to reduce noise.

It is worth mentioning that the formalism of Refs. 9 and 15, which treats the problem in discrete sine space (albeit only for $M = d = 1$, but the extension to d and M different from one is straightforward), results in identical numerical values for the parameter estimates and standard errors. However, our approach has the advantage that we obtain closed-form expressions for the standard errors [Eq. (20)], whereas these same quantities involve finite sums that grow with the trajectory length in discrete sine space. Working in real space can therefore be beneficial under certain circumstances, e.g., in the context of Bayesian inference involving priors based on the Fisher information matrix (see Sec. IV).

Finally, note that the estimators a^2 and ϕ (or, equivalently, a^2 and σ^2) are slightly correlated, as seen by the non-vanishing off-diagonal elements $I_{1,2} = I_{2,1}$ of the Fisher information matrix. Figure 1 explores the bias in estimates of σ^2 and the associated diffusion coefficient as a function of the trajectory length $N + 1$, using two-dimensional trajectories generated via Brownian dynamics simulations. The simulations were conducted using discretized versions of Eqs. (2)–(4) (see Appendix D 1) and analyzed both on a single-trajectory level and collectively. Our numerical results demonstrate that a bias of $\mathcal{O}(N^{-3/2})$ affects single-trajectory estimates of short trajectories but is virtually non-existent in global estimates, where the bias scales like $\mathcal{O}(\mathcal{N}_M^{-3/2})$, i.e., with the total number of increments. This is in good agreement with analytical results obtained in Ref. 14, which show that the estimators in discrete sine space are unbiased up to $\mathcal{O}(N^{-1})$. However, the authors of Ref. 14 claimed that next to the said above-mentioned bias, which originates from the asymmetry of the likelihood function, there should be an additional $\mathcal{O}(N^{-1/2})$ bias coming from the fact that the maximum likelihood approach requires a^2 and σ^2 to be positive. The diffusion

coefficient estimator in real space does not exhibit this second bias, at least not for the signal-to-noise values considered here.

C. Quality factor analysis

How can we test that we are actually observing free diffusion? According to Eq. (10), for a diffusive process, the elements of $\mathbf{\Sigma}_m^{-1/2}\vec{\Delta}_{m,n}$ for a fixed n are independent, uncorrelated normal random variables with zero mean and unit variance. The sum of their squares, given by

$$\chi_m^2 = \sum_{n=1}^d \vec{\Delta}_{m,n}^T \mathbf{\Sigma}_m^{-1} \vec{\Delta}_{m,n} > 0, \tag{21}$$

should therefore be distributed according to a χ^2-distribution with dN_m degrees of freedom. Note that we do not correct dN_m by the number of model parameters because the degrees of freedom associated with a single trajectory are just a tiny fraction of the overall number of unconstrained degrees of freedom, $d\mathcal{N}_M - 2$.

In principle, we could test whether a sample of M trajectories adheres to the diffusion model in Eqs. (2)–(4) by verifying the distribution of the corresponding χ_m^2-values, but only if they are all of equal length. For trajectories that differ in their lengths $N_m + 1$, the quadratic forms χ_m^2 follow different χ^2-distributions. To simplify the analysis, we focus on their corresponding cumulative distribution functions, which are all uniform on the interval $[0, 1]$. The associated statistic is given by

$$\text{Ϙ}_m = 1 - \frac{\gamma(dN_m/2, \chi_m^2/2)}{\Gamma(dN_m/2)}, \tag{22}$$

where $\gamma(a, z)$ and $\Gamma(a)$ denote the lower incomplete and ordinary Γ-functions, respectively, that are defined as follows:

$$\gamma(a, z) = \int_0^z dx\, x^{a-1} e^{-x}, \quad \Gamma(a) = \lim_{z \to \infty} \gamma(a, z).$$

This is reminiscent of the discussion in Ref. 13, where a quantity similar to the one in Eq. (22) is referred to as the quality factor Q, because its values are uniformly distributed whenever the elements of $\mathbf{\Sigma}_m^{-1/2}\vec{\Delta}_{m,n}$ are truly independent. The main difference between Eq. (22) and traditional quality factors, such as the one in Ref. 13, is that the latter consider χ^2-statistics computed by summing over the weighted squared differences between data and model predictions. Here, the model makes predictions about the distribution of χ_m^2 of individual trajectories m and, in turn, of Ϙ_m, into which the parameters a^2 and σ^2 enter via the covariance matrix. To highlight this distinction, we make use of the archaic greek letter *qoppa*, instead of Q, to denote this (somewhat unorthodox) quality factor.

To test whether a set $\{\text{Ϙ}_m\}_{m=1,\ldots,M}$ of quality factor values follows the uniform distribution on $[0, 1)$ expected for diffusive processes, we employ a variation of the Kolmogorov–Smirnov statistic called the Kuiper statistic.[38,39] Kuiper's variant measures the largest vertical deviations of the empirical distribution function above and below the cumulative distribution function and is defined as their sum. In our case, it thus reads

$$\kappa = \sqrt{M}\left(\max_m\left[\frac{m}{M} - \text{Ϙ}_m\right] + \max_m\left[\text{Ϙ}_m - \frac{m-1}{M}\right]\right) \tag{23}$$

for a list of Ϙ_m-values, $m = 1, 2, \ldots, M$, sorted in ascending order. Ideally, we have $\kappa \approx 1$, but significantly larger κ-values arise when the underlying assumption of $\mathbf{\Sigma}_m^{-1/2}\vec{\Delta}_{m,n}$ being normal distributed is violated. This is the case for heterogeneous data arising from subpopulations with differing diffusion coefficients, as discussed in Sec. II D.

Finally, it should be mentioned that one can construct a right-tailed p-value from the sampling distribution of κ. Asymptotically, it takes the form

$$p \overset{M\to\infty}{\sim} 2\sum_{m=1}^\infty \left(4m^2\kappa^2 - 1\right)e^{-2\,m^2\kappa^2}, \tag{24}$$

where the sum has to be truncated at some large value, e.g., at $m = 100$ or $m = 1000$, if used in practice. The associated probability p can be helpful to determine whether a κ-value is reasonable or not. For sparse datasets, tables can be used to look up the p-value.[40]

D. Analysis of mixtures of subpopulations

In situations where a single diffusion coefficient cannot describe the observed dynamics in a sample of trajectories, the single-population analysis of Sec. II B breaks down. Here, we extend the model by introducing K distinct subpopulations, each characterized by a set of parameters $\{a_k^2, \sigma_k^2\}$, $k = 1, 2, \ldots, K$. Every tracked particle is assumed to belong to one of these subpopulations, but the exact assignment is not known *a priori*. Particles are not allowed to switch between populations and therefore have a fixed diffusion coefficient, depending on which subpopulation k they belong to. The task of parameter fitting now becomes twofold: The parameters of each subpopulation have to be varied to find their optimal values, while the particle trajectories have to be assigned to the subpopulations they most likely belong to. Problems like these can be treated with the help of the EM algorithm,[41] as outlined below.

To analyze data from a mixture of K subpopulations with distinct diffusive dynamics, we consider the following likelihood function with latent mixing fractions P_k:

$$\ell\left(\{\vec{\Delta}_{m,n}\} \mid \{P_k, a_k^2, \sigma_k^2\}_{k=1,\ldots,K}\right) = \prod_{m=1}^M \sum_{k=1}^K P_k \prod_{n=1}^d \ell_k(\vec{\Delta}_{m,n} \mid a_k^2, \sigma_k^2). \tag{25}$$

The one-dimensional likelihoods $\ell_k(\vec{\Delta}_{m,n} \mid a_k^2, \sigma_k^2)$ are given by Eq. (10), where the subscript k indicates that the covariance matrix $\mathbf{\Sigma} = \mathbf{\Sigma}_m$ is evaluated using the subpopulation parameters a_k^2 and σ_k^2. Now, due to the sum appearing in $\ell(\{\vec{\Delta}_{m,n}\} \mid \{P_k, a_k^2, \sigma_k^2\})$, its negative log-likelihood remains intractable. This is where the EM algorithm comes in: Instead of minimizing $-\ln(\ell)$ explicitly, we consider the upper bound

$$L\left(\{\vec{\Delta}_{m,n}\} \mid \{P_k, a_k^2, \sigma_k^2\}\right)$$
$$= -\sum_{m=1}^M \sum_{k=1}^K T_{k,m} \ln\left(\frac{P_k}{T_{k,m}}\prod_{n=1}^d \ell_k(\vec{\Delta}_{m,n} \mid a_k^2, \sigma_k^2)\right)$$
$$\geq -\sum_{m=1}^M \ln\left(\sum_{k=1}^K T_{k,m}\frac{P_k}{T_{k,m}}\prod_{n=1}^d \ell_k(\vec{\Delta}_{m,n} \mid a_k^2, \sigma_k^2)\right)$$
$$= -\ln(\ell(\{\vec{\Delta}_{m,n}\} \mid \{P_k, a_k^2, \sigma_k^2\})), \tag{26}$$

which follows from Jensen's inequality. In principle, such upper bounds can be constructed for arbitrary coefficients $T_{k,m} > 0$ whose interpretation becomes clearer in what follows. If we were able to choose $T_{k,m}$ such that equality holds in Eq. (26) for all parameter tuples $\{P_k, a_k^2, \sigma_k^2\}_{k=1,\ldots,K}$, then it would not matter if we minimized $-\ln(\ell)$ or L, given a known set of $T_{k,m}$. This is exactly the case whenever

$$\frac{P_k}{T_{k,m}} \prod_{n=1}^{d} \ell_k(\vec{\Delta}_{m,n} \mid a_k^2, \sigma_k^2) = \text{const.},$$

so by requiring that the coefficients $T_{k,m}$ are also normalized, we finally obtain

$$T_{k,m} = \left[\sum_{k'=1}^{K} P_{k'} \prod_{n'=1}^{d} \ell_{k'}(\vec{\Delta}_{m,n'} \mid a_k^2, \sigma_k^2) \right]^{-1} P_k \prod_{n=1}^{d} \ell_k(\vec{\Delta}_{m,n} \mid a_k^2, \sigma_k^2).$$
$$(27)$$

It now becomes apparent that $T_{k,m}$ corresponds to the probability of trajectory m belonging to the kth subpopulation.

