Detecting directed motion and confinement in single-particle trajectories using hidden variables

François Simon*[1], Guillaume Ramadier[1], Inès Fonquernie[1,2], Janka Zsok[3], Sergiy Patskovsky[1], Michel Meunier[1], Caroline Boudoux[1], Elisa Dultz[3], Lucien E. Weiss*[1]

[1]Department of Engineering Physics, Polytechnique Montréal, Montréal, Québec, Canada; [2]École Polytechnique, Palaiseau, France; [3]Institute of Biochemistry, ETH Zürich, Switzerland

Correspondence: simon.francois@protonmail.com; lucien.weiss@polymtl.ca

Abstract

Single-particle tracking is a powerful tool for understanding protein dynamics and characterizing microenvironments. As the motion of unconstrained nanoscale particles is governed by Brownian diffusion, deviations from this behavior are biophysically insightful. However, the stochastic nature of particle movement and the presence of localization error pose a challenge for the robust classification of non-Brownian motion. Here, we present *aTrack*, a versatile tool for classifying track behaviors and extracting key parameters for particles undergoing Brownian, confined, or directed motion. Our tool quickly and accurately estimates motion parameters from individual tracks and determines their likely motion state. Further, our tool can analyze populations of tracks and determine the most likely number of motion states. We determine the working range of our approach on simulated tracks and demonstrate its application for characterizing particle motion in cells and for biosensing applications. Our tool is implemented as a stand-alone software package, making it simple to analyze tracking data.

Main

Single-particle tracking (SPT) is a powerful tool for characterizing protein activity (1–3). While Brownian diffusion is commonly observed inside cells, deviations from this behavior are of major interest and can provide key biophysical insight (4–6). Classically, non-Brownian behaviors have been identified by fitting a power law to the mean-squared displacement (MSD) curve, (MSD $= \Gamma \cdot t^{\alpha}$) (6), where subdiffusive motion has an α exponent less than 1 and superdiffusion has an exponent > 1. While fitting the MSD can be used for classifying motion, the finite number of measurements from stochastically moving particles creates uncertainty in the estimated parameters (7–9). This motivates the application of statistical tests to estimate the certainty of fitted parameters, such as the exponent α (7, 10, 11). Even so, fitting the MSD curve to a power law does not specify a particular underlying motion model. To extract further biophysical insight, various mathematical models have been proposed to describe the anomalous diffusion behaviors observed in measurements (4, 12–15). These models, in turn, have been used to train machine learning methods designed to select the most appropriate model and parameters (16–18).

Probabilistic models can also be used to describe the distribution of observed behaviors. In brief, these probabilistic models work by mathematically defining the probability of an observed track as a function of hidden variables. Next, the probability of observing a given track given the model is calculated by integrating over the possible values for the hidden variables. Finally, Maximum Likelihood Estimation (MLE) is used to determine the underlying parameter values. The integration step can be computationally expensive; however, when an analytical integration formula is used, this step becomes trivial (19–21). This framework can be used to model a variety of behaviors with informative hidden parameters, including diffusion measurements with finite localization precision (21), directed motion with a hidden velocity vector (22), and subdiffusion with a hidden potential well (23–28).

Here, we present aTrack, a probabilistic tool for efficient and robust identification of non-Brownian behavior in SPT data. Our approach uses a versatile motion model that accounts for localization error while representing a variety of behaviors including confinement, diffusion, and directed motion. Inputted SPT data is used to determine (1) the most likely states of motion, (2) whether tracks can be statistically categorized as confined or directed, and (3) the parameters that describe their behavior, for example, diffusion coefficient, radius of confinement, and speed. Further, the motion models we employ can be analytically integrated, making the analysis both rapid and accurate. We validate the approach on simulated data and demonstrate its versatility for analyzing a variety of experimental SPT data, including particle diffusion in an optical trap, detection of motile bacteria with nanoparticles, and motion characterization of spindle pole bodies in budding yeast.

Results

Accounting for hidden variables that characterize confined and persistent motion

Modeling non-Brownian motion

We model noisy tracks undergoing confined, Brownian, and directed motion by considering four relations at each time step. (1) there is a Brownian diffusion step followed by (2) an anomalous step (Fig. 1a). (3) localization error is incorporated as a Gaussian-distributed noise term added to the underlying real position to produce the observed positions. (4)

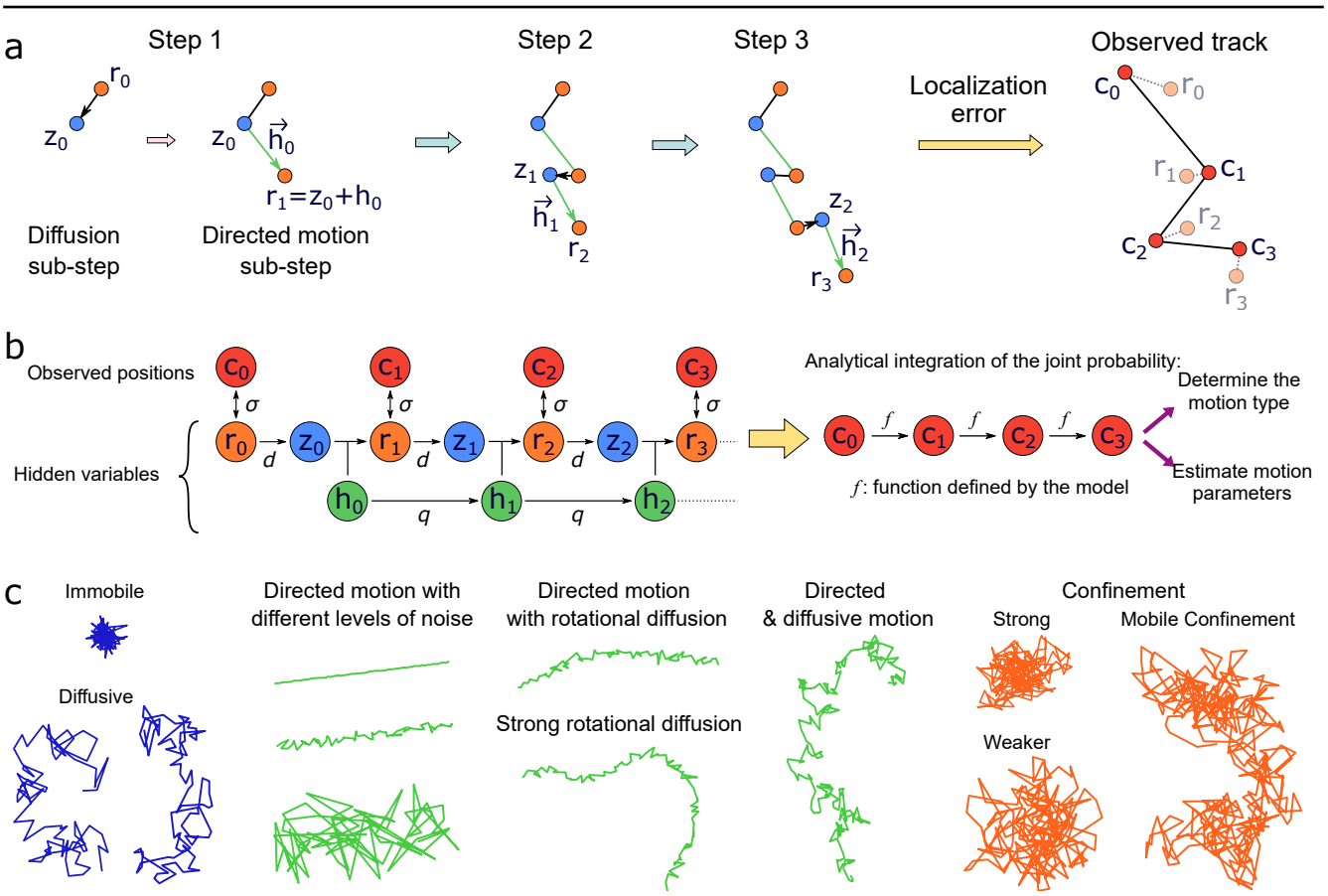


Figure 1. Principle of our method. a: The track-generation steps with our motion model shown here for directed motion. Each time step is decomposed into sub-steps, namely diffusion and anomalous motion (either directed or confining) with r_i the real positions, z_i an intermediate position, h_i the anomalous variable, and the generation of observed (measured) positions, c_i. The variables $c_i - r_i$ and $r_{i+1} - r_i$ follow Gaussian distributions with mean 0 and standard deviations σ and d, respectively. The sub-step from z_i to r_{i+1} is deterministic and the anomalous variable h_i can also evolve with a standard deviation q. **b**: Graph representation of aTrack showing the motion model (*left*), analytical integration (*middle*), and outputs (*right*). To compute the track probability, we integrate over the hidden variables. This results in an analytical recurrence formula that is used to determine the type of motion and to estimate the parameters of the motion. **c**: Examples of tracks that can be produced with our motion model.

the hidden anomalous variable, h, can evolve according to a Gaussian distribution. This model encompasses a variety of motion types depending on the model parameters, as illustrated in (Fig. 1c). For example, particles can be immobile, where the observed displacements are due to localization error; tracks can undergo Brownian motion, as well as anomalous super- and subdiffusion; or multiple motion mechanisms can also occur simultaneously, (e.g. diffusion and drift, or changes in directed motion direction and speed). The model can also account for confinement in an area of variable radius (26, 27, 29) as well as a diffusing potential well.

Determining the type of motion

To categorize the type of motion from a measured trajectory, we must first calculate the likelihood that a track belongs to each considered motion class (e.g. diffusion or confinement) and then perform a statistical comparison between the likelihoods. To do the former, we must integrate the joint probability density function of a track for all the modeled hidden variables, e.g. all the real positions. Since our motion models and localization error are expressed as the multiplication and convolution of Gaussian functions that can be rewritten as a single Gaussian, and the integral of a Gaussian process is a Gaussian distribution, we can employ a fast analytical recurrence formula (see Supplementary information) (20, 21).

With the calculated probabilities, we perform a likelihood ratio test (30, 31) between the maximum likelihood assuming Brownian diffusion (null hypothesis) and the maximum likelihood assuming confinement or directed motion as the alternative hypothesis. This likelihood ratio, $\rho = l_{\text{Brownian}}/l_{\text{confined}}$ or $\rho = l_{\text{Brownian}}/l_{\text{directed}}$, is analogous to a p-value 2, for example $\rho < 0.05$ can be considered statistically significant. If particle motion is Brownian, ρ is skewed toward 1. Conversely, ρ is skewed toward zero when applying the directed-diffusion test to directed tracks or the confined-diffusion test to confined tracks. Due to the separation of these distributions, this likelihood ratio test is less likely to produce false positives than the equivalent p-value test. In practice, to test both directed motion and confinement simultaneously, we calculate the difference between the log likelihoods of the directed motion and the confinement models, since both models intrinsically include the null hypothesis of pure Brownian motion with localization error.

To estimate the impact of the track length on the classification, we simulated confined or directed tracks of varying lengths (5-400 steps for confined, 4-100 steps for directed) and computed the likelihood ratios. As expected, the classification certainty increases with track length; where the inverse of log ρ increases with the number of steps (Fig 2a). Of course, the statistical certainty depends on the track parameters (Fig 2b). To determine the useful range, we varied the track length and either the confinement factor for confined motion or the velocity for directed motion, respectively (Fig 2c), where values <0.05 are considered successfully categorized.

Characterizing confinement

To characterize confined trajectories, aTrack estimates several parameters, namely the diffusion coefficient D and the diffusion length $\sqrt{2D\Delta t}$ where Δt is the single-frame time step; the confinement factor l, which is proportional to the spring constant of the potential well; the diffusion coefficient of the confinement area D_c; and the localization error σ. The confinement radius, which is proportional to $\sqrt{D\Delta t/l}$, can also be calculated.

To measure the precision of parameter estimation, we simulated tracks with different confinement factors, l, (Fig. 3a). We then used aTrack to estimate the parameters for each track of 200 time steps (Fig. 3b). The average diffusion coefficient, confinement factor, and confinement radius were accurate over the range of confinement factors, 0 to 1. Panel Fig. 3c shows the working range for calculating the confinement factor as a function of track length and confinement factor.Longer tracks result in better parameter estimation. Similarly, our method correctly estimates the confinement radius (Fig. S3d). In a second set of simulations, we tested the reliability of our predictions when the confinement area is moving (Fig. 3d, Fig. S3e).

Characterizing directed motion

Superdiffusive behavior, in particular directed motion, occurs in a variety of circumstances, such as molecular motor-mediated active transport (22, 32) and polymerase processivity (33). Localization error complicates velocity estimation, especially when the localization error is relatively large. We simulated noisy tracks undergoing linear motion at various speeds (Fig. 4a) with a localization error of 20 nm$\cdot\Delta t^{-1}$ and estimated the velocity per tracks. Fig. 4b shows the velocity estimates for tracks with 30 time points. By varying both the track length and the velocity, we found our method to be reliable for a wide range of velocities as long as tracks were long enough (Fig. 4c).

