SUPPORTING MATERIAL

"Bayesian approach to MSD-based analysis of particle motion in live cells"
N. Monnier, S. Guo, M. Mori, J. He, P. Lénárt, M. Bathe

SUPPLEMENTARY METHODS

Simulated particle trajectories. Diffusive single-particle trajectories were simulated in 3D by drawing random steps in each of the three Cartesian directions from a Gaussian distribution with zero mean and standard deviation equal to $\sqrt{2Ddt}$, where D is the diffusion coefficient and dt is the time interval for each step. Confinement of a diffusing particle was modeled as a reflecting spherical boundary of radius R_C centered at the initial particle position. Directed flow was modeled by adding a fixed displacement $\mathbf{v}dt$ to the diffusive motion at each time step, where $\mathbf{v}$ is the velocity vector.

Fluorescent markers and bead preparation. mEGFP-UtrCH (1) was subcloned into pGEMHE for *in vitro* transcription as described in Lenart et al (2). Capped mRNAs were synthesized from linearized templates using the mMessage mMachine kit (Ambion), and dissolved in 11 µl RNase-free water (typically at 1-2 µg/µl). Streptavidin-coated 0.2 µm diameter green fluorescent microspheres (Invitrogen) were incubated with biotin-PEG-OH (MW 3,000) (Iris Biotech GmbH) and 3% BSA in PBS overnight at RT to make them inert and electrostatically neutral (3).

Chromosome and bead imaging in starfish oocytes. Chromosomes were imaged as previously described (4). For bead imaging, we expressed mEGFP-UtrCH by injecting mRNA (1% of the oocyte volume) into oocytes and incubating overnight. After oocytes expressed sufficient levels of protein to inhibit actin network contraction, beads were injected into oocyte nuclei, and meiotic maturation was triggered by the addition of 10 µM 1-methyladenine (Sigma). Microscopy was done on a Zeiss LSM780 Axiovert confocal microscope using a 40x C-Apochromat 1.2 NA water immersion objective lens.

MSD calculation and error estimation. MSD curves were calculated for each particle trajectory according to Eq. 1. Estimation and regularization of the MSD error covariance matrix $\mathbf{C}$ was carried out as described in Supplementary Note 1 using multiple observations of the MSD values as a function of time lag τ. These multiple observations came either from MSD curves estimated from different particle trajectories or from MSD curves estimated from sub-trajectories within a single particle trajectory, as described in the main text and in Supplementary Note 1.2. The multiple observations of the MSD values were also used to calculate a mean MSD curve for the model fitting and model selection procedure.

Model fitting and model selection. Mean MSD curves obtained from simulated or experimental trajectories were fit with the models given in Eqs. 2-5 (as well as additive combinations of Eq. 5 with Eqs. 2-4) using the Bayesian approach described in the main text. Generalized least squares was implemented by transforming the fitting equation with

a transformation matrix $\mathbf{A}$ equal to the inverse of the Cholesky decomposition of $\mathbf{C}$, such that the resulting equation $\mathbf{Ay} = \mathbf{Af}(\mathbf{x};\boldsymbol{\beta}) + \mathbf{A}\boldsymbol{\epsilon}$ has uncorrelated errors. In this transformed coordinate system, ordinary least squares can be used to obtain the maximum likelihood estimate for the parameters. Estimation of $\widehat{\boldsymbol{\beta}}_{k,MLE}$ and $\widehat{\boldsymbol{\Sigma}}_{k,MLE}$ for each model f_k was carried out in MATLAB (The MathWorks, Inc., Natick, MA) using the function `lsqcurvefit`. The marginal likelihood and model probability calculations were implemented in MATLAB, with the prior probability of the parameters for each model assumed to be uniform over a range $\beta_{kj}^{(min)}$ and $\beta_{kj}^{(max)}$ for each parameter β_{kj} in the set $\boldsymbol{\beta}_k$, such that,

$$P(\boldsymbol{\beta}_k|M_k) = \frac{1}{\prod_j\left(\beta_{kj}^{(max)} - \beta_{kj}^{(min)}\right)}.$$

As in previous work, we chose the ranges $\beta_{kj}^{(min)}$ and $\beta_{kj}^{(max)}$ to be centered at the maximum likelihood estimate of each parameter and to span 200 times the uncertainty (standard deviation) in that parameter (5). Thus, parameters with higher uncertainties reduce the likelihood of a model more than parameters with smaller uncertainties (6, 7).

