

# MODELLING AND INFERENCE OF SPATIO-TEMPORAL PROCESSES IN SINGLE MOLECULE LOCALISATION MICROSCOPY

A THESIS PRESENTED FOR THE DEGREE OF  
DOCTOR OF PHILOSOPHY OF IMPERIAL COLLEGE LONDON

AND THE

DIPLOMA OF IMPERIAL COLLEGE

BY

LEKHA PATEL

DEPARTMENT OF MATHEMATICS

IMPERIAL COLLEGE

180 QUEEN'S GATE, LONDON SW7 2BZ

NOVEMBER 2019

I certify that this thesis, and the research to which it refers, are the product of my own work, and that any ideas or quotations from the work of other people, published or otherwise, are fully acknowledged in accordance with the standard referencing practices of the discipline.

Signed: \_\_\_\_\_

# COPYRIGHT

The copyright of this thesis rests with the author and is made available under a Creative Commons Attribution Non-Commercial No Derivatives licence. Researchers are free to copy, distribute or transmit the thesis on the condition that they attribute it, that they do not use it for commercial purposes and that they do not alter, transform or build upon it. For any reuse or redistribution, researchers must make clear to others the licence terms of this work.

## Modelling and Inference of Spatio-Temporal Processes in Single Molecule Localisation Microscopy

### ABSTRACT

Recent advancements in super-resolution microscopy have enabled cellular structures to be imaged beyond sub-diffraction limits. In order to do so, a widely used class of super resolution methods called single molecule localisation microscopy (SMLM) exploit the stochastic nature of fluorescent probes, or fluorophores, that move between bright and dark states until they permanently cease to transition. When observing a large number of fluorophores, this behaviour enables only a sparse subset of them to be detected at any one time, resulting in the ability to accurately record and accumulate their spatial measurements to produce a super-resolved image. While this stochastic behaviour has been heavily exploited, it induces multiple localisations per molecule which gives rise to misleading representations of the true structures of interest. Accurate quantification of the underlying photo-kinetic behaviour is therefore required before any spatial analysis can be conducted.

In this thesis, we model the photo-kinetic behaviour of a molecule as a continuous time Markov process that can transition between a photo-emitting On state, several (unknown) non-photon emitting dark states and an permanently dark state. From this, we develop the Photo-Switching Hidden Markov Model (PSHMM) which relates this underlying behaviour to an observed signal indicating whether or not a molecule is detected in a given frame. Under this model, we derive a maximum likelihood estimator which is used to estimate the unknown transition rates and photo-kinetic model. Under different experimental conditions, the statistical properties of this estimator are also investigated. When an unknown number of fluorescing molecules is filmed, the PSHMM set-up subsequently allows us to derive the distribution of the

Thesis advisor: Dr. Edward Cohen

Lekha Patel

total number of observed localisations in an experiment, from which an accurate molecular counting tool can be constructed. Finally, we formulate true molecular positions as a spatio-temporal hidden point process and describe the observation process it generates at each time step. The full Bayes filter is then derived, from which the point process and static parameters of the model can be inferred using Markov Chain Monte Carlo (MCMC).

*To my family.*

# ACKNOWLEDGMENTS

Firstly, I would like to thank my supervisor Ed for giving me this opportunity, supporting me through this journey and for always believing in me. To all of the statistics section, it has been an absolute pleasure being in everyone's company. In particular, I would like thank Andy, Nicola, Nick Heard, Nik Kantas, Niall, Dinand Heather for their overwhelming guidance and many useful discussions. To all of my friends I have met during the PhD, thank you for giving me some of best few years at Imperial. Special thanks also goes to all of the amazing people I had the pleasure of meeting in Texas A & M, I will never forget my few months there. I will miss you all.

Lastly, none of this work would have been made possible without the continuing love and support I have received from my family and Golo, especially over this past year. Thank you all for being there.

# LIST OF FIGURES

1	(Patel et al., 2019) (a) Illustration of the SMLM imaging process. When all fluorophores simultaneously stay in a photon emitting On state, diffraction renders structures unresolvable. Stochastically photo-switching fluorophores imaged over time across several frames give rise to a sequence of sparsely populated images, where each fluorophore can be isolated and localised with high precision. Aggregating these frames gives rise to a super-resolved image. Data from Sage et al. (2015). (b) Isolated fluorophores are localised by fitting the point spread function (PSF) to the diffraction limited spot. . . . .	2
1.1	Common models used to describe the continuous time photo-switching dynamics of a fluorophore with homogeneous transition rates. . . . .	8
1.2	(a) A simulated intensity signal of a fluorophore across time. Each measurement corresponds to the intensity in a frame. 7500 frames were recorded over 250 seconds at a rate of 30 frames per second. (b) Close up of the signal over the time window of 35s to 55s. In red is the observed signal $\{Y_n\}$ indicating if the fluorophore was detected in a particular frame. (c) A further close up of the signal showing intensity read-outs for independent frames. The true, hidden photon emitting On state of the molecule is also indicated, demonstrating how sub-frame length photon emitting events (such as the event at 38.37s) can be missed due to noise or the temporal resolution of the data acquisition. . . . .	11
1.3	General $d + 3$ state ( $d \in \mathbb{Z}_{\geq 0}$ ) model of a fluorophore. . . . .	18
1.4	Illustration of how the states for $Y_n$ derive from the process $X(t)$ . . . . .	19



1.5	Illustration of the HMM set-up. (a): Traditional HMM where the observed state is dependent on the current hidden. (b): Our HMM where the observed state depends on both the current and past hidden states. . . . .	21
1.6	Figures 1.6a and 1.6b consider two possible paths of $\{X(t)\}$ when $j \neq 1$ and show that all individual time pieces in state 1 (blue) are distinctly exponentially distributed. Figures 1.6c and 1.6d consider two paths of $\{X(t)\}$ when $j = 1$ and show that all but the last individual time pieces in state 1 (blue) are exponential. The final time piece suffers from right-censoring.	27
1.7	Highlighted holding times in each dark state: 0 (red), $0_1$ (orange), $0_2$ (brown) and $0_3$ (purple) from a possible path when $X(\Delta) = 1$ and $d = 3$ . All individual time pieces in each dark state are distinctly exponentially distributed. Subtracting the sum of all these (highlighted) time pieces from $\Delta$ gives the total time spent in state 1. . . . .	29
1.8	Estimates of $\log_{10}(\lambda_{01})$ and $\log_{10}(\lambda_{10})$ simulated from model $\mathcal{M}_{\{1\}}^0$ using both exponential fitting (1.8a) and PSHMM fitting (1.8b) are plotted in dark yellow and pink, respectively. True rates are plotted as black crosses. Measurement unit is per second. Estimates for the photo-bleaching rate $\mu_1$ , along with means, RMSEs and 2.5 and 97.5 empirical percentiles are given in Table 1.3 (see Appendix 1.G). . . . .	47
1.9	Top Left: Examples of single simulated frames at the indicated number of photons per frame (Appendix 1). Box-plots showing quantiles from estimates of $\lambda_{01}$ , $\lambda_{10}$ and $\mu_1$ from both exponential fitting (black) and PSHMM fitting (red) are plotted against increasing photons per frame. $N_F = 9872$ for all simulations. True rates given by the blue line. . . . .	48
1.10	Rate predictions and associated 95% bootstrap confidence sets are shown for $\lambda_{01}$ , $\lambda_{10}$ , $\mu_1$ , $\lambda_{0_1 1}$ , $\lambda_{0_1 0_2}$ , $\lambda_{0_1 1}$ and $\lambda_{0_2 1}$ , for 8 different values of laser intensity (see Table 1.2). Intervals in blue correspond to those from exponential fitting and those in red correspond to those gained from the PSHMM. Point estimates from each of the 27 datasets are given by diamond (PSHMM) or square (exponential). . . . .	52

1.11	Boxplots showing quantiles from estimates of $\lambda_{01}$ , $\lambda_{10}$ and $\mu_1$ from both exponential fitting (black) and PSHMM fitting (red) are plotted against $\log(\Delta)$ . True rates given by the blue line. . . . .	76
1.12	Boxplots showing quantiles from estimates of $\lambda_{01}$ , $\lambda_{10}$ and $\mu_1$ from both exponential fitting (black) and PSHMM fitting (red) are plotted against $\log(\text{threshold})$ . $N_F = 9872$ for all simulations. True rates given by the blue line. . . . .	77
1.13	Boxplots showing quantiles from estimates of $\lambda_{01}$ , $\lambda_{10}$ and $\mu_1$ from both exponential fitting (black) and PSHMM fitting (red) are plotted against $N_F$ . True rates given by the blue line. . . . .	77
1.14	Boxplots showing quantiles from estimates of $\lambda_{01}$ , $\lambda_{10}$ and varying $\mu_1$ from both exponential fitting (black) and PSHMM fitting (red) are plotted against $\log_{10}(\mu_1)$ . True rates given by the blue lines. . . . .	77
2.1	Contour plots showing the log-likelihood surface of $\lambda_{01}$ against $\delta$ when $d = 1$ under study 4 (of Table 2.1) with other rate parameters $\lambda_{00_1}$ , $\lambda_{10}$ and $\mu_1$ fixed at their maximum likelihood values. The left most plot shows the surface when $\frac{\delta}{\Delta} = 0.1$ , the middle shows the surface when $\frac{\delta}{\Delta} = 0.4$ and the rightmost when $\frac{\delta}{\Delta} = 0.7$ . Maximum likelihood estimates are shown as asterisks (black) with the true values shown as squares (red). The process $\{X(t)\}$ generating the data is the same for all three plots. . . . .	87
2.2	Contour plots showing the log-likelihood surface of $\lambda_{01}$ against $\delta$ when $d = 1$ under study 6 (of Table 2.1) with other rate parameters $\lambda_{00_1}$ , $\lambda_{10}$ and $\mu_1$ fixed at their maximum likelihood values. The left most plot shows the surface when $\frac{\delta}{\Delta} = 0.1$ , the middle shows the surface when $\frac{\delta}{\Delta} = 0.4$ and the rightmost when $\frac{\delta}{\Delta} = 0.7$ . Maximum likelihood estimates are shown as asterisks (black) with the true values shown as squares (red). The process $\{X(t)\}$ generating the data is the same for all three plots. . . . .	87
2.3	Rate estimates for $\lambda_{0_10_2} = 0.05$ under the slow switching scenario when $d = 2$ , estimates show multi-modality when $N_F = 10^4$ (left) and uni-modality when $N_F = 5 \times 10^4$ . . . . .	88