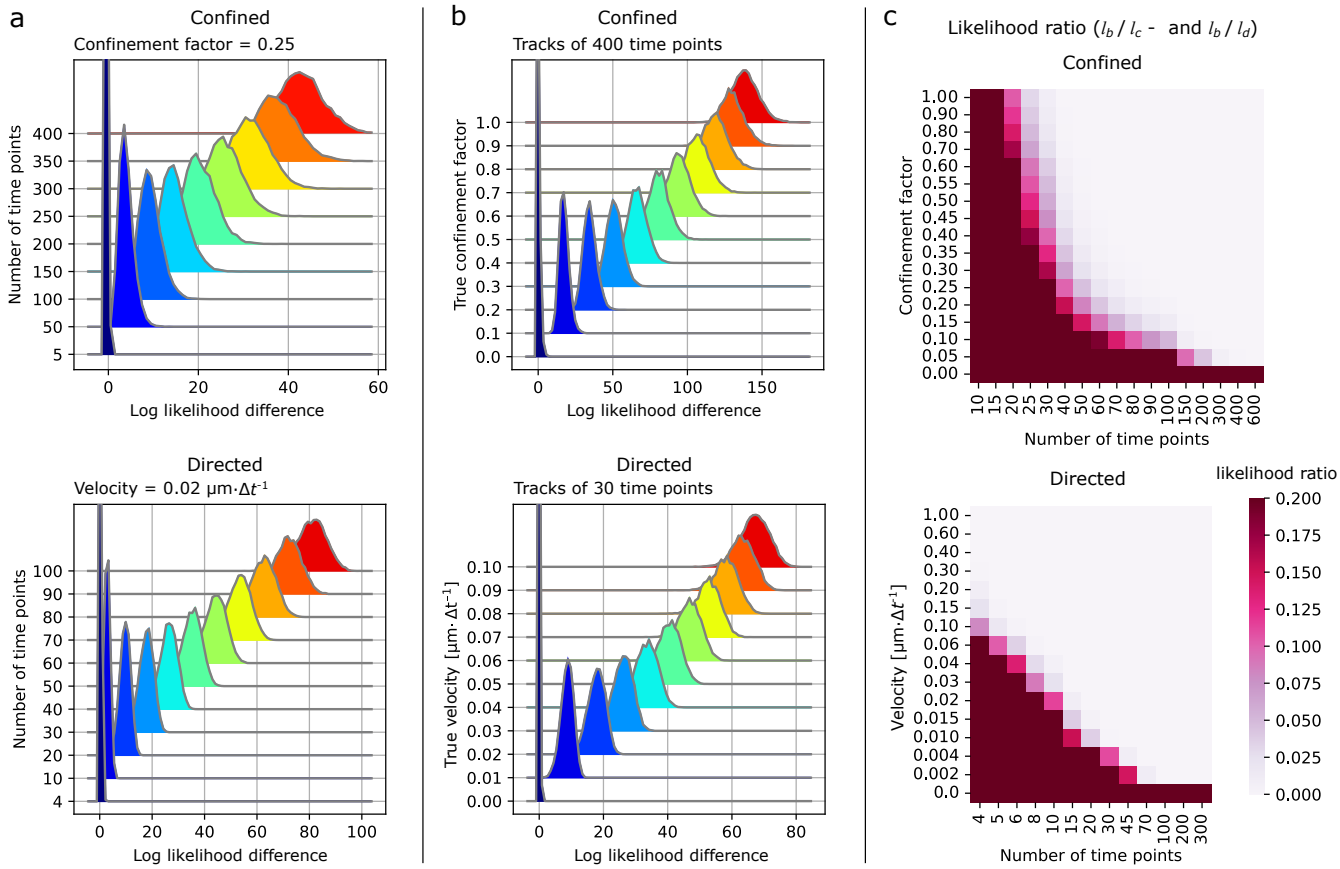


Figure 2. Determining the motion type using a likelihood-ratio test. a-b: Probability distributions of the difference between the log of the maximum likelihood of the alternative hypothesis (either confinement L_c or diffusion L_d) and the null hypothesis (Brownian diffusion L_b) for single tracks (10,000 tracks). Confinement factor $l = 0.25$ and velocity $v = 0.02$ μm·Δt^{-1}. **a**: Effect of the number of time points in a track on its log difference ($L_c - L_b$ for confined tracks) and ($L_d - L_b$ for directed tracks). **b**: The ability to distinguish confinement and directed motion from diffusion as a function of the confinement factor and particle velocity, respectively. **c**: heatmaps of the likelihood ratios l_b/l_c (confined) or l_b/l_d (diffusive) varying both the anomalous diffusion parameter and the track length. Mean of 10,000 tracks. **a-c**: When not stated otherwise, the track parameters were as following. Localization error $\sigma = 0.02$ μm. Confined tracks: diffusion length $d = 0.1$ μm. Directed tracks: $d = 0.0$ μm (no diffusion), constant speed and orientation.

In real experiments, persistent motion is rarely perfectly linear, that is changes in direction and speed are relatively common (22, 32–34). Naturally, characterizing directed motion with direction (orientation) changes is more difficult when localization error is non-negligible (35, 36). To verify our method's capacity to accurately quantify tracks with such behaviors, we simulated tracks varying the changes in orientation and velocities and estimated the track characteristics (Fig. 4d).

When orientation changes are fast, directed tracks appear diffusive, particularly in the presence of localization error. This manifests in our likelihood ratio heatmap (Fig. S5a), where tracks with high rotational diffusion angles are not significantly differentiable from Brownian motion. As the velocity increases relative to localization error, we can distinguish directed motion from Brownian motion for an increasing range of rotational diffusion angles (Fig. 4e). However, in this regime, it becomes practically impossible to distinguish directed motion with orientation changes from a combination of directed motion and diffusive motion. This effect is visible when representing the impact of the true rotational diffusion angle on the histogram of the estimated rotational diffusion angle: Rotational diffusion is well estimated for low values, but the estimate worsens as the rotational diffusion coefficient increases (Fig. 4e-f).

Sometimes, particles undergo diffusion and directed motion simultaneously, for example, particles diffusing in a flowing medium (37). Sample drift can also introduce a combination of diffusion and directed motion in single-molecule tracking. Indeed, the thermal expansion of instrument components like microscope stages can induce steady motion that masks the biologically relevant diffusive motion (38). We first tested that our method could correctly differentiate directed motion from diffusion (Fig. 4g) for a range of diffusion coefficients and track lengths at a fixed velocity of 0.1

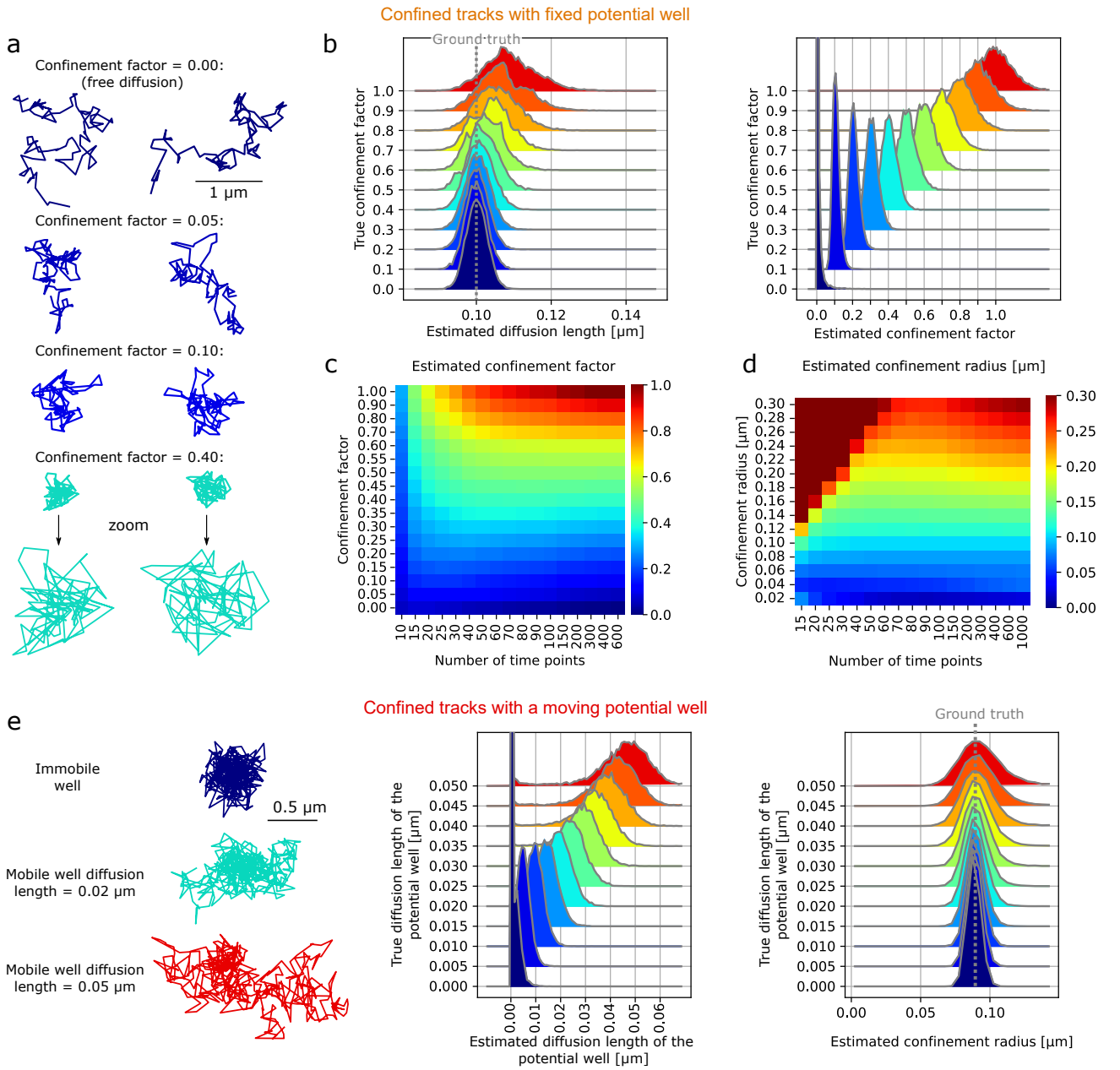


Figure 3. Characterizing confinement with aTrack. a-d: Confinement of tracks with a fixed potential well. **a**: Examples of simulated tracks with different confinement factors. **b**: Histograms of the estimated parameters for individual tracks of 200 time points varying the number of time points in tracks. **c-d**: Heatmaps of the mean estimated confinement factor and confinement radius ($\frac{d_{\text{est}}-0.1}{0.1}$) depending on the track length and the confinement factor or radius respectively. **d**: Confinement of tracks with a moving potential well (Brownian motion). Left: simulated tracks with different diffusion length of the potential well. Right: histograms of the estimated diffusion length of the the potential well and confinement radius = $\sqrt{D\Delta t/l}$ varying the actual diffusion length of the potential well. Confinement factor = 0.1. **a-d**: 10,000 tracks per condition. $d = 0.1$ µm, Localization error $\sigma = 0.02$ µm. See Fig. S3 for complementary results.

nm.Δt^{-1}. For mixed motion, our method accurately estimates the diffusion length and velocities for short tracks so long as the diffusion length was low compared to the velocity. When the diffusion length is large compared to the velocity, parameters can still be estimated, but longer tracks are needed for a reliable measurement.

Characterizing populations of tracks

The amount of information contained in an individual track is limited by its length; however, it is also of interest to consider a population of tracks moving with the same underlying parameters. We first simulated populations of tracks that

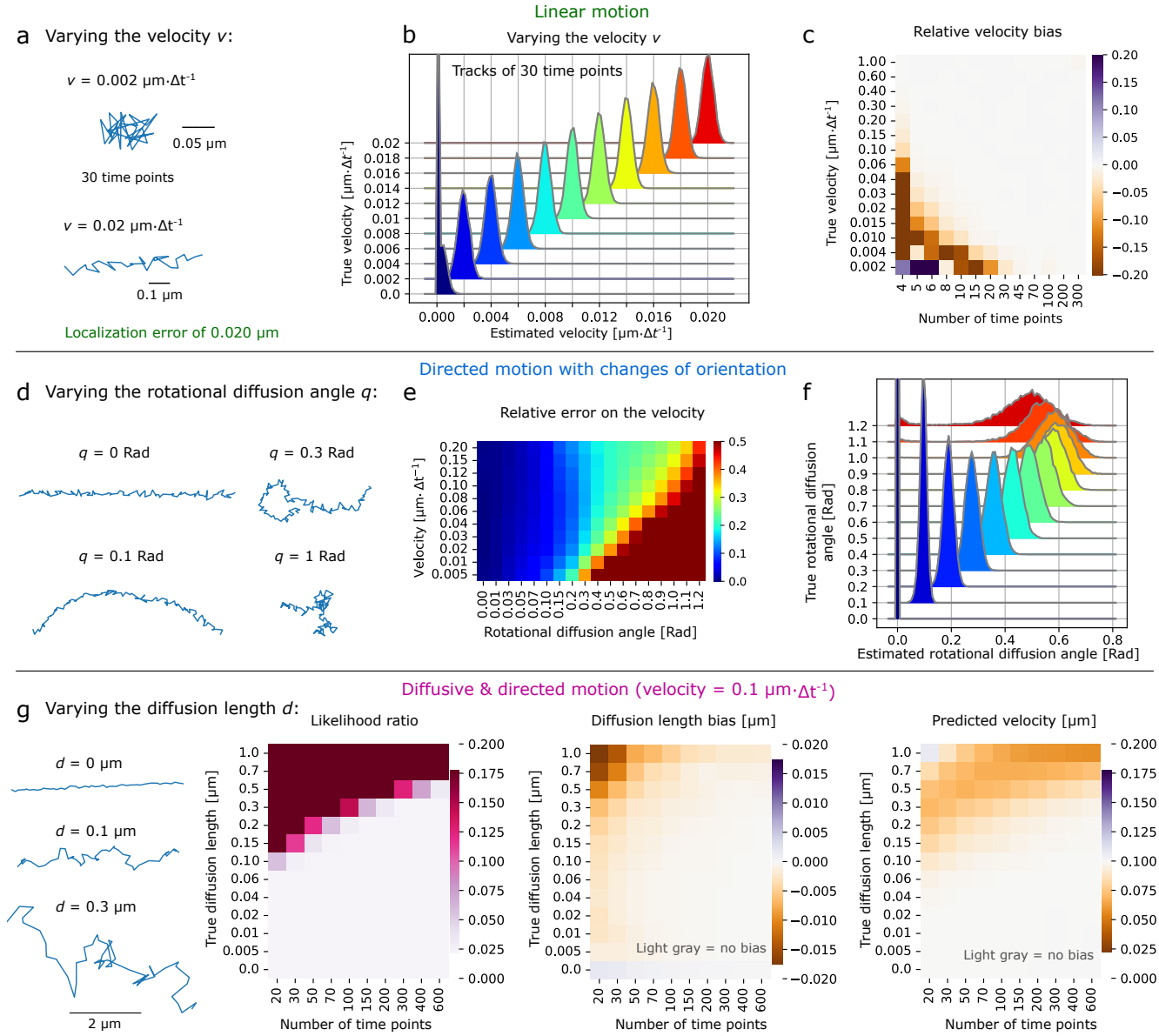


Figure 4. Characterizing directed motion with aTrack. a-c: Tracks with linear motion (constant speed and orientation). **a**: Examples of simulated tracks in directed motion. **b**: Histograms of the estimated velocity of individual tracks of 30 time points. 10,000 tracks per histogram. True parameters $d = 0.\ \mu\text{m}$, Localization error $\sigma = 0.02\ \mu\text{m}$ (fixed). The next panels use the same parameters unless specified otherwise. **c**: Heatmap of the relative biases on the estimated velocity ($\frac{v_{\text{est}} - v_{\text{true}}}{v_{\text{true}}}$). **d-f**: Tracks with constant speed but changing orientation. **d**: Simulated directed tracks with rotational diffusion. Here, the rotational diffusion angle coefficient is defined as the standard deviation of the change of orientation at each time step (analogous to the diffusion length). $v = 0.02\ \mu\text{m}{\cdot}\Delta\text{t}^{-1}$. **e**: Heatmap of the error on the rotational diffusion angle for a range of velocities and rotational diffusion angles. Tracks of 200 time points. **f**: Distributions of the estimated rotational diffusion angle for a range of rotational diffusion angles. Tracks of 200 time points with $v = 0.1\ \mu\text{m}{\cdot}\Delta\text{t}^{-1}$. **g**: Tracks simultaneously undergoing both linear motion and diffusion with varying levels of diffusion. Heatmaps of the likelihood ratio, bias on the diffusion length in μm, and estimated velocity depending on the number of time points per track and on the diffusion length d. **a-g**: Where not stated otherwise, the track parameters were as following. Localization error $\sigma = 0.02\ \mu\text{m}$, $d = 0.0\ \mu\text{m}$, velocity $v = 0.1\ \mu\text{m}{\cdot}\Delta\text{t}^{-1}$, constant speed and orientation.

are either in a confined state or in a diffusive state to demonstrate that using the data from more tracks results in higher certainty regarding the type of motion and produces better estimates of the motion parameters (Fig. S6). Next, we simulated populations of tracks with multiple states to determine if our method could correctly distinguish and characterize these states (Fig. 5a). Even under complex conditions, our method could reliably retrieve the parameters of each state even for relatively few tracks (Fig. 5b).

In many circumstances, the number of states is not known. It is thus important to have a method that determines the number of unique states in a population of tracks. The challenge is that adding more states inevitably better fits a dataset, albeit with a diminishing margin of return (39). In the case shown in (Fig. 5c), the likelihood significantly increases when the number of states from 4 to 5; however, for higher numbers of states, the likelihood is significantly less affected, see from 5 to 10 states. Using this effect, it is therefore possible to estimate the number of states. We verified this trend for populations composed of 50 tracks to 12,800 tracks. We found the Akaike Information Criterion (AIC) (40) and the Bayesian Information Criterion were not always suited for determining the number of states in this task (Fig. S7), most likely due to small discrepancies between our model and simulated tracks. Correcting the AIC or the BIC with an additional term proportional to the number of tracks and the number of parameters provides a robust criterion (Fig. S7).