The EM algorithm is a two-step algorithm, where in the first step, the current values of $\{P_k, a_k^2, \sigma_k^2\}$ are used to update the classification coefficients $T_{k,m}$ via Eq. (27). In the second step, the latter are plugged into the upper bound [Eq. (26)], which can be rewritten as

$$L(\{\vec{\Delta}_{m,n}\} \mid \{P_k, a_k^2, \sigma_k^2\})$$
$$= \sum_{m=1}^{M} \sum_{k=1}^{K} T_{k,m} \left[\sum_{n=1}^{d} \mathcal{L}(\vec{\Delta}_{m,n} \mid a_k^2, \sigma_k^2) - \ln\left(\frac{P_k}{T_{k,m}}\right) \right]. \quad (28)$$

Here, $\mathcal{L}(\vec{\Delta}_{m,n} \mid a_k^2, \sigma_k^2)$ is defined in Eq. (11) for $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_m$. Minimizing Eq. (28) with respect to $\{P_k, a_k^2, \sigma_k^2\}$ updates the estimate for the parameters, and the first step is repeated. The minimization can mostly be done analytically because Eq. (28) is separable and therefore susceptible to the methods applied in Sec. II B. For the mixing fractions, we get

$$P_k = \frac{1}{M} \sum_{m=1}^{M} T_{k,m}. \quad (29)$$

On the boundary, the solutions read

$$a_k^2 \big|_{\sigma_k^2 = 0} = \frac{1}{C_k} \sum_{m=1}^{M} T_{k,m} \sum_{n=1}^{d} \vec{\Delta}_{m,n}^T \boldsymbol{\Sigma}_m'^{-1} \vec{\Delta}_{m,n}, \quad (30)$$

$$\sigma_k^2 \big|_{a_k^2 = 0} = \frac{1}{C_k} \sum_{m=1}^{M} T_{k,m} \sum_{n=1}^{d} \vec{\Delta}_{m,n}^T \boldsymbol{\Sigma}_m''^{-1} \vec{\Delta}_{m,n}, \quad (31)$$

$$C_k = d \sum_{m=1}^{M} T_{k,m} N_m. $$

Using the parameter representation (a_k^2, ϕ_k), with $\phi_k = \sigma_k^2 / a_k^2$, Eq. (28) is minimized for

$$a_k^2 = \frac{1}{C_k} \sum_{m=1}^{M} T_{k,m} \sum_{n=1}^{d} \vec{\Delta}_{m,n}^T \tilde{\boldsymbol{\Sigma}}_m^{-1} \vec{\Delta}_{m,n}, \quad (32)$$

$$\phi_k = \arg\min_{\phi_k} \frac{C_k}{2} \ln\left(\sum_{m=1}^{M} T_{k,m} \sum_{n=1}^{d} \vec{\Delta}_{m,n}^T \tilde{\boldsymbol{\Sigma}}_m^{-1} \vec{\Delta}_{m,n} \right) + \frac{d}{2} \sum_{m=1}^{M} T_{k,m} \ln(|\tilde{\boldsymbol{\Sigma}}_m|). \quad (33)$$

The high-dimensional minimization problem therefore reduces to a series of analytic expressions [Eqs. (29)–(32)] and a handful of independent one-dimensional optimization problems [Eq. (33)], which can be solved numerically in an efficient way. The algorithm is implemented in a Julia data analysis package.[31]

Finally, we would like to mention that one could also consider a continuum of a^2 and σ^2 values as an alternative to discrete subpopulations. If only a^2 is allowed to vary between trajectories, a latent variable model could be considered. In general, a Bayesian formulation would naturally lend itself to a continuous-parameter
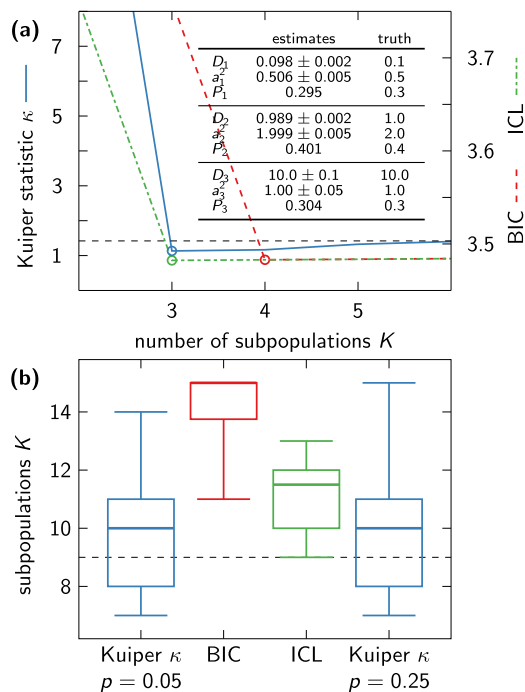

FIG. 2. Verifying the effectiveness of quality factor statistics as a selection criterion. (a) Determining the optimal number K of subpopulations. We generated $300 + 400 + 300 = 1000$ synthetic trajectories of different lengths (uniformly distributed on $[4, 101]$) via Brownian dynamics simulations, using the parameter values listed in the table under "truth." After analyzing the data with the EM algorithm for $K = 1, 2, \ldots, 6$, the corresponding Kuiper statistic [Eq. (23), blue solid line] was computed. For $K \geq 3$, κ remained below $\kappa \approx 1.42$ (black dashed line), corresponding to p-values greater than 0.25. Accordingly, we chose $K = 3$ (blue circle) as the optimal number of subpopulations. The associated parameter estimates are listed in the table under "estimates" and are in excellent agreement with ground truth. The red and green circles denote the corresponding BIC (red dashed line) and ICL (green dashed-dotted line) predictions. (b) Statistics of predictions for 20 distinct datasets, each containing $110 + 30 + 70 + 80 + 190 + 230 + 110 + 120 + 60 = 1000$ trajectories sampled from a complex mixture of nine subpopulations ($a_{k=1,2,\ldots,9}^2 = 0.04, 0.04, 0.05, 0.13, 0.06, 0.09, 0.20, 0.35, 0.99$ and $\sigma_{k=1,2,\ldots,9}^2 = 0.08, 0.0005, 0.009, 0.02, 0.18, 0.32, 0.10, 0.34, 0.72$) and following the same length distribution as in (a). The data were analyzed for $K = 1, 2, \ldots, 15$, and the predictions of BIC (red) and ICL (green) were determined by the positions of their minima. In contrast, the novel selection criterion (blue) relied on the first instance where κ went below a threshold value of either $\kappa \approx 1.75$ ($p = 0.05$) or $\kappa \approx 1.42$ ($p = 0.25$). In the event of the threshold not being reached, which happened more frequently for the lower threshold, the position where κ got minimized was chosen.

treatment, in which a prior distribution would be updated into a posterior distribution in light of the observed data.

E. Selection criterion for optimal number of subpopulations

As the number of subpopulations K is increased, the fit to heterogeneous data gradually gets better. Because the regularity conditions for conventional criteria, such as the Bayesian information criterion (BIC), do not hold for finite mixture models,[42] we propose to rely on a quality factor analysis for model selection. This is done as follows: After repeated fitting of a dataset via the EM algorithm for a fixed K, the classification coefficients $T_{k,m}$ that (along with the optimal parameter values $\{P_k, a_k^2, \sigma_k^2\}$) minimize Eq. (28) are used to assign each trajectory m to the subpopulation k it most likely belongs to, according to

$$k = \arg\max_i T_{i,m}. \tag{34}$$

The associated quality factors $\{Q_m^{(k)}\}$ are then determined using the subpopulation-specific parameters a_k^2 and σ_k^2, resulting in a set

$\{\{Q_m^{(1)}\}, \{Q_m^{(2)}\}, \ldots, \{Q_m^{(K)}\}\}$ that is finally plugged into Eq. (23) to evaluate the Kuiper statistic κ. Figure 2(a) serves as a proof of concept for this selection procedure, where κ is plotted as a function of K for a simulated heterogeneous dataset with three distinct subpopulations. The estimates for $K = 3$ are in excellent agreement with the input parameter values of our simulations for three subpopulations and result in $\kappa \approx 1.13$. Note that the slight increase in κ for $K > 3$ is due to stochasticity in the EM optimization, which becomes more pronounced as the number of subpopulations is increased (for details on our implementation of the EM algorithm, see Appendix D 2). Also note that we are not interested in the global minimum of κ with respect to K, but in the smallest K where $\kappa \approx 1$. In practice, this can be realized by a threshold value for κ, below which κ is considered sufficiently close to one to accept the associated K as the optimal number of subpopulations. We recommend practitioners to choose a threshold somewhere between $\kappa \approx 1.75$ and 1.42 (corresponding to the p-values 0.05 and 0.25, respectively) that reflects their confidence in the quality of the data.

For comparison, we show predictions of two established selection criteria, namely, the BIC and the integrated completed likelihood (ICL) criterion,[42] in Fig. 2(a). The former can be evaluated for
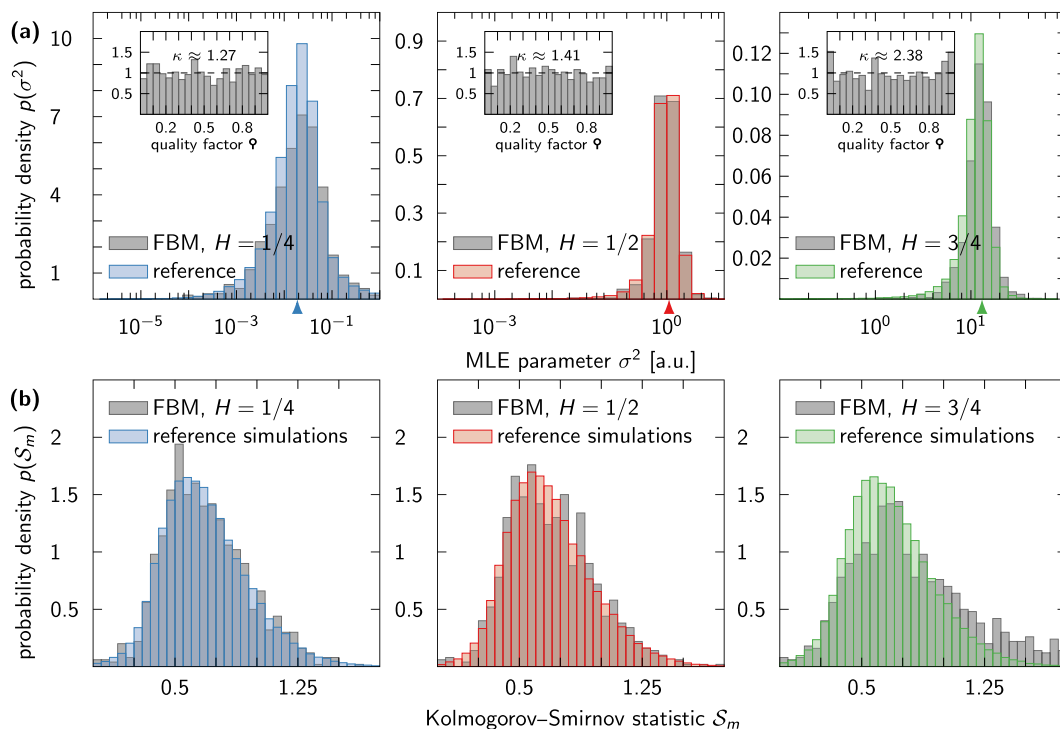

FIG. 3. Testing for diffusivity in FBM subject to static noise and dynamic motion blur. Using $B = 1/6$ and $a^2 = \sigma^2 = 1$, where σ^2 is the variance of the normal random variable used as input for the circulant method, we generated 1000 trajectories with lengths uniformly distributed on [4, 101] governed by subdiffusive (left panels), diffusive (center panels) and superdiffusive (right panels) dynamics and compared them to reference data made up of 10^5 diffusive trajectories. We considered the (a) distributions of single-trajectory σ^2-estimates, distributions of quality factors calculated for the FBM data (insets), and (b) distributions of KS statistics, which test for varying diffusivity within the trajectories. The gray histograms are associated with the FBM data, whereas the colored histograms belong to the reference data. The arrows at the bottom of the plots in (a) indicate the global σ^2-estimates. For the superdiffusive data, we saw clear deviations in all three distributions, thus confirming that the process under scrutiny was non-diffusive. However, for $H = 1/4$, the results were more subtle: we only detected clearly noticeable deviations in the distribution of σ^2, which, in combination with the fact that the other two distributions fit the reference data, hinted at non-diffusive behavior.

our model as follows:

$$\text{BIC} = \frac{2L\left(\left\{\vec{\Delta}_{m,n}\right\} \mid \left\{P_k, a_k^2, \sigma_k^2\right\}\right) + (3\,K - 1)\ln(d\mathcal{N}_M)}{\mathcal{N}_M}, \quad (35)$$

where the likelihood is given by Eq. (28). The ICL is formulated in an almost identical manner as the BIC, except that the classification coefficients $T_{k,m}$ in front of the square brackets in Eq. (28) are replaced by

$$\tilde{T}_{k,m} = \begin{cases} 1 & \arg\max_i T_{i,m} = k \\ 0 & \text{otherwise.} \end{cases}$$

Whereas the BIC overestimates the number of subpopulations, the ICL and the Kuiper statistics criteria produce reliable results for the mixture of three subpopulations. However, for more complex mixtures, also the ICL fails. Figure 2(b) visualizes the prediction statistics of BIC, ICL, and the Kuiper statistics criterion when applied to datasets composed of trajectories sampled from nine distinct subpopulations. Although the distributions of predicted K-values are fairly broad, probably due to the trajectories being so short and few, the biases of BIC and ICL are clear, thus confirming the claim that traditional selection criteria cannot be relied on in our particular case.