SUPPLEMENTARY NOTE 1. Noise correlations in MSD curves.

1.1) Origin and effects of correlated noise in MSD curves.

Particle motion typically contains a stochastic diffusive component, with each step in the particle trajectory representing a single observation of the stochastic process. The stochastic nature of diffusion and the finite observation time of the trajectory will result in a deviation of the observed MSD curve from the stationary MSD curve even in the absence of any error in the particle position measurements (localization error). When a trajectory is used to calculate mean square displacement values (as in Eq. 1 of the main text), the same set of steps—grouped into different sized windows—is used to calculate the MSD value at each time lag τ. In other words, the same observations are reused for each calculation. This process leads to strong correlations in the deviations of the calculated MSD values from their expected values (Figure 1A in the main text), with any individual MSD curve deviating from the analytical form of the MSD that is asymptotically correct only in the limit of infinite trajectory length. The deviation of the mean MSD curve from the analytical MSD curve is also correlated in τ because averaging MSD curves improves the estimate of the mean but does not eliminate correlations. For a purely diffusing particle in the absence of localization error, an analytical solution for the noise covariance along the MSD curve has been derived (8, 9) and is shown in Figure 1A in the main text. Note that the magnitude of the noise correlations increases with larger values of τ.

Over-fitting in the presence of correlated errors occurs because some of these correlated fluctuations around the analytical MSD values have shapes that can be fit by the extra parameters of models that are more complex than the true model. For example, the confined diffusion model can fit correlated noise in pure diffusion MSD curves that

causes the curve to trend downward at large time lags, and the diffusion plus flow model can fit correlated noise in pure diffusion MSD curves that causes the curve to trend upward at large time lags. Fitting by weighted least squares, which ignores correlations, rather than by generalized least squares as described in the main text results in a wide range of preferred models (Supplementary Figure 1A). Performing generalized least squares fitting using the theoretical MSD covariance structure for simple diffusion (8, 9) during the model selection process for these pure diffusion simulations eliminates this overfitting, resulting in consistent preference for the true model (pure diffusion) independent of how many single-trajectory MSD curves are averaged to obtain the mean curve (Supplementary Figure 1B). These results demonstrate that accounting for the effect of correlated noise is essential for properly analyzing and interpreting MSD curves, particularly when the true particle motion is a simple model such as pure diffusion. The degree to which correlated noise impacts model selection varies depending on the true type of motion and the set of models included in the model selection process.

1.2) Estimating the sample MSD noise covariance matrix.

The noise covariance matrix may be estimated from multiple independent observations of the MSD curve, yielding the sample covariance matrix **S**. The multiple MSD curves used to calculate **S** must be derived from independent, non-overlapping particle trajectories. As described in the main text, these independent MSD curves can be obtained either from multiple particle trajectories or from splitting a single particle trajectory into multiple non-overlapping and independent sub-trajectories. In the latter case, the full trajectory of $N-1$ steps (N position measurements) is divided into J non-overlapping sub-trajectories of $\lfloor (N-1)/J \rfloor$ steps each. In the case of splitting single particle trajectories there is a tradeoff between accurate estimation of the covariance matrix (which improves with the number of sub-trajectories J) and the time range spanned by the MSD curve, which is limited by the number of steps in the sub-trajectories. Given J independent observations $\{\mathbf{y}^{(j)}\}_{j=1}^{J}$ of MSD values over the same set of time lags, whether from multiple or single trajectories, the residuals between each individual MSD curve and the mean MSD curve are used to estimate the variance and the covariance of the noise in the MSD estimates. For each MSD curve j, the residual at each time lag τ_i is given by,

$$\epsilon_i{}^{(j)} = y_i{}^{(j)} - \bar{y}_i \ , \tag{S1}$$

where $\bar{y}_i = \sum_j y_i{}^{(j)}/J$. The entries in the sample covariance matrix **S** for the mean MSD curve are then equal to,

$$\mathrm{S}_{ii'} = \frac{1}{J(J-1)} \sum_{j=1}^{J} \epsilon_i{}^{(j)} \epsilon_{i'}{}^{(j)} \ . \tag{S2}$$

