Variational algorithms for analyzing noisy multi-state diffusion trajectories

Martin Lindén* and Johan Elf†

Department of Cell and Molecular Biology, Uppsala University, Sweden.

(Dated: March 8, 2018)

Single particle tracking offers a non-invasive high-resolution probe of biomolecular reactions inside living cells. However, efficient data analysis methods that correctly account for various noise sources are needed to realize the full quantitative potential of the method. We report new algorithms for hidden Markov-based analysis of single particle tracking data, which incorporate most sources of experimental noise, including heterogeneous localization errors and missing positions. Compared to previous implementations, the algorithms offer significant speed-ups, support for a wider range of inference methods, and a simple user interface. This will enable more advanced and exploratory quantitative analysis of single particle tracking data.

Experimental techniques to track the conformational and binding states of single biomolecules can offer unique mechanistic insights into life at the molecular level, but increasingly rely on statistical computing to extract quantitative and reproducible results. A simple example is superresolved single particle tracking (SPT) [1], where changes in diffusion constant, or between different modes of motion, offers a non-invasive probe of binding and unbinding reactions in living cells [2, 3].

Accurate quantitative analysis of this type of data requires a faithful account of localization noise, which come in the form of localization errors and motion blur, sometimes referred to as “static” and “dynamic” errors, respectively [4, 5]. In particular, live cell imaging often lead to heterogeneous and asymmetric localization errors, for example due to photobleaching, variability between and across cells, out-of-focus motion, or the dependence of localization errors on the diffusion constant [6, 7]. Several emerging techniques for 3D localization also give different precision in the axial and lateral directions [8].

A fundamental unknown in many live cell SPT studies is the number of underlying molecular states, which may differ in, e.g., diffusion constant. Counting diffusive states in SPT data presents a statistical model selection problem that has so far only been solved with simplified noise models [2], which may be inappropriate in many live cell applications [7, 9].

Here, we extend our previous HMM analysis [7] by deriving and implementing variational algorithms that increase computational speed by more than an order of magnitude, allow statistical model selection using Bayesian or information-theoretic methods, and can be generalized to a wider class of localization error models. The new methods are available in a user-friendly open source software suite.

I. RESULTS

A. Diffusive hidden Markov models with motion blur and localization uncertainty

The starting point for our analysis is a standard model for camera-based single particle tracking that includes a combination of averaging (motion blur) and localization errors, where the detected positions x_t are related to the underlying particle trajectory $y(t)$ through

$$x_t = \frac{1}{\Delta t} \int_t^{t+\Delta t} f(t')y(t')dt' + \sqrt{v_t}\xi_t^{(x)}. \quad (1)$$

Here, v_t is the localization error (variance) in frame t , and $\xi_t^{(\cdot)}$ are independent $N(0,1)$ -variables. The shutter function $f(t)$ describes how the image acquisition is distributed throughout the frame, e.g., $f(t) = 1/\Delta t$ for continuous acquisition [5].

We model the particle motion $y(t)$ as free diffusion, with a time-dependent diffusion constant governed by a hidden Markov process s_t with N discrete states. With variational approximations in mind, we need a model in the exponential family to get an analytically tractable algorithms [10]. This means that we model both the true hidden path and the true exposure-averaged positions (the integral in Eq. (1)) explicitly. In discrete time, this leads to the model

$$y_{t+1} = y_t + \sqrt{2D_{s_t}\Delta t}\xi_t^{(y)}, \quad (2)$$

$$z_t = (1 - \tau)y_t + \tau y_{t+1} + \sqrt{\beta 2D_{s_t}\Delta t}\xi_t^{(z)}, \quad (3)$$

$$x_t = z_t + \sqrt{v_t}\xi_t^{(x)}, \quad (4)$$

where τ and β are blur coefficients that depend on the shutter function [7] (see SI Sec. S1). Position coordinates in two- or three-dimensional trajectories are treated independently, which means that we neglect possible correlations between localization errors in different directions. As detailed below, and in the supplementary material, this model allows variational algorithms for both maximum likelihood estimation (MLE) and variational Bayes inference (VB) [11, 12].

