bioRxiv preprint doi: https://doi.org/10.1101/834572; this version posted November 7, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Blinking Statistics and Molecular Counting in direct Stochastic Reconstruction Microscopy (dSTORM)

Lekha Patel[a], Dylan M. Owen[b], and Edward A.K. Cohen[*a]

[a]Department of Mathematics, Imperial College London, South Kensington Campus, London, SW7 2AZ, U.K.

[b]Institute of Immunology & Immunotherapy and Department of Mathematics, University of Birmingham, Edgbaston, Birmingham, B15 2TT, U.K.

Abstract

Many recent advancements in single molecule localization microscopy exploit the stochastic photo-switching of fluorophores to reveal complex cellular structures beyond the classical diffraction limit. However, this same stochasticity makes counting the number of molecules to high precision extremely challenging. Modeling the photo-switching behavior of a fluorophore as a continuous time Markov process transitioning between a single fluorescent and multiple dark states, and fully mitigating for missed blinks and false positives, we present a method for computing the exact probability distribution for the number of observed localizations from a single photo-switching fluorophore. This is then extended to provide the probability distribution for the number of localizations in a dSTORM experiment involving an arbitrary number of molecules. We demonstrate that when training data is available to estimate photo-switching rates, the unknown number of molecules can be accurately recovered from the posterior mode of the number of molecules given the number of localizations.

Keywords:Molecular counting—Single Molecule Localization Microscopy—Fluorescence Imaging—STORM

[*]Corresponding author. Email: e.cohen@imperial.ac.uk

bioRxiv preprint doi: https://doi.org/10.1101/834572; this version posted November 7, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Single molecule localization microscopy (SMLM) approaches, such as photoactivated localization microscopy (PALM) [1, 8] and stochastic optical reconstruction microscopy (STORM) [7, 17], form some of the most celebrated advances in super-resolution microscopy. Using a fluorophore with stochastic photo-switching properties [6, 19] can provide an imaging environment where the majority of fluorophores are in a *dark* state, while a sparse number have stochastically switched into a transient photon-emitting state, from here on referred to as the *On* state. This results in the visible fluorophores being sparse and well separated in space. With the use of a high-performance camera the individual fluorophores in the On state can be identified and localized with nanometer scale precision by fitting point spread functions [14, 18].

One of the most common avenues to SMLM is direct STORM (dSTORM). As with the original implementation of STORM, dSTORM uses conventional immuno-staining strategies to label the cells with fluorophores i.e. the use of small molecule dyes and antibodies against the protein of interest. In dSTORM, imaging of isolated fluorophores is made possible by placing the majority of the dye molecules into a very long lived dark state e.g. a radical state or a very long lived triplet state. This is the purpose of the *STORM buffer*, of which there are many recipes, usually containing reducing and oxygen scavenging components. The dye is initially emissive but when rapidly excited by very high intensity excitation lasers, soon enters a dark state which is much longer lived than the emissive state, thus rendering the majority of fluorophores off. The dyes then cycle between dark and On states until photobleaching occurs, rendering the dye permanently off.

A key challenge that has persisted since the first SMLM methods were developed has been the characterization and quantification of this photo-switching behavior [4]. In particular, being able to accurately count the number of fluorescently labelled molecules from the recorded localizations would allow much greater insight into the cellular structures and processes under observation. This is a notoriously difficult task as deriving the probability distribution for the number of localizations per fluorophore is highly non-trivial due to complex photo-switching models and imperfect imaging systems.

Methods exist for recovering the number of imaged molecules in SMLM, however, these have primarily focused on photoactivated localization microscopy (PALM) and are not

bioRxiv preprint doi: https://doi.org/10.1101/834572; this version posted November 7, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

wholly applicable or adaptable for counting molecules that are imaged via dSTORM. For instance, the PALM methods of [5, 10, 13, 16] assume a four state kinetic model (inactive, photon-emitting/On, dark and bleached) for the photoactivatable fluorescent protein (PA-FP). Each PA-FP begins in the non-emissive inactive state before briefly moving into the photon-emitting On state. Then, there is the possibility of a small number of repeat transitions between this and a temporary dark state, before finally bleaching to a permanent off state. This kinetic model is inappropriate for dSTORM in which all fluorophores start in the On state, before stochastically moving back and forth between this and one or more transient dark states, before permanent bleaching. The analysis of [12] is applicable for dSTORM, however, it assumes the fluorophores can occupy only three states (On, dark and bleached), when in fact empirical evidence supports the existence of multiple dark states [11, 15].

Importantly, common to [5, 10, 12, 13] is the assumption that all blinks (transitions to the On state followed by a return to a dark state) are detected and hence the data is uncorrupted for statistical inference. In fact missed blinks occur in two different ways; (i) a PA-FP or fluorophore briefly transitions from the On state into a dark state and back again within a single camera frame; this transition will not be detected as a separate blink; (ii) a PA-FP or fluorophore may briefly transition from a dark state to the On state for such a short time that the number of emitted photons is insufficient to detect the event above background noise. Accounting for these missed transitions is key for precise molecular counting. Missed transitions will result in fewer blinks being recorded than actually occurred, which in turn will lead to fewer molecules being predicted than are in fact present. We note that in the four state PALM setting, [16] attempts to mitigate for missed transitions, however, to do so requires the exact extraction of dwell times from time-traces. This is not suitable for dSTORM, particularly in densely labeled environments, since the nuanced photo-switching behavior means we cannot be certain of a specific fluorophore's photo-kinetic state at any one time.

The method of molecular counting presented in this paper utilizes the photo-switching and observation model developed in [15]. Similar to [5, 10, 12, 13, 16], a continuous time Markov process is used to characterize the underlying photo-switching property of fluo-

bioRxiv preprint doi: https://doi.org/10.1101/834572; this version posted November 7, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

rophores in dSTORM. However, this model is very general; allowing any number of dark states which can either be set by the user or inferred via a model selection method (BIC). Furthermore, both missed blinks and false positives are fully accounted for in the modelling, something which has been absent from molecular counting methods thus far. By performing counting using just the localization count, it is highly scalable, being able to count thousands of molecules with computational ease.

We first summarize key statistics of the photo-switching fluorophore. In particular, we derive the *exact* form of the probability mass function for the number of localizations a single fluorophore produces during an imaging experiment. This distribution is specific to this application and highly non-standard, therefore we provide expressions for its mean and variance as derived via the probability generating function. This distribution, and its moments, is fully characterized by the unknown photo-switching imaging parameters, which are estimable through the PSHMM fitting method described in [15]. We then extend this distribution to give the probability mass function of the cumulative number of localizations obtained from M molecules, and demonstrate its validity through simulations. Using training data to estimate unknown photo-switching rates, we can compute the posterior distributions over the unknown number of fluorescing molecules, which is shown to recover M with high accuracy. We finally demonstrate the validity of our method on Alexa Fluor 647 data, providing both maximum a posteriori estimates of M from the resulting posterior distributions and their associated 95% credible intervals (a Bayesian interpretation of confidence intervals).

Problem Setup and Methods Overview

Modeling photoswitching kinetics

Following [15], we model the stochastic photo-switching behavior of a fluorophore as a continuous time Markov process $\{X(t) : t \in \mathbb{R}_{\geq 0}\}$ that moves between a discrete, finite set of states. In order to accommodate for the varying effects of different photo-physical models, it allows $\{X(t)\}$ to transition between an On state 1, $d+1$ dark states $0_0, 0_1, \ldots, 0_d$ (where $d \in \mathbb{Z}_{\geq 0}$ denotes the number of multiple dark states), and a photo-bleached state 2.

bioRxiv preprint doi: https://doi.org/10.1101/834572; this version posted November 7, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

As commonly referred to under the widely assumed $d = 0$ model consisting of a single dark state, we denote the state 0_0 as state 0 . The general model, as is illustrated in Figure 1, allows for transitions from the On state to multiple dark states through the first dark state 0, and further allows the photo-bleached state to be accessed by any other state. The state space of $\{X(t)\}$ is $\mathcal{S}_X = \{0, 0_1, \ldots, 0_d, 1, 2\}$. Under this Markovian model, the holding time in each state is exponentially distributed and parameterized by the *transition rates*. These are denoted as λ_{ij} for the transition rate from state i to j $(i, j = 0, 0_1, \ldots, 0_d, 1)$, and μ_i for the *photo-bleaching rate* from state i to 2 $(i = 0, 0_1, \ldots, 0_d, 1)$. These rates are summarized by the generator matrix for $\{X(t)\}$

$$G = \begin{pmatrix} -\sigma_0 & \lambda_{00_1} & 0 & 0 & 0 & 0 & \ldots & \lambda_{01} & \mu_0 \\ 0 & -\sigma_{0_1} & \lambda_{0_1 0_2} & 0 & 0 & 0 & \ldots & \lambda_{0_1 1} & \mu_{0_1} \\ 0 & 0 & -\sigma_{0_2} & \lambda_{0_2 0_3} & 0 & 0 & \ldots & \lambda_{0_2 1} & \mu_{0_2} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \ldots & -\sigma_{0_d} & \lambda_{0_d 1} & \mu_{0_d} \\ \lambda_{10} & 0 & 0 & 0 & 0 & 0 & \ldots & -\sigma_1 & \mu_1 \\ 0 & 0 & 0 & 0 & 0 & 0 & \ldots & 0 & 0 \end{pmatrix} \tag{1}$$

where $\sigma_{0_d} = \lambda_{0_d 1} + \mu_{0_d}$, $\sigma_1 = \lambda_{10} + \mu_1$ and when $d > 0$, $\sigma_{0_i} = \lambda_{0_i 0_{i+1}} + \lambda_{0_i 1} + \mu_{0_i}$, for $i = 0, \ldots, d-1$. In particular, for any $t \geq 0$, the transition probabilities $\mathbb{P}(X(t) = j | X(0) = i)$ can be recovered as the ith, jth elements of the matrix exponential e^{Gt}. The Markov process is further parameterized by $\boldsymbol{\nu}_X := \begin{pmatrix} \nu_0 & \nu_{0_1} & \ldots & \nu_{0_d} & \nu_1 & \nu_2 \end{pmatrix}^\top$ with $\sum_{j \in \mathcal{S}_X} \nu_j = 1$, which defines the probability distribution of $X(0)$ (the starting state of the Markov chain) over the possible states and is referred to as the *initial probability mass* of $\{X(t)\}$.

Modeling localizations from a fluorophore

The imaging procedure proceeds by sequentially exposing the fluorophore over N_F frames, each of length Δ. Following [15], the discrete time observed localization process $\{Y_n : n \in \mathbb{Z}_{>0}\}$ takes values in the set $\mathcal{S}_Y = \{0, 1\}$, indicating either no observation or a localization of the fluorophore within the time interval $[(n-1)\Delta, n\Delta)$. This observed process is formally

bioRxiv preprint doi: https://doi.org/10.1101/834572; this version posted November 7, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

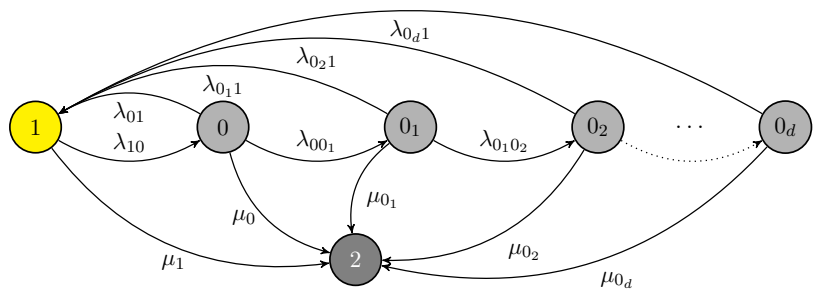


Figure 1: General $d+3$ state ($d \in \mathbb{Z}_{\geq 0}$) model of a fluorophore transitioning between an On state (1), $d+1$ temporary dark states $(0, 0_1, \ldots, 0_d)$ and a photo-bleached state (2).

defined as

$$Y_n = \mathbb{1}_{[\delta,\Delta)} \left(\int_{(n-1)\Delta}^{n\Delta} \mathbb{1}_{\{1\}}(X(t)) \ \ \mathrm{dt} \right),$$

where $\mathbb{1}_A(\cdot)$ is the indicator function such that $\mathbb{1}_A(x) = 1$ if $x \in A$ and is zero otherwise. This construction of $\{Y_n\}$ accounts for noise and the imaging system's limited sensitivity. A localization of a molecule in frame n is typically only recorded ($Y_n = 1$) when its continuous time process $\{X(t)\}$ reaches and remains in the On state for long enough to be detected. This minimum time is given by the unknown *noise parameter* $\delta \in [0, \Delta)$.