III. APPLICATION TO EMPIRICAL DATA

To illustrate the use and effectiveness of the above theory, we applied it to simulations and experimental data. Comparisons to the former are primarily intended to gauge the limits of the theoretical framework, while the analysis of experimental data is meant to demonstrate its use in practice.

A. Simulated diffusion data

As mentioned earlier, there are multiple modes of transport possible in microbiological systems, and a diffusion coefficient estimate is only reliable when the data satisfy the assumptions of the model. In this regard, we generated one-dimensional synthetic trajectories imitating fractional Brownian motion (FBM)[43] and diffusion in inhomogeneous media,[44] two processes that deviate from regular diffusion distinctly, and tested whether the theory was able to detect their non-diffusive behavior. This was realized by analyzing three well-defined quantities: the distribution of quality factors, the distribution of single-trajectory σ^2-estimates, and the distribution of statistics resulting from a two-sample Kolmogorov–Smirnov (KS) test, which estimates how likely it is for early parts of a trajectory to be governed by the same diffusion dynamics as its later parts. The distributions were extracted
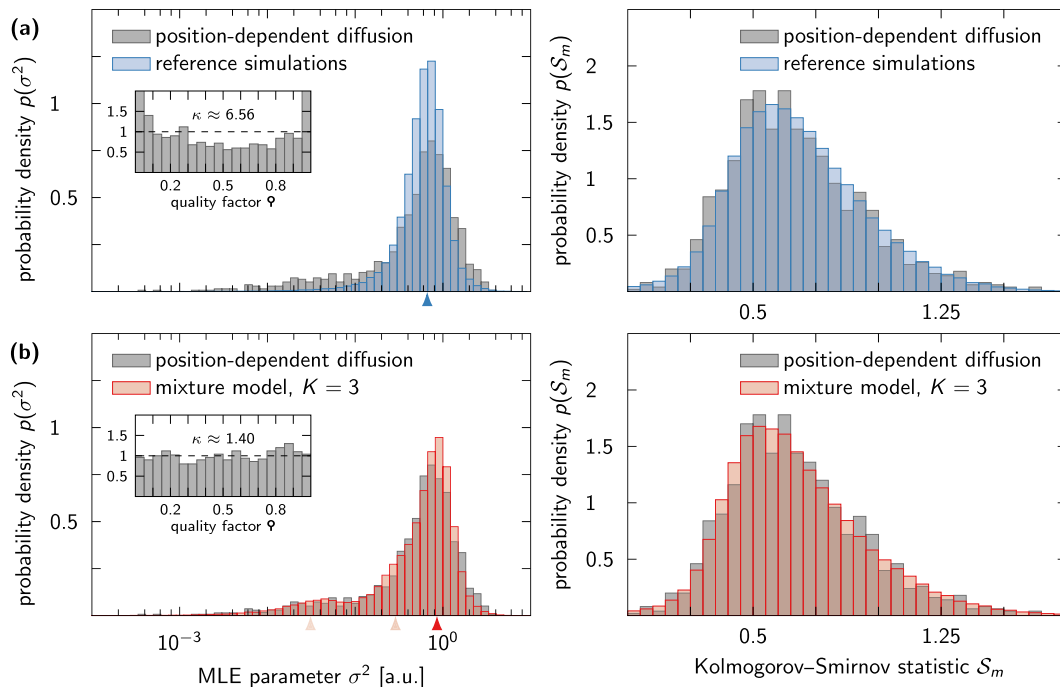

FIG. 4. Comparing position-dependent diffusion to mixture models. (a) Analogous to Fig. 3, we compared PDD data (gray histograms) to diffusive reference data (blue histograms) in terms of their distributions of single-trajectory σ^2-estimates (left panel), quality factors (inset), and KS statistics (right panel). The PDD data were generated via a stochastic process X satisfying Eqs. (36), (3), and (4), resulting in 1000 non-diffusive trajectories with lengths uniformly distributed on [4, 101]. The reference data consisted of 10^5 trajectories following the same length distribution as the non-diffusive data and were generated using $B = 1/6$, and a^2 and σ^2 extracted from a global analysis of the PDD data (the arrow indicates the global σ^2-estimate). The non-uniform quality factor distribution and strong deviations between the σ^2-distributions clearly indicate a non-diffusive behavior in the PDD data. (b) After analyzing the non-diffusive data via mixture models, we used the parameters for the minimal model best describing the data in (a) (arrows indicate the σ^2-estimates, and their relative opacities indicate the respective mixing fractions) to generate the 10^5 trajectories behind the red histograms. Insets: Quality factor distributions for the PDD data. The associated $\mathcal{Q}$-values [Eq. (22)] were calculated using (a) the parameters of a single-population fit and (b) the fit parameters of a mixture model with $K = 3$.

from the non-diffusive data and compared visually to their diffusive counterparts.

FBM is a Gaussian process with stationary increments, similar to the Wiener process in Eq. (2). However, the increments of FBM need not be independent and therefore give rise to a MSD that grows proportional to t^{2H}, with $H \in (0, 1)$ being the so-called Hurst index. For $H = 1/2$, we retrieve the Wiener process, whereas the cases $H < 1/2$ and $H > 1/2$ result in sub- and super-diffusion, respectively, due to negative and positive correlations. We generated realizations of FBM via the circulant method[45] and added static noise and dynamic motion blur by replacing $Y(t)$ in Eqs. (2)–(4) (or their discretized counterparts; see Appendix D 1) with the realizations. The resulting trajectories were analyzed both individually and globally using Eqs. (15), (16), (18), and (19), where the solution that resulted in the lowest negative log-likelihood value was chosen in each case. We used the global estimates for a^2 and σ^2, on the one hand, to calculate the corresponding Q_m-values [Eq. (22)] and, on the other hand, as input parameters for simulations of regular diffusion, i.e., Eqs. (2)–(4), which we considered as reference data. For further details on the Brownian dynamics simulations, see Appendix D 1. The above-mentioned KS test was achieved by splitting $\Sigma_m^{-1/2} \vec{\Delta}_m$ in half (dropping the last entry if the number of elements in $\vec{\Delta}_m$ was uneven) and forming two empirical distribution functions that were then compared. Note that $d = 1$, so we have dropped the index n in $\vec{\Delta}_{m,n}$. The covariance matrices Σ_m were evaluated using the global parameter estimates. The above procedures gave rise to the samples $\{\sigma_m^2\}$, $\{Q_m\}$, and $\{\mathcal{S}_m\}$ for $m = 1, 2, \ldots, M$ of single-trajectory σ^2-estimates, quality factors, and KS statistics, respectively, whose distributions are plotted in Fig. 3. Here and in the following, reference simulations were sampled extensively such that the statistical scatter in the reference histograms is negligible.

Overall, the data sparsity and strong static noise, which in most cases lead to a low effective signal-to-noise ratio σ^2/a^2, made the non-diffusive nature of FBM somewhat difficult to detect. For example, in the case of $H = 1/4$, the only significant discrepancy between the data and reference was found in the distribution of single-trajectory estimates of σ^2. Yet, this turned out to be a clear indication for non-diffusive dynamics in light of the fact that the other two quantities did not deviate from their diffusive counterparts. For $H = 3/4$, where discrepancies were found in all considered quantities, the situation was more definitive. A possible reason for the vastly different outcomes for $H = 1/4$ and $H = 3/4$ is that the latter case results in a slower decay of correlations between increments than $H = 1/4$. The two cases are therefore by no means symmetric, as one might naively assume.

The second non-diffusive process we considered allows for a position-dependent diffusion (PDD) profile $D(z) = \sigma(z)^2/2\Delta t$. One thereby has to replace Eq. (2) with the following stochastic differential equation:

$$\dot{Y}(t) = \frac{\sigma(Y)\sigma'(Y)}{\Delta t} + \frac{\sigma(Y)}{\sqrt{\Delta t}}\xi(t), \qquad (36)$$

which is to be interpreted in the sense of Itô. Here, $f'(z)$ denotes the derivative of $f(z)$ with respect to z. The resulting process is generally not Gaussian due to the presence of multiplicative noise. For

our simulations, we chose $\sigma(z) = \sigma_0[1 - 0.9e^{-(z/z_0)^2/2}]$ to mimic the diffusion around a diffusivity well, and the initial positions Y_0 were drawn from a uniform distribution on the interval $[-5z_0, 5z_0]$. Analogous to the FBM data, the lengths of the trajectories were uniformly distributed on $[4, 101]$ and a total of 1000 non-diffusive trajectories were considered.

Our results for $a^2 = \sigma_0^2 = 1$ and $z_0 = 3$ are given in Fig. 4(a). Again, we compared our results to reference data simulated using the global MLE estimates of a^2 and σ^2 obtained from an analysis of the PDD data. The fact that the quality factor distribution is anything but uniform clearly indicates that the process cannot be described by a single diffusion coefficient. This is also mirrored in the lack of overlap between the PDD data and the reference seen in the distribution of single-trajectory σ^2-estimates. Only the KS statistics seemed consistent between the two datasets. We also fitted the non-diffusive data with mixture models and found a decent match ($\kappa \approx 1.40$) for $K = 3$ and the parameters $a_{k=1,2,3}^2 \approx 0.97$, 1.03, 1.03, $\sigma_{k=1,2,3}^2 \approx 0.29$,
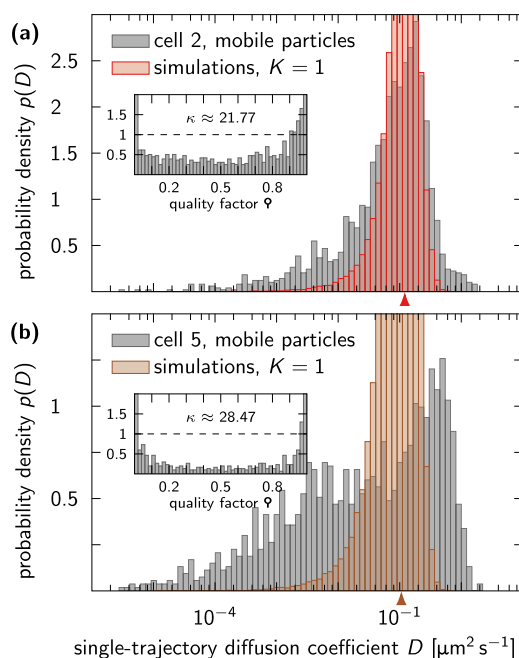
FIG. 5. Assessment of a model with a single diffusion coefficient against single-molecule localization microscopy experiments. Shown are the distributions of single-trajectory diffusion coefficients estimated from the experiment (gray histograms) and from synthetic trajectories generated via a single-population model (colored histograms). (a) The global fit parameters $a^2 = (3.51 \pm 0.03) \times 10^{-3}\ \mu m^2$ and $D = (1.21 \pm 0.01) \times 10^{-1}\ \mu m^2\ s^{-1}$ (red arrow) were used to simulate 10^5 independent trajectories following the same length distribution as the experimental data for cell 2. The distribution of single-trajectory diffusion coefficients (red histogram) turned out to be too narrow to explain the experimental observations. (b) Same as in (a), except that for cell 5, the parameters underlying the brown histogram are $a^2 = (3.09 \pm 0.03) \times 10^{-3}\ \mu m^2$ and $D = (1.062 \pm 0.009) \times 10^{-1}\ \mu m^2$ (brown arrow). Insets: Distributions of quality factors corresponding to the experimentally observed trajectories, evaluated using the respective global single-population fit parameters. Their horned shapes strongly deviate from the expected uniform distribution on $[0, 1)$ (black dashed lines), which highlights the fact that the experimental data cannot be properly described by a single-population model.