2.4	Box-plots showing rate estimates when $d = 0$ under study 3 of Table 2.1 when $\Delta^{-1} = 30$ (left), $\Delta^{-1} = 65$ (middle) and $\Delta^{-1} = 100$ (right). True rates indicated by the red dashed line.	90
2.5	Box-plots showing rate estimates when $d = 1$ under study 4 of Table 2.1 when $\Delta^{-1} = 30$ (left), $\Delta^{-1} = 65$ (middle) and $\Delta^{-1} = 100$ (right). True rates indicated by the red dashed line.	90
2.6	Box-plots showing rate estimates when $d = 1$ under study 4 of Table 2.1 when $\Delta^{-1} = 30$ (left), $\Delta^{-1} = 65$ (middle) and $\Delta^{-1} = 100$ (right). True rates indicated by the red dashed line.	91
2.7	Box-plots showing rate estimates when $d = 2$ under study 9 of Table 2.1 when $\Delta^{-1} = 30$ (left), $\Delta^{-1} = 65$ (middle) and $\Delta^{-1} = 100$ (right). True rates indicated by the red dashed line.	91
2.8	Box-plots showing rate estimates when $d = 2$ under study 9 of Table 2.1 when $\Delta^{-1} = 30$ (left), $\Delta^{-1} = 65$ (middle) and $\Delta^{-1} = 100$ (right). True rates indicated by the red dashed line.	92
2.9	Contour plots showing the log-likelihood surface of $\lambda_{01}$ against $\delta/\Delta$ when $d = 1$ under study 6 (of Table 2.1) with other rate parameters $\lambda_{00_1}$ , $\lambda_{10}$ and $\mu_1$ fixed at their maximum likelihood values. The left most plot shows the surface when $\Delta^{-1} = 30\text{s}^{-1}$ , the middle shows the surface when $\Delta^{-1} = 65\text{s}^{-1}$ and the rightmost when $\Delta^{-1} = 100\text{s}^{-1}$ . The true value for $\delta/\Delta$ is 0.7 for each study. Maximum likelihood estimates are shown as asterisks (black) with the true values shown as squares (red). The process $\{X(t)\}$ generating the data is the same for all three plots.	93
2.10	Box-plots showing rate estimates when $d = 0$ under study 3 of Table 2.1 when $M = 100$ (left), $M = 1000$ (middle) and $M = 5000$ (right). True rates indicated by the red dashed line.	95
2.11	Box-plots showing rate estimates when $d = 1$ under study 6 of Table 2.1 when $M = 100$ , $M = 1000$ and $M = 5000$ (right). True rates indicated by the red dashed line.	95
2.12	Box-plots showing rate estimates when $d = 1$ under study 6 of Table 2.1 when $M = 100$ , $M = 1000$ and $M = 5000$ (right). True rates indicated by the red dashed line.	96

2.13	Box-plots showing rate estimates when $d = 2$ under study 9 of Table 2.1 when $M = 100$ , $M = 1000$ and $M = 5000$ (right). True rates indicated by the red dashed line. . . . .	96
2.14	Box-plots showing rate estimates when $d = 2$ under study 9 of Table 2.1 when $M = 100$ , $M = 1000$ and $M = 5000$ (right). True rates indicated by the red dashed line. . . . .	97
3.1	Figure 3.1a shows the theoretical and histogram estimate (from $10^6$ simulations) of $p_{\theta_{\omega}^{(d)}}(S_{N_F} = s)$ under 3 photo-switching models: $d = 0$ (blue), $d = 1$ (red) and $d = 2$ (green). In all simulations, $\mu_1 > 0, \mu_0 = \dots = \mu_{0_d} = 0$ , $N_F = 1000$ , $\nu_0 = \nu_1 = 0.5$ , $\Delta = \frac{1}{30}\text{s}$ , $\delta = 10^{-3}\text{s}$ and $\omega = 10^{-6}$ ; rates chosen (where appropriately zero) are $\lambda_{00_1} = 0.35\text{s}^{-1}$ , $\lambda_{01} = 1\text{s}^{-1}$ , $\lambda_{0_1 0_2} = 0.2\text{s}^{-1}$ , $\lambda_{0_1 1} = 0.3\text{s}^{-1}$ , $\lambda_{0_2 1} = 0.1\text{s}^{-1}$ , $\lambda_{10} = 2.3\text{s}^{-1}$ , $\mu_1 = 0.05\text{s}^{-1}$ . Figure 3.1b shows the theoretical and histogram estimate (from $10^6$ simulations) of $p_{\theta_{\omega}^{(d)}}(S_{N_F} = s)$ when $d = 1$ with $\mu_1 = 0.5\text{s}^{-1}$ (blue), $\mu_1 = 0.2\text{s}^{-1}$ (red) and $\mu_1 = 0.05\text{s}^{-1}$ (green). . . . .	123
3.2	Simulation results from studies 1-3 in Table 3.1. Histograms represent counts of $\hat{M}$ under the slow (Figure 3.2a), medium (Figure 3.2b) and fast (Figure 3.2c) scenarios when $d = 0$ , from $10^4$ independently generated datasets with $M = 100$ and $N_F = 10^4$ . For each estimate, $\theta_{\omega}^{(0)}$ was determined using a training data set with $M = 250$ and $N_F = 10^4$ . . . . .	131
3.3	Simulation results from studies 4-6 in Table 3.1. Histograms represent counts of $\hat{M}$ under the slow (Figure 3.3a), medium (Figure 3.3b) and fast (Figure 3.3c) switching scenarios when $d = 1$ , from $10^4$ independently generated datasets with $M = 100$ and $N_F = 10^4$ . For each estimate, $\theta_{\omega}^{(1)}$ was determined using a training data set with $M = 250$ and $N_F = 10^4$ . . . . .	131
3.4	Simulation results from studies 1-3 in Table 3.1. Histograms represent counts of $\hat{M}$ under the slow (Figure 3.4a), medium (Figure 3.4b) and fast (Figure 3.4c) switching scenarios when $d = 2$ , from $10^4$ independently generated datasets with $M = 100$ and $N_F = 10^4$ . For each estimate, $\theta_{\omega}^{(2)}$ was determined using a training data set with $M = 250$ and $N_F = 10^4$ . . . . .	132

3.5	(a) Posterior distributions of $M$ given $\hat{\theta}_\omega^{(2)}$ and $N_l$ for the Alexa Fluor 647 datasets 1 and 2 (descriptions of which can be found in Table 3.3). For each study, $\hat{M}$ is given by the corresponding posterior mode plotted in cyan, with the true values of $M$ shown in magenta (dotted). 95% credible intervals for each $\hat{M}$ are shown in black (dotted). (b) Posterior estimates of $M$ given $\hat{\theta}_\omega^{(2)}$ and $N_l$ for the 27 Alexa Fluor 647 datasets (descriptions of which can be found in Table 3.2) with varying laser intensities ( $\text{kw}/\text{cm}^2$ ). For each study, $\hat{M}$ is given by the corresponding posterior mode plotted in blue (circle), with the true values of $M$ shown in red (crosses) and 95% credible intervals for each $\hat{M}$ are shown by blue error bars. . . . .	134
3.6	Posterior distributions of $M$ given $\hat{\theta}_\omega^{(2)}$ and $N_l$ for the 27 Alexa Fluor 647 datasets (descriptions of which can be found in Table 3.3). For each study, $\hat{M}$ is given by the corresponding posterior mode plotted in cyan, with the true values of $M$ shown in magenta (dotted). 95% credible intervals for each $\hat{M}$ are shown in black (dotted). . . . .	138
4.1	Figure 4.1a shows the true spatial locations of a set of fluorophores from one fluorescence microscopy experiment. Figure 4.1b shows the aggregated measurements of localised fluorophores detected from the first 3 frames; objects detected from frame 1 (blue), frame 2 (red) and frame 3 (cyan) are shown with respect to the true locations (black). . . . .	142
4.2	Superimposed image of the localisations (red) attained from seven (spatially close) fluorescing molecules with true positions indicated in blue (crosses). Measurement unit is nanometre ( $\text{nm}$ ) $\times 10^4$ . . . . .	142
4.3	An illustration of the RFS $C_n$ compared with the RFS $Z_n$ against time $n$ . Left: $C_1 = B_1 \cup B_0$ (with $B_0 = \emptyset$ ) is plotted in red at time $n = 1$ , $C_2 = C_1 \cup B_2$ is plotted with $B_2$ in cyan at time $n = 2$ and $C_3 = C_2 \cup B_3$ is plotted with $B_3$ in magenta at time $n = 2$ . Right: Observation sets $Z_1, Z_2, Z_3$ are shown, with false positive observations plotted in blue. . . . .	152

4.4	Plots showing how the probability of detection of a single fluorophore (left) and the the number of births in a single frame (right) changes over the course of the experiment. Blue shows the simulated (empirical) estimates and orange shows the true probabilities (calculated directly). Simulations carried out under model $M_{\{1\}}^1$ whereby $T^* = 0$ , $\nu_{0_1}^* = \nu_{0_1} = 1$ , $\Delta = \frac{1}{30}\text{s}$ , $\delta = 10^{-3}\text{s}$ and $\alpha = 0$ ; rates chosen are $\lambda_{00_1} = 0.35\text{s}^{-1}$ , $\lambda_{0_1} = 1\text{s}^{-1}$ , $\lambda_{0_11} = 0.3\text{s}^{-1}$ , $\lambda_{10} = 2.3\text{s}^{-1}$ , $\mu_1 = 0.05\text{s}^{-1}$ . . . . .	157
4.5	Posterior spatial analysis for study 1. 4.5a: Posterior intensity map of MCMC samples. 4.5b: Coordinate estimates from the Bayes filter (red crosses) plotted with the true positions (black stars) and the superimposition of offspring observation sets and false positive observations (red dots). Unit of measurement is $\text{nm} \times 10^4$ . . . . .	174
4.6	Posterior spatial analysis for study 2. 4.6a: Posterior intensity map of MCMC samples. 4.6b: Coordinate estimates from the Bayes filter (red crosses) plotted with the true positions (black stars) and the superimposition of offspring observation sets and false positive observations (red dots). Unit of measurement is $\text{nm} \times 10^4$ . . . . .	176
4.7	Posterior spatial analysis for study 3. 4.7a: Posterior intensity map of MCMC samples. 4.7b: Coordinate estimates from the Bayes filter (red crosses) plotted with the true positions (black stars) and the superimposition of offspring observation sets and false positive observations (red dots). Unit of measurement is $\text{nm} \times 10^4$ . . . . .	176
4.8	Trace plots of the transformed photo-switching rates $\log \lambda_{0_1}$ , $\log \lambda_{10}$ under study 1 with true values in green. . . . .	177
4.9	Trace plots of the transformed parameters $\log \mu_1$ , $\log \alpha$ , $-\log(\Delta/\delta - 1)$ under study 1 with true values in green. . . . .	177
4.10	Posterior histograms of transformed photo-switching rates $\log \lambda_{0_1}$ , $\log \lambda_{10}$ under study 1, with MAP values (green), true values (red) and 95% credible intervals (black, dotted). . . . .	178
4.11	Posterior histograms of the transformed parameters $\log \mu_1$ , $\log \alpha$ , $-\log(\Delta/\delta - 1)$ under study 1, with MAP values (green), true values (red) and 95% credible intervals (black, dotted). . . . .	178