The photo-switching hidden Markov model (PSHMM) is presented in [15] as a means of estimating the unknown parameters of the continuous time Markov process $\{X(t)\}$. By collecting observations of $\{Y_n\}$ from a known number of M individually identifiable fluorophores, the transition rates, initial probability mass and noise parameter δ can be estimated via a maximum likelihood procedure. In order to handle this complicated stochastic structure and mitigate for missed state transitions, the authors define *transmission matrices* $B_\Delta^{(0)}, B_\Delta^{(1)} \in \mathbb{R}^{(d+3)\times(d+3)}$. These characterize the probability of its hidden state *and* localizing a fluorophore at the end of a frame given its state at the beginning of a frame. These will play a key part in deriving the distribution for the number of localizations. For $i, j \in \mathcal{S}_X$ and $l \in \mathcal{S}_Y$, its elements are defined by

$$B_\Delta^{(l)}(i,j) := \mathbb{P}(Y_0 = l, X(\Delta) = j | X(0) = i),$$
$$B_\Delta^{(l)}(2,2) = \mathbb{1}_{\{0\}}(l).$$

bioRxiv preprint doi: https://doi.org/10.1101/834572; this version posted November 7, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

These are deterministic functions of the unknown photo-switching parameters G and δ. A procedure for computing them is given in Algorithm 3 of Section 1.4 in Appendix 1.

Incorporating false positives

Crucially, as well as accounting for missed transitions, this set-up also accounts for the random number of false positive localizations that occur during an experiment. Specifically, if $\alpha \in [0, 1]$ denotes the probability of falsely observing a fluorophore in any given frame (assumed independent of the observation process), then the updated transmission matrices take the form

$$B_{\Delta}^{*(0)} = (1-\alpha)B_{\Delta}^{(0)}$$
$$B_{\Delta}^{*(1)} = B_{\Delta}^{(1)} + \alpha B_{\Delta}^{(0)}.$$

When incorporated into the model, α can also be estimated with the PS-HMM procedure in [15]. An algorithm to compute transmission matrices $B_{\Delta}^{*(0)}, B_{\Delta}^{*(1)}$ adapted from [15], can be found in Algorithm 3 of Section 1.4 in Appendix 1.

Distribution of localizations

Given an *unknown* number of M independently fluorescing molecules, each with localization process $\{Y_{n,m}\}$ $(m = 1, ..., M)$, we now use this model to characterize the distribution of

$$N_l = \sum_{m=1}^{M} \sum_{n=1}^{N_F} Y_{n,m}, \tag{2}$$

the *cumulative* number of localizations obtained over an experiment of length N_F frames. In order to do so, we will firstly explicitly derive the density of N_l when $M = 1$ and explain how this can be used to computationally recover the density for when $M > 1$. We will then use this density, which will be seen as a function of M and the parameter set $\boldsymbol{\theta}_d := \{G, \delta, \boldsymbol{\nu}_X, \alpha\}$ to derive the posterior mass function of M given N_l and $\boldsymbol{\theta}_d$, thereby constructing a suitable approach to estimating M via its mode.

We define $\{S_n : n \in \mathbb{Z}_{>0}\}$ to be the non-decreasing discrete time series process denoting the *cumulative* number of localizations obtained from a single fluorophore up to and

bioRxiv preprint doi: https://doi.org/10.1101/834572; this version posted November 7, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

including frame $n \leq N_F$. This process takes values in the set $\mathcal{S}_{S_n} = \{0, 1, ..., n\}$ and is formally defined as

$$S_n = \sum_{i=1}^{n} Y_i,$$

where the sum is taken over the values $Y_1, \ldots, Y_n$ from the observed localization process $\{Y_n\}$. Ultimately, we will be looking to find the probability mass function for S_{N_F} when imaging is conducted over a known number of N_F frames.

For any $n \geq 1$, Proposition 1 outlines a method for computing the probability mass function for S_n recursively. Furthermore, an algorithm specifying the relevant steps for its implementation when $n = N_F$, is shown in Algorithm 5.

Proposition 1. *Fix $n \in \mathbb{Z}_{>0}$. For $k \in \mathcal{S}_{S_n}$, define $\mathbf{M}(k, n) \in \mathbb{R}^{1\times(d+3)}$ to be the vector*

$$\begin{pmatrix} M(0,k,n) & \ldots & M(0_d,k,n) & M(1,k,n) & M(2,k,n) \end{pmatrix},$$

whereby for each $j \in \mathcal{S}_X$

$$M(j, k, n) := \mathbb{P}_{\boldsymbol{\theta}_d}(X(n\Delta) = j, S_n = k). \tag{3}$$

By recursively computing

$$\begin{aligned}
\mathbf{M}(k,1) &= \boldsymbol{\nu}_X^\top B_\Delta^{*(k)} && k \in \{0,1\} \\
\mathbf{M}(0,n) &= \mathbf{M}(0,n-1)B_\Delta^{*(0)} && n > 1 \\
\mathbf{M}(k,n) &= \mathbf{M}(k,n-1)B_\Delta^{*(0)} && \\
&\quad + \mathbf{M}(k-1,n-1)B_\Delta^{*(1)} && 1 \leq k < n \\
\mathbf{M}(n,n) &= \mathbf{M}(n-1,n-1)B_\Delta^{*(1)} && k = n,
\end{aligned}$$

the probability mass function of S_n follows

$$p_{\boldsymbol{\theta}_d}(S_n = k) := \mathbb{P}_{\boldsymbol{\theta}_d}(S_n = k) = \mathbf{M}(k,n)\mathbf{1}_{d+3} \qquad k \in \mathcal{S}_{S_n}. \tag{4}$$

Proof. See Section 1.1.1 of Appendix 1. □

Figure 2(a) presents the exact distributions $p_{\boldsymbol{\theta}_d}(S_{N_F} = k)$ for $k \in \mathbb{Z}_{\geq 0}$ when compared with histograms for the simulated data under three photo-switching models, $d = 0, 1, 2$. The shape of the densities can be seen to be determined by d, the dwell times in dark

bioRxiv preprint doi: https://doi.org/10.1101/834572; this version posted November 7, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Algorithm 1 Compute probability mass function (PMF) for S_{N_F}

function PMF_S($\boldsymbol{\theta}_d, \Delta, N_F$)

  Compute $B_\Delta^{*(0)}$ and $B_\Delta^{*(1)}$ from $\boldsymbol{\theta}_d, \Delta$ ▷ Via Algorithm 2 (SI)

  $A_0, A_1 \leftarrow \mathbf{0}_{N_F+1}\mathbf{0}_{d+3}^\top$

  $A_0[1,:] \leftarrow \boldsymbol{\nu}_X^\top B_\Delta^{*(0)}$

  $A_0[2,:] \leftarrow \boldsymbol{\nu}_X^\top B_\Delta^{*(1)}$

  for $n = 2$ to N_F **do**

    $A_1[1,:] \leftarrow A_0[1,:]B_\Delta^{*(0)}$

    for $k = 2$ to n **do**

      $A_1[k,:] \leftarrow A_0[k,:]B_\Delta^{*(0)} + A_0[k-1,:]B_\Delta^{*(1)}$

    $A_1[n+1,:] \leftarrow A_0[n,:]B_\Delta^{*(1)}$

    $A_0 \leftarrow A_1$

  $\mathbf{p} \leftarrow A_0\mathbf{1}_{d+3}$ ▷ $\mathbf{p}[i] = \mathbb{P}_{\boldsymbol{\theta}_d}(S_{N_F} = i-1)$ for $i = 1, \ldots, N_F+1$

  return p ▷ Probability mass function for S_{N_F}

states and the photo-bleaching rates. Moreover, as is to be expected, the average number of localizations decreases as the number of dark states d increases. In particular, the slow growth to the mode of each distribution is related to the presence of the photo-bleached state, as seen in Figure 2(b), which compares the mass functions under the $d = 1$ model with $\mu_0 = 0$ when μ_1 varies. When μ_1 is close to zero (the expected time to move into the bleached state is long), a bell shaped curve is observed. This is sharply in contrast to when μ_1 is large and photo-bleaching is much more likely to occur at the beginning of the experiment, giving rise to a geometric decay. For values in between, a mixture of these two properties is detected. These simulations therefore provide strong evidence that photo-kinetic models incorporating a photo-bleached state are likely to give rise to mixture distributions (that are potentially multi-modal) for the number of localizations recorded per molecule.

The moments of the distribution $p_{\boldsymbol{\theta}_d}(S_n = k)$ are fully characterized by its probability generating function (pgf) $G_{S_n}(z) = \mathbb{E}_{\boldsymbol{\theta}_d}(z^{S_n})$, which can be shown to take the form given in (1) of Lemma 1 of Section 1.1 in Appendix 1. In particular, the expected value of S_n,

bioRxiv preprint doi: https://doi.org/10.1101/834572; this version posted November 7, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

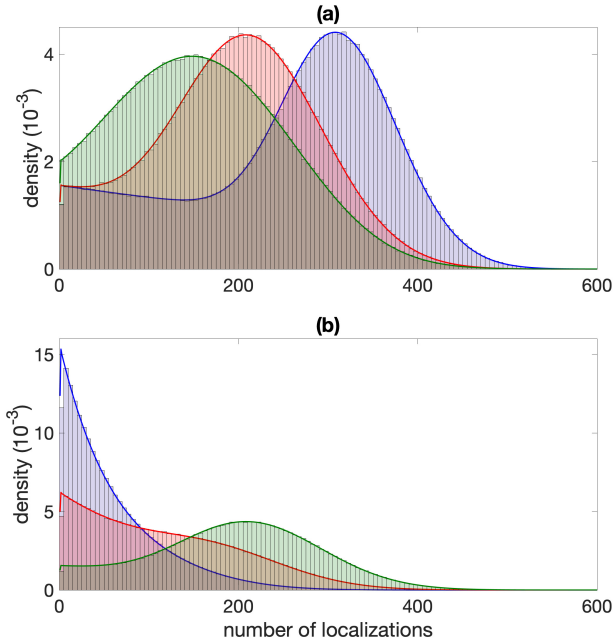


Figure 2: Figure (a) shows the theoretical and histogram estimate (from 10^6 simulations) of $p_{\boldsymbol{\theta}_d}(S_{N_F} = n)$ under 3 photo-switching models: $d = 0$ (blue), $d = 1$ (red) and $d = 2$ (green). In all simulations, $\mu_1 > 0, \mu_0 = \cdots = \mu_{0_d} = 0$, $N_F = 1000$, $\nu_0 = \nu_1 = 0.5$, $\Delta = \frac{1}{30}$, $\delta = 10^{-3}$ and $\alpha = 10^{-6}$; rates chosen (where appropriately zero) are $\lambda_{00_1} = 0.35, \lambda_{01} = 1, \lambda_{0_1 0_2} = 0.2, \lambda_{0_1 1} = 0.3, \lambda_{0_2 1} = 0.1, \lambda_{10} = 2.3, \mu_1 = 0.05$. Figure (b) shows the theoretical and histogram estimate (from 10^6 simulations) of $p_{\boldsymbol{\theta}_d}(S_{N_F} = n)$ when $d = 1$ with $\mu_1 = 0.5$ (blue), $\mu_1 = 0.2$ (red) and $\mu_1 = 0.05$ (green).

bioRxiv preprint doi: https://doi.org/10.1101/834572; this version posted November 7, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

denoted $\mathbb{E}_{\boldsymbol{\theta}_d}(S_n) = G'_{S_n}(1)$ and variance $\mathbb{V}\mathrm{ar}_{\boldsymbol{\theta}_d}(S_n) = G''_{S_n}(1) + \mathbb{E}_{\boldsymbol{\theta}_d}(S_n) - \mathbb{E}^2_{\boldsymbol{\theta}_d}(S_n)$ can be explicitly determined (see Proposition 2) by differentiating this pgf.