0.86, 0.03 and $P_{k=1,2,3} \approx 0.17, 0.71, 0.12$. As a test for consistency and to see if diffusion in inhomogeneous media can be distinguished from heterogeneous mixtures using samples of short trajectories, we then used the estimated parameters to generate new synthetic data. These data, drawn from the mixture model best describing the PDD, are compared to the PDD data in Fig. 4(b). Overall, the data from the mixture model and PDD process are fairly similar, thus making it impossible to distinguish between the processes, at least for the PDD profile considered here.

In conclusion, it can be challenging to distinguish between diffusive and non-diffusive processes when the samples are made up of extremely short trajectories. In addition, the presence of static noise and dynamic motion blur muddies the waters, especially when the (effective) signal-to-noise ratio is poor. We recommend that practitioners investigate the three above-mentioned distributions, compare their results to Brownian dynamics simulations, and even consider further tests, such as the goodness-of-fit test proposed in Ref. 14.

B. Experimental data

We also applied the above theory to single-molecule tracking data for human MET receptor tyrosine kinase, bound to labeled 3H3-Fab antibody fragments within the plasma membrane of HeLa cells.[30] In the original study, tracking in two dimensions was conducted for multiple cells using the super-resolution imaging protocol uPAINT[46] and imaging fluorescence correlation spectroscopy.[47] Next to "resting" MET, realized with the above-mentioned Fab antibodies, the authors of Ref. 30 also considered internalin B-bound MET. Here, we limit our analysis to the Fab-data recorded via uPAINT for ten randomly chosen cells. To account for the fact that each cell might be influenced by its local environment, we refrained from pooling together data measured in different cells. The temporal resolution of the experiment was $\Delta t = 0.02$ s, and we assumed a blurring coefficient of $B = 1/6$ for all trajectories.

Each trajectory of a given cell was first analyzed separately, analogous to the simulation data in Sec. III A. This gave rise to M diffusion coefficients per cell ($M = 1280, 1943, 1567, 1709, 1840, 874, 1165, 1582, 1178,$ and 1357 for cells 1 through 10), i.e., one for each trajectory, and allowed us to sieve out the trajectories of essentially immobile particles, which we defined as those having $D < 1 \times 10^{-6}$ μm^2 s^{-1}. This is in line with the original analysis of the data,[30] but, alternatively, one can explicitly account for the immobile fraction in the modeling.[48] We then analyzed globally the remaining trajectories to estimate a^2 and σ^2 for each cell. The normalized
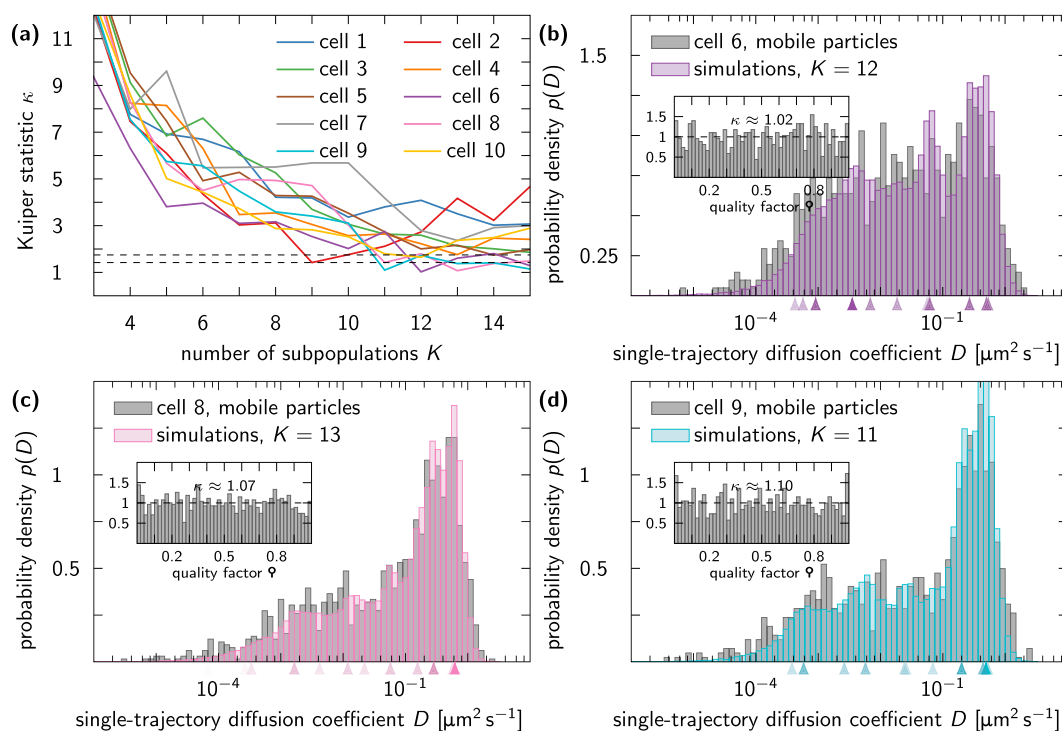

FIG. 6. Multi-mixture models compared to experimental data. (a) The Kuiper statistic [Eq. (23)] for each cell as a function of the number K of subpopulations in the mixture. The black dashed lines mark the thresholds $\kappa \approx 1.42$ and $\kappa \approx 1.75$, below which there is a > 25% and > 5% chance, respectively, to obtain quality factor results at least as extreme as actually observed. Three cells clearly cross the lower threshold, and a fourth one narrowly misses it. [(b)–(d)] Comparison between experimental single-trajectory diffusion coefficient distributions (gray histograms) and results from Brownian dynamics simulations. The 10^5 synthetic diffusion coefficients making up each of the colored histograms were obtained in almost the same way as in Fig. 5, except that different parameters a_k^2 and σ_k^2 were used for each mixture component, and the mixing fractions P_k determined their frequency of occurrence. The arrows indicate the diffusion coefficients behind the subpopulations and their relative opacities indicate the respective mixing fractions, both of which are tabulated in Table I. Insets: Distributions of quality factors, after sorting the trajectories according to the subpopulations they most likely belong to.

TABLE I. Fit parameters for specific cells obtained via the EM algorithm. The number of trajectories M, mean trajectory length $\langle N + 1 \rangle$, and standard deviation (SD) of the trajectory lengths are listed for each cell. The units of the diffusion coefficients D_k and static noise amplitudes a_k^2 are $\mu m^2 \ s^{-1}$ and μm^2, respectively.

	Cell 6 $M = 874$ $\langle N + 1 \rangle = 26$ SD $= 44$	Cell 8 $M = 1582$ $\langle N + 1 \rangle = 30$ SD $= 56$	Cell 9 $M = 1178$ $\langle N + 1 \rangle = 30$ SD $= 49$
D_1	$(5.4 \pm 0.4) \times 10^{-1}$	$(6.2 \pm 0.1) \times 10^{-1}$	$(5.3 \pm 0.5) \times 10^{-1}$
a_1^2	$(2.5 \pm 0.2) \times 10^{-2}$	$(4.3 \pm 0.3) \times 10^{-3}$	$(2.8 \pm 0.2) \times 10^{-2}$
P_1	0.09	0.223	0.06
D_2	$(5.0 \pm 0.2) \times 10^{-1}$	$(5.7 \pm 0.4) \times 10^{-1}$	$(5.0 \pm 0.2) \times 10^{-1}$
a_2^2	$(4.7 \pm 0.5) \times 10^{-3}$	$(2.0 \pm 0.1) \times 10^{-2}$	$(7.3 \pm 0.4) \times 10^{-3}$
P_2	0.11	0.052	0.17
D_3	$(2.7 \pm 0.1) \times 10^{-1}$	$(2.83 \pm 0.07) \times 10^{-1}$	$(4.4 \pm 0.2) \times 10^{-1}$
a_3^2	$(2.0 \pm 0.2) \times 10^{-3}$	$(3.4 \pm 0.1) \times 10^{-3}$	$(1.8 \pm 0.3) \times 10^{-3}$
P_3	0.11	0.184	0.12
D_4	$(6.2 \pm 0.4) \times 10^{-2}$	$(1.55 \pm 0.05) \times 10^{-1}$	$(2.01 \pm 0.06) \times 10^{-1}$
a_4^2	$(7.9 \pm 0.3) \times 10^{-3}$	$(2.3 \pm 0.1) \times 10^{-3}$	$(2.6 \pm 0.1) \times 10^{-3}$
P_4	0.11	0.089	0.16
D_5	$(6.1 \pm 0.3) \times 10^{-2}$	$(5.8 \pm 0.5) \times 10^{-2}$	$(6.8 \pm 0.3) \times 10^{-2}$
a_5^2	$(1.9 \pm 0.1) \times 10^{-3}$	$(1.42 \pm 0.05) \times 10^{-2}$	$(2.5 \pm 0.1) \times 10^{-3}$
P_5	0.06	0.046	0.05
D_6	$(5.6 \pm 0.9) \times 10^{-2}$	$(5.7 \pm 0.2) \times 10^{-2}$	$(2.6 \pm 0.2) \times 10^{-2}$
a_6^2	$(2.3 \pm 0.1) \times 10^{-2}$	$(1.58 \pm 0.05) \times 10^{-3}$	$(9.2 \pm 0.3) \times 10^{-3}$
P_6	0.03	0.065	0.04
D_7	$(1.8 \pm 0.1) \times 10^{-2}$	$(2.2 \pm 0.1) \times 10^{-2}$	$(2.4 \pm 0.1) \times 10^{-2}$
a_7^2	$(1.19 \pm 0.06) \times 10^{-3}$	$(5.1 \pm 0.1) \times 10^{-3}$	$(2.02 \pm 0.06) \times 10^{-3}$
P_7	0.07	0.053	0.06
D_8	$(6.9 \pm 0.7) \times 10^{-3}$	$(1.18 \pm 0.04) \times 10^{-2}$	$(5.8 \pm 0.3) \times 10^{-3}$
a_8^2	$(6.0 \pm 0.2) \times 10^{-3}$	$(1.51 \pm 0.03) \times 10^{-3}$	$(1.54 \pm 0.03) \times 10^{-3}$
P_8	0.08	0.080	0.10
D_9	$(3.5 \pm 0.3) \times 10^{-3}$	$(4.1 \pm 0.3) \times 10^{-3}$	$(2.6 \pm 0.3) \times 10^{-3}$
a_9^2	$(2.15 \pm 0.05) \times 10^{-3}$	$(3.77 \pm 0.08) \times 10^{-3}$	$(3.81 \pm 0.09) \times 10^{-3}$
P_9	0.13	0.045	0.07
D_{10}	$(9.2 \pm 0.7) \times 10^{-4}$	$(1.64 \pm 0.08) \times 10^{-3}$	$(5.8 \pm 0.4) \times 10^{-4}$
a_{10}^2	$(1.49 \pm 0.03) \times 10^{-3}$	$(1.38 \pm 0.02) \times 10^{-3}$	$(1.34 \pm 0.02) \times 10^{-3}$
P_{10}	0.12	0.103	0.12
D_{11}	$(5.6 \pm 0.8) \times 10^{-4}$	$(3.3 \pm 0.3) \times 10^{-4}$	$(3.8 \pm 0.3) \times 10^{-4}$
a_{11}^2	$(9.9 \pm 0.3) \times 10^{-4}$	$(9.2 \pm 0.2) \times 10^{-4}$	$(6.7 \pm 0.1) \times 10^{-4}$
P_{11}	0.05	0.044	0.05
D_{12}	$(4.2 \pm 0.8) \times 10^{-4}$	$(2.6 \pm 0.5) \times 10^{-4}$	
a_{12}^2	$(2.90 \pm 0.08) \times 10^{-3}$	$(5.7 \pm 0.2) \times 10^{-4}$	
P_{12}	0.04	0.012	
D_{13}		$(7 \pm 5) \times 10^{-5}$	
a_{13}^2		$(3.4 \pm 0.3) \times 10^{-4}$	
P_{13}		0.004	

histograms of single-trajectory diffusion coefficients for two distinct cells are depicted in Fig. 5, next to Brownian dynamics simulation results for a single-population model. The simulations were conducted using discretized versions of Eqs. (2)–(4) (see Appendix D 1), where a^2 and σ^2 were set equal to the global cell parameter estimates and the length of each trajectory was drawn from the observed length distribution of the respective cell.