4.12	Posterior histograms of the transformed photo-switching rates $\log \lambda_{01}$ , $\log \lambda_{10}$ under study 2, with MAP values (green), true values (red) and 95% credible intervals (black, dotted). . . . .	179
4.13	Posterior histograms of the transformed photo-switching rates $\log \lambda_{00_1}$ , $\log \lambda_{0_11}$ under study 2, with MAP values (green), true values (red) and 95% credible intervals (black, dotted). . . . .	179
4.14	Posterior histograms of the transformed parameters $\log \mu_1$ , $\log \alpha$ , $-\log(\Delta/\delta - 1)$ under study 2, with MAP values (green), true values (red) and 95% credible intervals (black, dotted). . . . .	180
4.15	Posterior histograms of the transformed photo-switching rates $\log \lambda_{01}$ , $\log \lambda_{10}$ under study 3, with MAP values (green), true values (red) and 95% credible intervals (black, dotted). . . . .	180
4.16	Posterior histograms of the transformed photo-switching rates $\log \lambda_{00_1}$ , $\log \lambda_{0_11}$ under study 3, with MAP values (green), true values (red) and 95% credible intervals (black, dotted). . . . .	181
4.17	Posterior histograms of the transformed photo-switching rates $\log \lambda_{0_10_2}$ , $\log \lambda_{0_21}$ under study 3, with MAP values (green), true values (red) and 95% credible intervals (black, dotted). . . . .	181
4.18	Posterior histograms of the transformed parameters $\log \mu_1$ , $\log \alpha$ , $-\log(\Delta/\delta - 1)$ under study 3, with MAP values (green), true values (red) and 95% credible intervals (black, dotted). . . . .	182
4.19	Superimposed localisations obtained from a dSTORM experiment observing chromosomes labelled with Alexa Fluor 488 fluorophores. Blue squares correspond to datasets 1 and 2 analysed. . . . .	183
4.20	Posterior spatial analysis for dataset 1 (Alexa Fluor 488). 4.20a: Posterior intensity map of MCMC samples. 4.20b: Coordinate estimates from the Bayes filter (blue crosses) plotted with the superimposed dataset. Measurement unit is nanometres. . . . .	184
4.21	Posterior spatial analysis for dataset 2 (Alexa Fluor 488). 4.21a: Posterior intensity map of MCMC samples. 4.21b: Coordinate estimates from the Bayes filter (blue crosses) plotted with the superimposed dataset. Measurement unit is nanometres. . . . .	184

4.22	Posterior histograms of transformed photo-switching rates $\log \lambda_{00_1}$ for the Alexa Fluor 488 datasets 1 (Figure 4.22a) and 2 (Figure 4.22b), plotted with MAP values (green) and 95% credible intervals (black, dotted). . . . .	193
4.23	Posterior histograms of transformed photo-switching rates $\log \lambda_{0_1}$ for the Alexa Fluor 488 datasets 1 (Figure 4.23a) and 2 (Figure 4.23b), plotted with MAP values (green) and 95% credible intervals (black, dotted). . . . .	193
4.24	Posterior histograms of transformed photo-switching rates $\log \lambda_{0_1 0_2}$ for the Alexa Fluor 488 datasets 1 (Figure 4.24a) and 2 (Figure 4.24b), plotted with MAP values (green) and 95% credible intervals (black, dotted). . . . .	193
4.25	Posterior histograms of transformed photo-switching rates $\log \lambda_{0_1 1}$ for the Alexa Fluor 488 datasets 1 (Figure 4.25a) and 2 (Figure 4.25b), plotted with MAP values (green) and 95% credible intervals (black, dotted). . . . .	194
4.26	Posterior histograms of transformed photo-switching rates $\log \lambda_{0_2 1}$ for the Alexa Fluor 488 datasets 1 (Figure 4.26a) and 2 (Figure 4.26b), plotted with MAP values (green) and 95% credible intervals (black, dotted). . . . .	194
4.27	Posterior histograms of transformed photo-switching rates $\log \lambda_{10}$ for the Alexa Fluor 488 datasets 1 (Figure 4.27a) and 2 (Figure 4.27b), plotted with MAP values (green) and 95% credible intervals (black, dotted). . . . .	194
4.28	Posterior histograms of transformed photo-switching rates $\log \mu_1$ for the Alexa Fluor 488 datasets 1 (Figure 4.28a) and 2 (Figure 4.28b), plotted with MAP values (green) and 95% credible intervals (black, dotted). . . . .	195
4.29	Posterior histograms of transformed photo-switching rates $-\log(\Delta/\delta-1)$ for the Alexa Fluor 488 datasets 1 (Figure 4.29a) and 2 (Figure 4.29b), plotted with MAP values (green) and 95% credible intervals (black, dotted). . . . .	195
4.30	Posterior histograms of transformed photo-switching rates $\log \alpha$ for the Alexa Fluor 488 datasets 1 (Figure 4.30a) and 2 (Figure 4.30b), plotted with MAP values (green) and 95% credible intervals (black, dotted). . . . .	195



A.1	Copyright agreement for Patel et al. (2019).	202
A.2	Copyright agreement for Patel and Cohen (2018).	203

# LIST OF TABLES

1.1	Confusion table showing the empirical percentage of models predicted from three candidates: $\mathcal{M}_{\{1\}}^0$ , $\mathcal{M}_{\{1\}}^1$ and $\mathcal{M}_{\{1\}}^2$ under simulation studies 16, 19 and 20 (see Tables 1.3, 1.4 and 1.5 in Appendix 1.G), with $M = 300$ , $\delta = \frac{1}{100}s$ and $\Delta = \frac{1}{50}s$ . 100 datasets from each study were generated and the BIC was used to select the best fitted model. . . . .	49
1.2	A description of the Alexa Fluor 647 datasets with reference to the laser intensities in kW/cm <sup>2</sup> and frames sampled per second (or $\Delta^{-1}$ ) measured in s <sup>-1</sup> used to characterise each of the 27 cell experiments. The $N_F \times M$ size of each cell's dataset is also included. . . . .	50
1.3	Caption next page. . . . .	74
1.4	Simulation results showing mean, bias, root mean squared error (RMSE) and the 2.5 and 97.5 empirical percentiles of the estimates of $\boldsymbol{\theta}^{(1)} = (\lambda_{00_1} \ \lambda_{01} \ \lambda_{0_1 0_2} \ \lambda_{0_1 1} \ \lambda_{10} \ \mu_1)^\top$ under model $\mathcal{M}_{\{1\}}^1$ for both the PSHMM and exponential fitting (Exp) methods across 100 repeat experiments. $\Delta = \frac{1}{30}s$ , $\delta = 0.01s$ , $\omega > 0$ (unknown), and $M = 100$ . . . . .	75
1.5	Simulation results showing mean, bias, root mean squared error (RMSE) and the 2.5 and 97.5 empirical percentiles of the estimates of $\boldsymbol{\theta}^{(2)} = (\lambda_{00_1} \ \lambda_{01} \ \lambda_{0_1 0_2} \ \lambda_{0_1 1} \ \lambda_{0_2 1} \ \lambda_{10} \ \mu_1)^\top$ under model $\mathcal{M}_{\{1\}}^2$ for both the PSHMM and exponential fitting (Exp) methods across 100 repeat experiments. $\Delta = \frac{1}{30}s$ , $\delta = 0.01s$ , $\omega > 0$ (unknown), and $M = 100$ . . . . .	75
2.1	Global parameter values for the stimulation studies conducted in this section. All studies have been conducted with the model format of $M_{\{1\}}^d$ for $d = 0, 1, 2$ , and with $T^* = 0$ . . . . .	83

2.2	Simulation results from $d = 0$ slow switching (study 1 of Table 2.1): Mean approximate correlations obtained from the Hessian matrices compared with the values calculated directly from 500 fits. . . . .	98
2.3	Simulation results from $d = 0$ medium switching (study 2 of Table 2.1): Mean approximate correlations obtained from the Hessian matrices compared with the values calculated directly from 500 fits. . . . .	99
2.4	Simulation results from $d = 0$ fast switching (study 3 of Table 2.1): Mean approximate correlations obtained from the Hessian matrices compared with the values calculated directly from 500 fits. . . . .	99
2.5	Simulation results from $d = 1$ slow switching (study 4 of Table 2.1): Mean approximate correlations obtained from the Hessian matrices compared with the values calculated directly from 500 fits. . . . .	100
2.6	Simulation results from $d = 1$ medium switching (study 5 of Table 2.1): Mean approximate correlations obtained from the Hessian matrices compared with the values calculated directly from 500 fits. . . . .	101
2.7	Simulation results from $d = 1$ fast switching (study 6 of Table 2.1): Mean approximate correlations obtained from the Hessian matrices compared with the values calculated directly from 500 fits. . . . .	102
2.8	Simulation results from $d = 2$ slow switching (study 7 of Table 2.1): Mean approximate correlations obtained from the Hessian matrices compared with the values calculated directly from 500 fits. . . . .	103
2.9	Simulation results from $d = 2$ slow switching (study 7 of Table 2.1) continued: Mean approximate correlations obtained from the Hessian matrices compared with the values calculated directly from 500 fits. . . . .	104
2.10	Simulation results from $d = 2$ medium switching (study 8 of Table 2.1): Mean approximate correlations obtained from the Hessian matrices compared with the values calculated directly from 500 fits. . . . .	105

2.11	Simulation results from $d = 2$ medium switching (study 8 of Table 2.1) continued: Mean approximate correlations obtained from the Hessian matrices compared with the values calculated directly from 500 fits. . . . .	106
2.12	Simulation results from $d = 2$ fast switching (study 9 of Table 2.1): Mean approximate correlations obtained from the Hessian matrices compared with the values calculated directly from 500 fits. . . . .	107
2.13	Simulation results from $d = 2$ fast switching (study 9 of Table 2.1) continued: Mean approximate correlations obtained from the Hessian matrices compared with the values calculated directly from 500 fits. . . . .	108
2.14	$d = 0$ slow switching (study 1 of Table 2.1): Means and Mean Squared Errors (MSE) from datasets with $M = 100$ (100) and $M = 1000$ (1000) are shown for all parameter estimates. . . .	109
2.15	$d = 0$ medium switching (study 2 of Table 2.1): Means and Mean Squared Errors (MSE) from datasets with $M = 100$ (100) and $M = 1000$ (1000) are shown for all parameter estimates. . . . .	109
2.16	$d = 0$ fast switching (study 3 of Table 2.1): Means and Mean Squared Errors (MSE) from datasets with $M = 100$ (100) and $M = 1000$ (1000) are shown for all parameter estimates. . . .	110
2.17	$d = 1$ slow switching (study 4 of Table 2.1): Means and Mean Squared Errors (MSE) from datasets with $M = 100$ (100) and $M = 1000$ (1000) are shown for all parameter estimates. . . .	110
2.18	$d = 1$ medium switching (study 5 of Table 2.1): Means and Mean Squared Errors (MSE) from datasets with $M = 100$ (100) and $M = 1000$ (1000) are shown for all parameter estimates. . . . .	111
2.19	$d = 1$ fast switching (study 6 of Table 2.1): Means and Mean Squared Errors (MSE) from datasets with $M = 100$ (100) and $M = 1000$ (1000) are shown for all parameter estimates. . . .	111
2.20	$d = 2$ slow switching (study 7 of Table 2.1): Means and Mean Squared Errors (MSE) from datasets with $M = 100$ (100) and $M = 1000$ (1000) are shown for all parameter estimates. . . .	112