* martin.linden@icm.uu.se

† johan.elf@icm.uu.se

Out focus in this work is the case where the localization variances v_t are input data, estimated from the localization of single spots [7]. However, one could also treat v_t as model parameters, for example as a single average error ($v_t = v$), dependent on the hidden state ($v_t = v_{s_t}$), and/or varying with coordinate dimension. These modified models remain in the exponential family, and thus allow similarly efficient variational algorithms that differ only in details compared to our main case.

B. Model selection

The number of diffusive states is often a biological unknown of great interest, but since different number of diffusive states correspond to statistical models with different number of parameters, counting states is a non-trivial problem of statistical model selection.

Bayesian reasoning, including model selection, is an extension of formal logic to uncertain statements, that yields unique and consistent results [13]. Assuming equal prior preference for a set of candidate models with uncertain parameters and unobserved degrees of freedom (latent variables), the Bayesian approach uses marginalization to select the model with the largest (log) evidence [11, 12], which in compact notation can be written

$$\ln L = \ln p(x|M) = \ln \int dn d\theta p(x, n|\theta, M) p_0(\theta|M). \quad (5)$$

Here, x denotes the observed data, M the model, $n = \{s_t, y_t, z_t\}$ the latent variables (summed or integrated out as appropriate), θ_M the unknown parameters with prior distribution $p_0(\cdot|M)$ for the different models M .

In our case (and many others), the evidence is analytically intractable, in which case a variational approximation (“variational Bayes”, VB) can be an attractive approximative approach [2, 14–17]. VB yields an approximate log evidence usable for Bayesian model selection as well as approximate posterior distributions of parameters and hidden states [10–12]. Moreover, since it involves direct optimization of the evidence in Eq. (5), VB algorithms have an intrinsic parsimony that depopulates superfluous states and can be utilized for efficient greedy model search algorithms [2, 17].

However, Bayesian inference may be statistically inefficient. In particular, the common practice of using uninformative priors to minimize bias in parameter estimates means that the prior likelihood for any particular parameter value is low. This in turn can lead to overly steep penalties against models with many parameters, a phenomenon known as Lindley’s paradox, which means that an unnecessarily large amount of data is needed in order to resolve some feature of interest [18–20]. An alternative approach is to rank competing models by their estimated predictive performance, which does not suffer from Lindley’s paradox [19, 20]. The most famous example of this is Akaike’s information criterion (AIC) [21], but this is only asymptotically valid for regular models

Figure 1. Statistical model selection. We generated a range of synthetic data sets with three diffusive states and z -dependent localization errors, and estimated the number of states using variational maximum evidence (top) and cross-validation using variational pseudo-Bayes factors.

and large data sets. For finite data sets, one could instead use cross-validation, where the data is divided into two parts, one for learning model parameters (“training”) and one for estimating predictive performance (“validation”). In practice, the performance is estimated from averaging over several such divisions. Here, we focus on a Bayesian variant, pseudo-Bayes factors (PBF) [20], which are easy to compute with variational methods (see SI S2.5).

Figure 1 compares VB and PBF model selection on synthetic test data with three diffusive states and parameters that resemble in vivo SPT experiments in bacteria [2, 22] (see Methods). Broadly, one expects predictive model selection to avoid Lindley’s paradox and penalize complex models less severely than Bayesian methods as the amount of data increases, possibly at the expense of consistency, i.e., there is no guaranteed convergence to the correct model [23]. These expectations are qualitatively borne out in Fig. 1, where the Bayesian VB criterion is more prone to select too few states for small data sets, but less prone to select too many states for large data sets.

There is no simple optimal rule for selecting training and validation subsets, and HMMs also suffer the additional complication that individual observations are correlated due to the hidden state dynamics. SPT experiments typically generate many short trajectories, which we use as atoms for constructing training and validation sets. Fig. 1b uses randomly sampled validation sets containing about 10% of the data. Some other choices, including AIC, are explored in SI Fig. Figure S1, but turn out to perform worse.