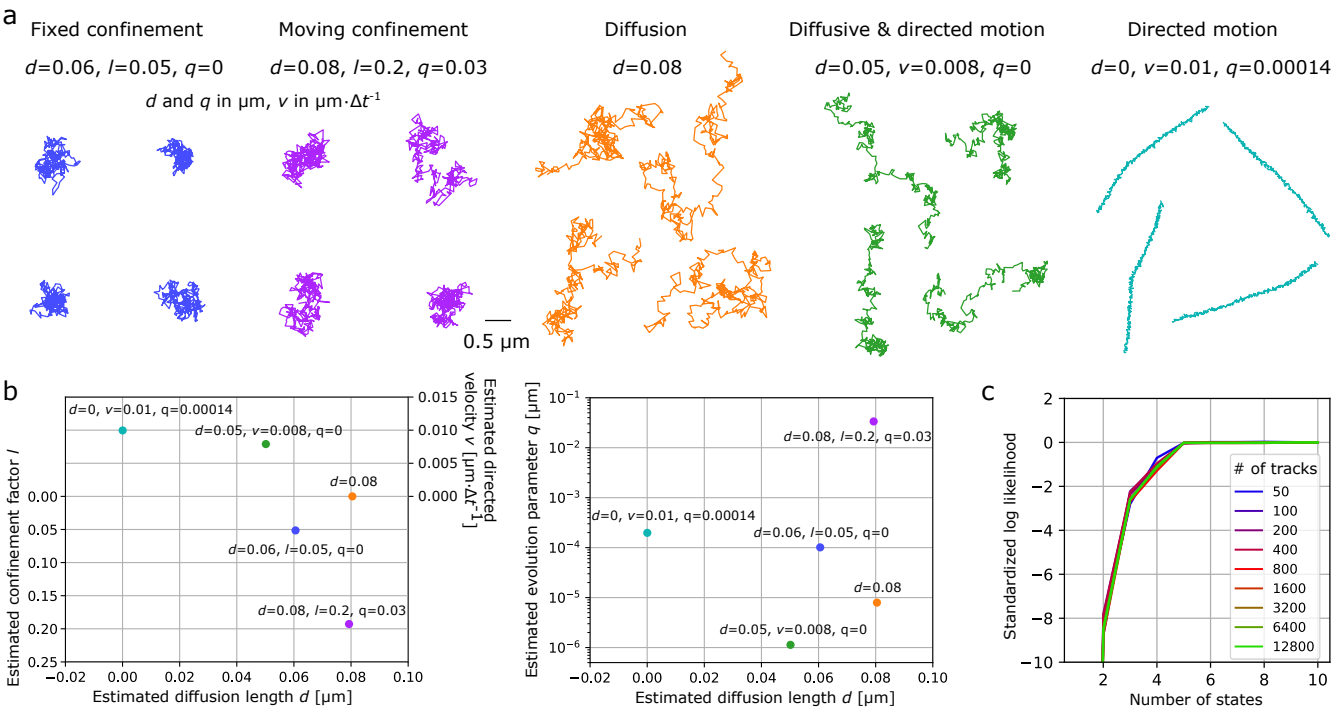


Figure 5. Characterizing populations of multiple states. Analysis of tracks with 5 sub-populations of set diffusion length d, confinement factor l, velocity v for directed tracks, and anomalous change parameter q (diffusion length of the potential well for confined tracks and changes of speed for directed tracks). Tracks are 300 time point long. **a**: Track examples from each of the 5 states with the corresponding state parameters. **b**: Estimated parameters for the 5 states (using a 5-state model). **c**: Log likelihood of the model depending on the number of states assumed by the model. The log likelihood was normalized by the number of tracks, offset by the log likelihood assuming 10 states.

Robustness to model mismatches

One of the most important features of a method is its robustness to deviations from the assumptions. Experimentally obtained tracks will inevitably not match the model assumptions to some degree, and models need to be resilient to these deviations. To test the generalizability of our approach to other types of motion, we simulated tracks following a subdiffusion motion model that does not fit our motion model, namely fractional Brownian motion (12). This was performed with three anomalous diffusion coefficients, 0.5, 1.0, and 1.5, corresponding to subdiffusion, Brownian, and super-diffusion, respectively Fig. 6a. The performance of aTrack was then measured by computing the difference between the likelihood of the directed motion model and the confined motion model. We then plotted a histogram of the log-likelihood differences for each type of motion and estimated the classification accuracy in detecting anomalous diffusion from Brownian diffusion. aTrack was 88% accurate when differentiating subdiffusive fractional Brownian motion (anomalous diffusion coefficient of $\alpha = 0.5$) from Brownian motion ($\alpha = 1$), and 95% accurate for differentiating superdiffusive fractional Brownian motion ($\alpha = 1.5$) from Brownian motion. To see how our approach compared to one of the leading methods for characterizing fractional Brownian motion, we performed the same test with Randi (41), a machine learning method

found to be the most powerful in a head-to-head comparison of available techniques (16). We applied Randi to the same dataset and found that both methods achieved similar accuracies for the fractional Brownian motion data, which Randi was designed to quantify.

We then used the same approach to compare the two methods on tracks that are based on our motion model (Fig. 6b). Here, we simulated tracks with parameters that result in a qualitatively similar motion to that observed with fractional Brownian motion. For those tracks, our model performs with high accuracy, at least 99%, while Randi shows a decreased classification accuracy. In particular, differentiating diffusive tracks from tracks with both diffusion and directed motion had an accuracy of 61% (only 11% better than random labeling). Identifying superdiffusive tracks with Randi had an accuracy of 83%, underscoring how model mismatches complicate track classification when the motion model is insufficiently flexible.

To test the robustness of our method in the context of spatial confinement with rigid boundaries, we simulated tracks of diffusive particles trapped on a disk with a specific radius and measured the effect of the confinement radius (the disk radius) on our predictions (Fig. 6c). While aTrack uses a potential well and not a hard boundary, the model nonetheless identifies particle confinement for a wide range of confinement radii. The lower radius bound of the operating range relates to the diffusion length per step while the upper bound is limited by the extent of the particle's exploration of the confinement area. Note that the exploration distance is determined by the track length and by the diffusion length). Within the working range, the confinement radius can be accurately estimated using the estimated standard deviation of the potential well, where the disk radius is three times the standard deviation of the well. This calculation was relatively independent of track length (Fig. 6d).

Implementation on experimental data

To test the usefulness of aTrack on experimental data, we performed classification, population characterization, and parameter-estimation experiments. First, we analyzed tracks from the spindle pole body (SPB) in Saccharomyces cerevisiae. The SPB is the microtubule organizing center in yeast and is embedded in the nuclear envelope. Long-range directed motion of the SPB is expected during mitosis when the nucleus orients towards the bud and the spindle elongates to separate the chromosome masses of mother and daughter cell (Fig. 7a), but some directed motion of the SPB may also occur in other phases of the cell cycle due to the forces exerted on it by microtubules (42, 43). When fitting individual tracks, we observe a significant fraction of persistently moving SPB (13% with a confidence interval of 95%, Fig. 7b). Tracks with a significant likelihood ratio show visible directed motion (Fig. 7c). We did not observe a significant fraction of confined motion with a single-track approach or population-level approach.

Next, we analyzed tracks of 100nm gold nanoparticle (AuNP) biomarkers that can attach to mobile *E. coli*, freely diffuse in the environment, or attach nonspecifically to the substrate. Our method identified the directed tracks, i.e., biomarkers adhered to live bacteria with very high certainty (likelihood ratio < 0.001, Fig. 7d)). 5.9% of the AuNPs were significantly directed according to aTrack, in agreement with a manual inspection of the data. Interestingly, AuNPs attached to bacteria exhibited nearly linear motion coupled to lateral oscillations, consistent with cell rotation (44). Next, we analyzed the tracks at the population level and computed the number of states as a function of likelihood (Fig. 7e). The likelihood function shows a noticeable difference between 4 and 5 states. A 5-state model, indicating the presence of multiple fractions within the populations of cell-bound AuNPs and free AuNPs. Using aTrack, we find two states with directed motion and a high velocity, representing 6.2% of the tracks. Assuming a different number of states, for instance, 3 states or 7 states, results in the same fraction of clearly directed tracks with similar parameters.

Finally, we tested our method's capability to detect confinement on particles stuck in an optical trap, an experimental system similar to that of our motion model. To do so, we measured 1-micron beads trapped using optical tweezers (Fig. 7f). Using tracks with 100 frames, we detected significant confinement, as expected (Fig. 7g).

Discussion

aTrack is a new tool for classifying and characterizing noisy tracks, which performs well in a diverse range of conditions. Specifically, the framework's flexibility is relevant for a wide variety of diffusive, confined, and directed motion types. our tool classifies the motion type as Brownian or not and quantifies the biologically relevant parameters of the motion, such as the diffusion coefficient, confinement area, and velocities. Importantly, this approach calculates how statistically robust a classification is regarding the likelihood difference from a Brownian diffusion model. The flexibility of our approach in

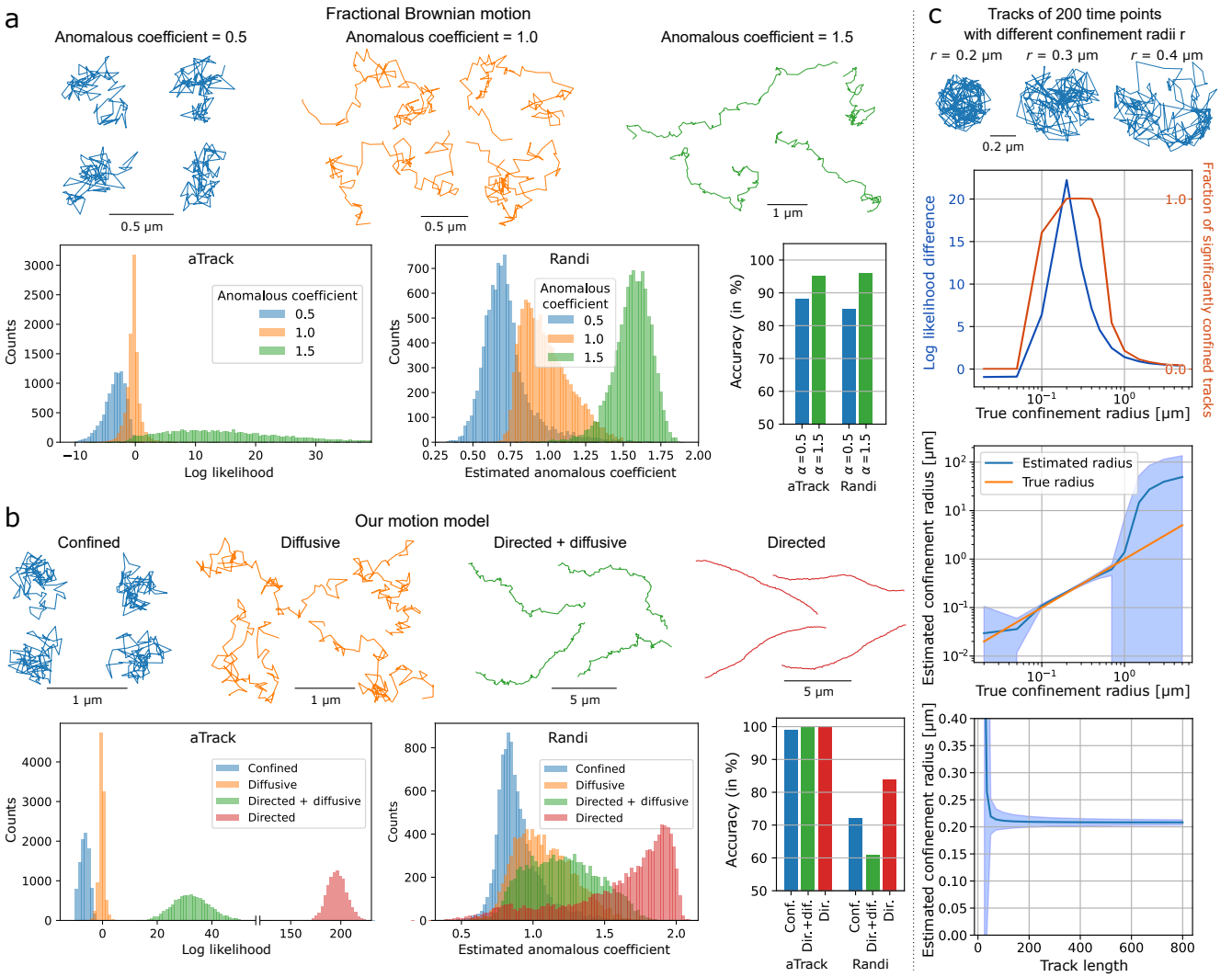


Figure 6. Model robustness with other motion types. a-b: Example tracks and corresponding distributions used to determine the type of motion for aTrack and Randi (41). aTrack uses the difference between the likelihood assuming super-diffusion and the likelihood assuming sub-diffusion (bottom-left). To classify tracks using Randi, we used the estimated anomalous diffusion coefficient. The accuracy is the fraction of correctly labeled tracks in a data set composed of 5,000 sub-diffusive or superdiffusive tracks and 5,000 Brownian tracks. Classifications were done using the thresholds that best divide the distributions. **a**: Analysis of tracks with 200 time steps following fractional Brownian motion with anomalous coefficients of 0.5 (subdiffusive), 1 (diffusive), and 1.5 (superdiffusive). **b**: Analysis of tracks with 200 time steps following our motion model. Confined tracks: diffusion length $d = 0.1$ µm, localization error $\sigma = 0.02$ µm, confinement force $l = 0.2$, fixed potential well. Brownian tracks: $d = 0.1$ µm, $\sigma = 0.02$ µm. tracks in both directed and diffusive motion: $d = 0.1$ µm, $\sigma = 0.02$ µm, directional velocity $v = 0.1$ µm·Δt^{-1}. Directed tracks: $d = 0.$ µm, $\sigma = 0.02$ µm, $v = 0.1$ µm·Δt^{-1}, angular diffusion coefficient 0.1 Rad^2·s^{-1}. **c**: Analyzing tacks confined by hard boundaries using aTrack. A simulated track with 200 time points diffusing on disks of different sizes. Top panel: Log likelihood difference $L_c - L_B$ and fraction of significantly confined tracks (likelihood ratio $l_B/l_c < 0.05$) depending on the confinement radius. Middle panel: Estimated confinement radius ($= 3\frac{d}{\sqrt{2l}}$) depending on the true confinement radius. Bottom: estimated confinement radius depending on the track length. Blue areas: standard deviations of the estimates.

capturing a variety of motion-type behaviors contains the usefulness of the catch-all approach of using an anomalous exponent; however, unlike the anomalous exponent that fits multiple motion models (6), our framework has the advantage of outputting interpretable parameters that describe the motion, e.g., velocity or confinement radii.

Using aTrack, we found that directed motion can be readily identified from short tracks due to the deterministic nature of this type of motion. As biological systems rarely contain exactly linear motion, we incorporated a parameter to account for changes in velocity, allowing curved trajectories to be robustly classified. Our method can also account for motion that contains a combination of directed and diffusive motions, namely, that found when particles diffuse in a flowing environment.