1.3) Regularizing the sample MSD noise covariance matrix.

When the number of available MSD curves J is less than the number of points in the MSD curve, the sample noise covariance matrix **S** is typically singular, requiring

regularization to obtain a non-singular covariance matrix $\mathbf{C}$ to carry out generalized least squares estimation as described in the main text. We tested multiple regularization methods based on an empirical Bayesian shrinkage approach (10-12). The shrinkage estimator is a linear combination of the sample covariance matrix and a shrinkage target $\mathbf{T}$,

$$\mathbf{S}^* = \lambda\mathbf{T} + (1-\lambda)\mathbf{S}\,, \tag{S3}$$

where λ is the shrinkage weight, which is calculated from the uncertainty in $\mathbf{S}$ as described in Schafer & Strimmer (11). As the number of independent curves J increases, the uncertainty in $\mathbf{S}$ decreases and the shrinkage weight also decreases so that $\mathbf{S}^*$ is closer to $\mathbf{S}$. The shrinkage estimator $\mathbf{S}^*$ is then used as the covariance matrix $\mathbf{C}$ in GLS fitting of the mean MSD curve.

We found that shrinkage to a target that is a diagonal matrix with the mean MSD variance along the diagonal ("Target B" in Schafer & Strimmer (11)) performs best when low numbers of observed MSD curves are available (Supplementary Figure 1C). This method performs nearly as well as using the analytical covariance matrix in the case of normal diffusion (compare Supplementary Figure 1B and 1C). It gives model preferences that are nearly indistinguishable from those obtained using the analytical covariance matrix when 10 or more MSD curves are used and on average continues to prefer the true pure diffusion model down to 4 independent MSD curves (Supplementary Figure 1C), which we found was the minimum number of curves that could be used to reliably obtain a non-singular covariance matrix.

SUPPLEMENTARY NOTE 2. Effect of particle heterogeneity on model selection.

2.1) Heterogeneity in velocity.

For particles undergoing directed motion, the contribution of the directed motion to the MSD curve is given by Eq. 5 in the main text: $MSD_V(\tau) = v^2\tau^2$. If there is heterogeneity in the value of v for each particle, such that each particle j of J total particles has a velocity v_j, then the mean MSD curve takes the form,

$$MSD_V(\tau) = \frac{1}{J}\sum_{j=1}^{J} v_j^2\tau^2 = \frac{1}{J}\tau^2\sum_{j=1}^{J} v_j^2\ . \tag{S4}$$

This mean MSD curve has the same quadratic dependence on τ as the original MSD curve for directed motion, with an effective v for the heterogeneous population given by,

$$v_{eff} = \sqrt{\frac{1}{J}\sum_{j=1}^{J} v_j^2}\ . \tag{S5}$$

In this case, the directed motion model will still be a good fit to the heterogeneous MSD curve, but the v_{eff} parameter is biased toward the larger magnitude velocities in the

particle population and will be greater than the mean velocity in the population (see Figure 3B in the main text). Sub-classification of particles is required to remove the effect of heterogeneity on the observed mean velocity, as described in the main text.

2.2) Heterogeneity in diffusion coefficient.