Figure 2. Speed of our variational algorithm (YZShmm) compared to that of Ref. [7] (EMhmm). Time per iteration vs. number of steps in the data (a trajectory with T positions contains $T - 1$ steps), with 3 diffusive states, was measured on a dual 6-core Intel Xeon 2.4GHz computer running Matlab R2017a. Scaling laws are guides to the eye.

C. Speed-up

In addition to more flexibility in modeling and inference methods, the new algorithms are also considerably faster compared to our previous implementation [7]. This is mainly because variational algorithms based on Eqs. (2-4) are analytically tractable, and hence avoid a costly numerical optimization step, but also due to an improved matrix inversion algorithms [24]. Figure 2 shows the time per iteration for a 3-state model on data sets of different sizes for the new algorithm compared to that of Ref. [7]. We see speedups of 1-2 orders of magnitude for experimentally relevant data set sizes of $10^4 - 10^5$ positions, as well as better scaling.

D. Finding the global optimum

Variational learning of a model and its parameters, diffusion constants and transition rates, involves finding the overall best fit to the data, but VB and other expectation-maximization type algorithms only converge towards local optima. An additional global search is needed.

The simplest approach is to converge multiple models from different starting points. To speed things up, we use the built-in parsimony of the VB algorithms to start from complex many-state models and then systematically search for simpler ones by removing un- or low-populated states [2]. However, the extra complexity of our model compared to standard HMMs [2] makes this approach more challenging to apply.

One attractive feature of our model is the ability to handle long trajectories with missing positions due to, e.g., fluorophore blinking, by setting $v_t = \infty$. However, when fitting high-dimensional models to data with missing positions, groups of superfluous states sometimes converge towards identical parameters and finite occupancy associated with the missing positions. Since this is clearly

unphysical, we choose to remove such state clusters before commencing normal model pruning.

Another challenge is related to the presence of two types of latent variables, for the discrete diffusive states (s_t) and uncertainty in true particle positions (y_t, z_t), respectively. Although hard to quantify, it seems reasonable to expect more latent variables to yield a more complex search landscape, with more local optima for the search to get trapped in, compared to ordinary HMMs where the particle positions are not latent variables [2]. More concretely, the variational treatment uses a three-fold factorization ansatz (parameters, hidden states, and hidden particle trajectories), and to initialize the local optimization iterations, two of three factors need to be initialized.

We use randomly selected parameter values, and explore different strategies to initialize either hidden states or hidden trajectories: uniform hidden state occupancy, hidden trajectories modeled directly on observed data (with no uncertainty or correlations between y_t and z_t), and hidden trajectory models generated by a running average where a pure diffusion model is fit to a small window of variable length. Generally, there does not seem to be a single silver bullet. Instead, different initialization methods do well on different types of data. Figure 3 shows an analysis of a data set from simulated images (see Methods) using 50 independent initializations of model parameters with 15 states, and 10 different initializations of latent variables with each parameter set. Fig. 3a shows the lower bounds of models originating from a single initial parameter set, with each line corresponding to models generated by the reductive search starting from one latent variable initialization. We see that the initialization with the largest number of non-spurious states does not lead to the best overall model, and that the search lines sometimes cross, meaning that relative ranking among the different reduction searches can change as states are removed. Looking at the best models from 50 independent parameter initializations (Fig. 3b), we again see search lines crossing, and note that the two most high-ranked model sizes originate from different initialization methods.

E. Application: simulated tRNA tracking

Quantitative live cell single particle tracking is complex, and errors may arise during measurement, spot detection, localization, trajectory building, and trajectory analysis. Comprehensive tests of the whole analysis chain are needed to validate quantitative interpretations of the experiments under particular conditions. To evaluate the capabilities of our new trajectory analysis only, we seek test data with known ground truth and sufficient realism to be experimentally relevant. We use simulated video microscopy [9] to produce realistic test data, run spot detection and localization using our standard methods (see Methods), but use our knowledge of the simulated ground

Figure 3. Model search with different initialization strategies. Each color/marker combination shows to the lower bounds $\ln L$ from the best model of each size found from different latent variable (hidden states s_t or position uncertainties y_t, z_t) initializations. (a) Model search from a single parameter initialization. (b) Best models from 50 independent restarts.