Proposition 2. *The expected value $\mathbb{E}_{\boldsymbol{\theta}_d}(S_n)$ and variance $\mathbb{V}ar_{\boldsymbol{\theta}_d}(S_n)$ of S_n are given as*

$$\mathbb{E}_{\boldsymbol{\theta}_d}(S_n) = \boldsymbol{\nu}_X^T \left[\sum_{i=1}^{n} \mathrm{e}^{G\Delta(n-i)} B_\Delta^{*(1)} \mathrm{e}^{G\Delta(i-1)} \right] \mathbf{1}_{d+3} \tag{5}$$

$$\mathbb{V}ar_{\boldsymbol{\theta}_d}(S_n) = G''_{S_n}(1) + \mathbb{E}_{\boldsymbol{\theta}_d}(S_n) - \mathbb{E}^2_{\boldsymbol{\theta}_d}(S_n), \tag{6}$$

where

$$G''_{S_n}(1) = \boldsymbol{\nu}_X^\top \left(\sum_{i=1}^{n-1} \sum_{j=1}^{n-i} \mathrm{e}^{G\Delta(n-i-j)} B_\Delta^{*(1)} \mathrm{e}^{G\Delta(j-1)} B_\Delta^{*(1)} \mathrm{e}^{G\Delta(i-1)} \right.$$
$$\left. + \sum_{j=1}^{i} \mathrm{e}^{G\Delta(n-i-1)} B_\Delta^{*(1)} \mathrm{e}^{G\Delta(i-j)} B_\Delta^{*(1)} \mathrm{e}^{G\Delta(j-1)} \right) \mathbf{1}_{d+3},$$

and e^G denotes the matrix exponential of the generator G defined in [1].

Proof. See Section 1.1.3 of Appendix 1. □

When M independent molecules are imaged, the total number of localizations N_l (which can take a minimum value of 0 and a maximum value of MN_F) can be written as

$$N_l = \sum_{m=1}^{M} S_{N_F,m} = \sum_{m=1}^{M} \sum_{n=1}^{N_F} Y_{n,m},$$

where $S_{N_F,m}$ denotes the total number of localizations made by the mth fluorophore over an experiment consisting of N_F frames. Specifically, the density of N_l follows

$$p_{\boldsymbol{\theta}_d,M}(N_l) = \sum_{\substack{k_1,\ldots k_M \\ :k_1+\cdots+k_M=N_l}} \prod_{i=1}^{M} p_{\boldsymbol{\theta}_d}(S_{N_F} = k_i), \tag{7}$$

which can be readily obtained by applying M convolutions of the mass function for S_{N_F}. This is most efficiently achieved via the Fast Fourier Transform (see Algorithm 4 of Section 1.4 in Appendix 1). The expected number and variance of *total* localizations are $\mathbb{E}_{\boldsymbol{\theta}_d,M}(N_l) = M\mathbb{E}_{\boldsymbol{\theta}_d}(S_{N_F})$ and $\mathbb{V}\mathrm{ar}_{\boldsymbol{\theta}_d,M}(N_l) = M\mathbb{V}\mathrm{ar}_{\boldsymbol{\theta}_d}(S_{N_F})$, which can be computed using [5] and [6].

bioRxiv preprint doi: https://doi.org/10.1101/834572; this version posted November 7, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Inference

The task of interest is to estimate M, the unknown number of molecules in a dSTORM experiment, from N_l, the number of localizations recorded across N_F frames. Our method first requires the use of training data to obtain at estimate of the photoswitching parameters $\boldsymbol{\theta}_d = \{G, \delta, \boldsymbol{\nu}_X, \alpha\}$. This training data consists of a set of observations of the localization process $\{Y_n\}$ from a known number of molecules. Here, we estimate $\boldsymbol{\theta}_d$ via the method of [15], however other methods exist [e.g. 11]. Once an estimate for $\boldsymbol{\theta}_d$ is obtained, inference on M can proceed for the dSTORM experiment under analysis.

After plugging in the estimate for $\boldsymbol{\theta}_d$ into $p_{\boldsymbol{\theta}_d,m}(N_l)$, the posterior distribution of M given N_l localizations follows as

$$p_{\boldsymbol{\theta}_d,m}(M = m|N_l) \propto p_{\boldsymbol{\theta}_d,m}(N_l)\pi_M(m), \tag{8}$$

where $\pi_M(m) := \mathbb{P}(M = m)$ denotes a suitable prior distribution on M. We here elect to use a uniform prior restricted to $M_{\min} \leq m \leq M_{\max}$. A discussion on choosing $M_{\min}$ and $M_{\max}$ can be found in Section 1.3 of Appendix 1. An efficient algorithm for computing $p_{\boldsymbol{\theta}_d,m}(N_l)$ can be found in Algorithm 4 of Section 1.4 in Appendix 1. Subsequently, the estimate $\hat{M}$ of the number of molecules is found by locating the mode of the posterior $p_{\boldsymbol{\theta}_d,m}(M = m|N_l)$, known as the maximum a posteriori (MAP).

Under this inference mechanism, 95% credible interval or *highest density region (HDR)* [9] can also be obtained. The upper and lower bounds of this credible interval inform us that M (under this distribution) lies within this region with probability 0.95, and is therefore a useful tool in analyzing the potential number of molecules that are truly imaged. Specifically, this region is chosen to be $I = \{m : p_{\boldsymbol{\theta}_d,m}(m|N_l) \geq k_{0.05}\}$, where $k_{0.05}$ is the largest value such that

$$p_I := \sum_{m \in I} p_{\boldsymbol{\theta}_d,m}(M = m|N_l) \geq 0.95.$$

We provide a detailed algorithm, which uses this method of inference to obtain $p_{\boldsymbol{\theta}_d,m}(M = m|N_l)$ in Algorithm 5 of Section 1.4 in Appendix 1.

bioRxiv preprint doi: https://doi.org/10.1101/834572; this version posted November 7, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Parameter / Study	d	λ_{00_1}	λ_{01}	$\lambda_{0_10_2}$	λ_{0_11}	λ_{0_21}	λ_{10}	μ_1	Δ^{-1}	δ	α	ν_0	ν_1	M	N_F
1 (SLOW)	0		0.3162				1	0.0333	30	0.0033	10^{-5}	0.2	0.8	100	10^4
2 (MEDIUM)	0		1				3.162	0.1054	30	0.0033	10^{-5}	0.2	0.8	100	10^4
3 (FAST)	0		3.162				10	0.333	30	0.0033	10^{-5}	0.2	0.8	100	10^4
4 (SLOW)	1	0.15	0.3		0.1		0.8	0.01	30	0.0033	10^{-5}	0.2	0.8	100	10^4
5 (MEDIUM)	1	0.35	1		0.3		2.3	0.1	30	0.0033	10^{-5}	0.2	0.8	100	10^4
6 (FAST)	1	2	10		0.7		10	0.333	30	0.0033	10^{-5}	0.2	0.8	100	10^4
7 (SLOW)	2	0.15	0.3	0.05	0.1	0.001	0.8	0.05	30	0.0033	10^{-5}	0.2	0.8	100	10^4
8 (MEDIUM)	2	0.8	4	0.1	0.4	0.005	8	0.1	30	0.0033	10^{-5}	0.2	0.8	100	10^4
9 (FAST)	2	2	10	0.2	0.7	0.01	10	0.333	30	0.0033	10^{-5}	0.2	0.8	100	10^4

Table 1: Global parameter values for the stimulation studies conducted in this section.

Validation

We validate our method on both simulated and Alexa Fluor 647 data to demonstrate its precision and accuracy in counting molecules.

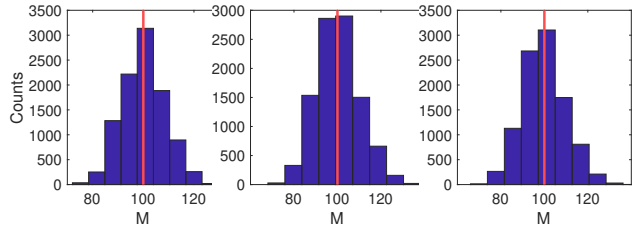
Validation with simulated data

Here we provide posterior estimates of M from nine simulation studies highlighting slow, medium and fast switching scenarios under photo-switching models with d, the number of dark states, equalling $0, 1$ and 2. For each simulation study, 10^4 independent datasets, each containing 350 molecules were simulated. From this, the localizations from 250 molecules were used to estimate $\boldsymbol{\theta}_d$. The number of localizations from the remaining 100 molecules were used to estimate M through the posterior mode of [8]. The true parameter values for each study can be found in Table 1, and in each case we use a uniform prior ($\pi_M(m) \propto 1$). Figures 3a - 5c show histograms of posterior modes $\hat{M}$ under each study and show that our estimation method can recover the true ($M = 100$) number of molecules from simulated data.

Validation with experimental data

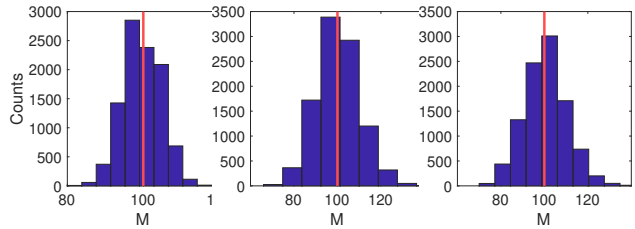
The data analysed in this section is taken from [11], in which detailed methods can be found. The original study examined the effect of laser intensity on the photo-switching rates of Alexa Fluor 647. Across a total of 27 experiments, 8 different laser intensities using 2 different frame rates were explored (see Table 2 for details). In each experiment, antibodies

bioRxiv preprint doi: https://doi.org/10.1101/834572; this version posted November 7, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



(a) (b) (c)

Figure 3: Simulation results from studies 1-3 in Table 1. Histograms represent counts of $\hat{M}$ under the slow (Figure 3a), medium (Figure 3b) and fast (Figure 3c) scenarios when $d = 0$, from 10^4 independently generated datasets with $M = 100$ and $N_F = 10^4$. For each estimate, $\boldsymbol{\theta}_0$ was determined using a training data set with $M = 250$ and $N_F = 10^4$.



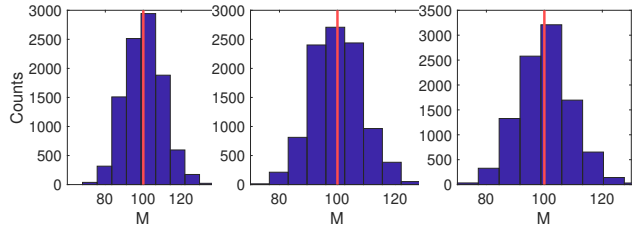
(a) (b) (c)

Figure 4: Simulation results from studies 4-6 in Table 1. Histograms represent counts of $\hat{M}$ under the slow (Figure 4a), medium (Figure 4b) and fast (Figure 4c) switching scenarios when $d = 1$, from 10^4 independently generated datasets with $M = 100$ and $N_F = 10^4$. For each estimate, $\boldsymbol{\theta}_1$ was determined using a training data set with $M = 250$ and $N_F = 10^4$.

labeled with Alexa Fluor 647 at a ratio of 0.13-0.3 dye molecules per antibody were imaged by total internal reflection fluorescence (TIRF) microscopy. The photo-emission time trace of each photo-switchable molecule detected was extracted. These were then used to estimate the photo-switching rates.

Here, we use these data for the purpose of validating the theory and counting method presented in this paper. In each experiment, the number of fluorophores present is known and therefore acts as a ground truth against which our estimate can be compared. For each dataset (labelled 1 - 27), each photo-switchable molecule detected has its discrete observation trace $\{Y_n\}$ extracted. 70% of these traces (the number of which we denote M_{tr}) are then used to create a *training set* with which to identify model parameters $\boldsymbol{\theta}_d$.

bioRxiv preprint doi: https://doi.org/10.1101/834572; this version posted November 7, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



(a) (b) (c)

Figure 5: Simulation results from studies 1-3 in Table 1. Histograms represent counts of $\hat{M}$ under the slow (Figure 5a), medium (Figure 5b) and fast (Figure 5c) switching scenarios when $d = 2$, from 10^4 independently generated datasets with $M = 100$ and $N_F = 10^4$. For each estimate, $\boldsymbol{\theta}_2$ was determined using a training data set with $M = 250$ and $N_F = 10^4$.

The remaining 30% (the test set) are used to validate the inference method outlined in this paper. Here, M (known) is the 30% of molecules that remain, and N_l is the number of localizations recorded from these M molecules. The $d = 2$ photo-kinetic model is assumed, as reasoned in [15].

For each experiment, the posterior modes (MAP values) $\hat{M}$ given N_l, along with the true values of M and corresponding 95% credible intervals are shown in Figure 6. With this are shown two examples of the posterior distribution of M given N_l (see [8]). The remaining figures can be found in Figure 7 of Section 1.6 in Appendix 1. The values of the laser intensity, frame rate Δ^{-1}, number of molecules in each dataset (M_{tr}, M), the number of frames over which they were imaged (N_F), the total number of localizations (N_l), the posterior mode $\hat{M}$, its 95% credible interval (I) and its corresponding value p_I is summarized in Table 2. The maximum likelihood estimates $\hat{\boldsymbol{\theta}}_2$ used for each study is presented in Table 3 of Section 1.5 in Appendix 1.