2.21	$d = 2$ medium switching (study 8 of Table 2.1): Means and Mean Squared Errors (MSE) from datasets with $M = 100$ (100) and $M = 1000$ (1000) are shown for all parameter estimates. . . . .	113
2.22	$d = 2$ fast switching (study 9 of Table 2.1) of Table 2.1): Means and Mean Squared Errors (MSE) from datasets with $M = 100$ (100) and $M = 1000$ (1000) are shown for all parameter estimates. . . . .	114
3.1	Global parameter values for the stimulation studies conducted in this section. . . . .	130
3.2	A description of the Alexa Fluor 647 datasets, with reference to the laser intensities in $\text{kW}/\text{cm}^2$ and frames $\Delta^{-1}$ sampled per second used to characterise each of the 27 experiments. For each dataset, a training set of size $M_{tr}$ was used to find the maximum likelihood estimate $\hat{\theta}_\omega^{(2)}$ . A hold out test set of size $M$ was used validate the inference procedure. . . . .	133
3.3	A description of the Alexa Fluor 647 datasets, with reference to the laser intensities in $\text{kW}/\text{cm}^2$ and frames $\Delta^{-1}$ sampled per second used to characterise 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 $\hat{\theta}_\omega^{(2)}$ via the PSHMM (estimated values shown). A hold out test set of size $N_F \times M$ (test) was used in the posterior computations of $M$ . . . . .	139
4.1	Global parameter values for the stimulation studies conducted in this section. All studies have been conducted with the model format of $M_{\{1\}}^d$ for $d = 0, 1, 2$ . . . . .	172
4.2	Parameter estimates and 95% credible intervals under study 1.	174
4.3	Parameter estimates and 95% credible intervals under study 2.	175
4.4	Parameter estimates and 95% credible intervals under study 3.	175
4.5	Parameter estimates and 95% credible intervals for the Alexa Fluor 488 datasets 1 and 2. . . . .	183

4.6	Table showing the computed DIC values ( $\times 10^5$ ) under simulation study 1, from three candidates $\mathcal{M}_{\{1\}}^0$ , $\mathcal{M}_{\{1\}}^1$ and $\mathcal{M}_{\{1\}}^2$ with varying values of $M$ . The predicted model is indicated in red. . . . .	190
4.7	Table showing the computed DIC values ( $\times 10^5$ ) under simulation study 2, from three candidates $\mathcal{M}_{\{1\}}^0$ , $\mathcal{M}_{\{1\}}^1$ and $\mathcal{M}_{\{1\}}^2$ with varying values of $M$ . The predicted model is indicated in red. . . . .	190
4.8	Table showing the computed DIC values ( $\times 10^5$ ) under simulation study 3, from three candidates $\mathcal{M}_{\{1\}}^0$ , $\mathcal{M}_{\{1\}}^1$ and $\mathcal{M}_{\{1\}}^2$ with varying values of $M$ . The predicted model is indicated in red. . . . .	191

# CONTENTS

0	INTRODUCTION	1
0.1	Motivation . . . . .	1
0.2	Thesis overview . . . . .	4
0.3	Publications . . . . .	6
1	PHOTO-SWITCHING MODELS	7
1.1	Introduction . . . . .	7
1.1.1	Chapter summary . . . . .	12
1.2	The Photo-Switching Hidden Markov Model . . . . .	14
1.2.1	Continuous time signal . . . . .	15
1.2.2	Discrete time observations . . . . .	18
1.2.3	Characterising photo-switching behaviour . . . . .	20
1.3	Derivation of transmission probabilities . . . . .	23
1.3.1	Overview . . . . .	24
1.3.2	Case $j \neq 1$ . . . . .	29
1.3.3	Case $j = 1$ . . . . .	34
1.3.4	Transmission probabilities . . . . .	38
1.4	Inference . . . . .	39
1.4.1	Likelihood function . . . . .	39
1.4.2	Model selection . . . . .	43
1.4.3	Bootstrapping . . . . .	44
1.5	Simulations . . . . .	44
1.5.1	Estimating rate parameters . . . . .	45

1.5.2	Model selection . . . . .	49
1.6	Application to Alexa Fluor 647 data . . . . .	49
1.7	Conclusions . . . . .	51
<b>APPENDIX 1</b>		<b>54</b>
Appendix 1.A	Proof of Theorem 1 . . . . .	54
Appendix 1.B	Algorithm to compute transmission matrices . . . . .	56
Appendix 1.C	Exact solution of transmission probabilities when there is a single dark state . . . . .	59
Appendix 1.D	Discussion on implementing the PSHMM algorithm	63
1.D.1	Convergence of transmission matrices . . . . .	63
1.D.2	Likelihood optimisation . . . . .	63
Appendix 1.E	Exponential fitting estimator . . . . .	67
1.E.1	Original method . . . . .	67
1.E.2	Extension to handling the photo-bleached state . . . . .	70
Appendix 1.F	Image simulation methods . . . . .	71
1.F.1	Imaging simulation . . . . .	71
1.F.2	Image analysis and trace idealisation . . . . .	72
1.F.3	Global parameter set for simulations . . . . .	72
Appendix 1.G	Rate estimates . . . . .	74
Appendix 1.H	Further results . . . . .	76
<b>2</b>	<b>PROPERTIES OF THE PHOTO-SWITCHING HIDDEN MARKOV MODEL ESTIMATOR</b>	<b>78</b>
2.1	Introduction . . . . .	78
2.1.1	Chapter summary . . . . .	80
2.2	Identification of model parameters . . . . .	81
2.2.1	Local identifiability . . . . .	82
2.2.2	Correlations . . . . .	84



2.2.3	Multi-modality . . . . .	87
2.2.4	The effect of frame length . . . . .	88
2.2.5	Experimental impact . . . . .	91
2.3	Consistency . . . . .	92
2.3.1	Experimental impact . . . . .	95
2.4	Conclusions . . . . .	96
<b>APPENDIX 2</b>		<b>98</b>
	Appendix 2.A Correlation tables . . . . .	98
	Appendix 2.B Consistency tables . . . . .	109
<b>3</b>	<b>MOLECULAR COUNTING</b>	<b>115</b>
3.1	Introduction . . . . .	115
3.1.1	Chapter summary . . . . .	117
3.2	Cumulative localisations . . . . .	118
3.2.1	Probability mass function . . . . .	119
3.2.2	Probability generating function . . . . .	122
3.2.3	Extension to $M$ molecules . . . . .	125
3.3	Inference . . . . .	128
3.4	Simulations . . . . .	130
3.5	Application to Alexa Fluor 647 data . . . . .	131
3.6	Conclusions . . . . .	134
<b>APPENDIX 3</b>		<b>136</b>
	Appendix 3.A Proof of Lemma 5 . . . . .	136
	Appendix 3.B Alexa Fluor 647 results . . . . .	138
<b>4</b>	<b>SPATIO-TEMPORAL MODELLING</b>	<b>140</b>
4.1	Introduction . . . . .	140

4.1.1	Chapter summary . . . . .	144
4.2	Random finite sets . . . . .	144
4.2.1	Probability density function . . . . .	146
4.3	Spatio-temporal model of a fluorophore . . . . .	150
4.3.1	Probability of detection . . . . .	152
4.3.2	Number of births . . . . .	154
4.3.3	Derivation of the Bayes filter . . . . .	158
4.3.4	Limitations . . . . .	160
4.4	Inference . . . . .	161
4.4.1	Update moves . . . . .	165
4.4.2	Shift moves . . . . .	166
4.4.3	Parameter estimation . . . . .	169
4.4.4	Model selection . . . . .	169
4.5	Simulations . . . . .	171
4.6	Application to Alexa Fluor 488 data . . . . .	176
4.7	Conclusions . . . . .	183
<b>APPENDIX 4</b>		<b>186</b>
Appendix 4.A	Proof of Proposition 9 . . . . .	186
Appendix 4.B	Imaging simulation . . . . .	188
Appendix 4.C	DIC outputs . . . . .	190
Appendix 4.D	Alexa Fluor 488 results . . . . .	192
4.D.1	Experimental set-up . . . . .	192
4.D.2	Posterior histograms . . . . .	192
<b>5</b>	<b>CONCLUSIONS AND FUTURE WORK</b>	<b>196</b>
<b>APPENDIX A</b>		<b>201</b>
A.1	Copyright statements . . . . .	201



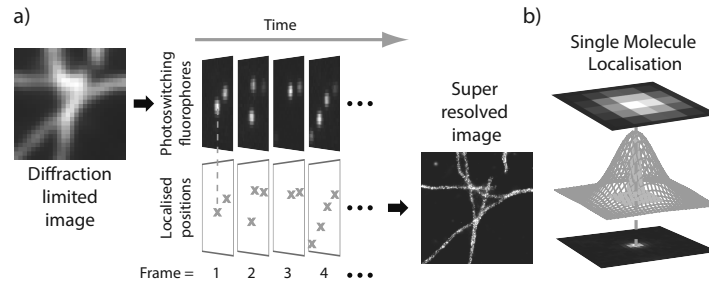


# INTRODUCTION

## 0.1 Motivation

Fluorescence microscopy is a collection of techniques that utilise the photon emitting properties of fluorescing molecules called *fluorophores*, to perform optical imaging, and has seen numerous applications in cell and medical biology. Recent years have seen the advent of a number of super-resolution microscopy techniques that have bypassed the classical resolution limits of fluorescence microscopy (Huang et al., 2009). Major contributions in this field include the successes of two main types of single molecule localisation microscopy (SMLM) approaches: photo-activated localisation microscopy (PALM) (Betzig et al., 2006, Hess et al., 2006) and stochastic optical reconstruction microscopy (STORM) (Rust et al., 2006, Heilemann et al., 2008). Each of these methods rely on the ability exhibited by some fluorophores to stochastically *photo-switch* between a photon emitting On state and non-photon emitting dark states (Van de Linde and Sauer, 2014, Ha and Tinnefeld, 2012). This results in the observation of fluorescent intermittency or *blinking* under a range of photo-kinetic scenarios.

Specifically, STORM (Rust et al., 2006, Heilemann et al., 2008) utilises con-



**Figure 1:** (Patel et al., 2019) (a) Illustration of the SMLM imaging process. When all fluorophores simultaneously stay in a photon emitting On state, diffraction renders structures unresolvable. Stochastically photo-switching fluorophores imaged over time across several frames give rise to a sequence of sparsely populated images, where each fluorophore can be isolated and localised with high precision. Aggregating these frames gives rise to a super-resolved image. Data from Sage et al. (2015). (b) Isolated fluorophores are localised by fitting the point spread function (PSF) to the diffraction limited spot.

ventional immuno-staining strategies to label proteins of interest with fluorescent chemicals; these are known as fluorophores. Molecules are placed in a *STORM buffer* which usually contains reducing and oxygen scavenging components, and are subject to an initial excitation from a high intensity laser. This excitation renders all fluorophores to initially be in the photon-emitting state, from which the STORM buffer subsequently allows transitions into a dark state, which is typically longer lived than the emissive state. In some experiments, fluorescent molecules can additionally be made to reach further dark states, for example, radical or very long triplet states, which are characterised by even longer lifetimes. The fluorophores then cycle between dark and On states until all are permanently unable to fluoresce. This is known as *photo-bleaching*. In the original implementation of STORM (Rust et al., 2006), proteins are labelled with a pair of activator-reporter fluorophores, whereby the activator is subject to the photo-switching and the reporter emits the detected signal. However, the requirement of dual labelling, and therefore the close proximity of the pairs of fluorophores, has subsequently motivated the development of direct STORM (dSTORM) (Heilemann et al., 2008). Since dSTORM experiments do not require the activator fluorophore, and can therefore be conducted using single labelling, this technique currently constitutes as the most widely used type of STORM imaging.