Figure 4. Model for simulated tracking of fluorescent tRNA molecules. (a) Cross-section of the simulation geometry, which consists of concentric cylinders with spherical end-caps, representing the nucleoid (blue) floating in the cytoplasm (red). Scale bar (black) 1 μm . (b) Kinetic model of the tRNA cycle. Two states B_1, B_2 with low diffusion coefficient represent ribosome bound states, and are excluded from the nucleoid, while the two unbound (U) and ternary complex (TC) states are free to roam the whole cell. (c) A simulated frame with two fluorophores in a single cell, with cell outline (red) and particle tracks (yellow) added. Pixels size 80 nm.

truth to produce trajectories free from false positives and linking errors which may lead to bad performance that do not reflect the intrinsic quality of the trajectory analysis. We allow at most 3 consecutive missing positions before starting an new trajectory.

As a test problem, we consider tracking of tRNA molecules in *E. coli* cells [22, 25], which should present several interesting difficulties. At least three discernible diffusive states may be expected: ribosome-bound (slow), free (fast diffusion), and a ternary complex (intermediate). The ribosome-bound states further display spatial structure in the form of nucleoid exclusion [25], as well as non-exponential waiting times [26], since tRNA

Figure 5. Analysis of simulated microscopy data. The different ground truth models are denoted by their total bound state dwell times, $\tau_B = 0.05\text{s}, 0.1\text{s}, 0.2\text{s}$, and 0.4s , respectively. (a) Number of states selected by the VB and PBF criteria. (b) Diffusion coefficients for VB-selected models. Dashed colored lines indicate the true diffusion constants of the U, TC, and $B_{1,2}$ states. (c) Kinetic scheme of the 4-state 0.05s model, with transition probabilities per time-step below $10^{-8}\Delta t^{-1}$ suppressed. States are named and colored according to the obvious similarity with the true scheme in Fig. 4b. For the true model, $k_u = 0.2\Delta t^{-1}$ and $k_{tc} = 0.067\Delta t^{-1}$. (d) Bound (B) and unbound (U) state mean dwell times, computed from the transition probability matrix. For the 0.05s model, we added the dwell times of the two B states. Dashed lines indicate the true mean dwell times. All precision indicators are bootstrap standard error of the mean, and all data sets contain about 16 000 steps.

goes through several reaction steps before dissociating from the ribosome [27]. We constructed a simplified kinetic and spatial model incorporating these features and generated synthetic fluorescent microscopy data with 200 Hz frame rate [9], as shown in Fig. 4. Rate constants are chosen to give a steady-state occupancy of 20/30/50 (bound/unbound/ternary-complex), and we vary the overall rates to get total bound state mean dwell time τ_B between 50ms and 400ms (see Methods).

Not all estimated model parameters can be directly compared to the simulated model parameters. For example, the nucleoid exclusion means that the $TC \rightarrow B$ reaction cannot take place in the nucleoid region, lowering the effective value of k_b . Nucleoid exclusion also distorts the state occupancy of the detected spot population, because defocused spots are more difficult to detect, and with our simulated focus in the cell mid-plane, the bound states are relatively enriched in defocused regions. In Fig. 5, we therefore focus on comparisons where these

complications are minimal.

The true model contains three diffusion constants but 4 kinetic states. Starting with the number of states (Fig. 5a), we see mostly four states, and note that the VB and PBF model selection agree half the time, and that the PBF favors more states in cases of disagreement. Plotting the diffusion constants from VB-selected models (Fig. 5b), we see that it finds diffusion constants close to the true values, although for the fastest kinetics ($\tau_B = 50\text{ms}$) the high-D states are biased towards each other, probably because the faster dynamics produce more short events that make it more difficult for the HMM to distinguish the two fast states correctly. There is also a general downward bias in the highest diffusion constant, perhaps because the HMM can only confuse it with slower states, or perhaps because of confinement effects.