The plots show that the modes of the posterior distributions $(\hat{M})$ can be used to accurately estimate the true number of imaged molecules, with all studies' 95% credible intervals containing the true values of M. Furthermore, the inference method shows a consistently strong performance, both in the MAP estimate and the width of the credible intervals, across the range of laser intensities and frame rates. This demonstrates its robustness to different experimental conditions and photo-switching rates.

bioRxiv preprint doi: https://doi.org/10.1101/834572; this version posted November 7, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Dataset	Laser intensity	Δ^{-1}	M_{tr}	M	N_F	N_l	$\hat{M}$	I	p_I
1	1.0	200	192	81	49796	4340	77	[62, 91]	0.951
2	1.9	200	180	77	49533	5300	81	[67, 94]	0.950
3	3.9	200	234	100	49815	2443	106	[87, 125]	0.955
4	3.9	200	295	110	39758	2834	112	[94, 130]	0.956
5	7.8	200	238	102	39721	2679	106	[88, 123]	0.954
6	7.8	800	171	72	29418	4648	75	[63, 87]	0.953
7	7.8	800	159	67	29257	4251	66	[54, 77]	0.956
8	7.8	800	121	51	29438	2760	54	[43, 65]	0.961
9	16	800	304	129	29467	3538	126	[108, 144]	0.953
10	16	200	201	86	39703	1609	89	[73, 104]	0.953
11	16	800	213	90	29074	3309	88	[74, 101]	0.952
12	16	800	201	85	29145	2977	84	[71, 97]	0.951
13	31	800	425	181	29059	4050	177	[157, 197]	0.955
14	31	800	374	159	29778	2845	156	[137, 174]	0.954
15	31	800	360	153	29179	3431	156	[136, 175]	0.954
16	31	800	343	147	29400	3013	140	[122, 158]	0.957
17	31	800	317	135	29071	4616	137	[120, 153]	0.950
18	62	800	385	164	29327	3160	165	[147, 183]	0.955
19	62	800	309	132	29107	2728	132	[116, 148]	0.950
20	62	800	294	126	29551	1935	124	[107, 141]	0.956
21	62	800	298	127	29426	3022	132	[116, 148]	0.952
22	62	800	279	119	28989	2842	121	[106, 136]	0.951
23	97	800	315	135	29191	1579	136	[117, 154]	0.955
24	97	800	307	131	29198	1659	138	[120, 156]	0.955
25	97	800	304	129	29270	2120	132	[115, 148]	0.954
26	97	800	295	126	29295	2280	124	[107, 140]	0.953
27	97	800	287	123	29218	1351	126	[106, 145]	0.954

Table 2: A description of the Alexa Fluor 647 datasets, with reference to the laser intensities in kW/cm^2 and frames sampled per second (or Δ^{-1}) measured in s^{-1} used to characterize each of the 27 experiments. For each dataset, a training set of size M_{tr} was used to find the maximum likelihood estimate $\hat{\boldsymbol{\theta}}_2$. A hold out test set of size M was used validate the inference procedure.

Discussion

We have derived the distribution of the number of localizations per fluorophore and for an arbitrary number of fluorophores is a dSTORM experiment. This has allowed us to present an inference procedure for estimating the unknown number of molecules, given an observed number of localizations. These results have been successfully validated on both simulated and experimental data across a range of different imaging conditions, thus demonstrating a robust and precise new tool for the quantification of biological structures and mechanisms imaged via SMLM methods.

This method separates out the rate estimation (training) procedure from the counting procedure. While the training procedure requires a separate experiment to estimate fluo-

bioRxiv preprint doi: https://doi.org/10.1101/834572; this version posted November 7, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

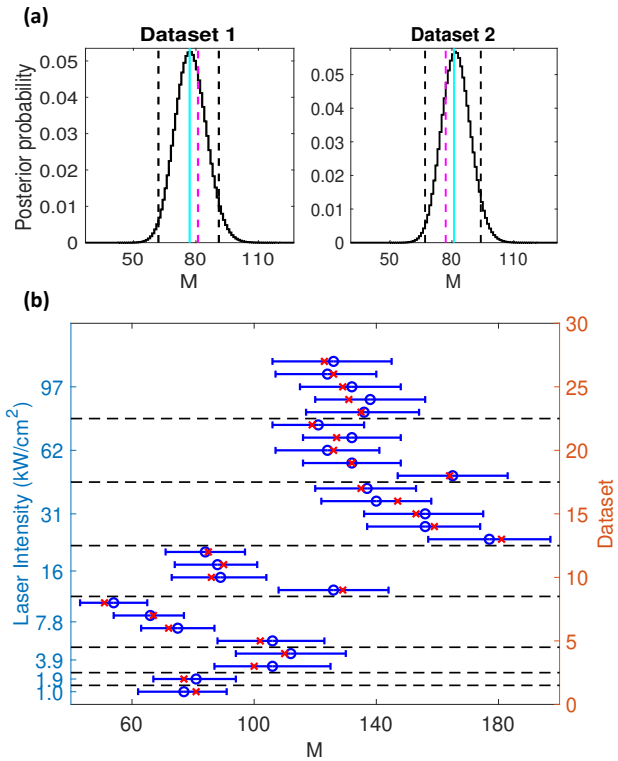


Figure 6: (a) Posterior distributions of M_{te} given $\hat{\boldsymbol{\theta}}_2$ and N_l for the Alexa FLuor 647 datasets 1 and 2 (descriptions of which can be found in Table 3). For each study, $\hat{M}$ is given by the corresponding posterior mode plotted in cyan, with the true values of M_{te} shown in magenta (dotted). 95% credible intervals for each $\hat{M}$ are shown in black (dotted).(b) Posterior estimates of M_{te} given $\hat{\boldsymbol{\theta}}_2$ and N_l for the 27 Alexa FLuor 647 datasets (descriptions of which can be found in Table 2) with varying laser intensities (kw/cm^2). For each study, $\hat{M}$ is given by the corresponding posterior mode plotted in blue (circle), with the true values of M_{te} shown in red (crosses) and 95% credible intervals for each $\hat{M}$ are shown by blue error bars.

rophore switching rates, it does mean that the counting process is computationally cheap and therefore highly scalable. This method can count several thousand molecules from tens of thousands of localizations with relative computational ease. In the PALM setting, [16] attempts to count and do rate estimation simultaneously. While having a single procedure avoids the problem of a separate training experiment, the computational burden of such a procedure is extreme and drastically limits the numbers of molecules that can be counted at

bioRxiv preprint doi: https://doi.org/10.1101/834572; this version posted November 7, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

any one time. Furthermore, it requires careful extraction of the time traces from crowded environments which is in itself problematic and challenging.

The counting procedure presented here assumes that localizations are acquired sparsely. In fact, if two or more molecules, within close enough proximity that their point spread functions sufficiently overlap, are in the On state simultaneously then it could be that only a single localization is obtained or the localization algorithm ignores them all together. This phenomenon is discussed in detail and quantified in [3]. They relate the frequency at which this occurs to the resolving capabilities of the algorithm used, the photo-kinetics of the fluorophores, and the unknown density and spatial distribution of the molecules being imaged. Incorporating this uncertainty in the density and spatial distribution of the molecules into this counting procedure is highly non-trivial and outside the scope of this paper. However, it is worth noting that [3] shows an imaging environment designed to minimise the number of fluorophores simultaneously in the On state can exponentially reduce this effect. Furthermore, recent developments in localization algorithms [e.g. 2] move ever closer to a satisfactory solution to this multi-emitter problem.

Acknowledgements

We would like to thank Prof Ricardo Henriques (UCL) for his valued input in earlier projects. We would also like to thank Prof Joerg Brewersdorf (Yale University) and Dr Yu Lin (European Molecular Biology Laboratory) for providing us with the Alexa Fluor 647 data used in this paper, and Dr. Nils Gustafsson for processing this data for our use.

bioRxiv preprint doi: https://doi.org/10.1101/834572; this version posted November 7, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

1 Appendix

In Section 1.1 we first detail proofs of Propositions 1 and 2. In Section 1.2, we describe the mathematical details needed to derive the probability mass function of the total number of localisations. In Section 1.3, we describe how to obtain the posterior distribution of M given N_l. In Section 1.4, we provide the necessarry algorithms required to compute this posterior. In Section 1.5, we provide a table with the parameters used when analysing the Alexa Fluor 647 data. In Section 1.6, we provide plots of these results.

1.1 Proofs

In this Section, we give detailed proofs of Propositions 1 and 2. Proposition 1 provides a method of computing the probability mass function of S_n, the cumulative number of localizations produced by a single molecule across n frames. Proposition 2 details its first and second moments, which uses the result of its probability generating function (pgf) derived in Lemma 1.

1.1.1 Proof of Proposition 1

Proof. Let $\mathbf{M}$ be as defined in equation (3).

Initializing with $n = 1$, we have (for $k \in \{0, 1\}$) that

$$\begin{aligned} M(j,k,1) &= \sum_{i \in \mathcal{S}_X} \mathbb{P}_{\boldsymbol{\theta}_d}(X(\Delta) = j, Y_0 = k | X(0) = i) \mathbb{P}_{\boldsymbol{\theta}_d}(X(0) = i) \\ &= \sum_{i \in \mathcal{S}_X} B_\Delta^{*(k)}(i,j) \mathbb{P}_{\boldsymbol{\theta}_d}(X(0) = i). \\ \implies \mathbf{M}(k,1) &= \boldsymbol{\nu}_X^\top B_\Delta^{*(k)}. \end{aligned}$$

For arbitrary $n > 1$, and for $k = 0$ we have

$$\begin{aligned} M(j,0,n) &= \sum_{i \in \mathcal{S}_X} \mathbb{P}_{\boldsymbol{\theta}_d}(X(n\Delta) = j, S_n = 0 | X((n-1)\Delta) = i, S_{n-1} = 0) M(i,0,n-1) \\ &= \sum_{i \in \mathcal{S}_X} B_\Delta^{*(0)}(i,j) M(i,0,n-1). \\ \implies \mathbf{M}(0,n) &= \mathbf{M}(0,n-1) B_\Delta^{*(0)}. \end{aligned}$$

bioRxiv preprint doi: https://doi.org/10.1101/834572; this version posted November 7, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

For $1 \leq k < n$ we have

$$\begin{aligned} M(j,k,n) &= \sum_{x=k-1}^{n} \sum_{i \in \mathcal{S}_X} \mathbb{P}_{\boldsymbol{\theta}_d}(X(n\Delta) = j, S_n = k | X((n-1)\Delta) = i, S_{n-1} = x) M(i, k-x, n-1) \\ &= \sum_{x=0}^{1} \sum_{i \in \mathcal{S}_X} B_{\Delta}^{*(x)}(i,j) M(i, k-x, n-1). \\ \implies \mathbf{M}(k,n) &= \mathbf{M}(k, n-1) B_{\Delta}^{*(0)} + \mathbf{M}(k-1, n-1) B_{\Delta}^{*(1)}. \end{aligned}$$

And finally for $k = n$, we have

$$\begin{aligned} M(j,n,n) &= \sum_{i \in \mathcal{S}_X} \mathbb{P}_{\boldsymbol{\theta}_d}(X(n\Delta) = j, S_n = n | X((n-1)\Delta) = i, S_{n-1} = n-1) M(i, n-1, n-1) \\ &= \sum_{i \in \mathcal{S}_X} B_{\Delta}^{*(1)}(i,j) M(i, n-1, n-1). \\ \implies \mathbf{M}(n,n) &= \mathbf{M}(n-1, n-1) B_{\Delta}^{*(1)}. \end{aligned}$$

Now since

$$\mathbb{P}_{\boldsymbol{\theta}_d}(S_n = k) = \sum_{j \in \mathcal{S}_X} \mathbb{P}_{\boldsymbol{\theta}_d}(X(n\Delta) = j, S_n = k),$$

we obtain

$$p_{\boldsymbol{\theta}_d}(S_n = k) := \mathbb{P}_{\boldsymbol{\theta}_d}(S_n = k) = \mathbf{M}(k,n) \mathbf{1}_{d+3} \qquad k \in \mathcal{S}_{S_n}.$$

□

1.1.2 Probability generating function (pgf)

In order to prove Proposition 2, we need a preliminary Lemma which derives the probability generating function (pgf) of S_n for $n \in \mathbb{Z}_{>0}$, since this result will be used in the main proof.