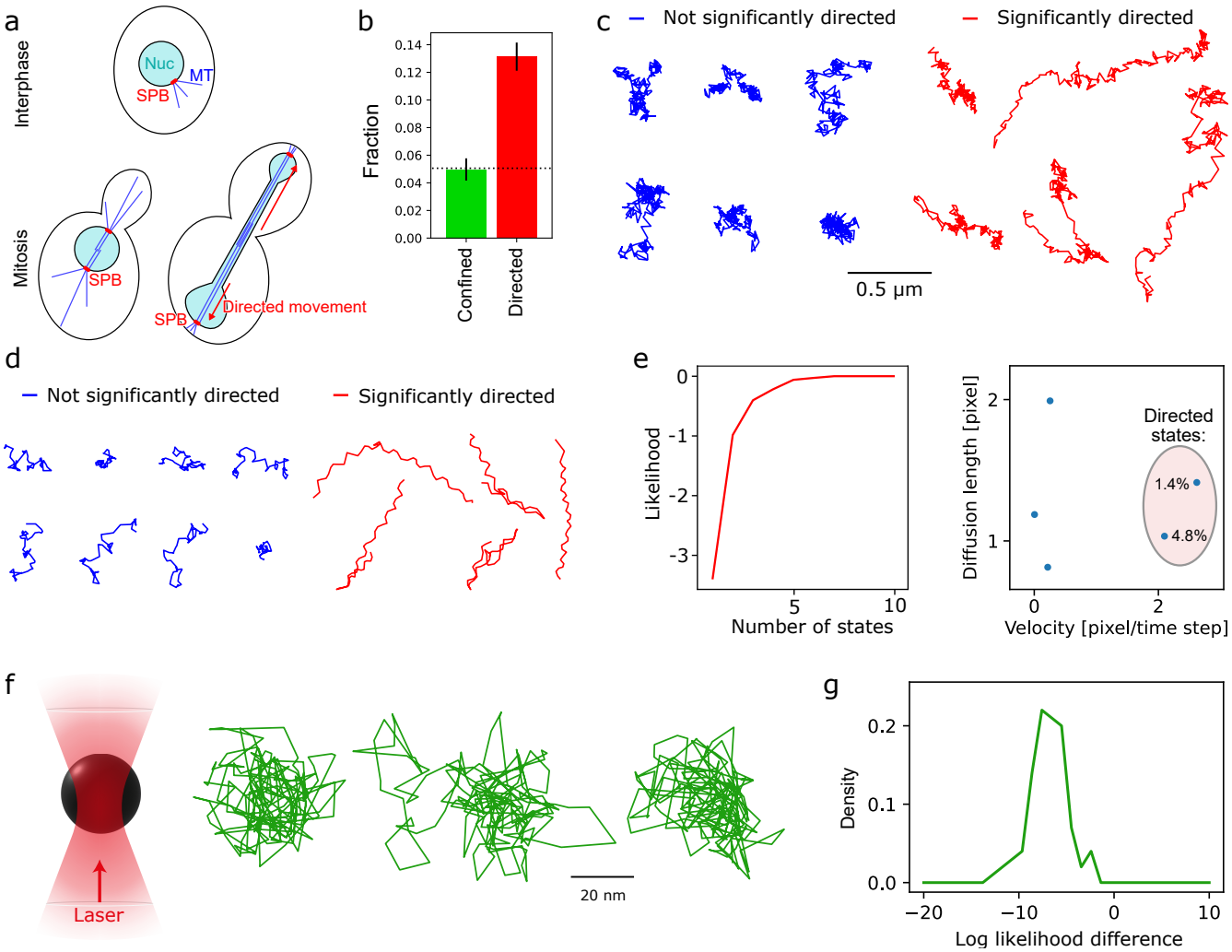


Figure 7. Experimental demonstrations. a-c: Analysis of spindle pole body (SPB) tracks. Each track has 200 time points. **a**: Illustration showing the role of the SPB in the cell cycle (MT: microtubule). During mitosis, the elongation of the interpolar microtubules induces a directed movement of the SPB that divides the nucleus. **b**: Fraction of tracks significantly directed or confined with a confidence interval of 95% according to our likelihood ratio test on single tracks. Mean and standard deviation of 3 biological replicates of at least 400 tracks. The 95% confidence interval implies that we expect 5% of false positives in a fully Brownian population. **c**: Example SPB tracks and state labeling from our individual-track analysis. Red tracks are significantly directed (95% confidence interval), and blue tracks are non-significantly directed. **d-e**: Analysis of nanoparticle (NP) tracks in the presence of motile bacteria (50 time points per track), where some NPs adhere to cells. **d**: NP tracks colored according to their state of motion classification using aTrack's single-track statistical test. **e**: Maximum likelihood (per track) of the population of tracks depending on the number of states (standardized by the likelihood assuming 10 states). **f-g**: Analysis of tracks for 1 μm beads trapped using optical tweezers. The data set is composed of 100 tracks of 100 time points. **f**: Track examples. **g**: Log likelihood difference ($L_{\text{directed}} - L_{\text{confined}}$).

Compared to directed motion, longer tracks are needed to classify confined motion with the same statistical certainty. Indeed, confined particles need to reach several boundaries of the confinement area to be distinguished from freely diffusive particles. In some cases, a confined particle may leave its local environment or hop between environments (45–47). To allow for this in our model, the center of the confinement well is modeled as a diffuser. We also require the diffusion coefficient of the well to be slower than the particle to avoid irrelevant behaviors. As with all classification tools, a key source of ambiguity arises when motion types resemble one another. For example, immobile particles and very tightly confined particles can sometimes be indistinguishable while having different underlying models. Nevertheless, determining whether a particle is strongly confined or immobile is not necessarily insightful. Indeed, immobile fluorophores can be modeled as confined to an area around a substrate to which they are bound. In an experimental context, adapting the framerate can be important to best observe the features of interest (48).

Relative to available deep-learning tools, our technique contains many fewer parameters. These values map onto the

stochastic physical behaviors of particle motion. In comparison to the deep-learning-based tool we tested, we found that aTrack to be more robust to model mismatches. This finding is consistent with the well-documented issues of machine-learning models generalizing poorly to new data. Of course, in the context of tracking, the generalizability of a model to new data is a key factor, as experimentally obtained data never perfectly match the model assumptions or the training data set. One solution to make a machine learning model that generalizes better is to use a physics-informed neural network (49). Such a network would use probabilistic relationships to efficiently learn the physical properties of unlabeled tracks and would contain far fewer parameters than classical networks.

It is often useful to assume a finite number of states with fixed parameters to model the various molecular states in a sample (50). Selecting the right number of states is difficult but can be automated by different methods. The criterion, e.g., AIC and BIC, used to determine the number of states is reliable when the model and the data are in perfect agreement. However, experimental data never matches the model assumptions perfectly, and even discrepancies between models and simulated data can affect reliability, such as continuous-time simulations and discrete models. This usually results in overestimating the number of states for large data sets (39). To avoid this flaw of classical criteria, we showed that a penalization term proportional to the number of states and to the log likelihood of the data can prevent the spurious increase of the number of states when increasing the number of tracks. The drawback of this approach is the addition of a tunable parameter that influences the number of estimated parameters of the model, and it may still be necessary to limit the number of states to the biologically relevant system.

An important limitation of our approach is that it presumes that a given track follows a unique underlying model with fixed parameters. In biological systems, particles often transition from one motion type to another; for example, a diffusive particle can bind to a static substrate or motor (51). In such cases, our model is not suitable. However, this limitation can be alleviated by implicitly allowing state transitions with a hidden Markov Model (21) or alternatives such as change-point approaches (41, 52, 53), and spatial approaches (24).

In conclusion, we have shown that aTrack can identify anomalous diffusion and parameterize the motion over a broad range of motion types using a robust probabilistic framework. As the motion-model parameters estimated by the method represent physical phenomena, these variables are readily interpretable, for example the diffusion coefficient, confinement radius, and velocities. Finally, the employed motion models were selected to permit analytical integration, which makes calculating the model parameters fast and accurate; of course, this integration strategy can be implemented for a variety of motion models with hidden states, further expanding the applicability of this approach to other motion types.

Methods

Modeling particle motion with observed and hidden variables

Probabilistic model for confined motion

At each time step i, our confined-motion model consists of (1) a Brownian motion step, (2) a confinement step, (3) an update to parameters, and (4) a localization error step. The Brownian motion step updates the particle's position r_i to an intermediate position z_i. The variable $z_i - r_i$ follows a Gaussian distribution centered at 0 with standard deviation d, where d is the diffusion length, $d = \sqrt{2D\Delta t}$, D is the diffusion coefficient, and Δt is the time step. Next, the confinement step is modeled by an attractive force between the particle with a potential well centered at h_i. More precisely, the particle moves toward the center of the potential well proportionally to a confinement factor l and to the distance $r_i - h_i$. The next real position r_{i+1} is thus determined by the following relationship $r_{i+1} = (1-l)\cdot z_i + l \cdot h_i$. To allow the potential well to move, h is updated at each step such that $h_{i+1} - h_i$ follows a Gaussian distribution of mean 0 and standard deviation q. Finally, to model localization error, the observed positions c_i and the real positions r_i are related by the localization precision, σ, where $c_i - r_i$ follows a Gaussian distribution of mean 0 and standard deviation σ.

The distribution of positions for a diffusive particle in a fixed potential well is a Gaussian distribution with standard deviation $\rho = \sqrt{\frac{D\cdot \mathrm{d}t}{l}} = \frac{d}{\sqrt{2l}}$. As we have no prior information about the initial center of the potential well, we assume it is positioned according to a Gaussian probability density function centered on the initial observed position. While it would be even better to consider a Gaussian f_{h_0} centered around r_0, we simplify it by approximating it to be centered around c_0 and of standard deviation $q_0 = \sqrt{\rho^2 + \sigma^2}$. The joint probability density function corresponding to this model is the product of Gaussian functions shown in equation 1, which is integrated over all hidden parameters to calculate the likelihood, $l_{confined}$.

$$
\begin{aligned}
& f_\sigma(r_0 - c_0)\, f_d(z_0 - r_0)\, f_{q_0}(h_0 - c_0) \\
& f_\sigma((1-l)z_0 + lh_0 - c_1)\, f_d((1-l)z_0 + lh_0 - z_1)\, f_q(h_1 - h_0) \\
& f_\sigma((1-l)z_1 + lh_1 - c_2)\, f_d((1-l)z_1 + lh_1 - z_2)\, f_q(h_2 - h_1) \\
& [\ldots] \\
& f_\sigma((1-l)z_{n-2} + lh_{n-2} - c_{n-1})\, f_d((1-l)z_{n-2} + lh_{n-2} - z_{n-1})\, f_q(h_{n-1} - h_{n-2}) \\
& f_\sigma((1-l)z_{n-1} + lh_{n-1} - c_n)
\end{aligned}
\tag{1}
$$

Where the real positions of the particle r_i, potential well h_i, and intermediate position z_i are hidden variables, and l is the confinement factor. By integrating this joint probability over the hidden variables, we can retrieve the probability of the track (the observed positions) given the model and its parameters.

While this integration step can be computed with a Monte Carlo approach (27), this is computationally expensive. Instead, we integrate using an analytical recurrence formula. This formula is allowed by the fact that the joint probability density function is a product of Gaussians and by the property that for two Gaussians, f and g, with means μ_f, μ_g and standard deviations σ_f, σ_g, the product is also Gaussian, $f(x)\cdot g(x) = \phi \cdot \eta(x)$ where ϕ and η are Gaussian distributions described by Equation 2.

$$
\phi = \frac{\exp\left(-\frac{(\mu_f - \mu_g)^2}{2\cdot(\sigma_f^2 + \sigma_g^2)}\right)}{\sqrt{2\pi\,(\sigma_f^2 + \sigma_g^2)}} \text{ and } \eta(x) \text{ a Gaussian of mean } \mu_{fg} = \frac{\sigma_f^2\mu_g + \sigma_g^2\mu_f}{\sigma_f^2 + \sigma_g^2} \text{ and variance } \mu_{fg} = \frac{\sigma_f^2\sigma_g^2}{\sigma_f^2 + \sigma_g^2}. \tag{2}
$$

See Supplementary information for more details.

Probabilistic model for directed motion

To consider directed motion, we use the same general framework as confinement. At each time step i, the real particle position r_i is first updated to an intermediate position z_i, where the variable $z_i - r_i$ follows a Gaussian distribution centered at 0 with standard deviation d. Next, we add the directed-motion component with a vector w_i, such that the next real position is the sum of the two steps, $r_{i+1} = z_i + w_i$. Combining these two substeps, we get $z_i - r_i = r_{i+1} - r_i - w_i$, simplifying the integration process of the probability density function expressed in equation 3. Analogous to our confinement model,

the velocity vector, w_i, is allowed to change over time, where the $w_{i+1} - w_i$ follows a Gaussian distribution with mean 0 and standard deviation q. Finally, we include the effect of localization error, where the observed position c_i is related to the real position r_i following a Gaussian distribution, where $c_i - r_i$ is Gaussian distributed with mean 0 and standard deviation σ.

During the first time step, the orientation and length of the directed motion vector (speed) need to be initialized. To do so, we assume w_0 follows a Gaussian distribution function with mean 0 and standard deviation v. The resulting joint probability density function is shown in Equation 3.

$$\begin{aligned}
&f_\sigma(r_0 - c_0)\, f_v(w_0)\, f_d(r_1 - r_0 - w_0)\, f_q(w_1 - w_0)\\
&f_\sigma(r_1 - c_1)\, f_d(r_2 - r_1 - w_1)\, f_q(w_2 - w_1)\\
&f_\sigma(r_2 - c_2)\, f_d(r_3 - r_2 - w_2)\, f_q(w_3 - w_2)\\
&[\ldots]\\
&f_\sigma(r_{n-2} - c_{n-2})\, f_d(r_{n-1} - r_{n-2} - w_{n-2})\, f_q(w_{n-1} - w_{n-2})\\
&f_\sigma(r_{n-1} - c_{n-1})\, f_d(r_n - r_{n-1} - w_{n-1})\\
&f_\sigma(r_n - c_n)
\end{aligned} \tag{3}$$

As with confinement, the probability of a directed track given the model parameters can be calculated using an analytical recurrence formula to integrate over the hidden positions, $r_{0\to n}$, and velocities $w_{0\to n-1}$. The parameter v can, in principle, be used to estimate the velocity of a particle with a constant speed but changing orientation in the imaging plane; however, to estimate the average speed of a particle, we use another metric, k, that appears in our integration process (see Supplementary information).

Modeling time-varying velocities and changes in direction

In our model, each axis, x and y, is treated separately, and the velocities can evolve according to a Gaussian distribution. This allows a particle's direction to change over time. To quantify how direction changes affect the analysis for particles with a constant speed, we simulated tracks with a fixed speed and time-dependent direction for a range of angular diffusion coefficients, D_θ. The model parameter, q, represents the standard deviation of the change in speed, which can be converted to an angular diffusion length using the following trigonometric relation and the estimated speed of the motion.

In the case of pure directed motion with direction changes, let us consider a single time step, i. We have A, the previous particle position r_{i-1}; B, the particle position assuming no changes of orientation or diffusion, $r_{i-1} + v_{i-1}$; and C, the actual particle position after a change of orientation, r_i. ABC forms an isosceles triangle, which can be split into 2 right triangles. We find the following relationship between the scalar distances BC, AB, and θ, $sin(\frac{\theta}{2}) = \frac{BC/2}{AB}$, where θ is the orientation angle change.