Regarding the kinetics, a detailed look at the 4-state model for the fastest kinetics (Fig. 4c) shows a striking resemblance to the true rate model in Fig. 4b. The unidirectional cycle is clearly visible, and the transition probabilities corresponding to k_u and k_{tc} closely resemble the underlying ground truth. The mean dwell times of this model is comparable to the average trajectory length of about 0.12s. For slower models, the bound state dwell time is well reproduced (5d, black), but otherwise the kinetics is not as well reproduced. Only one bound state is identified, the unbound dwell times (5d, magenta) are not close to the true values, and the transition matrices (not shown), do not resemble the cyclic pattern of the underlying model. However, even the more limited ability to measure the mean dwell time of a slow or immobile state when that dwell time exceeds the average trajectory length could be of biological interest, for example to study the interactions of small molecules such as tRNA or proteins interacting with larger structures such as ribosomes or DNA.

II. DISCUSSION AND CONCLUSION

Together with methods to extract both positions and position uncertainty from images of single spots [7], the variational algorithm we present here makes it possible to significantly decrease analysis artifacts due to variable localization quality, such as out-of-focus motion, gradual bleaching or stage drift, or fast fluorophore blinking. We expect this to become a powerful tool for quantitative analysis of diffusive *in vivo* single particle tracking data.

Compared to our previous implementation [7], the present algorithms are significantly faster, which makes exploratory analysis of large data sets practical, and support both maximum likelihood or variational Bayes inference, which allows a more nuanced statistical analysis. Here, we compared model selection by the Akaike criterion (AIC) [21], a variational implementation of cross-validation using pseudo-Bayes factors (PBF) [20], and the purely Bayesian variational Bayes (VB) approach

[10, 11]. While the AIC clearly overfitted in our example, the PBF showed a much weaker preference for more complex models compared to VB, and in light of the theoretical arguments that can be made against a purely Bayesian approach in certain situations [18, 19], we think this approach merits further study.

One can also imagine combining our approach to localization uncertainty with other types of heterogeneity in the data. Interesting examples include variability in the underlying diffusion constant or other model parameters [16], or the presence complex spatial structure [28].

Another interesting direction for future work is to consider more complex models. The simple diffusion models could potentially be extended in several useful ways within the exponential family of models that enable efficient variational algorithms[10]: Localization errors could be treated as model parameters rather than external observations, possibly depending on the chemical state or coordinate dimension (see SI Sec. S5). Combinations of directed motion and confinement in harmonic potentials that lead to Gaussian models have been proposed for analysis of particle tracking experiments [3, 29]. Introducing explicit termination rates could correct bias that arise from correlations between chemical states and trajectory termination [30], for example when fast-diffusing molecules move out of focus faster than slow-diffusing ones [31].

A. Software

Our algorithms are freely available as open source Matlab code from <https://github.com/bmelinden/vbSPTu>. The vbSPTu software suite includes a GUI to run a simple standard analysis, support for scripting large analysis tasks, and low-level tools for creating customized analysis.

METHODS

B. Simulated trajectories

For the model selection experiments in Figures 1 and Figure S1, we used synthetic trajectories simulated using the analysis model, Eqs. (2-4).

We simulated a 3-state model with the following parameters: Diffusion constants $D_1 = 0.1\mu\text{m}^2\text{s}^{-1}$, $D_2 = 6\mu\text{m}^2\text{s}^{-1}$, and $D_3 = 3\mu\text{m}^2\text{s}^{-1}$. The kinetic model is an irreversible cycle $D_1 \rightarrow D_2 \rightarrow D_3 \rightarrow D_1 \rightarrow \dots$ with mean waiting times of 100 ms, 100 ms, and 116 ms, respectively. The positions are simulated according to Eqs. (2-4), with motion blur corresponding to an exposure time of $t_E = 1.5\text{ms}$ ($\tau = 0.15$, $\beta = 0.0775$).