Lemma 1. *For any $n \in \mathbb{Z}_{>0}$, the probability generating function (pgf) of S_n, $G_{S_n}(z) = \mathbb{E}_{\boldsymbol{\theta}_d}(z^{S_n})$ is given by*

$$G_{S_n}(z) = \boldsymbol{\nu}_X^{\top} (B_{\Delta}^{*(0)} + z B_{\Delta}^{*(1)})^n \mathbf{1}_{d+3}. \tag{9}$$

Proof. By defining the vector quantity $\mathbf{G}_{S_n}(z) := \sum_{i=0}^{n} \mathbf{M}(i,n) z^i$, we have $G_{S_n}(z) = \mathbf{G}_{S_n}(z) \mathbf{1}_{d+3}$. We therefore need to equivalently show that $\mathbf{G}_{S_n}(z) = \boldsymbol{\nu}_X^{\top} (B_{\Delta}^{*(0)} + z B_{\Delta}^{*(1)})^n$.

bioRxiv preprint doi: https://doi.org/10.1101/834572; this version posted November 7, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

The statement in [9] is true for $n = 1$, since

$$\begin{aligned} G_{S_1}(z) &= \mathbb{P}(S_1 = 0) + z\mathbb{P}(S_1 = 1) \\ &= (\boldsymbol{\nu}_X^\top B_\Delta^{*(0)} + z\boldsymbol{\nu}_X^\top B_\Delta^{*(1)})\mathbf{1}_{d+3} \\ &= \boldsymbol{\nu}_X^\top (B_\Delta^{*(0)} + zB_\Delta^{*(1)})\mathbf{1}_{d+3}. \end{aligned}$$

Assuming that [9] is true for $n = k$, consider $n = k + 1$:

$$\begin{aligned} G_{S_{k+1}}(z) &= \sum_{i=0}^{k+1} \mathbb{P}(S_{k+1} = i) z^i \\ &= \left(\sum_{i=0}^{k+1} \mathbf{M}(i, k+1) z^i \right) \mathbf{1}_{d+3} \\ &= \left(\mathbf{M}(0,k) B_\Delta^{*(0)} + \left(\sum_{i=1}^{k} (\mathbf{M}(i,k) B_\Delta^{*(0)} + \mathbf{M}(i-1,k) B_\Delta^{*(1)}) z^i \right) + \mathbf{M}(k,k) B_\Delta^{*(1)} z^{k+1} \right) \mathbf{1}_{d+3} \\ &= \left(\left(\sum_{i=0}^{k} \mathbf{M}(i,k) z^i \right) B_\Delta^{*(0)} + z \left(\sum_{i=0}^{k} \mathbf{M}(i,k) z^i \right) B_\Delta^{*(1)} \right) \mathbf{1}_{d+3} \\ &= \mathbf{G}_{S_k}(z)(B_\Delta^{*(0)} + zB_\Delta^{*(1)})\mathbf{1}_{d+3} \\ &= \boldsymbol{\nu}_X^\top (B_\Delta^{*(0)} + zB_\Delta^{*(1)})^{k+1} \mathbf{1}_{d+3}. \end{aligned}$$

□

1.1.3 Proof of Proposition 2

Proof. The expected value of S_n, denoted $\mathbb{E}_{\boldsymbol{\theta}_d}(S_n) = G'_{S_n}(1)$ and variance $\mathbb{V}\mathrm{ar}_{\boldsymbol{\theta}_d}(S_n) = G''_{S_n}(1) + \mathbb{E}_{\boldsymbol{\theta}_d}(S_n) - \mathbb{E}^2_{\boldsymbol{\theta}_d}(S_n)$ can be explicitly determined by differentiating the pgf in (9) from first principles.

In the following, we utilize the following expansion

$$(C_z + hB_\Delta^{(1)})^n = C_z^n + hC_z^{n-1}B_\Delta^{(1)} + hC_z^{n-2}B_\Delta^{(1)}C_z + \ldots + hB_\Delta^{(1)}C_z^{n-1} + O(h^2),$$

which holds for the two square matrices C_z and $B_\Delta^{(1)}$.

bioRxiv preprint doi: https://doi.org/10.1101/834572; this version posted November 7, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

From the definition of a derivative, we have

$$
\begin{aligned}
G_{S_n}(z) &= \boldsymbol{\nu}_X^\top (B_\Delta^{*(0)} + zB_\Delta^{*(1)})^n \mathbf{1}_{d+3}. \\
\frac{dG_{S_n}}{dz} &= \lim_{dz\to 0} \frac{1}{dz}\left[\boldsymbol{\nu}_X^T (B_\Delta^{*(0)} + (z+dz)B_\Delta^{*(1)})^n \mathbf{1}_{d+3} - \boldsymbol{\nu}_X^T C_z^n \mathbf{1}_{d+3}\right] \\
&= \boldsymbol{\nu}_X^T \lim_{dz\to 0} \frac{(B_\Delta^{*(0)} + (z+dz)B_\Delta^{*(1)})^n - C_z^n}{dz} \mathbf{1}_{d+3} \\
&= \boldsymbol{\nu}_X^T \lim_{dz\to 0} \frac{(C_z + dzB_\Delta^{*(1)})^n - C_z^n}{dz} \mathbf{1}_{d+3} \\
&= \boldsymbol{\nu}_X^T \lim_{dz\to 0} \frac{C_z^n + C_z^{n-1}dzB_\Delta^{*(1)} + C_z^{n-2}dzB_\Delta^{*(1)}C_z + \cdots - C_z^n}{dz} \mathbf{1}_{d+3} \\
&= \boldsymbol{\nu}_X^T \left[C_z^{n-1}B_\Delta^{*(1)} + C_z^{n-2}B_\Delta^{*(1)}C_z + C_z^{n-3}B_\Delta^{*(1)}C_z^2 + \cdots + C_zB_\Delta^{*(1)}C_z^{n-2} + B_\Delta^{*(1)}C_z^{n-1}\right] \mathbf{1}_{d+3} \\
&\equiv \boldsymbol{\nu}_X^T \left[\sum_{i=1}^n C_z^{n-i}B_\Delta^{*(1)}C_z^{i-1}\right] \mathbf{1}_{d+3},
\end{aligned}
$$

defining $C_z := B_\Delta^{*(0)} + zB_\Delta^{*(1)}$.

When $z = 1$, $C_1 = B_\Delta^{*(0)} + B_\Delta^{*(1)} = \mathrm{e}^{G\Delta}$, giving

$$
\mathbb{E}_{\boldsymbol{\theta}_d}(S_n) = \boldsymbol{\nu}_X^T \left[\sum_{i=1}^n \mathrm{e}^{G\Delta(n-i)}B_\Delta^{*(1)}\mathrm{e}^{G\Delta(i-1)}\right] \mathbf{1}_{d+3}.
$$

Defining $D := \sum_{j=1}^{n-1} C_z^{n-1-j}B_\Delta^{*(1)}C_z^{j-1}$, we can now derive $G''_{S_n}(1)$ as follows

$$
\begin{aligned}
\frac{d^2G_{S_n}}{dz^2} &= \boldsymbol{\nu}_X^T \lim_{dz\to 0} \frac{1}{dz} \sum_{i=1}^n \left[C_{z+dz}^{n-i}B_\Delta^{*(1)}C_{z+dz}^{i-1} - C_z^{n-i}B_\Delta^{*(1)}C_z^{i-1}\right] \mathbf{1}_{d+3} \\
&= \boldsymbol{\nu}_X^T \left[DB_\Delta^{*(1)} + \lim_{dz\to 0} \frac{1}{dz} \sum_{i=2}^{n-1} C_{z+dz}^{n-i}B_\Delta^{*(1)}C_{z+dz}^{i-1} - C_z^{n-i}B_\Delta^{*(1)}C_z^{i-1} + B_\Delta^{*(1)}D\right] \mathbf{1}_{d+3} \\
&= \boldsymbol{\nu}_X^T \left[DB_\Delta^{*(1)} + \sum_{i=2}^{n-1}\left(\sum_{j=1}^{n-i} C_z^{n-(i+j)}B_\Delta^{*(1)}C_z^{j-1}B_\Delta^{*(1)}C_z^{i-1} + \right.\right. \\
&\quad \left.\left. \sum_{k=2}^{i} C_z^{n-i}B_\Delta^{*(1)}C_z^{i-k}B_\Delta^{*(1)}C_z^{k-2}\right) + B_\Delta^{*(1)}D\right] \mathbf{1}_{d+3}.
\end{aligned}
$$

bioRxiv preprint doi: https://doi.org/10.1101/834572; this version posted November 7, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

This gives

$$G''_{S_n}(1) = \boldsymbol{\nu}_X^T \left(\sum_{i=1}^{n-1} \sum_{j=1}^{n-i} \mathrm{e}^{G\Delta(n-i-j)} B_\Delta^{*(1)} \mathrm{e}^{G\Delta(j-1)} B_\Delta^{*(1)} \mathrm{e}^{G\Delta(i-1)} + \right.$$
$$\left. \sum_{j=1}^{i} \mathrm{e}^{G\Delta(n-i-1)} B_\Delta^{*(1)} \mathrm{e}^{G\Delta(i-j)} B_\Delta^{*(1)} \mathrm{e}^{G\Delta(j-1)} \right) \mathbf{1}_{d+3},$$

so that $\mathbb{E}_{\boldsymbol{\theta}_d}(S_n^2) = G''_{S_n}(1) + \mathbb{E}_{\boldsymbol{\theta}_d}(S_n)$ and therefore $\mathbb{V}\mathrm{ar}_{\boldsymbol{\theta}_d}(S_n) = G''_{S_n}(1) + \mathbb{E}_{\boldsymbol{\theta}_d}(S_n) - \mathbb{E}^2_{\boldsymbol{\theta}_d}(S_n)$.

□

1.2 Deriving the probability distribution of the total number of localizations

We defined the total number of localizations N_l detected from M fluorophores during an experiment (consisting of N_F frames) to be

$$N_l = \sum_{m=1}^{M} S_{N_F,m},$$

where $S_{N_F,m}$ denotes the cumulative number of localizations made by the mth fluorophore. The distribution S_{N_F} (for a single fluorophore) is carefully derived in Proposition 1, with Algorithm 1 providing the user with a scheme to computationally compute it given photo-switching parameters $\boldsymbol{\theta}_d$. Here, we describe how this can now be used to recover the probability mass function for N_l, given M.

Firstly, for any $u \in \mathbb{R}$, we define the *characteristic function* $\gamma_{S_{N_F}}(u)$ of the random variable S_{N_F} to be

$$\begin{aligned} \gamma_{S_{N_F}}(u) := \mathbb{E}_{\boldsymbol{\theta}_d}(\mathrm{e}^{\mathrm{i}uS_{N_F}}) &= \sum_{s=0}^{\infty} \mathbb{P}_{\boldsymbol{\theta}_d}(S_{N_F} = s)\mathrm{e}^{\mathrm{i}us} \\ &= \sum_{s=0}^{N_F} p_{\boldsymbol{\theta}_d}(S_{N_F} = s)\mathrm{e}^{\mathrm{i}us}, \end{aligned}$$

where $\mathrm{i} = \sqrt{-1}$. The characteristic function for $N_l = \sum_{m=1}^{M} S_{N_F,m}$ is then

$$\begin{aligned} \mathbb{E}_{\boldsymbol{\theta}_d,M}(\mathrm{e}^{\mathrm{i}uN_l}) &= \mathbb{E}_{\boldsymbol{\theta}_d,M}\left(\mathrm{e}^{\mathrm{i}u\left(S_{N_F,1}+\cdots+S_{N_F,M}\right)}\right) \\ &= \mathbb{E}_{\boldsymbol{\theta}_d,M}\left(\mathrm{e}^{\mathrm{i}uS_{N_F,1}} \ldots \mathrm{e}^{\mathrm{i}uS_{N_F,M}}\right) \\ &= \prod_{m=1}^{M} \mathbb{E}_{\boldsymbol{\theta}_d}\left(\mathrm{e}^{\mathrm{i}uS_{N_F,m}}\right) \qquad \text{(since } S_{N_F,1},\ldots,S_{N_F,M} \text{ are independent)} \\ &= \gamma^M_{S_{N_F}}(u) \qquad \text{(since } S_{N_F,1},\ldots,S_{N_F,M} \text{ are identically distributed).} \end{aligned} \tag{10}$$

bioRxiv preprint doi: https://doi.org/10.1101/834572; this version posted November 7, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