Track simulations

We subdivided each time step into 20 substeps to simulate approximately continuous tracks. We applied the Brownian diffusion and anomalous movements to each of these substeps. For confined motion, the latter step moves the particle toward the center of the potential well center proportionally to the distance multiplied by a scaled confinement factor, $l/20$. In simulations with a moving potential well, the center moves according to the well's diffusion coefficient, q. For directed motion, we applied a shift of constant velocity. The orientation of the directed motion could vary according to a rotational diffusion coefficient. The particle's position after the 20 substeps was set to the particle's real position, r_i, and the localization error was added to create the observed position c_i. This process is equivalent for undivided frames for directed motion, as the diffusion and the directed motion steps are independent. However, for the confined model, where the diffusion of the particle influences the anomalous step, thus there is a dependence between the diffusion and confinement steps.

Fractional Brownian motion was simulated using the Python package 'fbm' (https://pypi.org/project/fbm/) with the Davies and Harte method (54).

Experimental methods

Spindle-pole body

Yeast cells expressing Spc42-mCherry (KWY10328, (55)) were grown to exponential growth phase in synthetic complete medium with glucose (SCD) and imaged in Matrical 384-well glass bottom plates coated with Concanavalin A on a temperature-controlled inverted Nipkow spinning disk microscope equipped with the Yokogawa Confocal Scanner Unit CSU-W1-T2 controlled by the VisiVIEW Software (Visitron). It was used in spinning disk mode with a pinhole diameter of 50 µm combined with a 1.45 NA, 100x objective. Images were acquired on an EMCCD Andor iXon Ultra camera (1024x1024 pixel, 13x13um pixel size); for one of three biological replicates, the data was acquired with dual camera settings. Imaging was performed at 30 °C with 80% laser intensity of a Diode 561 nm, 200 mW laser. Timelapse data were acquired with 100 ms exposure time in stream mode for 300 frames. For all movies, tracks were obtained using the ImageJ (56) plugin TrackMate (57). Peak detection: LoG with radius = 0.3 µm and threshold = 60, linkage: simple lap tracker with a linking maxiumum distance of 0.4 µm and allowing gaps of one frame.

Bacteria detection and tracking

Bacteria detection with nanoparticles was performed following the protocol described in Zapata-Farfan et al. (2023). In brief, *E. coli* (strain 25922, ATCC) were cultured for 24 h in Trypticase soy broth (TSB) at 37°C and then incubated for 30 min with 100 nm spherical gold nanoparticles at 50 µg/mL (A11-100-CIT, Nanopartz) in 1x PBS. Samples were placed in a small chamber consisting of double-sided tape (5 µm thickness, Nitto) between a coverslip and glass slide. Image data was recorded in darkfield mode at 20 frames per second using an LED side-illumination system and microscope outfitted with a 40x objective (40x air, NA 0.7, Nikon) and monochrome CCD camera (QICAM Firewire camera, Teledyne Photometrics). Tracking was performed using Trackmate 7.0 (57) with the LoG detection method with a diameter of 0.70 µm (6 pixels) and quality threshold of 24.

Optical tweezers

The tracking of microspheres captured in an optical trap was achieved using a custom instrument constructed on a modular inverted microscope (MIM/RAMM, ASI Imaging), incorporating optical trapping and imaging paths. The trapping diode laser (SNP-06E-100, Teem) emitting at 1064nm (TEM_{00}) had an average output power of 60 milliwatts. The laser beam is expanded using a 1:3 telescope (Achromatic doublets, Thorlabs) to overfill the objective's back focal plane. This expanded beam was focused through a 100X, 1.45 NA objective (UPlanXApo 100X, Olympus) to trap dielectric particles. The same objective was used to collect bright-field illumination, which was imaged using a CMOS camera (ORCA - Flash4.0 LT3, Hamamatsu). A 3-axis piezo-driven stage (MicroScan SCXYZ100, Thorlabs) with a precision of 25 nm facilitated sample movement to capture particles in the trap. One µm polystyrene microbeads (Monodisperse fluorescent microspheres, Cromtech Research Center) were suspended in water. The solution was squeezed between two coverslips (No. 1.5, Thermo), and a single bead was brought into the laser trap by translating the sample. Data acquisition was performed at 315 frames per second for 2 minutes, these movies were then analyzed in shorter, 100-frame segments. Beads positions were tracked using TrackMate (57).

Data and code availability

Tracking data and the aTrack software is available online. aTrack is available as a stand-alone software for Windows and as a python package. Installation instructions are provided on the Github page https://github.com/FrancoisSimon/aTrack.

Acknowledgements

The authors thank Sven van Teeffelen for helpful discussions. This work was supported by the Natural Sciences and Engineering Research Council of Canada [NSERC Discovery grant to MM, (RGPIN-06404-2016) to CB, (RGPIN-2022-05142) to LEW], the Canada First Research Excellence Fund (TransMedTech Institute), and ETH research grant ETH-33 19-1 to ED.

Author contributions

Conceptualization: FS, LEW; Methodology: FS, LEW; Software: FS, IF; Validation: FS; Formal analysis: FS; Investigation: FS, GR, JZ, SP; Resources: MM, CB, ED, LEW; Data curation: FS, LEW; Writing - original draft: FS, LEW; Writing - review and editing: FS, GR, SP, ED, LEW; Visualization: FS, JZ, GR, ED; Supervision: LEW; Project administration: LEW; Funding acquisition: MM, CB, ED, LEW.

Competing interest statement

The authors declare no competing interest.

Bibliography

1. Hao Shen, Lawrence J Tauzin, Rashad Baiyasi, Wenxiao Wang, Nicholas Moringo, Bo Shuang, and Christy F Landes. Single particle tracking: from theory to biophysical applications. *Chemical reviews*, 117(11):7331–7376, 2017.
2. Achillefs N Kapanidis, Stephan Uphoff, and Mathew Stracy. Understanding protein mobility in bacteria by tracking single molecules. *Journal of molecular biology*, 430(22):4443–4455, 2018.
3. Zhao Wang, Xuejing Wang, Ying Zhang, Weili Xu, and Xiaojun Han. Principles and applications of single particle tracking in cell research. *Small*, 17(11):2005133.
4. Michael J Saxton. A biological interpretation of transient anomalous subdiffusion. i. qualitative model. *Biophysical journal*, 92(4): 1178–1191, 2007.
5. Felix Höfling and Thomas Franosch. Anomalous transport in the crowded world of biological cells. *Reports on Progress in Physics*, 76(4): 046602, 2013.
6. Ralf Metzler, Jae-Hyung Jeon, Andrey G Cherstvy, and Eli Barkai. Anomalous diffusion models and their properties: non-stationarity, non-ergodicity, and ageing at the centenary of single particle tracking. *Physical Chemistry Chemical Physics*, 16(44):24128–24164, 2014.
7. Vincent Briane, Charles Kervrann, and Myriam Vimond. Statistical analysis of particle trajectories in living cells. *Physical Review E*, 97(6): 062121, 2018.
8. Berglund AJ. Michalet X. Optimal diffusion coefficient estimation in single-particle tracking. *Phys Rev E Stat Nonlin Soft Matter Phys*, 2012.
9. Florian Rehfeldt and Matthias Weiss. The random walker's toolbox for analyzing single-particle tracking data. *Soft Matter*, 19:5206–5222, 2023.
10. Krzysztof Burnecki, Eldad Kepten, Joanna Janczura, Irena Bronshtein, Yuval Garini, and Aleksander Weron. Universal algorithm for identification of fractional brownian motion. a case of telomere subdiffusion. *Biophysical journal*, 103(9):1839–1847, 2012.
11. Maxime Woringer, Ignacio Izeddin, Cyril Favard, and Hugues Berry. Anomalous subdiffusion in living cells: bridging the gap between experiments and realistic models through collaborative challenges. *Frontiers in Physics*, 8:134, 2020.
12. Benoit B Mandelbrot and John W Van Ness. Fractional brownian motions, fractional noises and applications. *SIAM review*, 10(4):422–437, 1968.
13. Harvey Scher and Elliott W Montroll. Anomalous transit-time dispersion in amorphous solids. *Physical Review B*, 12(6):2455, 1975.
14. JW Haus and KW Kehr. Random walk model with correlated jumps: Self-correlation function and frequency-dependent diffusion coefficient. *Journal of Physics and Chemistry of Solids*, 40(12):1019–1025, 1979.
15. Pietro Massignan, Carlo Manzo, Juan A Torreno-Pina, Maria F García-Parajo, Maciej Lewenstein, and GJ Lapeyre Jr. Nonergodic subdiffusion from brownian motion in an inhomogeneous medium. *Physical review letters*, 112(15):150603, 2014.
16. Gorka Muñoz-Gil, Giovanni Volpe, Miguel Angel Garcia-March, Erez Aghion, Aykut Argun, Chang Beom Hong, Tom Bland, Stefano Bo, J Alberto Conejero, Nicolás Firbas, et al. Objective comparison of methods to decode anomalous diffusion. *Nature communications*, 12(1): 6253, 2021.
17. Naor Granik, Lucien E Weiss, Elias Nehme, Maayan Levin, Michael Chein, Eran Perlson, Yael Roichman, and Yoav Shechtman. Single-particle diffusion characterization by deep learning. *Biophysical journal*, 117(2):185–192, 2019.
18. Ziyuan Chen, Laurent Geffroy, and Julie Suzanne Biteen. Nobias: Analyzing anomalous diffusion in single-molecule tracks with nonparametric bayesian inference. *Frontiers in bioinformatics*, page 40, 2021.
19. Andrew J Berglund. Statistics of camera-based single-particle tracking. *Physical Review E*, 82(1):011917, 2010.
20. Peter K Relich, Mark J Olah, Patrick J Cutler, and Keith A Lidke. Estimation of the diffusion constant from intermittent trajectories with variable position uncertainties. *Physical Review E*, 93(4):042401, 2016.
21. François Simon, Jean-Yves Tinevez, and Sven van Teeffelen. Extrack characterizes transition kinetics and diffusion in noisy single-particle tracks. *Journal of Cell Biology*, 222(5):e202208059, 2023.
22. Nilah Monnier, Zachary Barry, Hye Yoon Park, Kuan-Chung Su, Zachary Katz, Brian P English, Arkajit Dey, Keyao Pan, Iain M Cheeseman, Robert H Singer, et al. Inferring transient particle transport dynamics in live cells. *Nature Methods*, 12(9):838, 2015.
23. Silvan Türkcan, Antigoni Alexandrou, and Jean-Baptiste Masson. A bayesian inference scheme to extract diffusivity and potential fields from confined single-molecule trajectories. *Biophysical journal*, 102(10):2288–2298, 2012.
24. Silvan Türkcan and Jean-Baptiste Masson. Bayesian decision tree for the classification of the mode of motion in single-molecule trajectories. *PloS one*, 8(12):e82799, 2013.
25. Christopher P Calderon. Data-driven techniques for detecting dynamical state changes in noisily measured 3d single-molecule trajectories. *Molecules*, 19(11):18381–18398, 2014.
26. Jason Bernstein and John Fricks. Analysis of single particle diffusion with transient binding using particle filtering. *Journal of theoretical biology*, 401:109–121, 2016.
27. Paddy J Slator and Nigel J Burroughs. A hidden markov model for detecting confinement in single-particle tracking trajectories. *Biophysical journal*, 115(9):1741–1754, 2018.

28. Amit Federbush, Amit Moscovich, and Yohai Bar-Sinai. Hidden markov modeling of single-particle diffusion with stochastic tethering. *Phys. Rev. E*, 109:034129, Mar 2024.
29. Felix Oswald, Aravindan Varadarajan, Holger Lill, Erwin JG Peterman, and Yves JM Bollen. Mreb-dependent organization of the e. coli cytoplasmic membrane controls membrane protein diffusion. *Biophysical journal*, 110(5):1139–1149, 2016.
30. Samuel S Wilks. The large-sample distribution of the likelihood ratio for testing composite hypotheses. *The annals of mathematical statistics*, 9(1):60–62, 1938.
31. Aad W Van der Vaart. *Asymptotic statistics*, volume 3. Cambridge university press, 2000.
32. Paolo Pierobon, Sarra Achouri, Sébastien Courty, Alexander R Dunn, James A Spudich, Maxime Dahan, and Giovanni Cappello. Velocity, processivity, and individual steps of single myosin v molecules in live cells. *Biophysical journal*, 96(10):4268–4275, 2009.
33. Sven Van Teeffelen, Siyuan Wang, Leon Furchtgott, Kerwyn Casey Huang, Ned S Wingreen, Joshua W Shaevitz, and Zemer Gitai. The bacterial actin mreb rotates, and rotation depends on cell-wall assembly. *Proceedings of the National Academy of Sciences*, 108(38): 15822–15827, 2011.
34. Aaron D Pilling, Dai Horiuchi, Curtis M Lively, and William M Saxton. Kinesin-1 and dynein are the primary motors for fast transport of mitochondria in drosophila motor axons. *Molecular biology of the cell*, 17(4):2057–2068, 2006.
35. Christopher P Calderon, Lucien E Weiss, and WE Moerner. Robust hypothesis tests for detecting statistical evidence of two-dimensional and three-dimensional interactions in single-molecule measurements. *Physical Review E*, 89(5):052705, 2014.
36. Cédric Bouzigues and Maxime Dahan. Transient directed motions of gabaa receptors in growth cones detected by a speed correlation index. *Biophysical journal*, 92(2):654–660, 2007.
37. Hong Qian, Michael P Sheetz, and Elliot L Elson. Single particle tracking. analysis of diffusion and flow in two-dimensional systems. *Biophysical journal*, 60(4):910–921, 1991.
38. Frank J. Fazekas, Thomas R. Shaw, Sumin Kim, Ryan A. Bogucki, and Sarah L. Veatch. A mean shift algorithm for drift correction in localization microscopy. *Biophysical Reports*, 1(1):100008, 2021. ISSN 2667-0747.
39. Martin Lindén and Johan Elf. Variational algorithms for analyzing noisy multistate diffusion trajectories. *Biophysical journal*, 115(2): 276–282, 2018.
40. Hirotugu Akaike. A new look at the statistical model identification. *IEEE transactions on automatic control*, 19(6):716–723, 1974.
41. Aykut Argun, Giovanni Volpe, and Stefano Bo. Classification, inference and segmentation of anomalous diffusion with recurrent neural networks. *Journal of Physics A: Mathematical and Theoretical*, 54(29):294003, 2021.
42. Jason A Kahana, Bruce J Schnapp, and Pamela A Silver. Kinetics of spindle pole body separation in budding yeast. *Proceedings of the National Academy of Sciences*, 92(21):9707–9711, 1995.
43. Kunalika Jain, Neha Khetan, Shivani A Yadav, Saravanan Palani, and Chaitanya A Athale. Collective dynein transport of the nucleus by pulling on astral microtubules during saccharomyces cerevisiae mitosis. *Yeast*, 38(6):352–366, 2021.
44. Thomas R. Powers. Role of body rotation in bacterial flagellar bundling. *Phys. Rev. E*, 65:040903, Apr 2002.
45. Ido Golding and Edward C Cox. Physical nature of bacterial cytoplasm. *Physical review letters*, 96(9):098102, 2006.
46. Aubrey V Weigel, Blair Simon, Michael M Tamkun, and Diego Krapf. Ergodic and nonergodic processes coexist in the plasma membrane as observed by single-molecule tracking. *Proceedings of the National Academy of Sciences*, 108(16):6438–6443, 2011.
47. Jens Ehrig, Eugene P Petrov, and Petra Schwille. Near-critical fluctuations and cytoskeleton-assisted phase separation lead to subdiffusion in cell membranes. *Biophysical journal*, 100(1):80–89, 2011.
48. Keith J Mickolajczyk and William O Hancock. Kinesin processivity is determined by a kinetic race from a vulnerable one-head-bound state. *Biophysical journal*, 112(12):2615–2623, 2017.
49. Maziar Raissi, Paris Perdikaris, and George E Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational physics*, 378:686–707, 2019.
50. François Simon, Lucien E Weiss, and Sven van Teffelen. Which way to go: a guide to single-particle tracking analysis methods. *Under review*, 2024.
51. Lucien E Weiss, Julia F Love, Joshua Yoon, Colin J Comerci, Ljiljana Milenkovic, Tomoharu Kanie, Peter K Jackson, Tim Stearns, and Anna-Karin Gustavsson. Single-molecule imaging in the primary cilium. In *Methods in Cell Biology*, volume 176, pages 59–83. Elsevier, 2023.
52. Vincent Briane, Myriam Vimond, Cesar Augusto Valades-Cruz, Antoine Salomon, Christian Wunder, and Charles Kervrann. A sequential algorithm to detect diffusion switching along intracellular particle trajectories. *Bioinformatics*, 36(1):317–329, 2020.
53. Carlo Manzo. Extreme learning machine for the characterization of anomalous diffusion from single trajectories (andi-elm). *Journal of Physics A: Mathematical and Theoretical*, 54(33):334002, jul 2021.
54. Robert B Davies and David S Harte. Tests for hurst effect. *Biometrika*, 74(1):95–101, 1987.
55. Janka Zsok, Francois Simon, Göksu Bayrak, Luljeta Isaki, Nina Kerff, Amy Wolstenholme, Lucien E Weiss, and Elisa Dultz. The nuclear basket regulates the distribution and mobility of nuclear pore complexes in budding yeast. *bioRxiv*, 2023.
56. Caroline A Schneider, Wayne S Rasband, and Kevin W Eliceiri. Nih image to imagej: 25 years of image analysis. *Nature methods*, 9(7): 671–675, 2012.
57. Jean-Yves Tinevez, Nick Perry, Johannes Schindelin, Genevieve M Hoopes, Gregory D Reynolds, Emmanuel Laplantine, Sebastian Y Bednarek, Spencer L Shorte, and Kevin W Eliceiri. Trackmate: An open and extensible platform for single-particle tracking. *Methods*, 115:80–90, 2017.
58. Jennyfer Zapata-Farfan, Morteza Hasanzadeh Kafshgari, Sergiy Patskovsky, and Michel Meunier. Dynamic multispectral detection of bacteria with nanoplasmonic markers. *Nanoscale*, 15(7):3309–3317, 2023.