On the other hand, PALM (Betzig et al., 2006, Hess et al., 2006) uses a low

laser intensity to illuminate a *small* sample of sparse photo-activatable fluorophores that are initially in an inactive (dark) state. Once activated, each type of fluorophore is excited by a different wavelength of light, typically through higher intensity lasers, and thereby releases a *continuous* emission burst of photons. Immediately after this excitation, fluorophores are most likely to photo-bleach, but can also be converted back to a dark state. This cycle of photo-activation and photo-bleaching therefore limits the photo-switching lifetime of each fluorophore, and is typically much shorter than from those imaged in STORM. Furthermore, since it is during an emission burst that fluorophores can be detected, the success of PALM imaging lies both in the random activation of smaller sets of fluorophores, and the low probability of photo-reactivation.

Under normal imaging conditions, a specimen that is decorated with a spatially dense number of molecules prevents accurate identification of individual fluorophores and resolution of structures smaller than the diffraction limit, as is depicted in Figure 1(a). Nevertheless, using a fluorophore with stochastic photo-switching properties provide an imaging environment where the vast majority of molecules are in a dark state, leaving only a sparse number to be in the photon-emitting On state, resulting in the visible fluorophores being sparse and well separated in space. Subsequently, with the use of a high-performance camera, the individual fluorophores occupying the On state can be identified by first utilising segmentations algorithms (Olivo-Marin, 2002, Henriques et al., 2010, Ovesný et al., 2014) to locate high intensity photon regions of isolated fluorophores. Secondly, by fitting point spread functions (PSFs) (Sage et al., 2015, Ober et al., 2015) to these areas, these molecules can then be localised in continuous space with nanometre scale precision; this is depicted in Figure 1(b). Through the acquisition of a video containing thousands of images, shown in Figure 1(a), numerous fluorophores can be localised in space and isolated in time. When aggregated and plotted, these localisations provide a detailed map of fluorophore positions, giving rise to its super-resolved image.

The utilisation of SMLM methods does not come without drawbacks, however. Firstly, the successes of STORM and PALM imaging rely on the fact

that photon emission bursts from two or more fluorophores separated within a diffraction limit region, are not likely to overlap at any one time. When imaging tens of thousands of molecules, however, this occurrence becomes more likely (Cohen et al., 2019), enabling some of the obtained localisations to be poorly representative of the underlying molecular configuration. Nevertheless, the ever improving precision of localisations outputted by more advanced localisation algorithms, most recently from those relying on deep learning (Boyd et al., 2018), indicates the dissipation of such image resolution limitations. Secondly, the stochastic photo-switching behaviour induced by these methods leads to fluorophores blinking multiple times during an experiment, and therefore gives rise to multiple localisations of the same molecule in the super-resolved image. This has motivated the development of image reconstruction algorithms (Cox et al., 2011, Marsh et al., 2018) that account for these unwanted artefacts to yield more enhanced super-resolution images. While reconstructed images provide us with a more accurate representation of cellular structures, poor understanding of the underlying stochastic nature of the fluorophores renders problems such as *molecular counting*, and therefore the extraction of true molecular positions, difficult. Depending on the imaging technique used, this issue may further be heightened by the uncertainty of the *exact* photo-kinetic scheme driving the On-dark cycle a fluorophore undergoes before photo-bleaching. To address these questions, accurate quantification of the photo-switching behaviour of imaged fluorophores (Dempsey et al., 2011, Lehmann et al., 2016) across different SMLM schemes is vital, and still remains a very active area of research in this field.

## 0.2 Thesis overview

In this thesis, we provide three important contributions, that are based upon the statistical modelling of fluorophores imaged in STORM and dSTORM experiments. However, for future applicability to other microscopy techniques such as PALM, in this work we endeavour (where possible) to keep all mathematical formulations general.

In Chapter 1, we formulate the underlying and unobserved photo-kinetic

behaviour of a fluorophore as a continuous time Markov process, which transitions between its photon emitting On state, an unknown number of dark states and the photo-bleached state. We subsequently define the discrete time observed process, which, subject to the inherent noise limitations of the imaging procedure, indicates whether or not a fluorophore is detected in any given frame. From this, we formulate the photo-switching hidden Markov model (PSHMM) and develop novel methodology that is used to compute the log-likelihood of imaging observations. Maximum likelihood estimation is subsequently applied to recover photo-switching rates and predict (through a model selection criterion) the unknown number of dark states a molecule transitions between during an experiment. Both simulation studies and an application to a real dataset will be presented to demonstrate the validity of this method.

In Chapter 2, we thoroughly investigate the statistical properties of the PSHMM estimator over a wide range of different experimental set-ups. In particular, we address potential parameter identification issues that may arise in some experimental situations. We do so by analysing the observed fisher information matrices and correlation structures between parameters. The log-likelihood surfaces resulting from the PSHMM are also explored. We then investigate the large sample properties, with particular reference to consistency, of the PSHMM parameter estimates. Simulation studies form the basis of the analyses presented in this chapter.

In Chapter 3, we use the PSHMM to derive the *exact* distribution of the number of localisations a single fluorophore makes during an experiment, and observe that this distribution is parametrised by the unknown photo-switching rates. Here, we also derive its probability generating function (pgf) and first two moments. Using this, the distribution of the total number of localisations  $N_l$  made from an unknown number of  $M$  independent fluorescing molecules can be (computationally) recovered by an application of the fast Fourier transform (FFT). Subsequently, given any prior knowledge of  $M$ , the (posterior) distribution of  $M$  given  $N_l$  can be easily obtained. We show through both simulated and real data, that the mode of the resulting distribution can be used as an accurate molecular counting tool.



In Chapter 4, we describe a pure birth spatio-temporal point process that characterises fluorescing molecules that appear and blink in an experiment. Here, we construct a state-space model relating the underlying and unknown set of molecular positions with observation sets collected across time, from which it is observed that the birth and detection probabilities of individual fluorophores in any one frame, can be calculated as deterministic functions of the underlying photo-switching parameters. We subsequently derive the full Bayes filter for this model which is necessary for inference. From this, we utilise a suitable MCMC algorithm which is used to recover the unknown number of molecules, their true positions and the photo-kinetic model that best represents their intrinsic photo-switching properties. The utility of this method will be presented through simulation studies and an application to a real dataset.

In Chapter 5, we conclude by discussing directions for future work and the applicability of the material presented in this thesis to fluorescence microscopy experiments that first give rise to much larger datasets than those considered here and, second, are conducted via other microscopy techniques.

### 0.3 Publications

This thesis contains work that has been published or submitted for publication. Specifically, the material in Chapters 1 and 2 have been published as a paper entitled “A hidden Markov model approach to characterizing the photo-switching behavior of fluorophores” [Patel et al. \(2019\)](#). Permission has been given to use the article in this thesis, as described in the copyright statement in Section A.1 of Appendix A. The material in Chapter 3, cited as [Patel et al. \(2019\)](#), is available as a preprint at <https://www.biorxiv.org/content/10.1101/834572v1>, and has been submitted for publication. The material in Chapter 4 is based around the publication entitled “Bayesian filtering for spatial estimation of photo-switching fluorophores imaged in Super-resolution fluorescence microscopy” [Patel and Cohen \(2018\)](#). Permission has also been given to use the article in this thesis, as described in the copyright statement in Section A.1 of Appendix A.

# 1

## PHOTO-SWITCHING MODELS

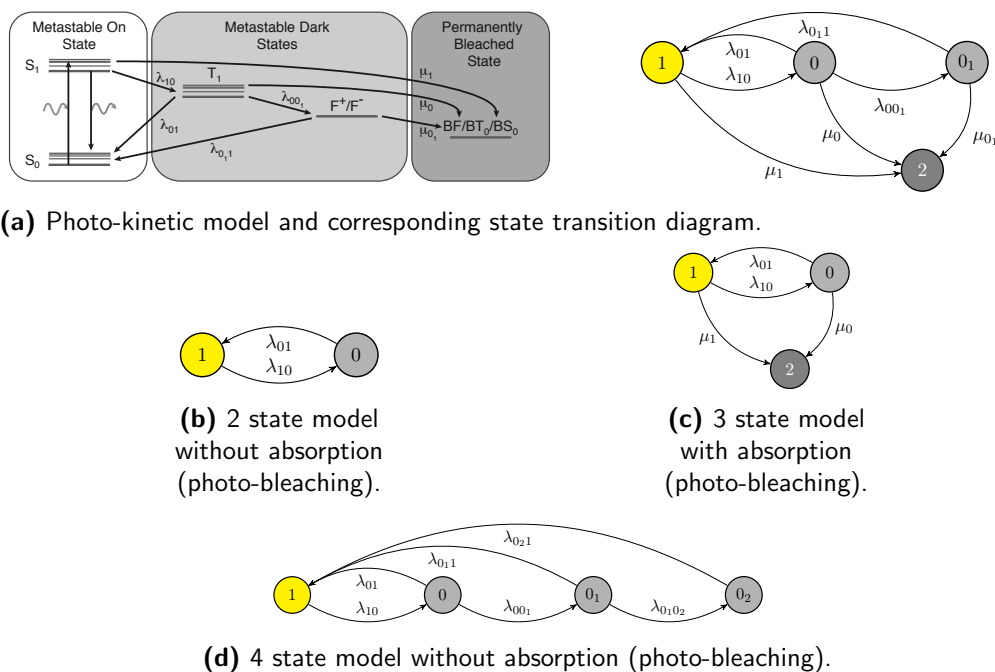
### 1.1 Introduction

Recent advances in Single Molecule Localisation Microscopy (SMLM) have enabled accurate resolvment of numerous biological samples' structures, many of which being imaged at lateral resolutions of between ten and thirty nanometres. The quality of the resolved images, however, is strongly dependent on the photo-switching properties of the fluorophore used. While longer On states provide a greater number of photons being recorded by the camera, which in turn leads to greater precision in localising spatially isolated fluorophores (Ober et al., 2004, Ram et al., 2012, Thompson et al., 2002, Rieger and Stallinga, 2014), the increased random occurrence of fluorophores simultaneously occupying the (photon emission) On state within a diffraction limited spot can lead to significant imprecision, missed events and unwanted artefacts (Van de Linde et al., 2010, Nieuwenhuizen et al., 2015). Thus, a careful choice of fluorophore and the environment used to promote photo-switching, which is controlled by the buffer solution and illumination intensity, must be made for the intended application. This is particularly important in live-cell applications, when considerations must be made for

temporal resolution and reduced laser intensities.

To inform the choice of fluorophore with its environment, and aid the development of novel fluorophores, accurate characterisation of the photo-kinetic model of the fluorophore, together with estimation of photo-switching rates (the rate at which fluorophores transition between On and dark states) is required (Dempsey et al., 2011, Lehmann et al., 2016). Further, accurate knowledge of the photo-switching characteristics could be employed to maximise resolutions achieved using advanced analytical methods (for example in 3B analysis (Cox et al., 2011) and DeconSTORM (Mukamel et al., 2012)), and improve the performance of molecular counting techniques (Rollins et al., 2014, Lee et al., 2012).