Figure 5 clearly illustrates that the experimentally determined distributions are too broad to be explained by a single set of diffusion parameters. To verify this conclusion, we computed the corresponding quality factors [Eq. (22)], one for each considered trajectory, and inspected visually, as well as with the help of the Kuiper statistic [Eq. (23)], whether they were uniformly distributed on the interval $[0, 1)$. Unsurprisingly, this was not the case, as seen in the insets of Fig. 5. We therefore proceeded to fit the data with a mixture of $K = 1, 2, \ldots, 15$ subpopulations via the EM algorithm of Sec. II D, starting from multiple different random initial values for $\{a_k^2, \sigma_k^2\}_{k=1,\ldots,K}$, and uniform mixing fractions $P_k = 1/K \, \forall k$. This was done to increase the chance of finding the global minimum of Eq. (28). Further details on our implementation of the EM algorithm can be found in Appendix D 2.

Although our results for the live-cell data were not entirely unambiguous, the general trend of κ decreasing with increasing K was still observed [see Fig. 6(a)]. Depending on the choice of threshold, we ended up with either six or three cells, for which the null hypothesis that their single-trajectory diffusion coefficient distributions originated from a finite mixture of fast and slowly diffusing particles could not be rejected. Furthermore, the three cells that crossed the more stringent threshold of $\kappa \approx 1.42$ all gave κ-values with associated p-values [Eq. (24)] well above 0.5. For cells 6 and 9, the same optimal K-values were predicted for both thresholds, whereas the optimum shifted somewhat upward for cell 8.

The results for the lower threshold suggest that the trajectories of cells 6, 8, and 9 should be sorted into 11 to 13 distinct subpopulations [see Figs. 6(b)–6(d)] whose relevant parameters are tabulated in Table I. However, it should be noted that some of the populations can be lumped together if we solely focus on the diffusion coefficient. In addition, the last two subpopulations of cell 8 are extremely underrepresented and can, in principle, be neglected. While this may not be apparent from the corresponding P_k-values, it becomes clear when the classification coefficients $T_{k,m}$ are inspected: It turns out that for the parameters presented here, only 9 and 6 trajectories out of a total of 1349 mobile ones get assigned to the 12th and 13th subpopulations, respectively. This might be an indication of an overly stringent threshold and that cell 8 is better described with $K = 11$ subpopulations. In conclusion, the effective number of subpopulations within each cell might therefore be somewhat smaller than the output of the EM algorithm seems to imply, but certainly not two or three.

For completeness, in Fig. 7, we plot the KS distributions for cells 6, 8, and 9, as well as for the multi-population simulation data of Fig. 6. In comparison to single-population reference data, we can see that both the cell data and multi-population simulation data are shifted and have a slightly broader tail than the reference. The qualitative agreement between the cell data and the simulation data strengthens our working hypothesis that the measured trajectories originate from a heterogeneous mixture of fast and slowly diffusing particles. However, in light of our results in Sec. III A, we cannot
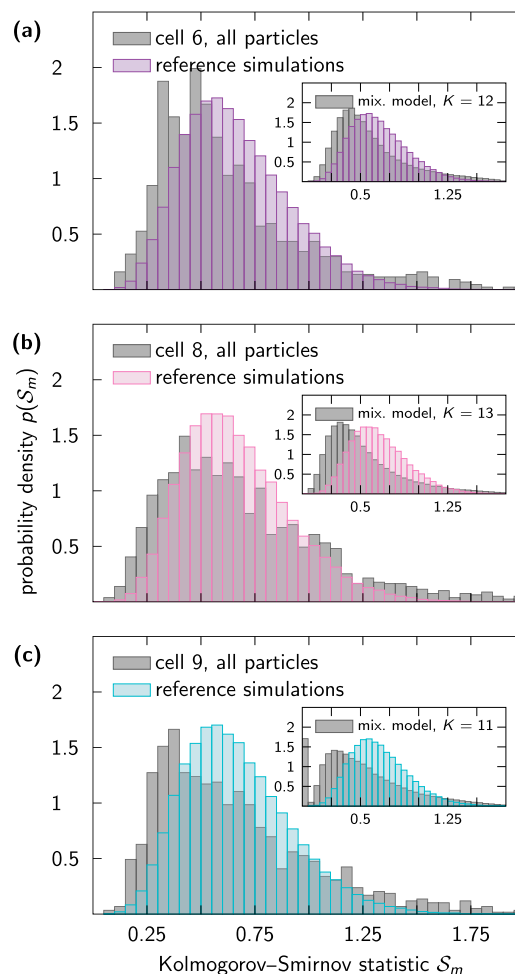
FIG. 7. Distributions of Kolmogorov–Smirnov statistics for the cells analyzed in Fig. 6. Using global single-population estimates for a^2 and σ^2, we estimated how likely it is for early parts of a trajectory to be governed by the same diffusion dynamics as its later parts via two-sample KS tests. The results for (a) cell 6, (b) cell 8, and (c) cell 9 (gray histograms) are consistent and show shifted weights toward smaller $\mathcal{S}_m$-values and broader tails than the corresponding reference distributions (colored histograms). Insets: Analogous plots for Brownian dynamics simulation data generated using the mixture model parameters in Figs. 6(b)–6(d) (see also Table I). We considered 10^5 trajectories with lengths distributed in the same way as in the experiments.

exclude the possibility that the experimental data originate from an underlying intricate PDD profile.

Note that there appears to be a weak negative correlation between the average trajectory length and the diffusion coefficient associated with each subpopulation. This means that the trajectories of fast diffusing populations are generally shorter than the trajectories of slow diffusing populations. We can rule out effects due to the trajectory-length bias of the MLEs (see Fig. 1) because there is virtually no bias in global estimates of trajectories of length $N + 1 \geq 5$ and the experimental datasets contain exclusively trajectories of length 8 or greater. The correlation is possibly
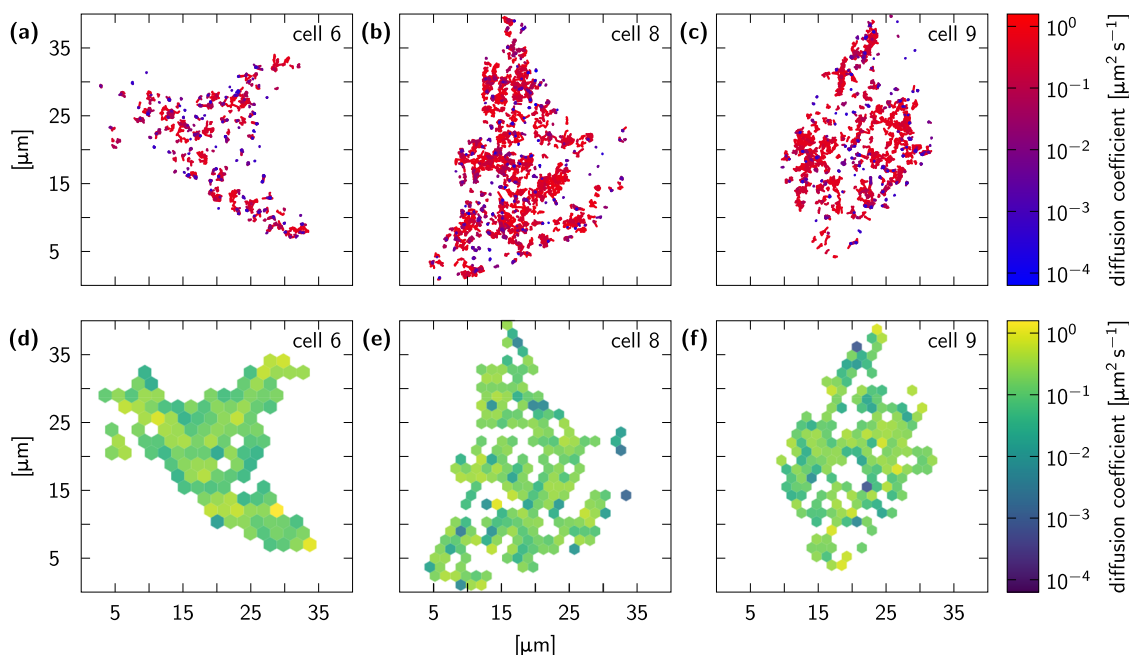
FIG. 8. Spatial distribution of fast and slowly diffusing MET receptor tyrosine kinases. Using the classification coefficients [Eq. (27)], we assign each trajectory to an appropriate subpopulation. Depending on their diffusion coefficients (see also Table I), the subpopulations have been color-coded from blue (slow) to red (fast). As in Fig. 6, we focus on the three cells whose diffusion coefficient distributions are well described with the mixture model, namely, (a) cell 6, (b) cell 8, and (c) cell 9. [(d)–(f)] Diffusivity landscapes inferred from mobile and immobile particles using the software platform *TRamWAy*. The localization uncertainty was estimated as $a \approx 0.05~\mu m$ from the values in Table I, and the results seemed independent of the value chosen for the regularization factor. Note that the resolution of the diffusivity landscape is lower than the trajectory representation in (a)–(c), which also results in the more moderate diffusion coefficient values.

explained by the fact that tracking algorithms have difficulties stitching together trajectories of fast diffusing particles with large displacements between frames and are therefore more likely to split them up into shorter trajectories.

Finally, we want to demonstrate how the information encoded in the classification coefficients $T_{k,m}$ can be employed, such as specifically picking out trajectories belonging to a certain subpopulation for further analysis. Figures 8(a)–8(c) visualize the trajectories of mobile particles in cells 6, 8, and 9. We indicate increasing receptor diffusivity with a color code from blue to red. The resulting map of confined blue and spread-out red trajectories is reminiscent of the diffusivity landscapes mapped out by Gaussian propagators in Refs. 20 and 23. Motivated by this observation, we applied the *TRamWAy* open-source software platform for analyzing single biomolecule dynamics[23] to the tracking data of our best-behaving cells. We thereby relied on the diffusion-only (`standard.d`) inference mode with the hyperparameter of the diffusivity prior (`diffusivity_prior`) and the minimum number of assigned (trans-)locations (`min_location_count`) both set to 1 and a nearest neighbor assignment by a count (`from_nearest_neighbors`) of 10. The localization precision (`localization_precision`) was chosen as the geometric average of the a^2-values reported in Table I, giving 0.05. The results are depicted in Figs. 8(d)–8(f). All in all, the inferred diffusivity landscapes do not have the same resolution as Figs. 8(a)–8(c), where immediately adjacent trajectories can have vastly different diffusion coefficients, and therefore do not cover the same range of

diffusivity as the mixture model does. Furthermore, *TRamWAy* does not consider a^2 as a parameter to infer, but a constant specified by the user. Whereas it could be insightful to analyze the data in Fig. 8 also in the context of corralled diffusion,[49] we leave a more definitive analysis of the spatial patterning of cell-surface receptor diffusion for future studies.

IV. OUTLOOK: TOWARD BAYESIAN INFERENCE

We want to point out that both the single-population log-likelihood [Eq. (14)] and the mixture log-likelihood [Eq. (28)] can be paired with priors $\Pi(a^2, \sigma^2)$ to construct posterior distributions via Bayes' theorem. This affects not only the numerical values of the parameter estimates but also their interpretation because in a Bayesian setting, the model parameters are assumed to be randomly distributed. A Bayesian prior can, e.g., be used to allow for variations in the localization uncertainty a^2 for each recorded particle individually.