Supplementary information

Derivations

Likelihood calculation for confined motion

To simplify the integration of the product of Gaussians described in the Method section, Eq. (), $f_\sigma((1-l)z_i + lh_i - c_{i+1})\, f_d((1-l)z_i + lh_i - z_{i+1})$ can be refactored to $K(z_{i+1} - c_{i+1})\, G'((1-l)z_i + lh_i - \frac{\sigma^2}{\sigma^2+d^2} z_{i+1} - \frac{d^2}{\sigma^2+d^2} c_{i+1})$ of variances $k^2 = \sigma^2 + d^2$ and $g'^2 = \sigma^2 d^2/(\sigma^2 + d^2)$. This can be further simplified into $\frac{1}{(1-l)}\, K(z_{i+1} - c_{i+1})\, G(z_i + \frac{l}{1-l} h_i - a z_{i+1} - b c_{i+1})$ with $a = \frac{\sigma^2}{(1-l)(\sigma^2+d^2)}$, $b = \frac{d^2}{(1-l)(\sigma^2+d^2)}$, K and G represent Gaussian distributions of mean 0 with variances $k^2 = \sigma^2 + d^2$ and $g^2 = \frac{\sigma^2 d^2}{(1-l)^2(\sigma^2+d^2)}$, respectively. Then, we use the annotation $N(x, V)$ to refer to other Gaussian probability density functions of mean 0 and of variance V.

Eq. (1) is equivalent to:

$$
\begin{aligned}
& f_\sigma(r_0 - c_0)\, f_d(z_0 - r_0)\, f_{h_0}(h_0 - c_0) \\
& \frac{1}{(1-l)}\, K(z_1 - c_1)\, G(z_0 + \frac{l}{1-l} h_0 - a z_1 - b c_1)\, f_q(h_1 - h_0) \\
& \frac{1}{(1-l)}\, K(z_2 - c_2)\, G(z_1 + \frac{l}{1-l} h_1 - a z_2 - b c_2)\, f_q(h_2 - h_1) \\
& [\ldots] \\
& \frac{1}{(1-l)}\, K(z_{n-1} - c_{n-1})\, G(z_{n-2} + \frac{l}{1-l} h_{n-2} - a z_{n-1} - b c_{n-1})\, f_q(h_{n-1} - h_{n-2}) \\
& \frac{1}{(1-l)}\, f'_\sigma(z_{n-1} + \frac{l}{1-l} h_{n-1} - \frac{c_n}{1-l})
\end{aligned}
\tag{4}
$$

During the integration process, we obtain the following set of recurrence formulas:
- Initialization:

$$
\begin{aligned}
& \int \text{ over } r_0: \\
& \qquad \int f_\sigma(r_0 - c_0)\, f_d(z_0 - r_0)\, \mathrm{d}r_0 = K(z_0 - c_0) \\
& \int \text{ over } z_0: \\
& \qquad s_0^2 = \left(\frac{1-l}{l}\right)^2 (k^2 + g^2), \quad \zeta_0 = \frac{a}{l}, \quad \eta_0 = \frac{b}{l}\, c_1 - \frac{1-l}{l}\, c_0 \\
& \int \text{ over } h_0: \\
& \qquad x_0^2 = q_0^2 + q^2, \quad \alpha_0 = \frac{q_0^2}{\zeta_0\, x_0^2}, \quad \gamma_0 = \frac{-\eta_0 + \frac{q^2}{x_0^2}\, c_0}{\zeta_0} \\
& \qquad w_0^2 = \frac{s_0^2\, q^2 + q^2\, q_0^2 + q_0^2\, s_0^2}{\zeta_0^2\, x_0^2}, \quad u_0 = c_0 \\
& P_0 = \frac{1-l}{l\, \zeta_0}
\end{aligned}
\tag{5}
$$

- Recurrence:

$$\int \text{ over } z_i:$$

$$m_i'^2 = k^2 + w_{i-1}^2,\ \tau_i = \frac{l}{1-l} + \frac{k^2}{m_i'^2}\alpha_{i-1}$$

$$s_i^2 = \frac{k^2\ w_{i-1}^2 + w_{i-1}^2\ g^2 + g^2\ k^2}{m_i'^2\ \tau_i^2}$$

$$\zeta_i = \frac{a}{(1-l)\ \tau_i}$$

$$\eta_i = \frac{1}{\tau_i}\left(\frac{b}{1-l}\ c_{i+1} - \frac{k^2}{m_i'^2}\ \gamma_{i-1} - \frac{w_{i-1}^2}{m_i'^2}\ c_i\right)$$

$$m_i^2 = \frac{m_i'^2}{\alpha_{i-1}^2}$$

$$\int \text{ over } h_i: \tag{6}$$

$$e_i^2 = x_{i-1}^2 + m_i^2$$

$$u_i = \frac{x_{i-1}^2}{\alpha_{i-1}\ e_i^2}\ (c_i - \gamma_{i-1}) + \frac{m_i^2}{e_i^2}\ u_{i-1},\ \phi_i^2 = \frac{x_{i-1}^2\ m_i^2}{e_i^2}$$

$$x_i^2 = \phi_i^2 + q^2$$

$$w_i^2 = \frac{\phi_i^2\ q^2 + q^2\ s_i^2 + s_i^2\ \phi_i^2}{x_i^2\ \zeta_i^2}$$

$$\alpha_i = \frac{\phi_i^2}{x_i^2\ \zeta_i},\ \gamma_i = \frac{\frac{q^2}{x_i^2}\ u_i - \eta_i}{\zeta_i}$$

$$P_i = P_{i-1} \cdot \frac{1}{\zeta_i\ \alpha_{i-1}\ (1-l)\ \tau_i} \cdot N\left(\frac{c_i - \gamma_{i-1}}{\alpha_{i-1}} - u_{i-1},\ e_i^2\right)$$

- Final step:

$$\int \text{ over } z_{n-1}:$$

$$m_{n-1}'^2 = k^2 + w_{n-2}^2$$

$$m_{n-1}^2 = m_{n-1}'^2/\alpha_{n-2}^2$$

$$\tau_{n-1} = \frac{l}{1-l} + \alpha\ \frac{k^2}{m_{n-1}'^2}$$

$$\eta_{n-1} = \frac{1}{\tau_{n-1}}\left(\frac{c_n}{1-l} - \frac{k^2}{m_{n-1}'^2}\ \gamma_{n-1} - \frac{w^2}{m_{n-1}'^2}\ c_{n-1}\right)$$

$$s_{n-1}^2 = \frac{k^2\ w_{i-1}^2 + w_{i-1}^2\ \frac{\sigma^2}{(1-l)^2} + \frac{\sigma^2}{(1-l)^2}\ k^2}{m_{n-1}'^2\ \tau_{n-1}^2} \tag{7}$$

$$\int \text{ over } h_{n-1}:$$

$$v^2 = x_{n-1}^2 + s_{n-1}^2$$

$$o^2 = \frac{x_{n-1}^2\ s_{n-1}^2 + s_{n-1}^2\ m_{n-1}^2 + m_{n-1}^2\ x_{n-1}^2)}{v^2}$$

$$\mu = \frac{s_{n-1}^2\ u_{n-2} + x_{n-1}^2\ \eta_{n-1}}{v^2}$$

$$P_{n-1} = P_{n-2} \cdot N\left(u_{n-2} - \eta_{n-1},\ v^2\right) \cdot N\left(\frac{c_{n-2} - \gamma_{n-2}}{\alpha_{n-2}} - \mu,\ o^2\right) \cdot \frac{1}{\alpha_{n-2}\ (1-l)\ \tau_{n-1}}$$

Likelihood calculation for directed motion

Similarly to the confined motion formula, the track probability can be expressed using a recurrence formula that depends on the parameters of the directed motion model, the localization error σ, the diffusion length d, the initial velocity v and the speed of the velocity change q.

- Initialization:

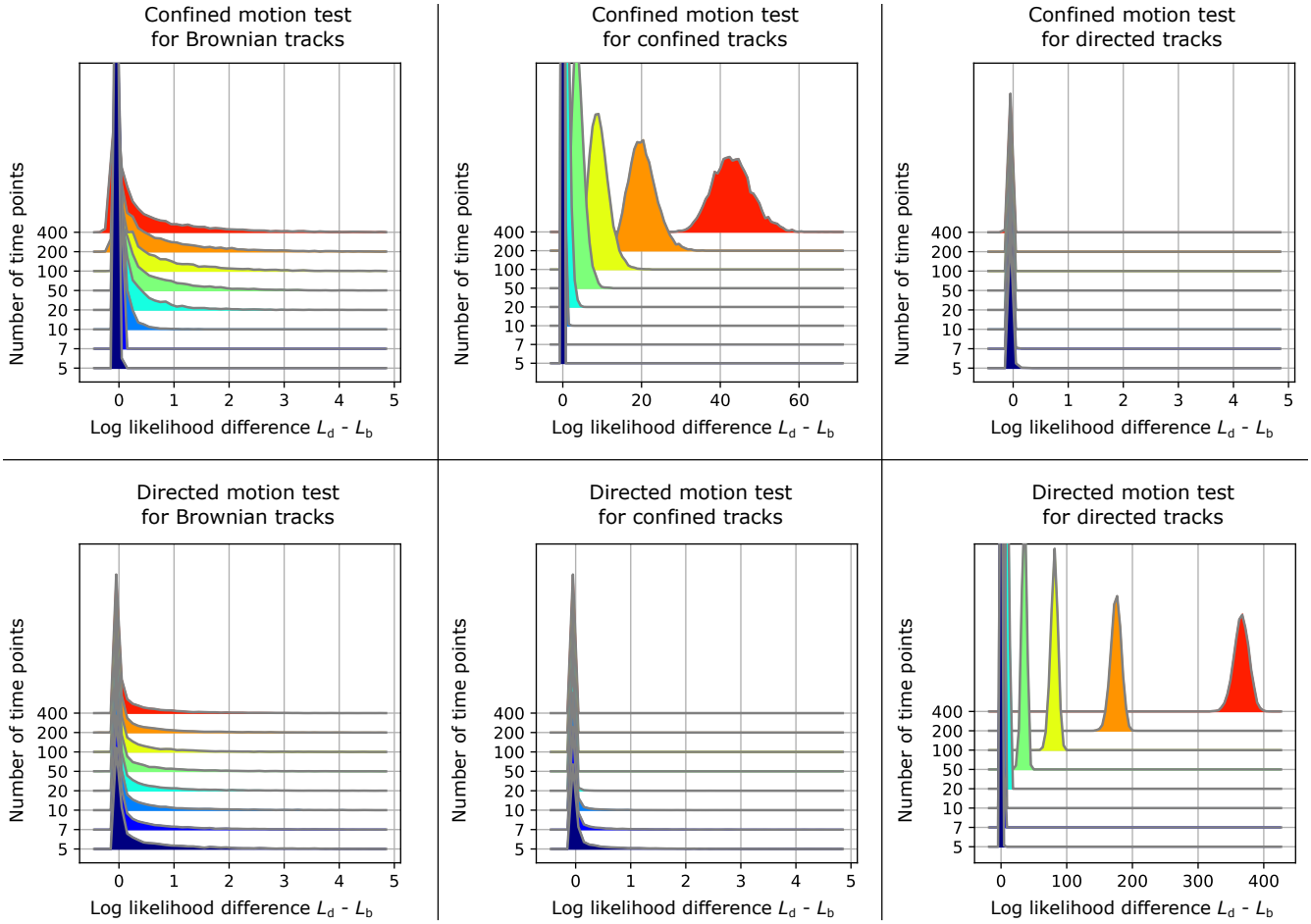
$$
\begin{aligned}
&\text{Step 0:}\\
&\quad s_0^2 = \sigma^2 + d^2\\
&\quad x_0^2 = v^2 + q^2\\
&\quad g_0^2 = \frac{q^2 s_0^2 + s_0^2 * v^2 + v^2 * q^2}{x_0^2}\\
&\quad u_0^2 = \sigma^2 + g_0^2\\
&\quad \alpha_0 = v^2 / x_0^2\\
&\quad \gamma_0 = c_0\\
&\quad h_0^2 = v^2\\
&\text{Step 1:}\\
&\quad k_1 = \frac{c_1 - \gamma_0}{\alpha_0}\\
&\quad h_1^2 = \frac{g_0^2 + \sigma^2}{\alpha_0^2}\\
&\quad a_1 = \frac{\sigma^2 \alpha_0}{u_0^2} + 1\\
&\quad b_0 = \frac{g_0^2 c_1 + \sigma^2 \gamma_0}{u_0^2}\\
&\quad s_1^2 = \frac{\sigma^2 g_0^2 + g_0^2 d^2 + d^2 \sigma^2}{u_0^2 a_i^2}\\
&\quad \beta_1 = \frac{x_0^2}{x_0^2 + q^2}\\
&\quad t_1^2 = \beta_1 q^2\\
&\quad x_1^2 = x_0^2 + q^2\\
&\quad \alpha_1 = \frac{a_1 \beta_1 h_0^2}{h_0^2 + t_1^2}\\
&\quad \gamma_1 = b_0 + \frac{a_1 t_1^2 k_0}{h_0^2 + t_1^2}\\
&\quad g_1^2 = a_i^2 \frac{t_1^2 s_1^2 + s_1^2 h_0^2 + h_0^2 t_1^2}{h_0^2 + t_1^2}\\
&\quad o_1 = \frac{k_1}{\beta_1}\\
&\quad q_1^2 = \frac{h_1^2 + t_1^2}{\beta_1^2}\\
&\quad P_1 = \frac{1}{\alpha_0 \beta_1}
\end{aligned}
\tag{8}
$$