### Modelling photo-switching



**Figure 1.1:** Common models used to describe the continuous time photo-switching dynamics of a fluorophore with homogeneous transition rates.

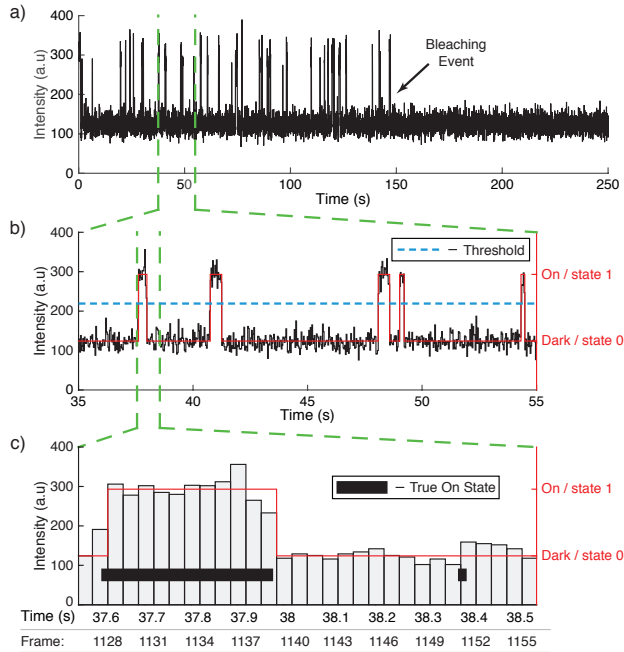
Several attempts have been made to model the kinetic schemes of fluorophore photo-switching and estimate the corresponding photo-switching

rates. These kinetic schemes, as is common across single molecule biophysics, are characterised by Markovian transitions between a finite set of discrete states and are therefore ideally suited to being modelled as continuous time Markov processes. In Figure 1.1 are four models for photo-switching fluorophores. The first, 1.1a, depicts a typical kinetic model, accompanied by the state-space diagram we will adopt in this chapter. This model contains a photon emitting On state 1 (involving rapid transitions between the excited state  $S_1$  and ground state  $S_0$  via the emission of a photon), two temporary dark states 0 and  $0_1$  (the triplet state,  $T_1$  and the redox states,  $F^+$  and  $F^-$ ) and an absorption state 2 (BF/BT<sub>0</sub>/BS<sub>0</sub>) which in this application is known as the *photo-bleached* state. Then in Figures 1.1b-1.1d are three further common state space models. Figure 1.1b portrays a photo-switching model with a simple two state {On(1) Dark(0)} structure. Models of this type are suitable for super-resolution methods including point accumulation for imaging in nano-scale topography (PAINT) and DNA-PAINT (Sharonov and Hochstrasser, 2006, Jungmann et al., 2010). Figure 1.1c depicts a model that incorporates an absorbing state 2. This form of photo-switching followed by photo-bleaching describes a first approximation to the behaviour that occurs spontaneously in a number of organic fluorophores and post-activation of photo-activatable proteins (Van de Linde and Sauer, 2014, Ha and Tinnefeld, 2012, Vogelsang et al., 2010). Figure 1.1d considers a model in which three distinct dark states are hypothesised, which in some cases is a necessary extension to the single dark state model (c), for instance when very rapid imaging is used (Lin et al., 2015).

The challenge comes in selecting the correct model and estimating the transition rates of the continuous time Markov process  $\{X(t) : t \in \mathbb{R}_{\geq 0}\}$  from an observed discrete-time random process  $\{Y_n : n \in \mathbb{Z}_{>0}\}$ . Here,  $\mathbb{R}_{\geq 0}$  and  $\mathbb{Z}_{>0}$  denote the non-negative real numbers and positive integers, respectively. Typically,  $\{Y_n\}$  is derived from a sequence of images (frames) with  $Y_n$  corresponding to the *observed* state of the molecule in the  $n$ th frame. This is formed by an exposure of the continuous time process  $\{X(t)\}$  over the time-interval  $[(n-1)\Delta, n\Delta)$ , where  $\Delta$  is the frame length. Process  $\{Y_n\}$  can either be a sequence of photon fluxes associated with that molecule for

each frame (Figure 1.2a), or a simple sequence of 1s and 0s indicating if the molecule was detected in the frame or not (Figure 1.2b). In all cases, the observations are subject to the effects of noise and instrument limitations. Essential to the subsequent analysis, therefore, is the ability to account for missed state transition events due to noise and the temporal resolution of the data acquisition, as well as the detection threshold used to determine the state of the system (Figure 1.2c). Similar problems occur in other areas of biophysics, where estimating transition rates of an underlying continuous time Markov process must be inferred from an observed discrete time signal. In particular, ion-channels have formed the focus of much work (Colquhoun and Hawkes, 1981, Qin and Li, 2004, Rief et al., 2000), including methods that attempt to account for missed events (Qin et al., 1996, Colquhoun et al., 1996, Hawkes et al., 1990, 1992, Epstein et al., 2016). However, the mechanism by which the observed signal is obtained and processed from the raw signal is fundamentally different to that of fluorescence microscopy imaging.

Up until now, methods for estimating photo-switching transition rates in fluorescence microscopy are limited. The method in Lin et al. (2015) involves defining  $\{Y_n\}$  to be the sequence of 1s and 0s and extracting the dwell times, namely the durations when  $Y_n$  is in the On state and when it is in its dark states. Assuming these dwell times to be exponentially distributed (or equal in distribution to a sum of exponentially distributed random variables in the case of multiple dark states), maximum likelihood estimates of the transition times are then computed. This method, termed here as *exponential fitting* and given a detailed discussion in Section 1.E of Appendix 1, has two limitations. Firstly, it does not correctly account for the effect of the imaging procedure on the stochastic structure of the discrete time process. Secondly, it does not allow for the photo-bleached (absorbing) state, which must be identified and accounted for by truncation of the data to the last observed On state. This is especially troublesome as, to an observer, it is indistinguishable from a temporary dark state. This method therefore results in the absence of estimates for the photo-bleaching rate(s) and can lead to significantly biased estimates of the transition rates between On and dark states.



**Figure 1.2:** (a) A simulated intensity signal of a fluorophore across time. Each measurement corresponds to the intensity in a frame. 7500 frames were recorded over 250 seconds at a rate of 30 frames per second. (b) Close up of the signal over the time window of 35s to 55s. In red is the observed signal  $\{Y_n\}$  indicating if the fluorophore was detected in a particular frame. (c) A further close up of the signal showing intensity read-outs for independent frames. The true, hidden photon emitting On state of the molecule is also indicated, demonstrating how sub-frame length photon emitting events (such as the event at 38.37s) can be missed due to noise or the temporal resolution of the data acquisition.

Hidden Markov models (HMMs) are used widely across scientific and engineering disciplines to relate a sequence of observations, called emissions, to the states of an unobserved (hidden) Markov process, the target of inference. Their use is particularly prevalent in image processing, where the observations are a sequence of images in time and it is commonly assumed that each image is dependent only on the state of the hidden process at the time at which it is observed. Such an approach has been proposed for this problem in Greenfeld et al. (2015), where the hidden process is a discretised version of  $\{X(t)\}$ . Here, they let  $\{Y_n\}$  be the sequence of photon-fluxes such that it is a standard (first-order) HMM with Poisson emissions. They then implement the Baum-Welch algorithm (Baum and Petrie, 1966, Baum and Eagon, 1967, Baum and Sell, 1968, Baum et al., 1970) to estimate the transition proba-

bilities of the discretised process and use an approximation to obtain the transition rates of the continuous time process  $\{X(t)\}$ . In doing so, they acknowledge that missed events will heavily bias rate estimates. Furthermore, their model is also unable to deal with the photo-bleached state.

### 1.1.1 Chapter summary

In this chapter, we provide two important contributions. Firstly, in Section 1.2, by considering a general model for  $\{X(t)\}$  that includes  $d + 1$  (with  $d \in \mathbb{Z}_{\geq 0}$ ) dark states and a photo-bleached state, we rigorously formulate the discrete time stochastic process  $\{Y_n\}$  that indicates whether a molecule is detected in each frame. A crucial part of this formulation is recognising that an image is not formed from an instantaneous sampling of the true state, as is usually assumed in image processing, but is instead formed by exposing a camera sensor over a time interval of length  $\Delta$ . That is to say,  $Y_n$  is not dependent on just  $X(n\Delta)$ , but instead on the integral (i.e. all values) of  $\{X(t)\}$  over the interval  $[(n - 1)\Delta, n\Delta)$ . Taking consideration of noise and instrument sensitivity, we fully account for missed events and give important results on the stochastic structure of  $\{Y_n\}$ , including showing it is non-Markovian. In Section 1.2.3, we develop the Photo-switching HMM (PSHMM) for  $\{Y_n\}$ , where we first implement a time discretisation scheme on the hidden Markov process  $\{X(t)\}$ . Crucially, as discussed above, correct understanding of the imaging procedure dictates two key properties. Firstly,  $Y_n$  depends on both the current (end of frame) and previous (beginning of frame) hidden states,  $X((n - 1)\Delta)$  and  $X(n\Delta)$ , respectively. Secondly, this HMM possesses emission probabilities that are dependent on the static parameters of the hidden process state transition rates that we ultimately wish to estimate. This *coupled* behaviour renders traditional expectation maximisation (EM)-type methods designed for parameter estimation in decoupled HMMs (e.g. Baum et al. (1970)) inappropriate. We therefore make the novel step of introducing what we call *transmission* (transition-emission) matrices that incorporate this coupling between transition and emission probabilities by capturing all the dependencies in the model.

In Section 1.3, we take time to carefully derive the form of the transmission matrices that are required for the PSHMM, under any model  $d \in \mathbb{Z}_{\geq 0}$ . This section begins with a brief overview outlining the key mathematical concepts, with special reference to multivariate counting processes, which will be used in the derivation. We then describe the methodology to compute transmission probabilities when the hidden signal  $X(\Delta) \neq 1$  and  $X(\Delta) = 1$ , respectively, with subsequent unification of these two cases to provide expressions for the transmission matrices. An algorithm to computationally recover these matrices is also provided. Although it is shown that in general, the form of these matrices cannot be provided in closed form, in Section 1.C of Appendix 1, we derive and provide explicit forms for the transmission matrices under the special case of a single dark state  $d = 0$ .