For particles undergoing pure diffusion, the contribution of the diffusive motion to the MSD curve is given by Eq. 2 in the main text: $MSD_D(\tau) = 6D\tau$. If there is heterogeneity in the value of D for each particle, such that each particle j of J total particles has a diffusion coefficient D_j, then the mean MSD curve takes the form,

$$MSD_D(\tau) = \frac{1}{J}\sum_{j=1}^{J} 6D_j\tau = \frac{1}{J}6\tau\sum_{j=1}^{J} D_j \ . \tag{S6}$$

This mean MSD curve has the same linear dependence on τ as the original MSD curve for pure diffusion, with an effective D for the heterogeneous population given by,

$$D_{eff} = \frac{1}{J}\sum_{j=1}^{J} D_j \ . \tag{S7}$$

In this case, the pure diffusion model will still be a good fit to the heterogeneous MSD curve, and the D_{eff} parameter is equal to the mean diffusion coefficient of the population.

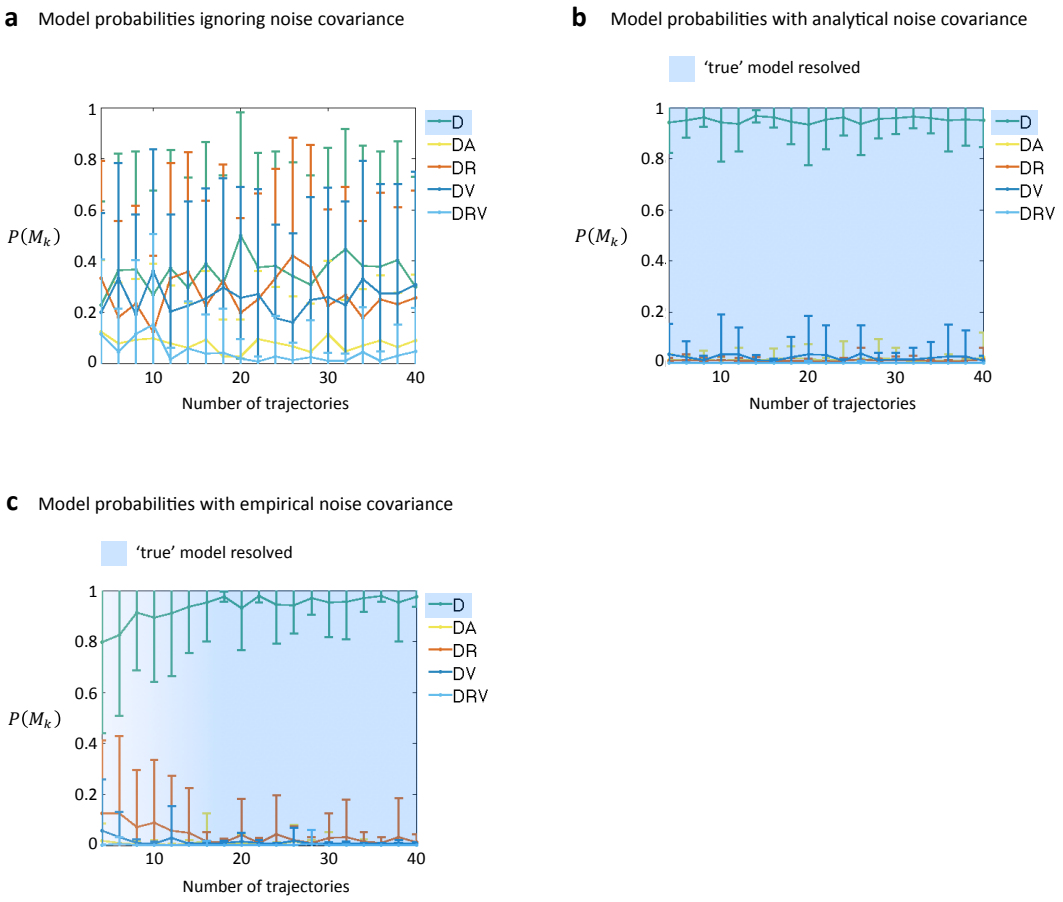
2.3) Heterogeneity in confinement radius.