Recurrence:

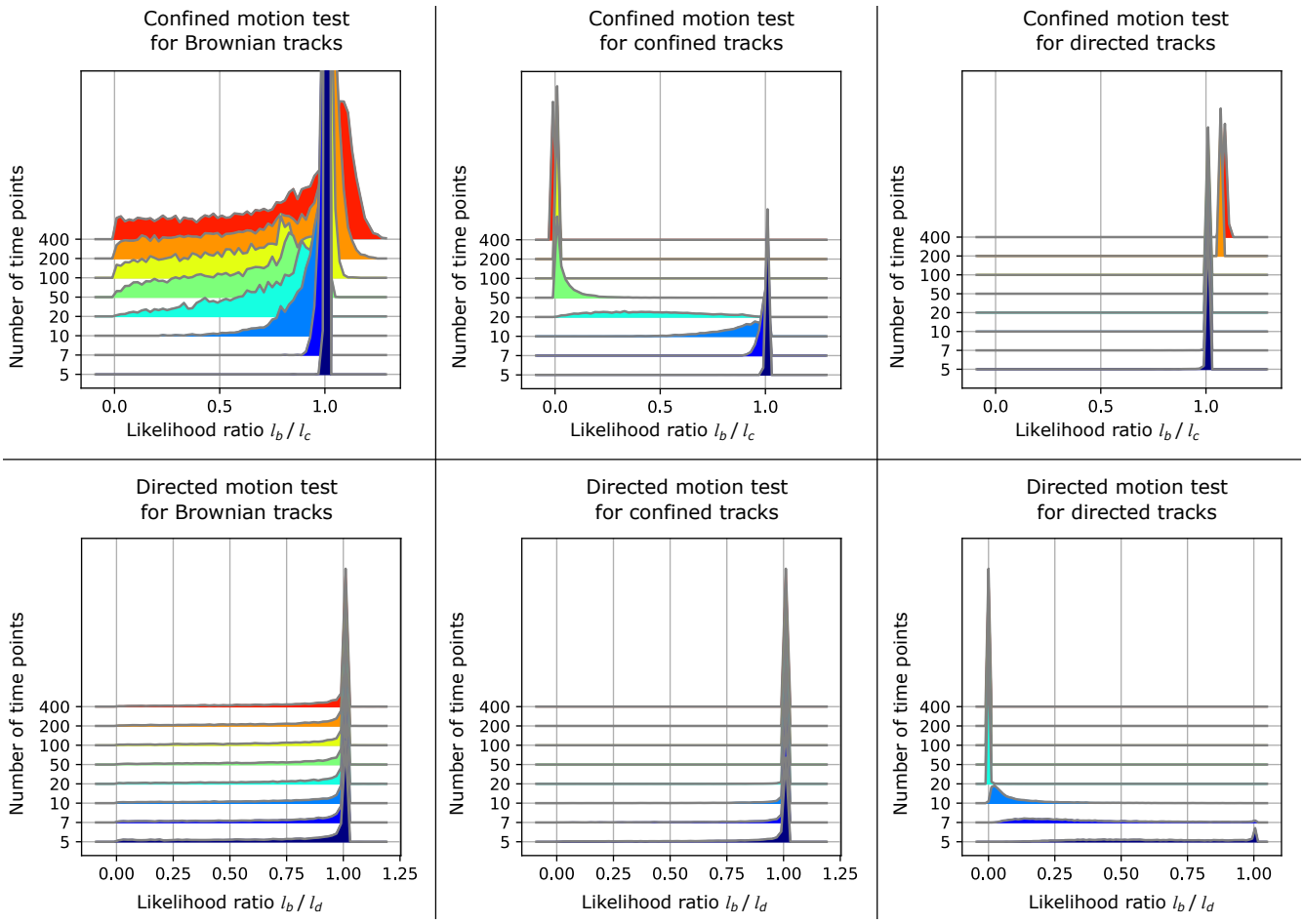
$$
\begin{aligned}
u_i &= \frac{c_i - \gamma_{i-1}}{\alpha_{i-1}} \\
h_i^2 &= \frac{g_{i-1}^2 + \sigma^2}{\alpha_{i-1}^2} \\
a_i &= \frac{\sigma^2 \alpha_{i-1}}{\sigma^2 + g_{i-1}^2} + 1 \\
b_i &= (g_{i-1}^2 c_i + \sigma^2 \gamma_{i-1}) / (\sigma^2 + g_{i-1}^2) \\
s_i^2 &= \frac{\sigma^2 g_{i-1}^2 + g_{i-1}^2 d^2 + d^2 \sigma^2}{a_i^2 (\sigma^2 + g_{i-1}^2)} \\
k_i &= \frac{h_i^2 o_{i-1} + q_{i-1}^2 u_i}{h_i^2 + q_{i-1}^2} \\
w_i^2 &= \frac{h_i^2 q_{i-1}^2}{h_i^2 + q_{i-1}^2} \\
x_i^2 &= x_{i-1}^2 + q^2 \\
\beta_i &= \frac{x_{i-1}^2}{x_i^2} \\
t_i^2 &= \beta_i q^2 \\
\alpha_i &= \frac{a_i \beta_i w_i^2}{w_i^2 + t_i^2} \\
\gamma_i &= b_i + \frac{a_i t_i^2 ki}{w_i^2 + t_i^2} \\
g_i^2 &= a_i^2 \frac{t_i^2 s_i^2 + s_i^2 w_i^2 + w_i^2 t_i^2}{w_i^2 + t_i^2} \\
o_i &= k_i / \beta_i \\
q_i^2 &= \frac{w_i^2 + t_i^2}{\beta_i^2} \\
P_i &= \frac{P_{i-1}}{\alpha_{i-1} \beta_i} N(u_i - o_{i-1},\ h_i^2 + q_{i-1}^2)
\end{aligned}
\tag{9}
$$

Final steps (steps $i = n-2$ and $i = n-1$):

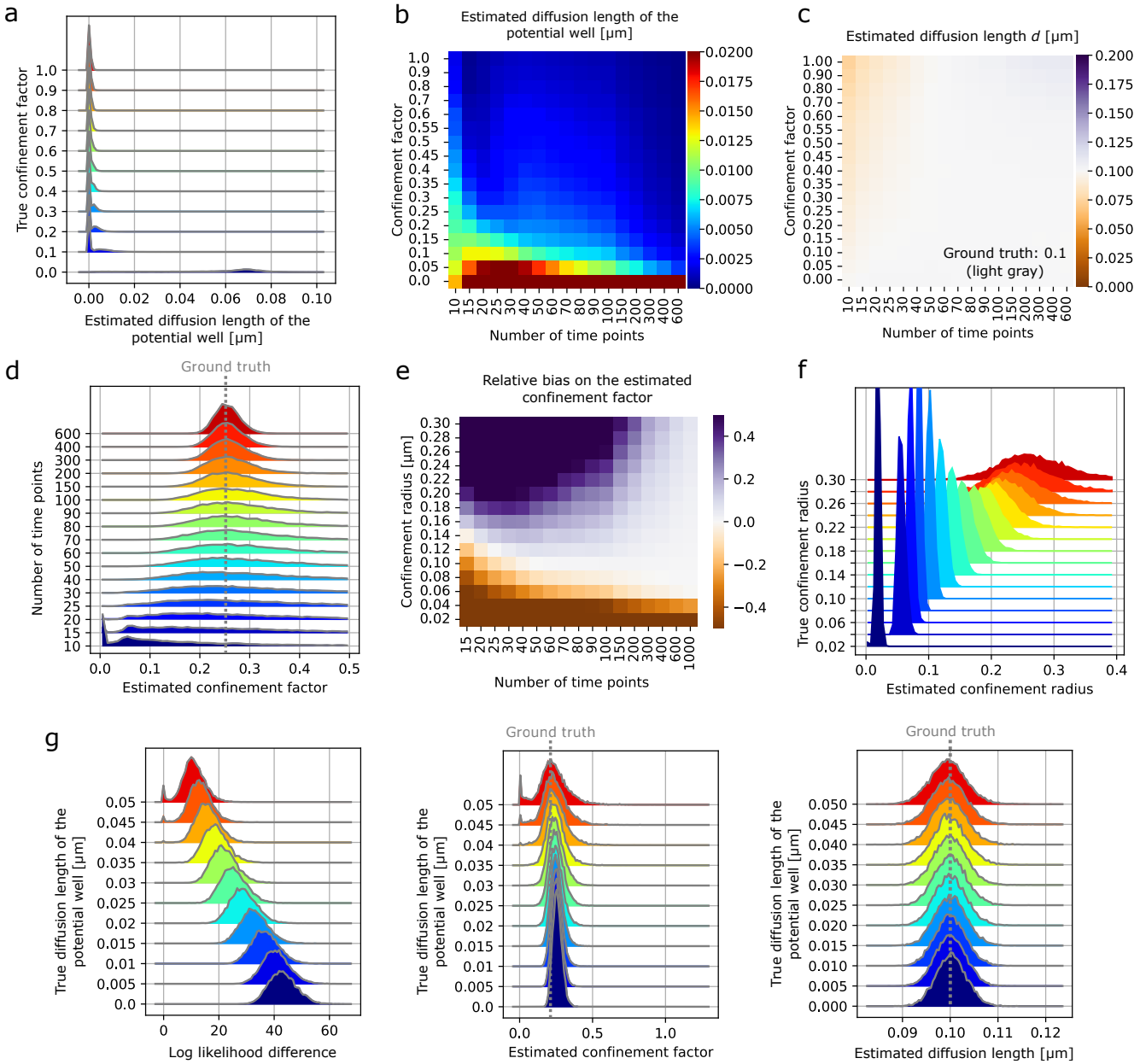
$$
\begin{aligned}
u_{n-2} &= \frac{c_{n-2} - \gamma_{n-3}}{\alpha_{n-3}} \\
h^2_{n-2} &= \frac{\sigma^2 + g^2_{n-3}}{\alpha^2_{n-2}} \\
a_{n-2} &= \frac{\sigma^2 \alpha_{n-3}}{\sigma^2 + g^2_{n-3}} + 1 \\
b_{n-2} &= \frac{g^2_{n-3} c_{n-2} + \sigma^2 \gamma_{n-3}}{\sigma^2 + g^2_{n-3}} \\
s^2_{n-2} &= \frac{\sigma^2 g^2_{n-3} + g^2_{n-3} d^2 + d^2 \sigma^2}{a^2_{n-2}(\sigma^2 + g^2_{n-3})} \\
k_{n-2} &= \frac{h^2_{n-2} o_{n-2} + q^2_{n-3} u_{n-2}}{h^2_{n-1} + q^2_{n-3}} \\
w^2_{n-2} &= \frac{h^2_{n-2} * q^2_{n-3}}{h^2_{n-2} + q^2_{n-3}} \\
y^2_{i-1} &= w^2_{n-2} + x^2_{n-3} \\
P_{n-2} &= \frac{P_{n-3}}{\alpha_{n-2}} \; N(u_{n-2} - o_{n-3},\; h^2_{n-2} + q^2_{n-3}) \; N(k_{n-1},\; w^2_{n-2} + x^2_{n-3}) \\
\gamma_{n-2} &= b_{n-2} + \frac{a_{n-2} x^2_{n-3} k_{n-2}}{w^2_{n-2} + x^2_{n-3}} \\
g^2_{n-2} &= a^2_{n-2} \frac{x^2_{n-3} s^2_{n-2} + s^2_{n-2} w^2_{n-2} + w^2_{n-2} x^2_{n-3}}{w^2_{n-2} + x^2_{n-3}} \\
P_{n-1} &= P_{n-2} \; N(c_{n-1} - \gamma_{n-1},\; \sigma^2 + g^2_{n-1})
\end{aligned}
\tag{10}
$$



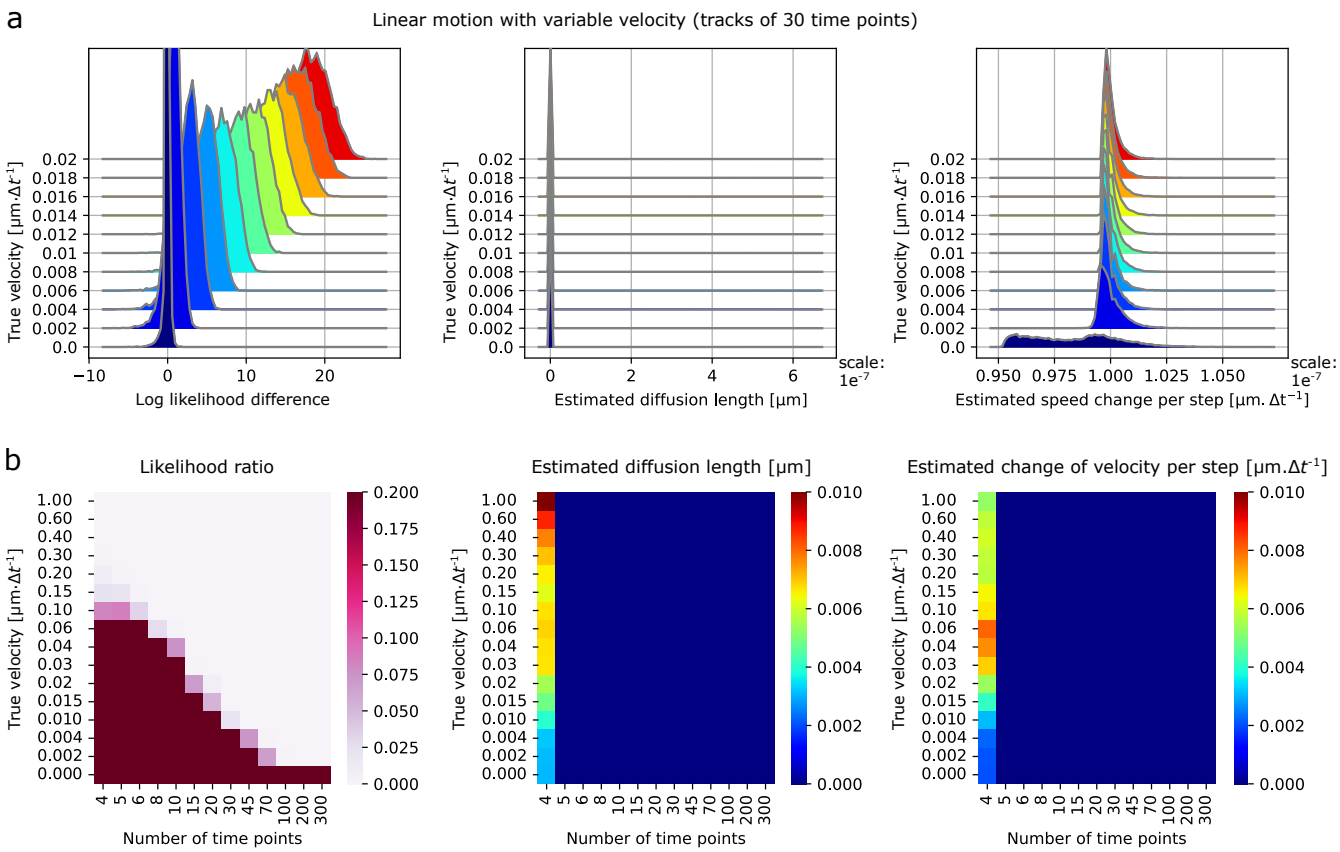
Supplementary Figure 1. Distributions of the log likelihood differences ($L_c - L_b$ and $L_d - L_b$) for tracks in Brownian motion (column 1), confined motion (column 1), or linear motion (column 3) using the confinement motion test (row 1) or the directed motion test (row 2) for tracks with different number of time points. 10,000 tracks per distribution. The simulated track parameters were as following: localization error $\sigma = 0.02$ μm; confined tracks: diffusion length per step $d = 0.1$ μm, confinement factor $l = 0.25$; linear tracks: $d = 0.0$ μm, velocity $v = 0.02$ μm·Δt^{-1}, constant speed and orientation.