The priors can be chosen as either (weakly) informative, where definite information about a parameter is taken into consideration, or uninformative. We expect informative prior information (if any) to exist for a^2, which can often be estimated empirically, e.g., via auxiliary experiments involving immobilized labeled particles. Its estimated mean a_{emp}^2 and uncertainty δa_{emp}^2 can be incorporated into an informative prior, such as

$$\Pi(a^2) \propto \exp\left(-\frac{(a^2 - a_{emp}^2)^2}{2(\delta a_{emp}^2)^2}\right),$$

which then enters the joint prior $\Pi(a^2, \sigma^2)$ as follows:

$$\Pi(a^2, \sigma^2) = \Pi(\sigma^2 \mid a^2)\Pi(a^2).$$

The remaining prior $\Pi(\sigma^2 \mid a^2)$ should be chosen as uninformative if no substantial information on σ^2 is available. The Jeffreys prior[50] is a classic uninformative prior, which is proportional to the square root of (the determinant of) the Fisher information and therefore invariant under coordinate transformations. Using the coordinates a^2 and ϕ, the Jeffreys prior for a fixed a^2 corresponds to the square root of the $I_{2,2}$-element of the matrix $\mathbf{I}(a^2, \phi)$ (see Appendix C), namely,

$$\Pi(\phi \mid a^2) = \sqrt{-\frac{d}{2}\sum_{m=1}^{M}\frac{d^2 \ln(|\bar{\Sigma}_m|)}{d\phi^2}} \equiv \Pi(\phi),$$

and is, in fact, independent of a^2. Alternatively, one can assume a uniform prior for ϕ, i.e.,

$$\Pi(\phi \mid a^2) = \text{const}.$$

If no information is available on any of the two parameters, the priors can be chosen to either differ between parameters or be of the same type. The two-parameter Jeffreys prior is given by

$$\Pi(a^2, \phi) = \sqrt{\det(\mathbf{I}(a^2, \phi))},$$

where an analytic expression for the Fisher information matrix $\mathbf{I}(a^2, \phi)$ can be found in Appendix C. While it is usually not recommended to use Jeffreys prior on anything other than single-parameter models, it turns out that for the present problem, the preferred alternative, the so-called reference prior,[51,52] coincides with the bivariate Jeffreys prior. This might be due to the fact that the Jeffreys prior becomes separable when treated in the (a^2, ϕ)-coordinates or because both parameters are scale parameters.

Bayesian estimates are, in comparison to MLEs, often more sharply distributed and thus have smaller uncertainties. However, poorly chosen priors introduce biases. For uninformative priors, the bias vanishes as the sample size goes to infinity, but different priors result in different biases. In addition, the specific choice of the estimator affects the size of the bias and the way it scales with the sample size. We therefore recommend that practitioners gauge which uninformative priors work best with their estimator of choice.

V. CONCLUSIONS

In this paper, we have derived a numerically efficient maximum likelihood estimator to extract diffusion coefficients from single-molecule tracking experiments subject to static noise and dynamic motion blur. To estimate diffusion coefficients from molecular dynamics simulations, one sets the blur coefficient to $B = 0$. The estimator is based on the statistics of trajectory increments in real space and therefore complements the theory presented in Refs. 9 and 15, where an estimator operating in discrete sine space was developed. Fourier representations are computationally efficient and naturally lead, e.g., to a power spectral analysis of the diffusion process.[14] Our approach has the benefit of delivering closed-form expressions for some quantities, such as the accompanying Fisher information

matrix. Compared to other diffusion coefficient estimators, the maximum likelihood estimator has several advantages, e.g., it allows for the inclusion of prior knowledge in a Bayesian manner, makes the error analysis for trajectories of different lengths straightforward, and naturally leads to a quality factor analysis, which helps us to assess whether the data satisfy the underlying assumptions of the diffusion model or not.

In the case of heterogeneous data, the theory can be extended to a mixture model, where the overall numerical efficiency is retained for each subpopulation. Minimizing the corresponding negative log-likelihood is then achieved with the help of the expectation–maximization algorithm. To demonstrate its applicability and effectiveness, we first considered sparse quantities of non-diffusive synthetic data to test whether the theory was able to detect any deviations from regular diffusion. Our results highlight the difficulty in distinguishing between diffusive and non-diffusive processes, especially when the trajectories are overwhelmingly short and corrupted by static noise and dynamic motion blur. We then used the framework to analyze single-molecule tracking data for "resting" MET receptor tyrosine kinase, recorded in ten different cells.[30] As the fit gradually got better with the increasing number of subpopulations K, we relied on the distribution of quality factors, which are a measure of how well a dataset conforms to the assumptions of the underlying diffusion model, to determine the optimal value of K. Although our results varied significantly between the cells, we found solutions for four of the ten cells, where the null hypothesis that the tracking data originated from a mixture model could not be rejected. All these fits involved fairly large K-values, so we ruled out models containing only a few subpopulations. However, the effective number of subpopulations can often be reduced, e.g., by neglecting sparsely represented subpopulations or by lumping together subpopulations with similar diffusion coefficients. Overall qualitative agreement between the experimental data and simulations suggests that the broad distributions of diffusion coefficients observed in the experiment are a result of heterogeneous mixtures, although position-dependent diffusion as an underlying process cannot be ruled out. In fact, for short trajectories that probe only local regions, position dependence and heterogeneity in diffusion are intertwined.

We believe that our results provide practitioners of single-molecule tracking techniques with valuable tools to analyze their data. To facilitate their use, we have implemented the theoretical formulations in a Julia data analysis package.[31] The general formulation and efficient evaluation of the likelihood functions should also pave the way for more elaborate analysis methods in the future, e.g., involving Bayesian statistics.

SUPPLEMENTARY MATERIAL

See the supplementary material for detailed information on the trajectory-length distribution for the ten cells analyzed, the model parameters and their uncertainties from fits with different numbers K of subpopulations, and the results of the quality factor analysis in terms of the Kuiper statistic κ and the associated p-value.

ACKNOWLEDGMENTS

providing us with the experimental tracking data. This research was supported by the Max Planck Society and by the LOEWE (Landes-Offensive zur Entwicklung Wissenschaftlich-ökonomischer Exzellenz) program of the state of Hesse conducted within the framework of the MegaSyn Research Cluster.

APPENDIX A: ANALYTIC CALCULATION OF DETERMINANT

Consider an $N \times N$ tridiagonal matrix with main, super-, and sub-diagonal elements $\{\alpha_1, \alpha_2, \ldots, \alpha_N\}$, $\{\beta_1, \beta_2, \ldots, \beta_{N-1}\}$, and $\{\gamma_1, \gamma_2, \ldots, \gamma_{N-1}\}$, respectively. Its determinant can be computed using the following three-term recurrence relation:[33]

$$f_n = \alpha_n f_{n-1} - \beta_{n-1}\gamma_{n-1}f_{n-2},$$
$$f_{-1} = 0, f_0 = 1. \tag{A1}$$

The covariance matrix $\boldsymbol{\Sigma}$ of the main text [Eq. (7)] is not only tridiagonal but also a symmetric Toeplitz matrix, which is why we have $\alpha_i = \alpha$ and $\gamma_i = \beta_i = \beta \forall i$. Multiplying both sides of Eq. (A1) with z^n and summing over n results in the generating function

$$w(z) = \sum_{n=0}^{\infty} f_n z^n = \frac{1}{1 - \alpha z + \beta^2 z^2} \tag{A2}$$

whose coefficients f_n satisfy Eq. (A1). The denominator of Eq. (A2) can be factored, and the resulting expression is rewritten as follows via partial fraction decomposition:

$$w(z) = \frac{1}{r_+ - r_-}\left[\frac{r_+}{1 - r_+ z} - \frac{r_-}{1 - r_- z}\right],$$

where $r_{\pm} = (\alpha \pm \sqrt{\alpha^2 - 4\beta^2})/2$. Each term can be further expanded into a geometric series of z to finally reveal that f_n has the functional form

$$f_n = \alpha^n \frac{(1+q)^{n+1} - (1-q)^{n+1}}{2^{n+1}q}, \tag{A3}$$

with $q = \sqrt{1 - 4\beta^2/\alpha^2}$. The determinant of an $N \times N$ symmetric tridiagonal Toeplitz matrix is given by f_N, which finally results in Eq. (12) of the main text.

APPENDIX B: POSITIVE DEFINITENESS OF THE COVARIANCE MATRICES

An $N \times N$ matrix $\mathbf{M}$ is said to be positive definite if $\vec{x}^T\mathbf{M}\vec{x} > 0 \ \forall \vec{x} \in \mathbb{R}^N\backslash\vec{0}$, which implies that all of its eigenvalues λ_n must be strictly positive. The eigenvalues of a symmetric tridiagonal Toeplitz matrix with diagonal and off-diagonal elements $\alpha, \beta \in \mathbb{R}$, respectively, are given by

$$\lambda_n = \alpha - 2|\beta|\cos\left(\frac{n\pi}{N+1}\right) > \alpha - 2|\beta|$$

for $n = 1, 2, \ldots, N$. The covariance matrices are therefore positive definite if they are diagonally dominant, i.e., $\alpha - 2|\beta| \geq 0$. This is obviously the case for $\boldsymbol{\Sigma}'$ and $\boldsymbol{\Sigma}''$ [Eqs. (8) and (9)] because $0 \leq B \leq 1/4$ must hold. Furthermore, because of

$$1 + \phi(1 - 2B) - 2\left|-\frac{1}{2} + \phi B\right| = \begin{cases} 2 + (1 - 4B)\phi, & \phi > (2B)^{-1} \\ \phi, & \phi \leq (2B)^{-1}, \end{cases}$$

the matrices $\boldsymbol{\Sigma}$ and $\tilde{\boldsymbol{\Sigma}}$ are also positive definite.

Finally, it should be mentioned that the inverse of a positive definite (nonsingular) symmetric matrix $\mathbf{M}$ is also positive definite because its eigenvalues are the reciprocals of those of the original matrix. These results guarantee the positivity of $\vec{\Delta}^T\boldsymbol{\Sigma}'^{-1}\vec{\Delta}$, $\vec{\Delta}^T\boldsymbol{\Sigma}''^{-1}\vec{\Delta}$, and $\vec{\Delta}^T\tilde{\boldsymbol{\Sigma}}^{-1}\vec{\Delta}$ and the associated parameter estimates.

APPENDIX C: THE FISHER INFORMATION AND ERROR ESTIMATES

By definition, the Fisher information matrix

$$\mathbf{I}(a^2, \phi) = \begin{pmatrix} I_{1,1} & I_{1,2} \\ I_{2,1} & I_{2,2} \end{pmatrix}$$

equals the ensemble-averaged Hessian of $\mathcal{L}(\{\vec{\Delta}_m\} \mid a^2, \phi)$ [Eq. (17)] with the following elements:

$$I_{1,1} = \left\langle \frac{\partial^2 \mathcal{L}}{(\partial a^2)^2}\right\rangle, \quad I_{2,2} = \left\langle \frac{\partial^2 \mathcal{L}}{\partial \phi^2}\right\rangle,$$
$$I_{1,2} = I_{2,1} = \left\langle \frac{\partial^2 \mathcal{L}}{\partial a^2 \partial \phi}\right\rangle.$$

On the boundary, the Fisher information is a scalar function given by

$$I(\theta) = \left\langle \frac{\partial^2 \mathcal{L}(\{\vec{\Delta}_m\} \mid \theta)}{\partial \theta^2}\right\rangle,$$

where $\mathcal{L}(\{\vec{\Delta}_m\} \mid \theta)$ is defined as $\mathcal{L}(\{\vec{\Delta}_m\} \mid a^2, \sigma^2)$ of Eq. (14) with one of the two parameters set to zero and $\theta \in \{a^2, \sigma^2\}$ denotes the remaining non-zero parameter.