For particles undergoing confined diffusion, the contribution of the diffusive motion to the MSD curve is given by Eq. 4 in the main text: $MSD_{DR}(\tau) = R_C^2\left(1 - e^{-6D\tau/R_C^2}\right)$. If there is heterogeneity in the value of R_C for each particle, such that each particle j of J total particles has a confinement radius $R_{C\,j}$, then the mean MSD curve takes the form,

$$MSD_{DR}(\tau) = \frac{1}{J}\sum_{j=1}^{J} R_{C\,j}^2\left(1 - e^{-6D\tau/R_{C_j}^2}\right) . \tag{S8}$$

Unlike the models above, this mean MSD curve does not have the same τ dependence as the original MSD curve for confined diffusion and can no longer be described by the confined diffusion model with a single effective confinement radius. At low heterogeneity in R_C, the confined diffusion model may still describe this behavior better than the other tested models if the long-time plateau region in the MSD is still present, in which case this model may still be preferred by Bayesian inference (see Figure 3B in the main text), but with large variability in model parameters due to the increased apparent noise in the mean MSD.

SUPPLEMENTARY FIGURE 1. Accounting for noise covariance in MSD curves.

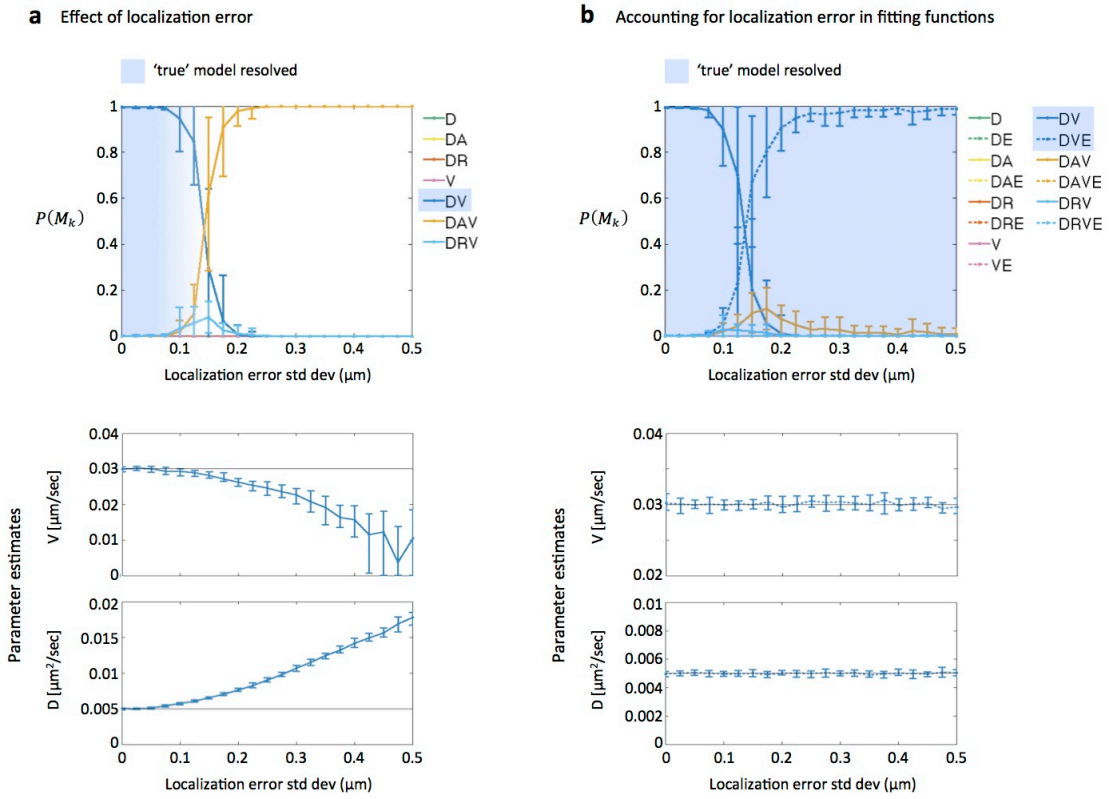


(A) Model probabilities obtained using weighted least squares fitting (ignoring noise covariance). Mean MSD curves with 100 points were calculated by averaging a variable number (shown on the x-axis) of independently simulated trajectories undergoing pure diffusion with the same parameters as in Figure 1A in the main text. The resulting model probabilities are shown as means and standard deviations over 40 repetitions of the simulations and inference procedure. The model probabilities can also be converted into a misclassification rate (bottom panel), where incorrect classification is defined as occurring when the probability of the true D model is less than 0.8 for a given mean MSD curve.