The second contribution of this chapter is to propose novel methodology for estimating the state transition rates of  $\{X(t)\}$  under this correct treatment of the imaging procedure. In Section 1.4, for a given photo-switching kinetic model, we adapt the well known forward-backward algorithm [Levinson et al. \(1983\)](#) to compute the likelihood of observations of the PSHMM, which relies on the transmission matrices previously constructed. Through numerical optimisation procedures, we are able to compute maximum likelihood estimates of the transition rate parameters for the continuous time process  $\{X(t)\}$  that we wish to draw inference on. In the case of an unknown kinetic model, we propose the use of the Bayesian information criterion (BIC) for selecting the best suited model from a set of proposals, thus also providing a powerful tool for chemists wishing to infer the number of quantum states a particular fluorophore can exist in. In this section, we also provide extensive empirical analysis of the proposed method. In Section 1.5.1, simulation studies that compare this new estimation scheme to the exponential fitting method on a range of photo-switching models demonstrate significant improvements in both the bias and the variance of the rate estimates. We further show the BIC performs accurate model selection when presented with a range of model proposals. Finally, in Section 1.6, the estimation scheme presented in this chapter is applied to the Alexa Fluor 647 data originally analysed by the exponential fitting method in [Lin et al. \(2015\)](#), consistently selecting the



hypothesised three temporary off-state model (Figure 1.1d) and revealing a clear dependence between laser intensity and key transition rate parameters.

## 1.2 The Photo-Switching Hidden Markov Model

The true photo-switching behaviour of the fluorophore is a continuous time stochastic phenomenon. However, an experimenter can only ever observe a discretised manifestation of this by imaging the fluorophore in a sequence of frames. These frames are regarded as a set of sequential exposures of the fluorophore and the observed discrete time signal indicates whether the fluorophore has been observed in a particular frame. It is the continuous time process on which we wish to draw inference, based on the observed discrete-time process indicating whether the fluorophore was observed in a frame. In this section we first present the continuous time Markov model of the true (hidden) photo-switching behaviour, and then derive the observed discrete time signal, together with key results on its statistical properties. Using a state-space representation relating observations with this hidden signal, we then formulate the PSHMM.

Hidden Markov models, first presented in [Baum and Petrie \(1966\)](#), relate a sequence of observations to the states of an unobserved or *hidden* Markov chain. The aim of building a hidden Markov model (HMM) is to allow inference on the hidden process using these observations. In its simplest form, an HMM assumes the propagation of both state and observed sequences to be in discrete time, and a general first order HMM assumes that the observation process  $\{Y_n : n \in \mathbb{Z}_{>0}\}$  is related to a hidden first order Markov Chain  $\{X_n : n \in \mathbb{Z}_{>0}\}$  via an emission probability distribution  $B := (B)_{i,j} = \mathbb{P}(Y_n = j | X_n = i)$ , considered to be fully independent of the static parameters that characterise the probability distribution of state transitions  $P := (P)_{i,j} = \mathbb{P}(X_n = j | X_{n-1} = i)$ . In this setting we say  $B$  and  $P$  are *decoupled*. For a sequence  $y_1, \dots, y_N$  of observations from this model, the Baum-Welch re-estimation algorithm ([Baum and Petrie, 1966](#), [Baum and Eagon, 1967](#), [Baum and Sell, 1968](#), [Baum et al., 1970](#)) is an Expectation Maximisation (EM) type method that utilises the forward-backward algorithm

(see [Levinson et al., 1983](#), for details) to optimise the likelihood function and compute maximum likelihood estimates of  $\nu_X$  (the probability mass of  $X_0$ ),  $B$  and  $P$ . This in turn can be used to estimate parameters of the emission and state transition probabilities. When the hidden Markov process and/or the observation process are of higher order, the HMM can be transformed to a general first order process ([Du Preez, 1998](#), [Lee and Lee, 2006](#), [Ching et al., 2003](#)) and Baum-Welch can be applied in the usual way. Readers are directed to [MacDonald and Zucchini \(1997\)](#) for a comprehensive review.

Whilst standard HMMs have been extensively studied and are most frequently used in applications, the rigid framework of being in discrete time with emission probabilities decoupled from state transition probabilities is not always suitable, as we will now show is the case for images formed by exposures over a time interval. Here, we take time to carefully formulate the HMM suitable for this application, presenting what we call *transmission* (transition-emission) matrices to capture the dependencies in the model.

### 1.2.1 Continuous time signal

We model the true photo-switching effect of the fluorophore as a continuous time Markov process,  $\{X(t) : t \in [-T^*, \infty)\}$  with discrete state space  $\mathcal{S}_X$ . This is a stochastic process which satisfies the Markov property

$$\mathbb{P}(X(t_n) = i_n | X(t_{n-1}) = i_{n-1}, \dots, X(t_0) = i_0) = \mathbb{P}(X(t_n) = i_n | X(t_{n-1}) = i_{n-1}),$$

for any sequence of times  $-T^* \leq t_0 < t_1 < \dots < t_n < \infty$  and any sequence of states  $i_j \in \mathcal{S}_X$  for  $j = 0, \dots, n$ .

**Remark 1.** *In (d)STORM, the experiment begins at time  $-T^*$ . At this time, all fluorophores are in the photon-emitting state 1 and a buffer period of time  $T^*$  is needed to allow the majority of these molecules to photo-switch to a metastable dark state. While images are collected within the period  $[-T^*, 0)$ , these are subject to wide-field image diffraction limitations due to extensively high signals produced by molecules occupying the On state. For this reason, all experiments conducted under these techniques discard the first*

images taken within this buffer period.

In this thesis, we consider a general model for  $\{X(t)\}$  that can accommodate the numerous mechanisms of photo-switching utilised in (d)STORM. Specifically, this model consists of a photon emitting (On) state 1,  $d + 1$  non photon emitting (dark/temporary off) states  $0_0, 0_1, \dots, 0_d$ , where  $d \in \mathbb{Z}_{\geq 0}$ , and a photo-bleached (absorbing) state 2. We denote the state  $0_0 \equiv 0$  for the  $d = 0$  case of a single dark state. The model, illustrated in Figure 1.3, allows for transitions from state 1 to the multiple dark states (from a photochemical perspective, these can include triplet, redox and quenched states). These dark states are typically accessed via the first dark state 0 (reached as a result of inter-system crossing of the excited  $S_1$  electron to the triplet  $T_1$  state; see Figure 1.1a). Further dark states  $0_{i+1}$ ,  $i = 0, \dots, d - 1$ , are accessible by previous dark states  $0_i$  (by, for example, the successive additions of electrons forming radical anions (Van de Linde et al., 2010)). We allow the On state 1 to be accessible by any dark state and we consider the most general model in which the photo-bleaching state 2 is accessible from any combination of other states (Vogelsang et al., 2010, Van de Linde and Sauer, 2014, Ha and Tinnefeld, 2012).

The state space of  $\{X(t)\}$  is  $\mathcal{S}_X = \{0, 0_1, \dots, 0_d, 1, 2\}$  and is of cardinality  $d + 3$ . We denote  $\lambda_{ij}$  to be the transition rate between states  $i$  and  $j$  and  $\mu_i$  to be the photo-bleaching rate from state  $i$  to 2, where  $i, j \in \bar{\mathcal{S}}_X := \mathcal{S}_X \setminus \{2\}$ .

The generator matrix for  $\{X(t)\}$  is therefore given as

$$G = \begin{pmatrix} -\sigma_0 & \lambda_{00_1} & 0 & 0 & 0 & 0 & \dots & \lambda_{01} & \mu_0 \\ 0 & -\sigma_{0_1} & \lambda_{0_1 0_2} & 0 & 0 & 0 & \dots & \lambda_{0_1 1} & \mu_{0_1} \\ 0 & 0 & -\sigma_{0_2} & \lambda_{0_2 0_3} & 0 & 0 & \dots & \lambda_{0_2 1} & \mu_{0_2} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \dots & -\sigma_{0_d} & \lambda_{0_d 1} & \mu_{0_d} \\ \lambda_{10} & 0 & 0 & 0 & 0 & 0 & \dots & -\sigma_1 & \mu_1 \\ 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 \end{pmatrix}, \quad (1.1)$$

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  $i = 0, \dots, d - 1$ . For brevity, we will specifically refer to the positive

elements of  $G$  as  $\lambda_G$ , the vector of photo-switching and photo-bleaching rates.

For full characterisation, we define the initial probability mass of  $\{X(t)\}$  at the beginning of imaging (this is the probability mass over the states of  $X(0)$ ) as  $\nu_X := (\nu_0 \ \nu_{0_1} \ \dots \ \nu_{0_d} \ \nu_1 \ \nu_2)^\top$  with  $\sum_{j \in \mathcal{S}_X} \nu_j = 1$ , and is in general **unknown**.

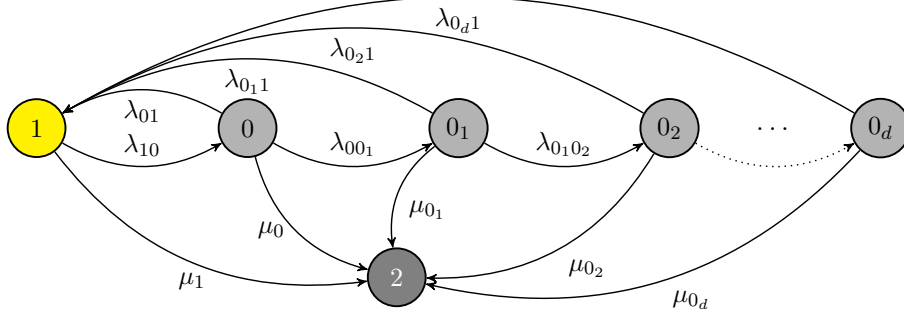
**Remark 2.** *It is important here to clarify that  $\nu_X$  is different from the initial probability mass of  $\{X(t)\}$  at the beginning of the experiment (this is the probability mass over the states of  $X(-T^*)$ ), which is denoted as*

$$\nu_X^* := (\nu_0^* \ \nu_{0_1}^* \ \dots \ \nu_{0_d}^* \ \nu_1^* \ \nu_2^*)^\top,$$

with  $\sum_{j \in \mathcal{S}_X} \nu_j^* = 1$ . By Remark 1, all fluorophores imaged via (d)STORM receive an initial excitation to the On state 1, enabling the most commonly occurring probability mass vector  $\nu_X^*$  in practice to be such that  $\nu_1^* = 1$ .

**Remark 3.** *The formulation of  $\{X(t)\}$  in (1.1) is also suitable for modelling the photo-switching behaviour of molecules imaged in PALM. In these experiments, all fluorophores initially occupy the inactive state  $0_1$  before transitioning into the On-dark cycle between states 1 and 0, and finally photo-bleaching into state 2. In this situation, the  $d = 1$  model can be applied, with  $\nu_{0_1}^* = 1$  and the rate  $\lambda_{00_1}$  set to zero.*

In this thesis, we will refer to specific sub-models (e.g. those presented in Figure 1.1b - 1.1d), of the full model shown in Figure 1.3 in the form  $\mathcal{M}_A^d$ . Here,  $d$  is the number of multiple dark states beyond the  $0_0$  state that is present in all models, and  $A \subseteq \bar{\mathcal{S}}_X$  (with cardinality  $|A|$ ) denotes the set of states from which the photo-bleaching state 2 is accessible. For the three classical models presented in Figure 1.1: model (b) is  $\mathcal{M}_\emptyset^0$ : the  $d = 0$  case where  $\mu_0 = \mu_1 = 0$ , model (c) is  $\mathcal{M}_{\{0,1\}}^0$ : the  $d = 0$  case where  $\mu_0, \mu_1 > 0$ , and model (d) is  $\mathcal{M}_\emptyset^2$ : the  $d = 2$  case where  $\mu_0 = \mu_{0_1} = \mu_1 = 0$ .