For the static localization errors v_t , we used a phenomenological model of defocus-widening. First, we confined $z_t^{(z)}$ (the z component of z_t) to $|z_t^{(z)}| < 400\text{nm}$ using

the method of images. Then, we defined an effective spot width

$$\sigma(z, D)^2 = \sigma_0^2 \left(1 + \left(\frac{z}{L_z} \right)^2 \right) + \frac{a^2}{12} + \frac{1}{3} D t_E, \quad (6)$$

with $\sigma_0 = 135\text{nm}$, $L_z = 300\text{nm}$, $a = 80\text{nm}$. The a^2 term approximates the effect of finite pixel size [32], and the $D t_E$ term spot-widening due to motion blur [6]. We then computed v_t using the approximate Cramer-Rao lower bound expression of Mortensen et al. [32],

$$v_t = 2 \frac{\sigma(z_t^{(z)}, D_{s_t})^2}{N_{\text{phot.}}} \left(\frac{16}{9} + \frac{8\pi b^2 \sigma(z, D)^2}{N_{\text{phot.}} a^2} \right). \quad (7)$$

for use in Eq. (4). We analyzed the x and y components of x_t , and chose trajectories to be exponentially distributed with mean length 25, but discarded trajectories with length below 5. For Fig. 3a,b, we used a data set of 100 trajectories. For the statistical model selection study (Figure 1 and Figure S1), we sampled data sets of various sizes from a large data set of several hundred thousand positions, such that all model selection techniques used the same set of trajectories.

C. Simulated microscopy

Simulated video-microscopy images for tRNA tracking was generated using the SMeagol simulation software [9]

with the spatial reaction-diffusion model illustrated in Fig. 4. We simulated uniform exposure during 1.5 ms of the 5 ms sampling time. Camera noise was generated using a high-gain approximation of EMCCD noise [32] with offset 200, EM gain 77, and Gaussian readout noise with standard deviation 20. We used 80 nm pixels and a uniform background fluorescence that decayed from two to one photon/pixel with a decay rate of 2 s^{-1} . For the optics, we used a Gibson-Lanni point-spread function (PSF) model [33] generated by PSFgenerator [34], with $\lambda = 680\text{nm}$ and $NA = 1.49$. This is a spherically symmetric PSF, suitable for isotropic emitters or fluorophores with high rotational mobility. Fluorescent spot intensity was set to give on average of 200 photons per frame, and the average bleaching time was 20 frames. Using custom Matlab scripts, we simulated 200-frame movies with about 30 cells spread evenly across a 512×130 pixel field of view, with a few active fluorescent spots per cell. An example of one such cell is shown in Fig. 4c.

D. Spot detection and localization

We use the fast radial symmetry transform[35] for spot detection, and estimated spot positions and localization uncertainty using a symmetric Gaussian spot model and maximum a posteriori estimates on 9-by-9 regions of interests, as described by Lindén et al. [9]. Spots with $\sqrt{v_t} > 80\text{nm}$ were discarded from the analysis.