(B) Model probabilities obtained using generalized least squares (GLS) fitting (accounting for noise covariance) of the simulated trajectories undergoing pure diffusion as in (A), using the analytical form of the noise covariance (8, 9) (as shown in Figure 1A of the main text) as the covariance matrix **C**. Light blue shading indicates the range over which the true model (pure diffusion, D) can be resolved given the simulated experimental parameters.

(C) Model probabilities obtained using GLS fitting as in (B), but using the regularized sample noise covariance matrix (see Supplementary Note 1.2 and 1.3) as the covariance matrix $\mathbf{C}$.

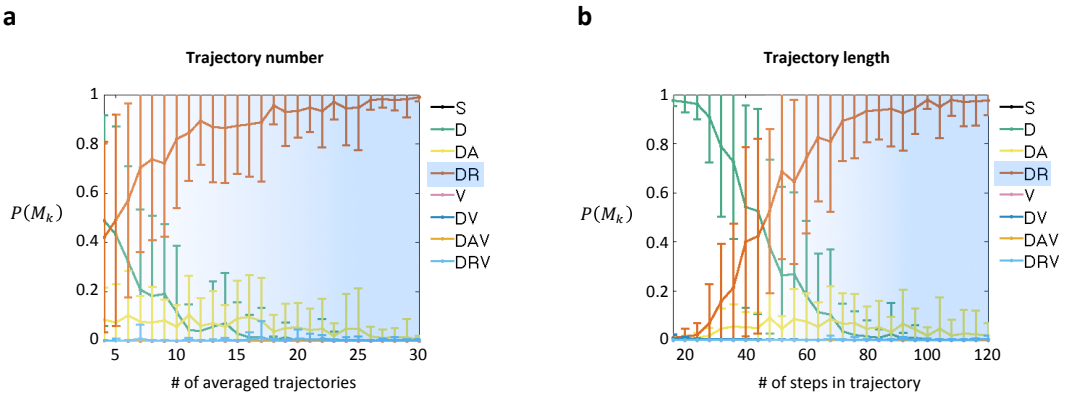
SUPPLEMENTARY FIGURE 2. Effect of localization error on model inference.



(A) Model probabilities for simulated trajectories undergoing diffusion plus flow with time sampling interval dt = 2.5 s, total time T = 300 s, D = 0.005 μm^2/s, and v = 0.03 μm/s. Localization error is also simulated at each time step by drawing an 'observed' particle position from a Gaussian distribution centered on the simulated 'true' position, with standard deviation σ_e as shown along the x-axis. MSD curves with 30 points (up to a $\tau_{\max}$ of 75 s) were calculated for datasets of 30 independently simulated trajectories, the resulting model probabilities are shown as means and standard deviations over 50 repetitions of the simulations and inference procedure, and misclassification rate (bottom panel) is defined using a 0.8 probability threshold as in Figure 2A of the main text. Light blue shading indicates the range of localization errors over which the true model (diffusion plus flow, DV) can be resolved given the simulated experimental parameters. Here only models that do not include a constant term were included in the fitting process to illustrate the effect of ignoring localization error. Note that localization error does not have a significant effect on model probabilities until its contribution to the analytical MSD curve, $MSD_{DVE}(\tau) = 6D\tau + v^2\tau^2 + 6\sigma_e^2$, for diffusion plus flow with localization error (see main text) becomes comparable to the contributions from diffusive motion and directed motion, which occurs when $6\sigma_e^2 \sim 6D\tau + v^2\tau^2$. At higher levels of localization error, an overly-complex model (DAV) is selected because the appropriate models accounting for localization error are not present.