Considering first the scalar case, we have $\boldsymbol{\Sigma} = \theta\boldsymbol{\Sigma}_\theta$, with $\boldsymbol{\Sigma}_{a^2} = \boldsymbol{\Sigma}'$ and $\boldsymbol{\Sigma}_{\sigma^2} = \boldsymbol{\Sigma}''$, and therefore,

$$\frac{\partial^2 \mathcal{L}}{\partial \theta^2} = \frac{1}{\theta^3}\sum_{m=1}^{M}\sum_{n=1}^{d}\vec{\Delta}_{m,n}^T\left(\boldsymbol{\Sigma}_\theta^{-1}\right)_m \vec{\Delta}_{m,n} - \frac{d\mathcal{N}_M}{2\theta^2}.$$

Ensemble averages only affect terms dependent on $\vec{\Delta}$, which are, in this case, all of quadratic form and can therefore be computed using $\langle \Delta_i \Delta_j \rangle = \Sigma_{i,j} = \theta (\Sigma_\theta)_{i,j}$. When evaluating the scalar Fisher information, we encounter a sum of terms of the form

$$\left\langle \vec{\Delta}^T\boldsymbol{\Sigma}_\theta^{-1}\vec{\Delta}\right\rangle = \sum_{i,j=1}^{N}(\Sigma_\theta^{-1})_{i,j}\langle\Delta_i\Delta_j\rangle = \theta\sum_{i,j=1}^{N}\delta_{i,j} = \theta N,$$

which simply reduces $I(\theta)$ to

$$I(\theta) = \frac{1}{\theta^2}\sum_{m=1}^{M}\sum_{n=1}^{d}N_m - \frac{d\mathcal{N}_M}{2\theta^2} = \frac{d\mathcal{N}_M}{2\theta^2}$$

and thus gives rise to the following lower bound for the standard error $\delta\theta = \sqrt{\text{var}(\theta)}$:

$$\delta\theta \geq \sqrt{\frac{1}{I(\theta)}} = \theta\sqrt{\frac{2}{d\mathcal{N}_M}}.$$

The general case makes use of the Hessian 2×2 matrix whose components are given by

$$\frac{\partial^2 \mathcal{L}}{(\partial a^2)^2} = \frac{1}{(a^2)^3} \sum_{m=1}^{M} \sum_{n=1}^{d} \vec{\Delta}_{m,n}^T \tilde{\mathbf{\Sigma}}_m^{-1} \vec{\Delta}_{m,n} - \frac{d\mathcal{N}_M}{2(a^2)^2},$$

$$\frac{\partial^2 \mathcal{L}}{\partial a^2 \partial \phi} = \frac{1}{2(a^2)^2} \sum_{m=1}^{M} \sum_{n=1}^{d} \vec{\Delta}_{m,n}^T \tilde{\mathbf{\Sigma}}_m^{-1} \mathbf{\Sigma}_m'' \tilde{\mathbf{\Sigma}}_m^{-1} \vec{\Delta}_{m,n},$$

$$\frac{\partial^2 \mathcal{L}}{\partial \phi^2} = \frac{1}{a^2} \sum_{m=1}^{M} \sum_{n=1}^{d} \vec{\Delta}_{m,n}^T \tilde{\mathbf{\Sigma}}_m^{-1} \mathbf{\Sigma}_m'' \tilde{\mathbf{\Sigma}}_m^{-1} \mathbf{\Sigma}_m'' \tilde{\mathbf{\Sigma}}_m^{-1} \vec{\Delta}_{m,n}$$
$$+ \frac{d}{2} \sum_{m=1}^{M} \frac{d^2 \ln(|\tilde{\mathbf{\Sigma}}_m|)}{d\phi^2},$$

according to the identity $d\mathbf{A}^{-1}/d\theta = -\mathbf{A}^{-1}(d\mathbf{A}/d\theta)\mathbf{A}^{-1}$. The first diagonal element of the Fisher matrix is calculated in an analogous way to the scalar Fisher information above, giving

$$I_{1,1} = \frac{d\mathcal{N}_M}{2(a^2)^2}.$$

For the summands of the non-diagonal elements, we find that

$$\left\langle \vec{\Delta}^T \tilde{\mathbf{\Sigma}}^{-1} \mathbf{\Sigma}'' \tilde{\mathbf{\Sigma}}^{-1} \vec{\Delta} \right\rangle = \sum_{i,j,k,l=1}^{N} \tilde{\Sigma}_{i,j}^{-1} \Sigma_{j,k}'' \tilde{\Sigma}_{k,l}^{-1} \langle \Delta_i \Delta_l \rangle$$
$$= a^2 \sum_{i,j=1}^{N} \tilde{\Sigma}_{i,j}^{-1} \Sigma_{j,i}'' = a^2 \mathrm{tr}\left(\tilde{\mathbf{\Sigma}}^{-1} \mathbf{\Sigma}'' \right).$$

With the help of $\mathbf{\Sigma}'' = d\tilde{\mathbf{\Sigma}}/d\phi$ and Jacobi's formula, which states that

$$\frac{d|\mathbf{A}|}{d\theta} = |\mathbf{A}| \mathrm{tr}\left(\mathbf{A}^{-1} \frac{d\mathbf{A}}{d\theta} \right)$$

must hold for an invertible matrix $\mathbf{A}$, we can rewrite the trace as follows:

$$\mathrm{tr}(\tilde{\mathbf{\Sigma}}^{-1} \mathbf{\Sigma}'') = \mathrm{tr}\left(\tilde{\mathbf{\Sigma}}^{-1} \frac{d\tilde{\mathbf{\Sigma}}}{d\phi} \right) = \frac{1}{|\tilde{\mathbf{\Sigma}}|} \frac{d|\tilde{\mathbf{\Sigma}}|}{d\phi} = \frac{d \ln(|\tilde{\mathbf{\Sigma}}|)}{d\phi}.$$

We therefore get

$$I_{1,2} = I_{2,1} = \frac{d}{2a^2} \sum_{m=1}^{M} \frac{d \ln(|\tilde{\mathbf{\Sigma}}_m|)}{d\phi}.$$

The last diagonal element of the Fisher matrix can be computed with the help of

$$\left\langle \vec{\Delta}^T \tilde{\mathbf{\Sigma}}^{-1} \mathbf{\Sigma}'' \tilde{\mathbf{\Sigma}}^{-1} \mathbf{\Sigma}'' \tilde{\mathbf{\Sigma}}^{-1} \vec{\Delta} \right\rangle$$
$$= \sum_{i,j,k,l,m,n=1}^{N} \tilde{\Sigma}_{i,j}^{-1} \Sigma_{j,k}'' \tilde{\Sigma}_{k,l}^{-1} \Sigma_{l,m}'' \tilde{\Sigma}_{m,n}^{-1} \langle \Delta_i \Delta_n \rangle$$
$$= a^2 \sum_{i,j,k,l=1}^{N} \tilde{\Sigma}_{i,j}^{-1} \Sigma_{j,k}'' \tilde{\Sigma}_{k,l}^{-1} \Sigma_{l,i}''$$
$$= a^2 \mathrm{tr}\left(\tilde{\mathbf{\Sigma}}^{-1} \mathbf{\Sigma}'' \tilde{\mathbf{\Sigma}}^{-1} \mathbf{\Sigma}'' \right),$$

where the identity $d \, \mathrm{tr}(\mathbf{A})/d\theta = \mathrm{tr}(d\mathbf{A}/d\theta)$ can be used to rewrite the trace in terms of derivatives, giving

$$\mathrm{tr}\left(\tilde{\mathbf{\Sigma}}^{-1} \frac{d\tilde{\mathbf{\Sigma}}}{d\phi} \tilde{\mathbf{\Sigma}}^{-1} \frac{d\tilde{\mathbf{\Sigma}}}{d\phi} \right) = -\frac{d}{d\phi} \mathrm{tr}\left(\tilde{\mathbf{\Sigma}}^{-1} \frac{d\tilde{\mathbf{\Sigma}}}{d\phi} \right)$$
$$= -\frac{d^2 \ln(|\tilde{\mathbf{\Sigma}}|)}{d\phi^2}.$$

The element $I_{2,2}$ thus takes the form

$$I_{2,2} = -\frac{d}{2} \sum_{m=1}^{M} \frac{d^2 \ln(|\tilde{\mathbf{\Sigma}}_m|)}{d\phi^2}.$$

The derivatives of $\ln(|\tilde{\mathbf{\Sigma}}_m|)$ read

$$\frac{d \ln(|\tilde{\mathbf{\Sigma}}_m|)}{d\phi} = \frac{N_m}{\tilde{\alpha}} (1 - 2B_m) - F(\tilde{q}),$$

$$\frac{d^2 \ln(|\tilde{\mathbf{\Sigma}}_m|)}{d\phi^2} = -\frac{N_m}{\tilde{\alpha}^2} (1 - 2B_m)^2 - \frac{dF(\tilde{q})}{d\phi},$$

where $\tilde{q} = (1 - 4\tilde{\beta}^2/\tilde{\alpha}^2)^{1/2} \geq 0$, $\tilde{\alpha} = 1 + \phi(1 - 2B_m)$, and $\tilde{\beta} = -1/2 + \phi B_m$. Here, we introduced the functions

$$F(\tilde{q}) = \frac{2\tilde{\beta}\left(N_m \tilde{q} - 1 + (N_m + 1)[\tilde{\alpha}(1 - \tilde{q})/2]^{N_m}/|\tilde{\mathbf{\Sigma}}_m| \right)}{\tilde{\alpha}^3 \tilde{q}^2 (1 + \tilde{q})},$$

$$F(0) = \lim_{\tilde{q} \to 0} F(\tilde{q}) = \frac{2\tilde{\beta} N_m (N_m - 1)}{3\tilde{\alpha}^3}$$

and its gradients

$$\frac{dF(\tilde{q})}{d\phi} = -\frac{2}{\phi}\left[F(\tilde{q}) - C_1 F(\tilde{q}) + C_2 F(0) - \frac{\phi}{2} F(\tilde{q})^2 \right],$$

$$C_1 = \frac{1}{4\tilde{\beta}} + \frac{3}{2\tilde{\alpha}} + \frac{\tilde{\beta}\phi}{\tilde{\alpha}^3}\left[\frac{3}{\tilde{q}^2} + \frac{\tilde{\alpha}^2 N_m}{2\tilde{\beta}^2} \right], C_2 = \frac{3\phi}{4\tilde{\alpha}\tilde{\beta}\tilde{q}^2},$$

to condense our expressions.

In the general case, the standard errors emerge from the diagonal elements of the inverse Fisher information matrix $\mathbf{I}^{-1}(a^2, \sigma^2)$. This inverse is related to $\mathbf{I}^{-1}(a^2, \phi)$, which is simply given by

$$\mathbf{I}^{-1}(a^2, \phi) = \frac{1}{I_{1,1} I_{2,2} - I_{1,2}^2} \begin{pmatrix} I_{2,2} & -I_{1,2} \\ -I_{1,2} & I_{1,1} \end{pmatrix}.$$

A coordinate transformation then gives us

$$\mathbf{I}^{-1}(a^2, \sigma^2) = \mathbf{J}^{-1} \mathbf{I}^{-1}(a^2, \phi) (\mathbf{J}^{-1})^T \big|_{\phi = \sigma^2/a^2},$$

with an inverse Jacobian

$$\mathbf{J}^{-1} = \begin{pmatrix} \frac{\partial a^2}{\partial a^2} & \frac{\partial a^2}{\partial \sigma^2} \\ \frac{\partial \phi}{\partial a^2} & \frac{\partial \phi}{\partial \sigma^2} \end{pmatrix}^{-1} = \begin{pmatrix} 1 & 0 \\ \phi & a^2 \end{pmatrix}.$$

This gives rise to Eq. (20) of the main text.