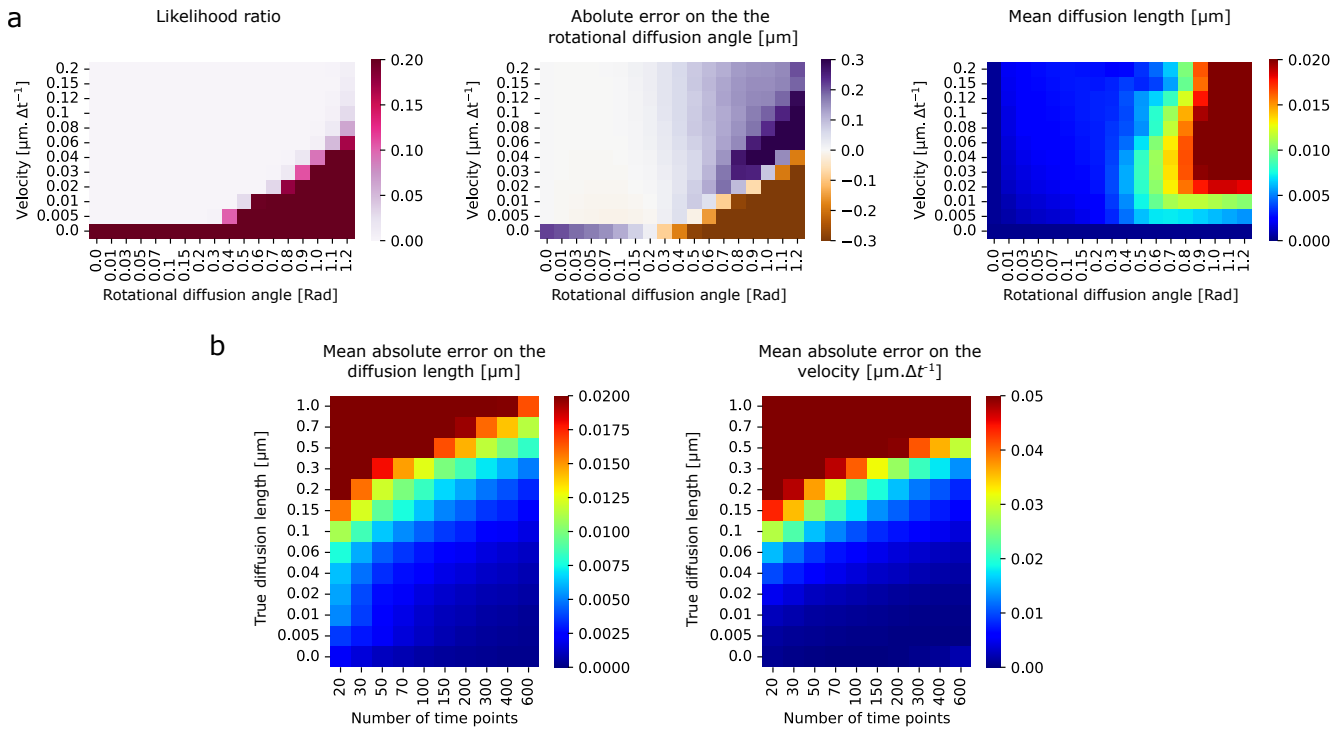
Supplementary Figure 2. Distributions of the likelihood ratios l_b/l_c and l_b/l_d corresponding to Fig. S1. As expected, the distributions are skewed toward 0 only when the proper test is applied.



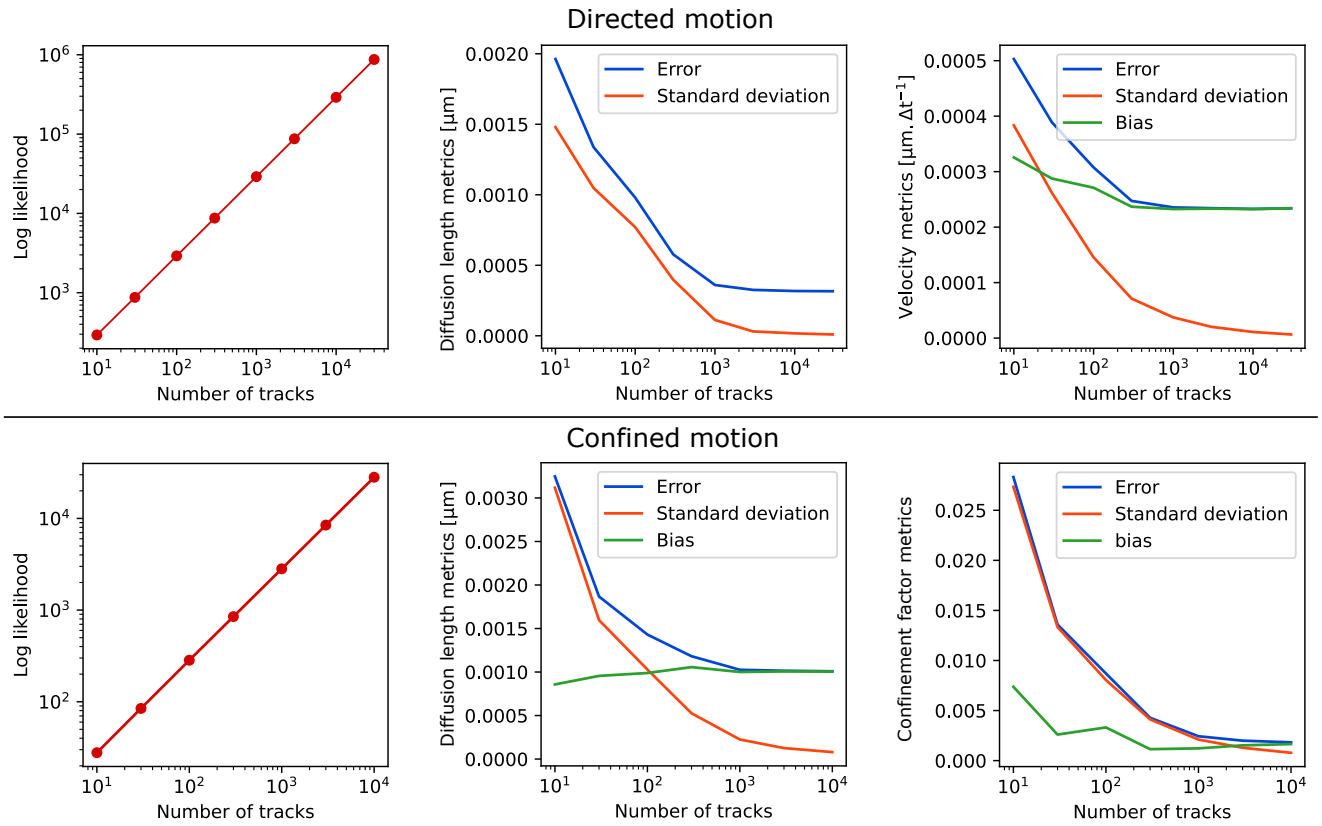
Supplementary Figure 3. **a**: Histograms of the estimated diffusion length per step of the potential well depending the confinement factor corresponding to Fig 3b. **b-c**: Heatmaps of the estimated diffusion lengths per step of the potential well (b) and of the estimated diffusion lengths or the particle depending on the track length and on the confinement factor corresponding to Fig 3c. **d**: Distributions of the estimated confinement factor depending on the track length (same conditions than Fig 3c with a confinement factor of 0.25). **e**: Heatmap of the relative biases on the estimated confinement factors depending on the confinement radius and track length corresponding to Fig 3d. **f**: Distributions of the estimated confinement radius depending on the true confinement radius (same conditions as in Fig 3d with tracks of 150 time points). **g**: Distributions of the log likelihood difference ($L_c - L_b$), estimated confinement factor and estimated diffusion length of the particle depending on the diffusion length of the potential well corresponding to Fig 3d.



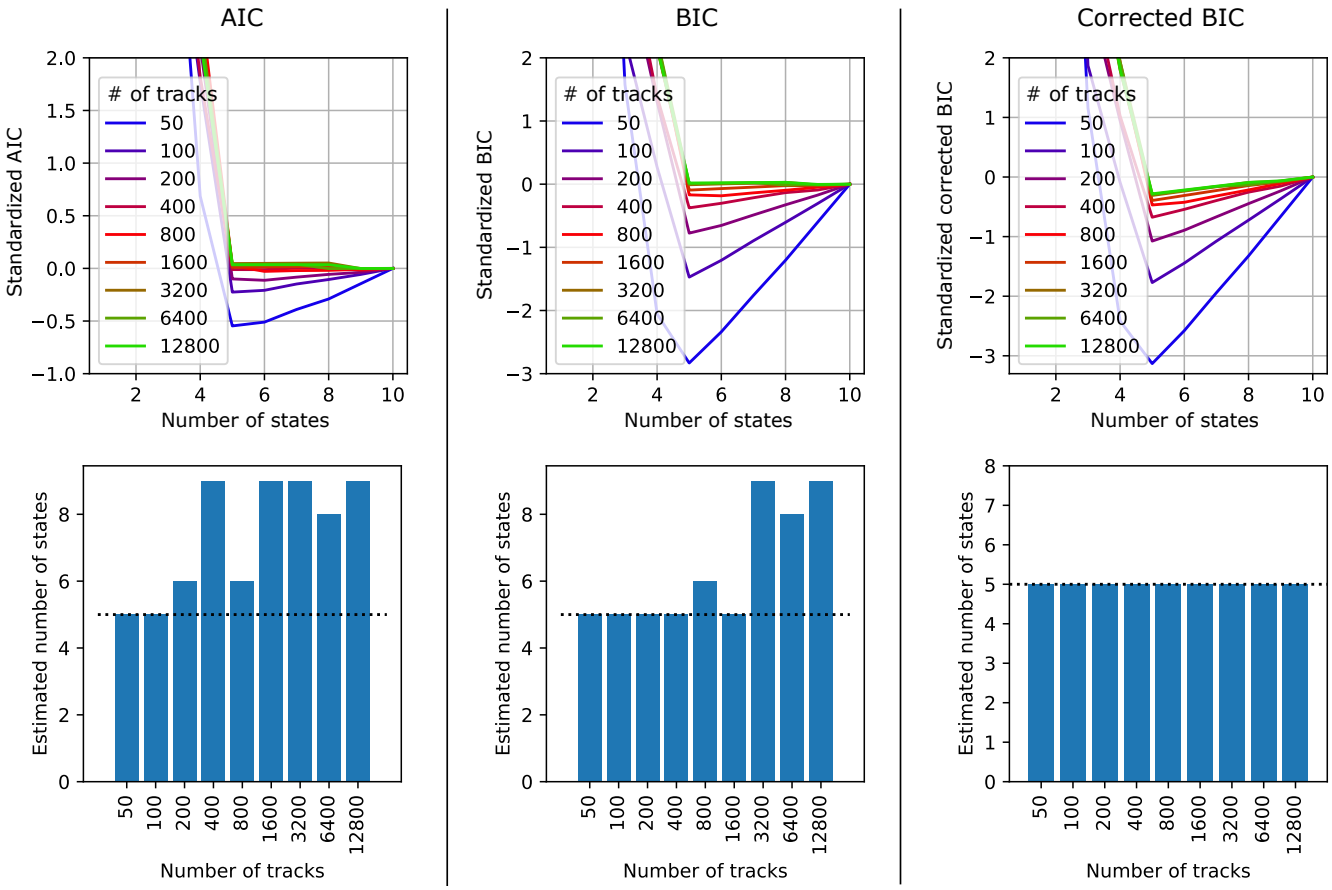
Supplementary Figure 4. **a**: (Complement of Fig 4b) Rainbow plots of the log likelihood difference, estimated diffusion length, and estimated change of velocity for tracks in perfect linear motion (with localization error) for different linear motion velocities. 10,000 tracks per condition. localization error: $\sigma = 0.02$ µm, tracks of 30 time points. **b**: (Complement of Fig 4c) Heatmaps of the likelihood ratio, estimated diffusion length and change of velocity (average) for tracks in perfect linear motion varying the track length and the directed motion velocity. $\sigma = 0.02$ µm.



Supplementary Figure 5. **a**: (Complement of Fig 4e) Study of the impact of the rotational diffusion of tracks in directed motion with changing orientation. Heatmaps of the likelihood ratio, of the absolute error on the rotational diffusion angle, and of the estimated diffusion length when varying the directed motion velocity and the rotational diffusion angle. $d = 0.0$ μm, $\sigma = 0.02$ μm. **b**: (Complement of Fig 4g) Characterization of the motion parameters of particles with both diffusive and directed motion. Mean absolute error on the diffusion length and on the velocity of the linear motion varying the number of time points in each track. Directed motion velocity: $v = 0.1$ μm·Δt^{-1}, $\sigma = 0.02$ μm. **a-b**: mean values from 10,000 tracks.



Supplementary Figure 6. Effect of the number of tracks on the different parameters of the tracks: the likelihood, the root mean squared error, the standard deviation and the bias on the estimates of the diffusion length and anomalous parameter (velocity or confinement factor) for both directed motion and confined motion. All tracks were composed of 50 time points and 50 replicates were performed to estimate the error for each number of tracks. Directed tracks: persistent motion velocity $v = 0.02$ µm·Δt^{-1}, angular diffusion coefficient : 0.1 Rad^2·Δt^{-1}, $d = 0.0$ µm, $\sigma = 0.02$ µm. Confined tracks: confinement factor 0.2, $d = 0.1$ µm, $\sigma = 0.02$ µm.



Supplementary Figure 7. AIC, BIC, and corrected BIC corresponding to the log likelihood shown in Fig 5c depending on the number of states assumed by the model and on the number of tracks per data set. The corrected BIC corresponds to the BIC with an additional penalization term of $0.02kn$ with n the number of tracks and k the number of parameters. Under the AIC, BIC, and corrected BIC curves, we plotted the optimal number of states (the one that minimizes the criterion) for each data set.