For any $N \geq 0$, we can define $t_N := \frac{2\pi}{N+1}$ and $u_N = -t_N k$, where k can take any value in the set $\{0, \ldots, N\}$. When $N = N_F$, this enables

$$\mathcal{F}_{s\to k}(p_{\boldsymbol{\theta}_d}(S_{N_F})) := \gamma_{S_{N_F}}(-u_{N_F}) = \sum_{s=0}^{N_F} p_{\boldsymbol{\theta}_d}(S_{N_F} = s)\mathrm{e}^{-\mathrm{i}t_{N_F}ks}$$

to be seen as the Discrete Fourier Transform (DFT) of the probability mass $p_{\boldsymbol{\theta}_d}(S_{N_F} = s)$, where $\mathcal{F}_{s\to k}(\cdot)$ denotes the discrete Fourier operator. The inverse DFT can then recover the probabilities via

$$\begin{aligned}\mathcal{F}^{-1}_{k\to s}(\gamma_{S_{N_F}}(-t_{N_F}k)) &= \frac{1}{N_F+1}\sum_{k=0}^{N_F}\gamma_{S_{N_F}}(-t_{N_F}k)\mathrm{e}^{\mathrm{i}t_{N_F}ks} \\ &\equiv p_{\boldsymbol{\theta}_d}(S_{N_F} = s).\end{aligned}$$

Using the characteristic function of N_l from [10], it now follows that probability mass $p_{\boldsymbol{\theta}_d,M}(N_l = s) := \mathbb{P}_{\boldsymbol{\theta}_d,M}(N_l = s)$ (where N_l takes values in the set $\{0, \ldots, MN_F\}$), can be recovered via

$$p_{\boldsymbol{\theta}_d,M}(N_l = s) = \frac{1}{MN_F+1}\sum_{k=0}^{MN_F}\gamma^M_{S_{N_F}}\left(-t_{MN_F}k\right)\mathrm{e}^{\mathrm{i}t_{MN_F}ks}, \tag{11}$$

so that $p_{\boldsymbol{\theta}_d,M}(N_l = s) = \mathcal{F}^{-1}_{k\to s}(\gamma^M_{S_{N_F}}(-t_{MN_F}k)) = \mathcal{F}^{-1}_{k\to s}(\mathcal{F}^M_{s\to k}(p_{\boldsymbol{\theta}_d}(S_{N_F})))$. It should be noted here that a computational implementation would require one to apply the DFT to the MN_F+1 vector of probabilities $\mathbf{p}$, whose $(s+1)$th element is defined as $p_{\boldsymbol{\theta}_d}(S_{N_F} = s)$. The first $N_F + 1$ elements of $\mathbf{p}$ are therefore those outputted by Algorithm 1 and the remaining $N_F(M-1)$ elements are zeros. Algorithm 4 of this supplement provides the user with a scheme to compute the probability distribution of N_l using this reasoning.

1.3 Deriving the posterior distribution of M

We defined the posterior distribution of M given the number of observed localizations N_l in test data $\mathcal{D}_{te} = \{N_l, \Delta, N_F\}$ and $\boldsymbol{\theta}_d$ the set of photo-switching parameters learned from training data $\mathcal{D}_{tr}$.

We choose $M_{\min} = \max\left(\left\lceil\frac{N_l}{N_F}\right\rceil, 1\right)$ and while it should be clear that $M_{\max} = \infty$, one may choose to pre-specify a large value for $M_{\max}$ to avoid unnecessarily large computations. For example, we let $\hat{m} = \left\lceil\frac{N_l}{\mathbb{E}_{\boldsymbol{\theta}_d}(S_{N_F})}\right\rceil$ and $M_{\max} = \hat{m} + \left\lceil 4\sqrt{\hat{m}\mathbb{V}\mathrm{ar}_{\boldsymbol{\theta}_d}(S_{N_F})}\right\rceil$ and consider the

bioRxiv preprint doi: https://doi.org/10.1101/834572; this version posted November 7, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

range $[M_{\min}, M_{\max}]$ suitable for inference. Here, $\mathbb{E}_{\boldsymbol{\theta}_d}(S_{N_F})$ and $\mathbb{V}\mathrm{ar}_{\boldsymbol{\theta}_d}(S_{N_F})$ can be computed using equations (5) and (6). For the studies conducted, we chose $M_{\min}$ and $M_{\max}$ using this reasoning. For a given prior distribution π_M, Algorithm 5 computes $p_{\boldsymbol{\theta}_d,m}(M = m|N_l)$ using this described method.

bioRxiv preprint doi: https://doi.org/10.1101/834572; this version posted November 7, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

1.4 Algorithms

In this Section, we provide two additional algorithms to supplement the material presented in this paper. Firstly, Algorithm 3 presents the algorithm to compute transmission matrices $B_{\Delta}^{*(0)}$ and $B_{\Delta}^{*(1)}$ given any parameter set $\boldsymbol{\theta}_d$. This algorithm has been taken from [15], and presented here for convenience. Secondly, we provide an algorithm to compute the probability mass function (distribution) of the total number of localizations N_l as previously described and in equation (11).

A small note on the notation used in Algorithm 3. We denote $\mathbf{0}_n$ and $\mathbf{1}_n$ to be the $n \times 1$ vectors of zeros and ones respectively and I_n to be the $n \times n$ identity matrix. Moreover, $\mathbf{e}_n^p$ denotes the pth canonical (standard) basis vector of $\mathbb{R}^n$. We denote $A[i_1 : i_2, j_1 : j_2]$ to be the matrix filled with rows i_1 to i_2 and columns j_1 to j_2 of any matrix A, and $A[i_1, j_1]$ to be the (i_1, j_1)th entry of A. We use the $\odot$ notation to denote the Hadamard (element wise) product between two matrices. Moreover, the Laplace transform of a scalar-valued function $q_{ij}(\mathbf{k}, t)$ with respect to its arguments $i, j \in \mathbb{Z}_{>0}, \mathbf{k} \in \mathbb{R}^n$ and $t \geq 0$, is defined as $\mathcal{L}_{t \to s}[q_{ij}(\mathbf{k}, t)](s) =: f_{ij}(\mathbf{k}, s) = \int_0^{\infty} \mathrm{e}^{-st} q_{ij}(\mathbf{k}, t)\mathrm{dt}$. The Laplace operator on a matrix-valued function is applied element wise to create a matrix output of the same dimension as the input.

bioRxiv preprint doi: https://doi.org/10.1101/834572; this version posted November 7, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Algorithm 5 Compute posterior distribution $p_{\boldsymbol{\theta}_d,m}(M = m|N_l)$

function COMPUTE_POSTERIOR($\mathcal{D}_{tr}, \mathcal{D}_{te}, \pi_M$)

Use $\mathcal{D}_{tr}$ to obtain $\boldsymbol{\theta}_d$ ▷ E.g. via the method in [15]

$\mathbf{p} \leftarrow \text{PMF_S}(\boldsymbol{\theta}_d, \Delta, N_F)$ ▷ From Algorithm 1

Compute $\mathbb{E}_{\boldsymbol{\theta}_d}(S_{N_F}), \mathbb{V}\text{ar}_{\boldsymbol{\theta}_d}(S_{N_F})$ ▷ From (5) and (6)

$M_{\min} \leftarrow \max\left(\left\lceil \frac{N_l}{N_F} \right\rceil, 1\right)$

$\hat{m} \leftarrow \left\lceil \frac{N_l}{\mathbb{E}_{\boldsymbol{\theta}_d}(S_{N_F})} \right\rceil$

$M_{\max} \leftarrow \hat{m} + \left\lceil 4\sqrt{\hat{m}\mathbb{V}\text{ar}_{\boldsymbol{\theta}_d}(S_{N_F})} \right\rceil$

$\mathbf{p}^* \leftarrow \mathbf{0}_{M_{\max}}$

for $i = M_{\min}$ to $M_{\max}$ **do**

$\mathbf{p_2} \leftarrow \text{PMF_NL}(\mathbf{p}, i)$ ▷ From Algorithm 4

$\mathbf{p}^*[i] \leftarrow \mathbf{p_2}[N_l + 1]\pi_M(i)$

$\mathbf{p}^* \leftarrow \frac{\mathbf{p}^*}{\mathbf{p}^*\mathbf{1}_{M_{\max}}}$ ▷ Normalize probabilities

return $\mathbf{p}^*$ ▷ $\mathbf{p}^*[m] = \mathbb{P}_{\boldsymbol{\theta}_d,m}(M = m|N_l)$

bioRxiv preprint doi: https://doi.org/10.1101/834572; this version posted November 7, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Algorithm 3 Compute transmission matrices $B_\Delta^{*(0)}$ and $B_\Delta^{*(1)}$

function COMPUTE_TRANSMISSION_MATRICES($\boldsymbol{\theta}_d, \Delta$)

Compute G from $\boldsymbol{\theta}_d$ using equation (1)

$G_{S,\mathcal{R}^0} \leftarrow \mathbf{0}_{d+2}\mathbf{0}_{d+2}^\top$

$G_S \leftarrow G[1:d+2, 1:d+2]$ ▷ To avoid numerical overflow in the computation of inverse Laplace transforms, one can (for some small tolerance $\epsilon > 0$), replace all such $(G)_{p,p}$ with $(G)_{q,q}$, when $|(G)_{p,p} - (G)_{q,q}| < \epsilon$; $p \neq q = 1, \ldots, d+2$.

$\boldsymbol{\mu} \leftarrow G[1:d+2, d+3]$

$\sigma_1 \leftarrow -G[d+2, d+2]$

$\boldsymbol{\sigma} \leftarrow -\mathrm{diag}(G[1:d+1, 1:d+1])$

for $i = 1$ to $d+1$ **do**

$G_{S,\mathcal{R}^0}[i, d+2] \leftarrow G_S[i, d+2]$

$G_{S,\bar{\mathcal{R}}^0} \leftarrow G_S - G_{S,\mathcal{R}^0}$

▷ //Compute initializations for transmission matrices

$$A_1 \leftarrow \begin{bmatrix} -G_{S,\bar{\mathcal{R}}^0}^\top & I_{d+2} \\ \mathbf{0}_{d+2}\mathbf{0}_{d+2}^\top & -G_{S,\bar{\mathcal{R}}^0}^\top \end{bmatrix}$$

$$A_2 \leftarrow \begin{bmatrix} G_{S,\bar{\mathcal{R}}^0} & I_{d+2} \\ \mathbf{0}_{d+2}\mathbf{0}_{d+2}^\top & \mathbf{0}_{d+2}\mathbf{0}_{d+2}^\top \end{bmatrix}$$

$$A \leftarrow \begin{bmatrix} A_1 & \mathbf{0}_{2(d+2)}\mathbf{0}_{2(d+2)}^\top \\ \mathbf{0}_{2(d+2)}\mathbf{0}_{2(d+2)}^\top & A_2 \end{bmatrix}$$

$Q_\Delta^0(0) \leftarrow \mathrm{e}^{G_{S,\bar{\mathcal{R}}^0}\Delta}$

$\bar{\mathbf{Q}}_\Delta^0(0) \leftarrow \mathrm{e}^{A\Delta}[i_1 : i_2, i_2+1 : i_3]\boldsymbol{\mu}$ ▷ $i_1 = 2d+5, i_2 = 3(d+2)$ and $i_3 = 4(d+2)$

$c \leftarrow \frac{1-e^{-\sigma_1\delta}}{1-e^{-\sigma_1\Delta}}$

$\Xi_\Delta^0(0) \leftarrow \begin{bmatrix} \mathbf{1}_{d+1}\mathbf{1}_{d+1}^\top & c\mathbf{1}_{d+1} \end{bmatrix}^\top$

$\Xi_\Delta^1(0) \leftarrow \mathbf{1}_{d+2}\mathbf{1}_{d+1}^\top - \Xi_\Delta^0(0)$

$\bar{\boldsymbol{\Xi}}_\Delta^0(0) \leftarrow \begin{bmatrix} \mathbf{1}_{d+1}^\top & c \end{bmatrix}^\top$

$\bar{\boldsymbol{\Xi}}_\Delta^1(0) \leftarrow \mathbf{1}_{d+2} - \bar{\boldsymbol{\Xi}}_\Delta^0(0)$

$$B_\Delta^{(0)} \leftarrow \begin{bmatrix} (Q_\Delta^0(0))_{(1:d+2),(1:d+1)} \odot \Xi_\Delta^0(0) & \mathbf{0}_{d+2} & \bar{\mathbf{Q}}_\Delta^0(0) \odot \bar{\boldsymbol{\Xi}}_\Delta^0(0) \\ \mathbf{0}_{d+1}^\top & 0 & 1 \end{bmatrix}$$

bioRxiv preprint doi: https://doi.org/10.1101/834572; this version posted November 7, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

$B_\Delta^{(1)} \leftarrow \begin{bmatrix} (Q_\Delta^0(0))_{(1:d+2),(1:d+1)} \odot \Xi_\Delta^1(0) & [\mathbf{0}_{d+1}^\top \quad \mathrm{e}^{-\sigma_1\Delta}]^\top & \bar{\mathbf{Q}}_\Delta^0(0) \odot \bar{\mathbf{\Xi}}_\Delta^1(0) \\ \mathbf{0}_{d+1}^\top & 0 & 0 \end{bmatrix}$