(B) Model probabilities for trajectories simulated as in (A), but now including models with an additional constant term $6\sigma_e^2$ to model localization error (as described in the main text) in the set of fit models. For example, "DVE" refers to a model including a diffusion term, a velocity term, and a localization error term with MSD equation as given in the legend for (A) above. Testing this full set of models with and without the localization error terms restores the ability to resolve the correct physical process (diffusion plus flow) even at high levels of localization error.

SUPPLEMENTARY FIGURE 3. Resolving confined diffusion with sampling limitations.



Additional sampling limitation tests to accompany Figure 3A in the main text. Trajectories are simulated as in Figure 3A (left panel) but at a fixed confinement radius R_C = 1.5 μm and systematically varying one of the sampling parameters. "S" represents a stationary-particle model including only a constant term, as described in the main text.

(A) The number of trajectories per dataset n is varied from 30 down to 4, as in Figure 2B (left panel) in the main text.

(B) The total trajectory time T is varied from 300 s down to 40 s (from 120 steps to 16 steps per trajectory), as in Figure 2B (center panel) in the main text. The number of points in the MSD curves is held constant at 1/4 of the number of steps in the trajectory.

SUPPORTING REFERENCES

1. Burkel, B.M., G. Von Dassow, and W.M. Bement. 2007. Versatile fluorescent probes for actin filaments based on the actin-binding domain of utrophin. Cell Motil Cytoskeleton. 64(11):822-832.
2. Lenart, P., G. Rabut, N. Daigle, A.R. Hand, M. Terasaki, and J. Ellenberg. 2003. Nuclear envelope breakdown in starfish oocytes proceeds by partial NPC disassembly followed by a rapidly spreading fenestration of nuclear membranes. J Cell Biol. 160(7):1055-1068.
3. Daniels, B.R., B.C. Masi, and D. Wirtz. 2006. Probing single-cell micromechanics in vivo: the microrheology of C. elegans developing embryos. Biophys J. 90(12):4712-4719.
4. Mori, M., N. Monnier, N. Daigle, M. Bathe, J. Ellenberg, and P. Lenart. 2011. Intracellular transport by an anchored homogeneously contracting F-actin meshwork. Curr Biol. 21(7):606-611.
5. He, J., S.M. Guo, and M. Bathe. 2012. Bayesian approach to the analysis of fluorescence correlation spectroscopy data I: Theory. Anal Chem. 84(9):3871-3879.
6. Sivia, D.S. and J. Skilling. 2006. Data Analysis: A Bayesian Tutorial. 2nd ed. Oxford University Press, Oxford.
7. Tejedor, V., O. Benichou, R. Voituriez, R. Jungmann, F. Simmel, C. Selhuber-Unkel, L.B. Oddershede, and R. Metzler. 2010. Quantitative analysis of single particle trajectories: Mean maximal excursion method. Biophysical Journal. 98(7):1364-1372.
8. Michalet, X. 2010. Mean square displacement analysis of single-particle trajectories with localization error: Brownian motion in an isotropic medium. Phys Rev E Stat Nonlin Soft Matter Phys. 82:041914.
9. Qian, H., M.P. Sheetz, and E.L. Elson. 1991. Single particle tracking: Analysis of diffusion and flow in two-dimensional systems. Biophys J. 60(4):910-921.
10. Ledoit, O. and M. Wolf. 2004. A well-conditioned estimator for large-dimensional covariance matrices. J Multivariate Anal. 88(2):365-411.
11. Schafer, J. and K. Strimmer. 2005. A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. Stat Appl Genet Mo B. 4(1):32.
12. Guo, S.M., J. He, N. Monnier, G. Sun, T. Wohland, and M. Bathe. 2012. Bayesian approach to the analysis of fluorescence correlation spectroscopy data II: Application to simulated and in vitro data. Anal Chem. 84(9):3880-3888.