-
- [1] Suliana Manley, Jennifer M. Gillette, George H. Patterson, Hari Shroff, Harald F. Hess, Eric Betzig, and Jennifer Lippincott-Schwartz. High-density mapping of single-molecule trajectories with photoactivated localization microscopy. *Nat. Meth.*, 5(2):155–157, 2008. doi: 10.1038/nmeth.1176.
 - [2] Fredrik Persson, Martin Lindén, Cecilia Unoson, and Johan Elf. Extracting intracellular diffusive states and transition rates from single-molecule tracking data. *Nat. Meth.*, 10(3):265–269, 2013. doi:10.1038/nmeth.2367.
 - [3] 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, and Mark Bathe. Inferring transient particle transport dynamics in live cells. *Nat. Meth.*, 12(9):838, 2015. doi: 10.1038/nmeth.3483.
 - [4] Thierry Savin and Patrick S. Doyle. Static and dynamic errors in particle tracking microrheology. *Biophys J*, 88(1):623–638, 2005. doi:10.1529/biophysj.104.042457.
 - [5] Andrew J. Berglund. Statistics of camera-based single-particle tracking. *Phys. Rev. E*, 82(1):011917, 2010. doi: 10.1103/PhysRevE.82.011917.
 - [6] Hendrik Deschout, Kristiaan Neyts, and Kevin Braeckmans. The influence of movement on the localization precision of sub-resolution particles in fluorescence microscopy. *J. Biophotonics*, 5(1):97–109, 2012. doi: 10.1002/jbio.201100078.
 - [7] Martin Lindén, Vladimir Ćurić, Elias Amselem, and Johan Elf. Pointwise error estimates in localization microscopy. *Nat Commun*, 8:15115, 2017. doi: 10.1038/ncomms15115.
 - [8] Bernd Rieger and Sjoerd Stallinga. The lateral and axial localization uncertainty in super-resolution light microscopy. *ChemPhysChem*, 15(4):664–670, 2014. doi: 10.1002/cphc.201300711.
 - [9] Martin Lindén, Vladimir Ćurić, Alexis Boucharin, David Fange, and Johan Elf. Simulated single molecule microscopy with SMeagol. *Bioinformatics*, 32(15):2394–2395, 2016. doi:10.1093/bioinformatics/btw109. <http://smeagol.sourceforge.net>.
 - [10] Matthew Beal. *Variational algorithms for approximate Bayesian inference*. PhD thesis, University of Cambridge, UK, 2003. URL <http://www.cse.buffalo.edu/faculty/mbeal/thesis/>.
 - [11] David MacKay. *Information theory, inference, and learning algorithms*. Cambridge University Press, 2003.
 - [12] Christopher Bishop. *Pattern recognition and machine learning*. Springer, New York, 2006.
 - [13] R. T. Cox. Probability, frequency and reasonable expectation. *American Journal of Physics*, 14(1):1–13, 1946. doi:10.1119/1.1990764.
 - [14] Jonathan E. Bronson, Jingyi Fei, Jake M. Hofman, Ruben L. Gonzalez Jr., and Chris H. Wiggins. Learning rates and states from biophysical time series: A