$k \leftarrow 1$ ▷ //Start convergence of transmission matrices via computations of different k

while $B_\Delta^{(0)}$ and $B_\Delta^{(1)}$ have not converged **do**

$Q_\Delta^0(k) \leftarrow \mathcal{L}_s^{-1}[(sI_{d+2} - G_{S,\bar{\mathcal{R}}^0})^{-1} \left(G_{S,\mathcal{R}^0}(sI_{d+2} - G_{S,\bar{\mathcal{R}}^0})^{-1}\right)^k](\Delta)$ ▷ Compute inverse Laplace transform matrix

$\bar{\mathbf{Q}}_\Delta^0(k) \leftarrow \left(\int_0^\Delta Q_s^0(k)\mathrm{ds}\right)\boldsymbol{\mu}$

for $i = 1$ to $d+1$ **do**

for $j = 1$ to $d+1$ **do**

▷ $\Upsilon \sim \mathrm{Erlang}(k, \sigma_1)$ and $F_\Upsilon(u, k, \sigma_1) = \mathbb{P}(\Upsilon \leq u)$

$\Xi_\Delta^0(k)[i,j], \bar{\mathbf{\Xi}}_\Delta^0(k)[i] \leftarrow \frac{F_\Upsilon(\delta,k,\sigma_1)}{F_\Upsilon(\Delta,k,\sigma_1)}$

$\Xi_\Delta^1(k)[i,j] \leftarrow 1 - (\Xi_\Delta^0(k))[i,j]$

$\bar{\mathbf{\Xi}}_\Delta^1(k)[i] \leftarrow 1 - \bar{\mathbf{\Xi}}_\Delta^0(k)[i]$

$\Xi_\Delta^0(k)[d+2,j], \bar{\mathbf{\Xi}}_\Delta^0(k)[d+2] \leftarrow \frac{F_\Upsilon(\delta,k+1,\sigma_1)}{F_\Upsilon(\Delta,k+1,\sigma_1)}$

$\Xi_\Delta^1(k)[d+2,j] \leftarrow 1 - \Xi_\Delta^0(k)[d+2,j]$

$\bar{\mathbf{\Xi}}_\Delta^1(k)[d+2] \leftarrow 1 - \bar{\mathbf{\Xi}}_\Delta^0(k)[d+2]$

$B_\Delta^{(0)} \leftarrow B_\Delta^{(0)} + \begin{bmatrix} Q_\Delta^0(k)[1:d+2, 1:d+1] \odot \Xi_\Delta^0(k) & \mathbf{0}_{d+2} & \bar{\mathbf{Q}}_\Delta^0(k) \odot \bar{\mathbf{\Xi}}_\Delta^0(k) \\ \mathbf{0}_{d+1}^\top & 0 & 0 \end{bmatrix}$

$B_\Delta^{(1)} \leftarrow B_\Delta^{(1)} + \begin{bmatrix} Q_\Delta^0(k)[1:d+2, 1:d+1] \odot \Xi_\Delta^1(k) & \mathbf{0}_{d+1} & \bar{\mathbf{Q}}_\Delta^0(k) \odot \bar{\mathbf{\Xi}}_\Delta^1(k) \\ \mathbf{0}_{d+1}^\top & 0 & 0 \end{bmatrix}$

for $i = 1$ to $d+2$ **do**

Find all vectors $\mathbf{k} = \begin{pmatrix} k_0 & k_1 & \ldots & k_d \end{pmatrix}^\top$ that belong to the set $\mathcal{C}_k^{0_{i-1}}$ ▷ $\mathcal{C}_k^{0_{i-1}} := \left\{\mathbf{k} : \mathbf{k}^\top \mathbf{1}_{d+1} = k, k_{i-1} > 0, k_0 \geq \ldots \geq k_{i-1} - 1 \geq \ldots \geq k_d - 1\right\}$

$\mathcal{C}_k^{0_{d+1}} \leftarrow \mathcal{C}_k^0$

For each $\mathbf{k}$, $f_{0_{i-1}1}(\mathbf{k}, s) \leftarrow \frac{\lambda_{10}}{s+\sigma_1} \sum_{p=0}^d \frac{\lambda_{0_p 1} \prod_{q=0}^{p-1} \lambda_{0_q 0_{q+1}}}{\prod_{q=0}^p (s+\sigma_{0_q})} f_{0_{i-1}1}\left(\mathbf{k} - \sum_{r=0}^p \mathbf{e}_{d+1}^{r+1}, s\right)$ ▷ Compute $f_{0_{i-1}1}(\mathbf{k}, s)$ recursively via the initializations $f_{0_{i-1}1}(\mathbf{0}_{d+1}, s) = \frac{\mathbb{1}_{\{d+2\}}(i)}{s+\sigma_1}$, $f_{0_p 1}(\mathbf{e}_{d+1}^{p+1}, s) = \frac{\lambda_{0_p 1}}{(s+\sigma_{0_p})(s+\sigma_1)}$ for $p = 0, \ldots, d$, and $f_{0_{d+1}1}(\mathbf{e}_{d+1}^1, s) = \frac{\lambda_{10}\lambda_{01}}{(s+\sigma_0)(s+\sigma_1)^2}$.

For each $\mathbf{k}$, compute $q_{0_{i-1}1}^1(\mathbf{k}, \Delta) = \mathcal{L}_s^{-1}(f_{0_{i-1}1}(\mathbf{k}, s))(\Delta)$ ▷ Compute inverse Laplace transforms

bioRxiv preprint doi: https://doi.org/10.1101/834572; this version posted November 7, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

$\xi^1_{0_{i-1}1}(0, \mathbf{k}, \Delta) \leftarrow \frac{F_\Phi(\Delta|\mathbf{k},\boldsymbol{\sigma}) - F_\Phi(\Delta-\delta|\mathbf{k},\boldsymbol{\sigma})}{F_\Phi(\Delta|\mathbf{k},\boldsymbol{\sigma})}$ ▷ $F_\Phi(\phi|\mathbf{k}, \boldsymbol{\sigma}) = \mathbb{P}(\Phi \leq \phi)$, where $\Phi = \sum_{p=0}^{m} W_p,\ W_p \overset{indep}{\sim} \text{Erlang}(k_p, \sigma_{0_p})$

$\xi^1_{0_{d+1}1}(0, \mathbf{k}, \Delta) \leftarrow \xi^1_{01}(0, \mathbf{k}, \Delta)$

$B^{(0)}_\Delta[i, d+2] \leftarrow B^{(0)}_\Delta[i, d+2] + \sum_{\mathbf{k}\in\mathcal{C}^{0_{i-1}}_k} q^1_{0_{i-1}1}(\mathbf{k}, \Delta)\xi^1_{0_{i-1}1}(0, \mathbf{k}, \Delta)$

$B^{(1)}_\Delta[i, d+2] \leftarrow B^{(1)}_\Delta[i, d+2] + \sum_{\mathbf{k}\in\mathcal{C}^{0_{i-1}}_k} q^1_{0_{i-1}1}(\mathbf{k}, \Delta)(1 - \xi^1_{0_{i-1}1}(0, \mathbf{k}, \Delta))$

$k \leftarrow k + 1$

▷ //Include the addition of false positives to transmission matrices

$B^{*(0)}_\Delta \leftarrow (1-\alpha)B^{(0)}_\Delta$

$B^{*(1)}_\Delta \leftarrow B^{(1)}_\Delta + \alpha B^{(0)}_\Delta$

return $B^{*(0)}_\Delta, B^{*(1)}_\Delta$ ▷ Output transmission matrices

Algorithm 4 Compute probability mass function (PMF) for N_l from M fluorophores

function PMF_NL($\mathbf{p}_1, M$) ▷ $\mathbf{p}_1 \leftarrow$ PMF_S($\boldsymbol{\theta}_d, N_F$) from Algorithm 1

$\mathbf{p}_2 \leftarrow [\mathbf{p}_1^\top\ \mathbf{0}^\top_{N_F(M-1)}]^\top$

$\mathbf{f} \leftarrow \mathcal{F}(\mathbf{p}_2)$ ▷ Apply Discrete Fourier Transform (DFT) to $\mathbf{p}_2$ to get $\mathbf{f}$

$\mathbf{f}_M \leftarrow \mathbf{f}^M$ ▷ $\mathbf{f}_M[i] = \mathbf{f}[i]^M$ for $i = 1, \ldots, MN_F + 1$

$\mathbf{p} \leftarrow \mathcal{F}^{-1}(\mathbf{f}_M)$ ▷ Apply inverse DFT to $\mathbf{f}_M$ to get $\mathbf{p}$, where $\mathbf{p}[i] = \mathbb{P}_{\boldsymbol{\theta}_d,M}(N_l = i-1)$ for $i = 1, \ldots, MN_F + 1$

return $\mathbf{p}$ ▷ Probability mass function for N_l

bioRxiv preprint doi: https://doi.org/10.1101/834572; this version posted November 7, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

1.5 Tables

In this Section, we provide a Table to detail the imaging parameters $\boldsymbol{\theta}_2$ used when deriving the posterior distribution of M_{te} given $\boldsymbol{\theta}_2$ for the 27 Alexa FLuor 647 datasets studied. As explained, for each study, a training set of size $N_F \times M_{tr}$ from the whole dataset was used to determine $\boldsymbol{\theta}_2$ via the PSHMM method [15]. A model with $d = 2$ was used when learning $\boldsymbol{\theta}_2$, further reasoned in [15]. Table 3 provides the number of each study, the Laser intensity used, Δ, M_{tr}, M_{te}, N_F and the maximum likelihood parameter estimates in $\boldsymbol{\theta}_2$.

1.6 Figures

In this section, we provide the posterior distributions of M given N_l from the Alexa Fluor 647 datasets studied in Section . Speficially, Figure 7 shows the posterior distributions of M given N_l, along with the true values and MAP estimates from the 27 experiments. Moreover, each distribution's 95% credible interval (under a uniform prior on M) is given.