APPENDIX D: NUMERICS

1. Brownian dynamics simulations

Interpreting Eqs. (2)–(4) in the sense of Itô, we discretize them as follows for $\Delta t = N\delta t$:

$$
\begin{aligned}
Y_{i+1} &= Y_i + \frac{\sigma}{\sqrt{N}} R'_i, \\
Z_i &= \sum_{n=1}^{N} \delta t\, s_n Y_{n+iN}, \qquad\qquad \text{(D1)} \\
X_i &= Z_i + \frac{a}{\sqrt{2}} R_i.
\end{aligned}
$$

Here, R_i and R'_i denote independent, normal distributed random variables with zero mean and unit variance, and the s_n are normalized, i.e.,

$$
\sum_{n=1}^{N} \delta t\, s_n = 1.
$$

In the case of position-dependent diffusion, the resulting multiplicative noise produces a spurious drift term,[44] so the first line in Eq. (D1) must be replaced with

$$
Y_{i+1} = Y_i + \frac{\sigma(Y_i)\sigma'(Y_i)}{N} + \frac{\sigma(Y_i)}{\sqrt{N}} R'_i.
$$

This is the discretized version of Eq. (36) of the main text.

In this paper, we exclusively chose $N = 100$ and $s_n = (N\delta t)^{-1}$. All two-dimensional trajectories were generated from two one-dimensional trajectories of equal length under the assumption of isotropic diffusion.

2. Implementation of EM algorithm

Starting from $P_k^{(0)} = 1/K \ \forall k$ and a set $\{a_k^{2\,(0)}, \sigma_k^{2\,(0)}\}_{k=1,\dots,K}$ of the initial parameter values, chosen randomly on a logarithmic scale, the classification coefficients $T_{k,m}^{(0)}$ are initialized according to Eq. (27). We then proceed to update the parameters via Eq. (29) and either Eqs. (30) and (31) or Eqs. (32) and (33), depending on which solution results in the lowest negative log-likelihood value [Eq. (28)]

$$
L^{(i)} = L\left(\left\{\vec{\Delta}_{m,n}\right\} \mid \left\{P_k^{(i)}, a_k^{2\,(i)}, \sigma_k^{2\,(i)}\right\}\right).
$$

The two steps of the EM algorithm are repeated for a fixed number N_{local} of iteration steps or broken off prematurely if the following inequality is satisfied for the tolerance threshold ε:

$$
\frac{L^{(i)} - L^{(i-1)}}{\mathcal{N}_M} < \varepsilon.
$$

The corresponding solution is saved in the "high score," and the algorithm is reinitialized with a new set of initial parameters. The "high score" solution is replaced by the new iteration if the latter results in a lower negative log-likelihood. The EM algorithm is run in total for N_{global} times.

We used $N_{\text{local}} = 500$ and $N_{\text{global}} = 50$ with a tolerance threshold of $\varepsilon = 10^{-10}$ when analyzing the data in Figs. 2, 4, and 6.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request. Requests for the tracking data should be directed to Professor Dr. Mike Heilemann, corresponding author of the original publication.

REFERENCES

[1]J. Gelles, B. J. Schnapp, and M. P. Sheetz, "Tracking kinesin-driven movements with nanometre-scale precision," Nature **331**, 450–453 (1988).

[2]A. Yildiz, J. N. Forkey, S. A. McKinney, T. Ha, Y. E. Goldman, and P. R. Selvin, "Myosin V walks hand-over-hand: Single fluorophore imaging with 1.5-nm localization," Science **300**, 2061–2065 (2003).

[3]J. C. M. Gebhardt, D. M. Suter, R. Roy, Z. W. Zhao, A. R. Chapman, S. Basu, T. Maniatis, and X. S. Xie, "Single-molecule imaging of transcription factor binding to DNA in live mammalian cells," Nat. Methods **10**, 421–426 (2013).

[4]M. J. Saxton and K. Jacobson, "Single-particle tracking: Applications to membrane dynamics," Annu. Rev. Biophys. Biomol. Struct. **26**, 373–399 (1997).

[5]V. Levi and E. Gratton, "Exploring dynamics in living cells by tracking single particles," Cell Biochem. Biophys. **48**, 1–15 (2007).

[6]N. Bobroff, "Position measurement with a resolution and noise-limited instrument," Rev. Sci. Instrum. **57**, 1152–1157 (1986).

[7]D. S. Martin, M. B. Forstner, and J. A. Käs, "Apparent subdiffusion inherent to single particle tracking," Biophys. J. **83**, 2109–2117 (2002).

[8]T. Savin and P. S. Doyle, "Static and dynamic errors in particle tracking microrheology," Biophys. J. **88**, 623–638 (2005).

[9]A. J. Berglund, "Statistics of camera-based single-particle tracking," Phys. Rev. E **82**, 011917 (2010).

[10]H. Qian, M. P. Sheetz, and E. L. Elson, "Single particle tracking. Analysis of diffusion and flow in two-dimensional systems," Biophys. J. **60**, 910–921 (1991).

[11]S. Wieser and G. J. Schütz, "Tracking single molecules in the live cell plasma membrane—Do's and don't's," Methods **46**, 131–140 (2008).

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

[13]J. T. Bullerjahn, S. von Bülow, and G. Hummer, "Optimal estimates of self-diffusion coefficients from molecular dynamics simulations," J. Chem. Phys. **153**, 024116 (2020).

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

[15]X. Michalet and A. J. Berglund, "Optimal diffusion coefficient estimation in single-particle tracking," Phys. Rev. E **85**, 061916 (2012).

[16]C. L. Vestergaard, "Optimizing experimental parameters for tracking of diffusing particles," Phys. Rev. E **94**, 022401 (2016).

[17]G. Hummer, "Position-dependent diffusion coefficients and free energies from Bayesian analysis of equilibrium and replica molecular dynamics simulations," New J. Phys. **7**, 34 (2005).

[18]J.-B. Masson, D. Casanova, S. Türkcan, G. Voisinne, M. R. Popoff, M. Vergassola, and A. Alexandrou, "Inferring maps of forces inside cell membrane microdomains," Phys. Rev. Lett. **102**, 048103 (2009).

[19]D. Kleinhans, "Estimation of drift and diffusion functions from time series data: A maximum likelihood framework," Phys. Rev. E **85**, 026705 (2012).

[20]M. El Beheiry, M. Dahan, and J.-B. Masson, "InferenceMAP: Mapping of single-molecule dynamics with Bayesian inference," Nat. Methods **12**, 594–595 (2015).

[21]N. Hoze and D. Holcman, "Recovering a stochastic process from super-resolution noisy ensembles of single-particle trajectories," Phys. Rev. E **92**, 052109 (2015).

[22]J. Krog and M. A. Lomholt, "Bayesian inference with information content model check for Langevin equations," Phys. Rev. E **96**, 062106 (2017).

[23]F. Laurent, C. Floderer, C. Favard, D. Muriaux, J.-B. Masson, and C. L. Vestergaard, "Mapping spatio-temporal dynamics of single biomolecules in living cells," Phys. Biol. **17**, 015003 (2020).

[24] A. Frishman and P. Ronceray, "Learning force fields from stochastic trajectories," Phys. Rev. X **10**, 021009 (2020).

[25] F. Persson, M. Lindén, C. Unoson, and J. Elf, "Extracting intracellular diffusive states and transition rates from single-molecule tracking data," Nat. Methods **10**, 265–269 (2013).

[26] M. Ott, Y. Shai, and G. Haran, "Single-particle tracking reveals switching of the HIV fusion peptide between two diffusive modes in membranes," J. Phys. Chem. B **117**, 13308–13321 (2013).

[27] N. Monnier, Z. Barry, H. Y. Park, K.-C. Su, Z. Katz, B. P. English, A. Dey, K. Pan, I. M. Cheeseman, R. H. Singer, and M. Bathe, "Inferring transient particle transport dynamics in live cells," Nat. Methods **12**, 838–840 (2015).

[28] M. Lindén and J. Elf, "Variational algorithms for analyzing noisy multistate diffusion trajectories," Biophys. J. **115**, 276–282 (2018).

[29] R. C. Falcao and D. Coombs, "Diffusion analysis of single particle trajectories in a Bayesian nonparametrics framework," Phys. Biol. **17**, 025001 (2020).

[30] M.-L. I. E. Harwardt, P. Young, W. M. Bleymüller, T. Meyer, C. Karathanasis, H. H. Niemann, M. Heilemann, and M. S. Dietz, "Membrane dynamics of resting and internalin B-Bound MET receptor tyrosine kinase studied by single-molecule tracking," FEBS Open Bio **7**, 1422–1440 (2017).

[31] See https://github.com/bio-phys/DiffusionMLE for a Julia implementation of our results.

[32] A. Quarteroni, R. Sacco, and F. Saleri, *Numerical Mathematics* (Springer-Verlag, New York, 2000).

[33] M. E. A. El-Mikkawy, "On the inverse of a general tridiagonal matrix," Appl. Math. Comput. **150**, 669–679 (2004).

[34] F. Qi and A.-Q. Liu, "Alternative proofs of some formulas for two tridiagonal determinants," Acta. Univ. Sapientiae, Matem. **10**, 287–297 (2018).

[35] R. P. Brent, *Algorithms for Minimization Without Derivatives* (Prentice-Hall, Englewood Cliffs, NJ, 1973).

[36] J. Kiefer, "Sequential minimax search for a maximum," Proc. Am. Math. Soc. **4**, 502–506 (1953).

[37] P. K. Mogensen and A. N. Riseth, "Optim: A mathematical optimization package for Julia," J. Open Source Software **3**, 615 (2018).

[38] N. H. Kuiper, "Tests concerning random points on a circle," Nederl. Akad. Wetensch. Proc. Ser. A **63**, 38–47 (1960).

[39] M. Tygert, "Statistical tests for whether a given set of independent, identically distributed draws comes from a specified probability density," Proc. Natl. Acad. Sci. U. S. A. **107**, 16471–16476 (2010).

[40] H. Arsham, "Kuiper's *P*-value as a measuring tool and decision procedure for the goodness-of-fit test," J. Appl. Stat. **15**, 131–135 (1988).

[41] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the *EM* algorithm," J. R. Stat. Soc. Series B Stat. Methodol. **39**, 1–22 (1977).

[42] C. Biernacki, G. Celeux, and G. Govaert, "Assessing a mixture model for clustering with the integrated completed likelihood," IEEE Trans. Pattern Anal. Mach. Intell. **22**, 719–725 (2000).

[43] B. B. Mandelbrot and J. W. van Ness, "Fractional Brownian motions, fractional noises and applications," SIAM Rev. **10**, 422–437 (1968).

[44] N. G. van Kampen, "Diffusion in inhomogeneous media," J. Phys. Chem. Solids **49**, 673–677 (1988).

[45] A. T. A. Wood and G. Chan, "Simulation of stationary Gaussian processes in $[0,1]^d$," J. Comput. Graphical Stat. **3**, 409–432 (1994).

[46] G. Giannone, E. Hosy, F. Levet, A. Constals, K. Schulze, A. I. Sobolevsky, M. P. Rosconi, E. Gouaux, R. Tampé, D. Choquet, and L. Cognet, "Dynamic super-resolution imaging of endogenous proteins on living cells at ultra-high density," Biophys. J. **99**, 1303–1310 (2010).

[47] N. Bag and T. Wohland, "Imaging fluorescence fluctuation spectroscopy: New tools for quantitative bioimaging," Annu. Rev. Phys. Chem. **65**, 225–248 (2014).

[48] G. J. Schütz, H. Schindler, and T. Schmidt, "Single-molecule microscopy on model membranes reveals anomalous diffusion," Biophys. J. **73**, 1073–1080 (1997).

[49] M. J. Saxton, "Single-particle tracking: Effects of corrals," Biophys. J. **69**, 389–398 (1995).

[50] H. Jeffreys, "An invariant form for the prior probability in estimation problems," Proc. R. Soc. London, Ser. A **186**, 453–461 (1946).

[51] J. M. Bernardo, "Reference posterior distributions for Bayesian inference," J. R. Stat. Soc.: Ser. B **41**, 113 (1979).

[52] J. O. Berger, J. M. Bernardo, and D. Sun, "The formal definition of reference priors," Ann. Stat. **37**, 905–938 (2009).