- Bayesian approach to model selection and single-molecule FRET data. *Biophys. J.*, 97(12):3196–3205, 2009. doi:10.1016/j.bpj.2009.09.031.
- [15] Kenji Okamoto and Yasushi Sako. Variational Bayes analysis of a photon-based hidden Markov model for single-molecule FRET trajectories. *Biophys. J.*, 103(6):1315–1324, 2012. doi:10.1016/j.bpj.2012.07.047.
- [16] Jan-Willem van de Meent, Jonathan E. Bronson, Chris H. Wiggins, and Ruben L. Gonzalez. Empirical Bayes methods enable advanced population-level analyses of single-molecule FRET experiments. *Biophys. J.*, 106(6):1327–1337, 2014. doi:10.1016/j.bpj.2013.12.055.
- [17] Stephanie Johnson, Jan-Willem van de Meent, Rob Phillips, Chris H. Wiggins, and Martin Lindén. Multiple LacI-mediated loops revealed by Bayesian statistics and tethered particle motion. *Nucl. Acids Res.*, 42(16):10265–10277, 2014. doi:10.1093/nar/gku563.
- [18] Robert D. Cousins. The JeffreysLindley paradox and discovery criteria in high energy physics. *Synthese*, pages 1–38, 2014. doi:10.1007/s11229-014-0525-z.
- [19] Colin H. LaMont and Paul A. Wiggins. The Lindley paradox: The loss of resolution in Bayesian inference. *arXiv:1610.09433 [math, stat]*, 2016. arXiv: 1610.09433.
- [20] A. E. Gelfand and D. K. Dey. Bayesian model choice: Asymptotics and exact calculations. *J. Roy. Stat. Soc. B Met.*, 56(3):501–514, 1994. doi:10.2307/2346123.
- [21] Kenneth P. Burnham and David R. Anderson. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. Springer, 2013.
- [22] Ivan L. Volkov, Martin Lindén, Javier Aguirre Rivera, Ka-Weng Ieong, Mikhail Metelev, Johan Elf, and Magnus Johansson. Direct measurements of protein synthesis kinetics in live cells using fluorescently labeled trnas. 2018. (in revision).
- [23] Jun Shao. Linear model selection by cross-validation. *J. Am. Stat. Assoc.*, 88(422):486–494, 1993. doi:10.1080/01621459.1993.10476299.
- [24] G. Meurant. A review on the inverse of symmetric tridiagonal and block tridiagonal matrices. *SIAM. J. Matrix Anal. & Appl.*, 13(3):707–728, 1992. doi:10.1137/0613045.
- [25] Anne Plochowitz, Ian Farrell, Zeev Smilansky, Barry S. Cooperman, and Achillefs N. Kapanidis. *In vivo* single-RNA tracking shows that most tRNA diffuses freely in live bacteria. *Nucleic Acids Res.*, page gkw787, 2016. doi:10.1093/nar/gkw787.
- [26] P. Kienker. Equivalence of aggregated Markov models of ion-channel gating. *Proc. R. Soc. Lond. B*, 236(1284):269–309, 1989.
- [27] Thomas A. Steitz. A structural understanding of the dynamic ribosome machine. *Nat Rev Mol Cell Biol*, 9(3):242–253, 2008. doi:10.1038/nrm2352.
- [28] Mohamed El Beheiry, Silvan Türkcan, Maximilian U. Richly, Antoine Triller, Antigone Alexandrou, Maxime Dahan, and Jean-Baptiste Masson. A primer on the Bayesian approach to high-density single-molecule trajectories analysis. *Biophys. J.*, 110(6):1209–1215, 2016. doi:10.1016/j.bpj.2016.01.018.
- [29] Christopher P. Calderon. Motion blur filtering: A statistical approach for extracting confinement forces and diffusivity from a single blurred trajectory. *Phys. Rev. E*, 93:053303, 2016. doi:10.1103/PhysRevE.93.053303.
- [30] Anatoly B. Kolomeisky and Michael E. Fisher. Periodic sequential kinetic models with jumping, branching and deaths. *Physica A*, 279(1-4):1–20, 2000. doi:10.1016/S0378-4371(99)00586-5.
- [31] Thorsten Kues and Ulrich Kubitschek. Single molecule motion perpendicular to the focal plane of a microscope: Application to splicing factor dynamics within the cell nucleus. *Single Mol.*, 3(4):218–224, 2002. doi:10.1002/1438-5171(200208)3:4<218::AID-SIMO218>3.0.CO;2-C.
- [32] Kim I. Mortensen, L. Stirling Churchman, James A. Spudich, and Henrik Flyvbjerg. Optimized localization analysis for single-molecule tracking and super-resolution microscopy. *Nat. Meth.*, 7(5):377–381, 2010. doi:10.1038/nmeth.1447.
- [33] S F Gibson and F Lanni. Experimental test of an analytical model of aberration in an oil-immersion objective lens used in three-dimensional light microscopy. *J Opt Soc Am A*, 9(1):154–166, 1992. doi:10.1364/JOSAA.9.000154.
- [34] H. Kirshner, F. Aguet, D. Sage, and M. Unser. 3-D PSF fitting for fluorescence microscopy: implementation and localization application. *J. Microsc.-Oxford*, 249(1):13–25, 2013. doi:10.1111/j.1365-2818.2012.03675.x. <http://bigwww.epfl.ch/algorithms/psfgenerator/>.
- [35] Gareth Loy and Alexander Zelinsky. Fast radial symmetry for detecting points of interest. *IEEE Trans. Pattern Anal. Mach. Intell.*, 25(8):959–973, 2003. doi:10.1109/TPAMI.2003.1217601.

Acknowledgements

This project was funded by grants to J.E. from the Knut and Alice Wallenberg Foundation and the European Research Council.

Author contributions

M.L. and J.E. planned research and wrote the paper. M.L. designed and implemented new analysis methods, generated data, and analyzed data.

Conflicts of interest

The authors declare no conflicts of interest.