bioRxiv preprint doi: https://doi.org/10.1101/834572; this version posted November 7, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Dataset	Laser intensity	Δ^{-1}	M_{tr}	M_{te}	N_F	λ_{00_1} $\times$ Δ	λ_{01} $\times$ Δ	$\lambda_{0_1 0_2}$ $\times$ Δ	$\lambda_{0_1 1}$ $\times 10$ $\times\Delta$	$\lambda_{0_2 1}$ $\times 10^4$ $\times\Delta$	λ_{10} $\times$ Δ	μ_1 $\times 10^2$ $\times\Delta$	$\frac{\delta}{\Delta}$	α $\times 10^5$	$\boldsymbol{\nu}_X$
1	1.0	200	192	81	49796	0.10	0.55	0.01	0.22	1.24	0.65	1.04	0.78	1.48	(0.21, 0.00, 0.65, 0.13, 0)
2	1.9	200	180	77	49533	0.23	0.73	0.02	0.46	1.43	0.92	1.37	0.32	1.13	(0.00, 0.46, 0.34, 0.20, 0)
3	3.9	200	234	100	49815	0.12	0.46	0.02	0.21	0.58	0.55	2.44	0.65	0.80	(0.10, 0.07, 0.70, 0.13, 0)
4	3.9	200	295	110	39758	0.28	0.67	0.03	0.42	1.22	0.55	2.53	0.69	0.98	(0.02, 0.12, 0.61, 0.24, 0)
5	7.8	200	238	102	39721	0.14	0.39	0.02	0.14	1.42	0.55	2.98	0.57	0.27	(0.10, 0.06, 0.72, 0.12, 0)
6	7.8	800	171	72	29418	0.03	0.15	1.35	6.08	1.39	0.52	0.65	0.56	1.17	(0.52, 0.00, 0.00, 0.47, 0)
7	7.8	800	159	67	29257	0.25	0.58	0.02	0.47	1.12	0.81	0.61	0.37	1.60	(0.50, 0.03, 0.00, 0.47, 0)
8	7.8	800	121	51	29438	0.13	0.40	0.01	0.23	0.68	0.54	0.00	0.66	0.09	(0.71, 0.00, 0.00, 0.29, 0)
9	16	800	304	129	29467	0.38	0.70	0.02	0.57	0.81	0.59	1.18	0.77	0.72	(0.23, 0.03, 0.00, 0.74, 0)
10	16	200	201	86	39703	0.19	0.42	0.01	0.08	1.25	0.57	3.10	0.73	0.83	(0.00, 0.01, 0.46, 0.53, 0)
11	16	800	213	90	29074	0.21	0.46	0.03	0.37	0.73	0.54	0.00	0.64	0.48	(0.54, 0.00, 0.00, 0.46, 0)
12	16	800	201	85	29145	0.12	0.35	0.02	0.19	0.72	0.57	0.00	0.61	0.00	(0.13, 0.00, 0.00, 0.87, 0)
13	31	800	425	181	29059	0.21	0.41	0.03	0.28	0.75	0.58	0.01	0.72	0.93	(0.33, 0.07, 0.04, 0.56, 0)
14	31	800	374	159	29778	0.25	0.50	0.04	0.30	0.71	0.70	0.01	0.75	0.95	(0.26, 0.00, 0.00, 0.74, 0)
15	31	800	360	153	29179	0.13	0.32	0.02	0.11	0.70	0.61	0.00	0.63	0.34	(0.50, 0.00, 0.09, 0.41, 0)
16	31	800	343	147	29400	0.17	0.38	0.03	0.20	0.68	0.65	0.00	0.67	0.35	(0.25, 0.00, 0.00, 0.75, 0)
17	31	800	317	135	29071	0.21	0.47	0.03	0.34	0.75	0.59	0.00	0.68	1.18	(0.09, 0.00, 0.00, 0.91, 0)
18	62	800	385	164	29327	0.22	0.37	0.04	0.21	0.87	0.69	0.17	0.61	1.35	(0.26, 0.00, 0.00, 0.73, 0)
19	62	800	309	132	29107	0.25	0.47	0.04	0.26	0.87	0.69	0.23	0.66	1.10	(0.54, 0.00, 0.00, 0.46, 0)
20	62	800	294	126	29551	0.18	0.36	0.03	0.15	0.60	0.75	0.00	0.63	1.20	(0.14, 0.04, 0.00, 0.81, 0)
21	62	800	298	127	29426	0.16	0.39	0.03	0.14	0.77	0.65	0.05	0.67	1.68	(0.06, 0.00, 0.00, 0.94, 0)
22	62	800	279	119	28989	0.17	0.37	0.03	0.16	0.85	0.67	0.00	0.60	1.35	(0.39, 0.00, 0.00, 0.61, 0)
23	97	800	315	135	29191	0.21	0.36	0.04	0.19	0.95	0.79	3.50	0.60	0.75	(0.45, 0.00, 0.00, 0.55, 0)
24	97	800	307	131	29198	0.17	0.30	0.02	0.08	0.75	0.77	1.10	0.67	1.11	(0.36, 0.00, 0.00, 0.64, 0)
25	97	800	304	129	29270	0.30	0.48	0.04	0.27	1.17	0.75	2.47	0.61	1.97	(0.00, 0.00, 0.00, 1.00, 0)
26	97	800	295	126	29295	0.18	0.42	0.02	0.10	1.04	0.62	1.35	0.82	1.14	(0.17, 0.00, 0.00, 0.82, 0)
27	97	800	287	123	29218	0.26	0.51	0.04	0.34	0.96	0.71	4.22	0.79	0.93	(0.51, 0.00, 0.00, 0.48, 0)

Table 3: A description of the Alexa Fluor 647 datasets, with reference to the laser intensities in kW/cm^2 and frames sampled per second (or Δ^{-1}) measured in s^{-1} used to characterize each of the 27 experiments. For each dataset, a training set of size $N_F \times M_{tr}$ (train) was used to find the maximum likelihood estimate $\boldsymbol{\theta}_2$ via the PSHMM (estimated values shown). A hold out test set of size $N_F \times M_{te}$ (test) was used in the posterior computations of M.

bioRxiv preprint doi: https://doi.org/10.1101/834572; this version posted November 7, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

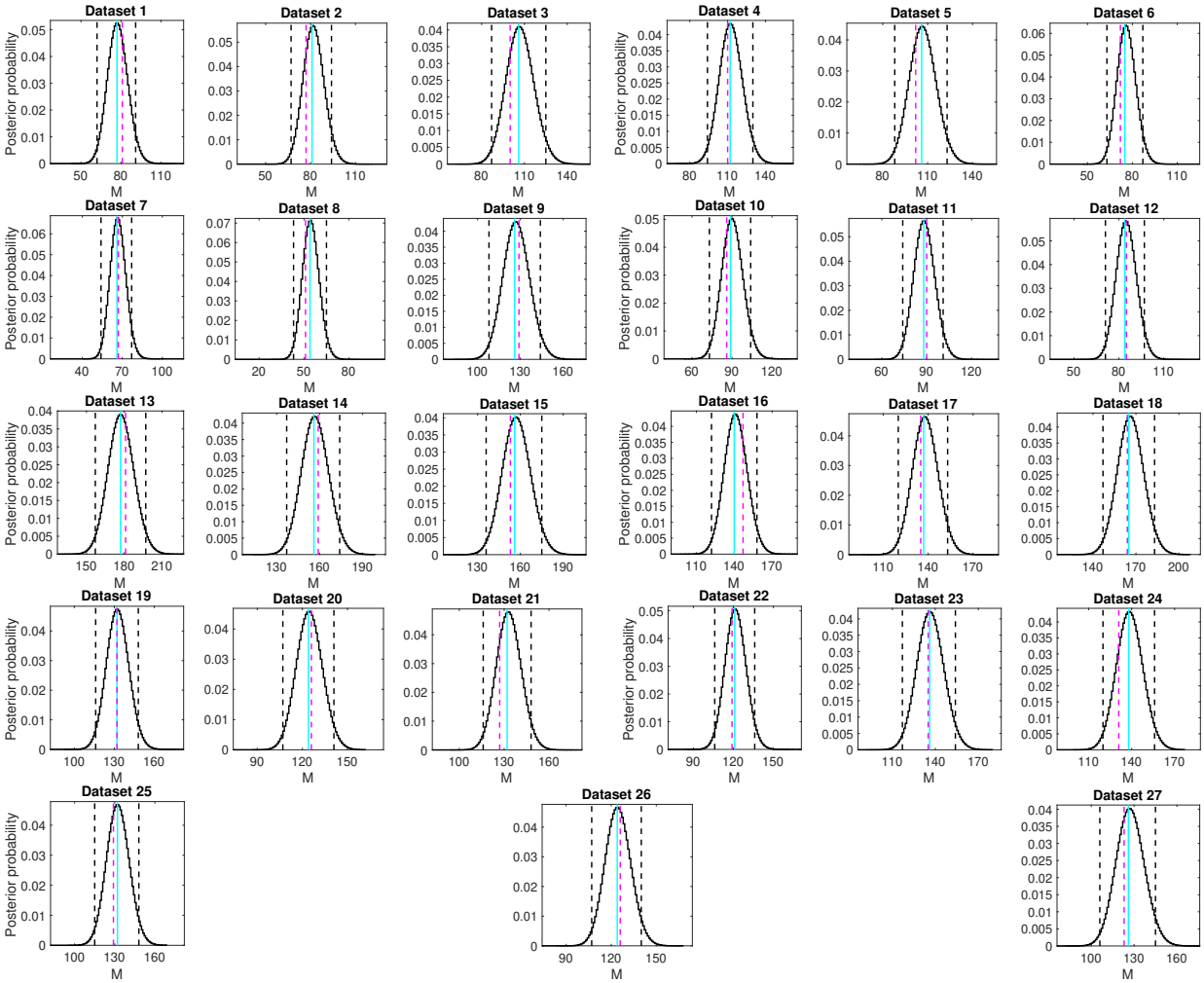


Figure 7: Posterior distributions of M_{te} given $\hat{\boldsymbol{\theta}}_2$ and N_l for the 27 Alexa FLuor 647 datasets (descriptions of which can be found in Table 3). For each study, $\hat{M}$ is given by the corresponding posterior mode plotted in cyan, with the true values of M_{te} shown in magenta (dotted). 95% credible intervals for each $\hat{M}$ are shown in black (dotted).

bioRxiv preprint doi: https://doi.org/10.1101/834572; this version posted November 7, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

References

[1] E. Betzig, G. H. Patterson, R. Sougrat, O. W. Lindwasser, S. Olenych, J. S. Bonifacino, M. W. Davidson, J. Lippincott-Schwartz, and H. F. Hess. Imaging Intracellular Fluorescent Proteins at Nanometer Resolution. *Science*, 313(5793):1642–1645, 2006.

[2] N. Boyd, E. Jonas, H. Babcock, and B. Recht. DeepLoco: Fast 3D localization microscopy using neural networks. *bioRxiv*, https://doi.org/10.1101/267096, 2018.

[3] E.A.K. Cohen, A.V. Abraham, S. Ramakrishnan, and R.J. Ober. Resolution limit of image analysis algorithms. *Nature Communications*, 10:793, 2019.

[4] Graham T Dempsey, Joshua C Vaughan, Kok Hao Chen, Mark Bates, and Xiaowei Zhuang. Evaluation of fluorophores for optimal performance in localization-based super-resolution imaging. *Nature Methods*, 8(12):1027–1036, 2012.

[5] R. Eils M. Heilemann F. Fricke, J. Beaudouin. One, two or three? probing the stoichiometry of membrane proteins by single-molecule localization microscopy. *Scientific Reports*, 14072(5), 2015.

[6] T. Ha and P. Tinnefeld. Photophysics of Fluorescent Probes for Single-Molecule Biophysics and Super-Resolution Imaging. *Annual Review of Physical Chemistry*, 63(1): 595–617, 2012.

[7] M. Heilemann, S. Van de Linde, M. Schüttpelz, R. Kasper, B. Seefeldt, A. Mukherjee, P. Tinnefeld, and M. Sauer. Subdiffraction - Resolution Fluorescence Imaging with Conventional Fluorescent Probes. *Angewandte Chemie International Edition*, 47(33): 6172–6176, 2008.

[8] S. T. Hess, T. P. K. Girirajan, and M. D. Mason. Ultra-high resolution imaging by fluorescence photoactivation localization microscopy. *Biophysical journal*, 91(11): 4258–4272, 2006.

[9] R. J. Hyndman. Computing and graphing highest density regions. *The American Statistician*, 50(2):120–126, 1996.

bioRxiv preprint doi: https://doi.org/10.1101/834572; this version posted November 7, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

[10] S. H. Lee, J. Y. Shin, A. Lee, and C. Bustamante. Counting single photoactivatable fluorescent molecules by photoactivated localization microscopy (PALM). *Proceedings of the National Academy of Sciences of the United States of America*, 109(43):17436–17441, 2012.

[11] Y. Lin, J. J. Long, F. Huang, W. C. Duim, S. Kirschbaum, Y. Zhang, L. K. Schroeder, A. A. Rebane, M. G. M. Velasco, A. Virrueta, D. W. Moonan, J. Jiao, S. Y. Hernandez, Y. Zhang, and J. Bewersdorf. Quantifying and Optimizing Single-Molecule Switching Nanoscopy at High Speeds. *Plos One*, 10(5):e0128135, 2015.

[12] R. P. J. Nieuwenhuizen, M. Bates, A. Szymborska, K. A. Lidke, B. Rieger, and S. Stallinga. Quantitative Localization Microscopy: Effects of Photophysics and Labeling Stoichiometry. *PLoS ONE*, 10(5):e0127989, 2015.

[13] D. Nino, N. Rafiei, Y. Wang, A. Zilman, and J. N. Milstein. Molecular counting with localization microscopy: A bayesian estimate based on fluorophore statistics. *Biophysical Journal*, 112(9):1777–1785, 2017.

[14] R.J. Ober, A. Tahmasbi, S. Ram, Z. Lin, and E.S. Ward. Quantitative Aspects of Single-Molecule Microscopy: Information-theoretic analysis of single-molecule data. *IEEE Signal Processing Magazine*, 32(1):58–69, 2015.

[15] L. Patel, N. Gustafsson, Y. Lin, R. Ober, R. Henriques, and E. Cohen. A hidden Markov model approach to characterizing the photo-switching behavior of fluorophores. *Annals of Applied Statistics*, 13(1), 2019.

[16] G. C. Rollins, J. Y. Shin, C. Bustamante, and S. Pressé. Stochastic approach to the molecular counting problem in superresolution microscopy. *Proceedings of the National Academy of Sciences of the United States of America*, 112(2):110–118, 2014.

[17] M. J. Rust, M. Bates, and X. Zhuang. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nature methods*, pages 793–795, 2006.

[18] D. Sage, H. Kirshner, T. Pengo, N. Stuurman, J. Min, S. Manley, and M. Usher. Quan-

bioRxiv preprint doi: https://doi.org/10.1101/834572; this version posted November 7, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

titative evaluation of software packages for single-molecule localization microscopy. *Nature Methods*, 12(8):717–724, 2015.

[19] S. Van de Linde and M. Sauer. How to switch a fluorophore: from undesired blinking to controlled photoswitching. *Chemical Society reviews*, 43(4):1076–1087, 2014.