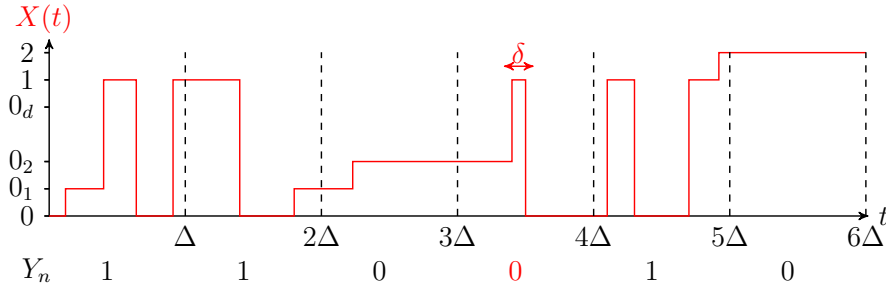


**Figure 1.3:** General  $d + 3$  state ( $d \in \mathbb{Z}_{\geq 0}$ ) model of a fluorophore.

### 1.2.2 Discrete time observations

Having presented the continuous time model for the true photo-switching behaviour, we will now introduce the model for the observed discrete time process and show how the transition rates given in (1.1) are not amenable to direct estimation.

The imaging procedure requires taking a series of successive frames. Frame  $n$  is formed by an exposure over the time interval  $[(n - 1)\Delta, n\Delta)$ , where  $n \in \mathbb{Z}_{>0}$ . The constant  $\Delta$  corresponds to the exposure time for a single frame, also known as the frame length. We define the *discrete* time observed process  $\{Y_n : n \in \mathbb{Z}_{>0}\}$ , with state space  $\mathcal{S}_Y = \{0, 1\}$ , as  $Y_n = 1$  if the fluorophore (characterised by  $\{X(t)\}$ ) is observed in frame  $n$  and equal to 0 otherwise. For the fluorophore to be observed in the time interval  $[(n - 1)\Delta, n\Delta)$  it must be in the On state 1 for a time greater than  $\delta \in [0, \Delta)$ . The value of  $\delta$  is unknown and is a result of background noise and the imaging system's limited sensitivity. We note that if  $\{X(t)\}$  exhibits multiple jumps to state 1 within a frame, then a sufficient condition for observing the fluorophore is that the total time spent in the On state exceeds  $\delta$ . The  $\delta = 0$  case is the idealistic scenario of a noiseless system and perfect sensitivity such that the fluorophore is detected if it enters the On state for any non-zero amount of time during the exposure time  $\Delta$ .



**Figure 1.4:** Illustration of how the states for  $Y_n$  derive from the process  $X(t)$

We formally define the (deterministic) observed process as

$$Y_n = \mathbb{1}_{[\delta, \Delta)} \left( \int_{(n-1)\Delta}^{n\Delta} \mathbb{1}_{\{1\}}(X(t)) \, dt \right), \quad (1.2)$$

where  $\mathbb{1}_A(\cdot)$  is the indicator function such that  $\mathbb{1}_A(x) = 1$  if  $x \in A$  and is zero otherwise. Figure 1.4 illustrates the manifestation of the discrete time signal  $\{Y_n\}$  from the continuous time signal  $\{X(t)\}$ .

Importantly, Theorem 1 shows that the observation process  $\{Y_n\}$  does not exhibit the Markov property (of any order) for any  $d \in \mathbb{Z}_{\geq 0}$ , and for any  $\Delta$  and  $\delta$  such that  $\Delta > \delta \geq 0$ . The non-Markovianity excludes classical inference methods and motivates the use of a Hidden Markov Model (HMM), with a likelihood based approach for estimating the unknown parameters of the model.

**Theorem 1.** *Consider the set of processes  $\{X(t) : t \in \mathbb{R}_{\geq 0}\}$  defined from all models  $\mathcal{M}_A^d$ , where  $d$ , the number of multiple dark states, takes any value in  $\mathbb{Z}_{\geq 0}$  and  $A$  is any subset of  $\bar{\mathcal{S}}_X := \mathcal{S}_X \setminus \{2\} = \{0, 0_1, \dots, 0_d, 1\}$ , denoting the set of states the photo-bleaching state 2 is accessible from.*

*Then fixing  $\Delta > 0$  and any  $\delta \in [0, \Delta)$ , the process  $\{Y_n : n \in \mathbb{Z}_{> 0}\}$  generated by  $\{X(t) : t \in \mathbb{R}_{\geq 0}\}$  as defined in (1.2) from all models  $\mathcal{M}_A^d$ , is not a Markov Chain of any order.*

*Proof.* See Section 1.A of Appendix 1. □

## The inference problem

The inference problem is two-fold. Firstly for a given model (with a fixed number of multiple dark states  $d$ ) that yields a finite length realisation of  $\{Y_n\}$ , the aim is to estimate the vector of unknown parameters

$$\boldsymbol{\theta}^{(d)} = \left( \boldsymbol{\lambda}_G \quad \boldsymbol{\nu}_X \quad \delta \right)^\top. \quad (1.3)$$

Here, under model  $\mathcal{M}_A^d$ ,  $\boldsymbol{\theta}^{(d)}$  takes values in the parameter space

$$\Theta^{(d)} = \mathbb{R}_{>0}^{2(d+1)+|A|} \times \mathbb{S}^{d+3} \times [0, \Delta),$$

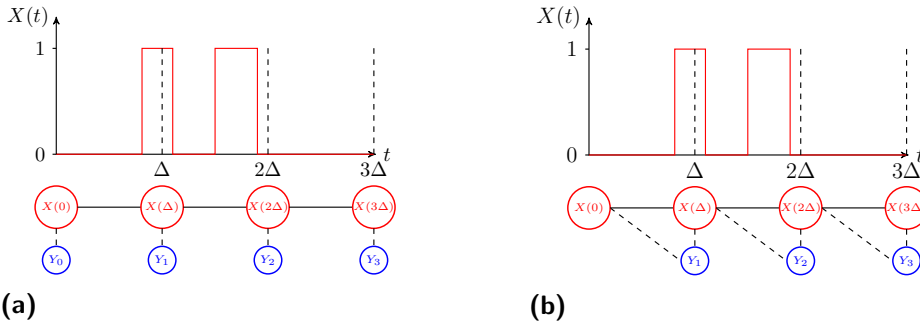
where  $\mathbb{S}^{d+3}$  denotes the  $d + 3$ -dimensional simplex of probability vectors.

Beyond this, it may be the case that the true model (characterised by its number of dark states) is unknown and may need to be selected in addition to estimating the unknown parameters. In order to tackle these problems, we formulate the likelihood function of imaging observations given  $\boldsymbol{\theta}^{(d)}$ , which can be derived from a HMM construction that we now present.

### 1.2.3 Characterising photo-switching behaviour

Here we build an HMM for our observation process  $\{Y_n\}$ , which we call the Photo-switching hidden Markov model (PSHMM). The first immediate reason as to why the standard set-up outlined above is inappropriate for this application is because the hidden Markov process  $\{X(t)\}$  evolves in continuous time. To deal with this, we need to adopt a time-discretisation scheme for the hidden process. Analogously to [Liu et al. \(2015\)](#), we state that  $\{X(t)\}$  propagates in  $\Delta$ -separated discrete time steps according to the transition probability matrix  $P_\Delta = e^{G\Delta}$ , where  $G$  is given in (1.1). Our hidden process is therefore now represented by the discrete time Markov chain  $\{X(n\Delta) : n \in \mathbb{Z}_{\geq 0}\}$ .

When  $Y_n$  depends solely on  $X(n\Delta)$  (see Figure 1.5a) and the corresponding emission matrix  $B := (B)_{i,j} = \mathbb{P}(Y_n = j | X_n = i)$  is decoupled from  $P$ , a con-



**Figure 1.5:** Illustration of the HMM set-up. (a): Traditional HMM where the observed state is dependent on the current hidden. (b): Our HMM where the observed state depends on both the current and past hidden states.

tinuous time EM algorithm (Liu et al., 2015) analogous to the Baum-Welch can be used to estimate  $\nu_X$ ,  $B$  and  $P$ . However, this will be inappropriate in our setting for two related reasons. Firstly, we have shown in Section 1.2, specifically Equation (1.2), that exposing images over a non-zero length of time means  $Y_n$  depends on the full path of  $\{X(t)\}$  within the interval  $[(n-1)\Delta, n\Delta)$ . To correctly deal with this it is necessary to construct the HMM to consider dependence between  $Y_n$  and both  $X((n-1)\Delta)$  and  $X(n\Delta)$  (see Figure 1.5b). Secondly, this construction of  $\{Y_n\}$  in (1.2) means the emission probabilities are clearly dependent on the static parameters  $\theta^{(d)}$  of the hidden process and are therefore coupled with  $P$ . The EM procedure highlighted in Liu et al. (2015) requires decoupled  $B$  and  $P$  so that at each step the quasi-likelihood can be optimised separately. To the best of our knowledge, methods for dealing with coupled systems have not been dealt with in the literature. While an EM algorithm could be used for a coupled system, analytic forms for the update steps would in general be intractable, leading to numerical maximisation procedures at each iteration, thereby increasing computational complexity. We will now formally characterise the PSHMM and provide a novel method for estimating the unknown static parameters in the case of a coupled system.



## Formal characterisation of the PSHMM

Formally, we characterise our PSHMM with

1. an initial probability vector  $\boldsymbol{\nu}_X = (\nu_0 \ \nu_{0_1} \ \dots \ \nu_{0_d} \ \nu_1 \ \nu_2)^\top$  where  $\nu_i := \mathbb{P}(X(0) = i)$  for  $i \in \mathcal{S}_X$ ;
2. Transmission matrices

$$B_\Delta^{(l)} = \begin{pmatrix} b_{00,\Delta}^{(l)} & b_{00_1,\Delta}^{(l)} & \dots & b_{00_d,\Delta}^{(l)} & b_{01,\Delta}^{(l)} & b_{02,\Delta}^{(l)} \\ b_{0_1 0,\Delta}^{(l)} & b_{0_1 0_1,\Delta}^{(l)} & \dots & b_{0_1 0_d,\Delta}^{(l)} & b_{0_1 1,\Delta}^{(l)} & b_{0_1 2,\Delta}^{(l)} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ b_{0_d 0,\Delta}^{(l)} & b_{0_d 0_1,\Delta}^{(l)} & \dots & b_{0_d 0_d,\Delta}^{(l)} & b_{0_d 1,\Delta}^{(l)} & b_{0_d 2,\Delta}^{(l)} \\ b_{10,\Delta}^{(l)} & b_{10_1,\Delta}^{(l)} & \dots & b_{10_d,\Delta}^{(l)} & b_{11,\Delta}^{(l)} & b_{12,\Delta}^{(l)} \\ 0 & 0 & 0 & 0 & \dots & b_{22,\Delta}^{(l)} \end{pmatrix}, \quad (1.4)$$

where

$$\begin{aligned} b_{ij,\Delta}^{(l)} &:= \mathbb{P}_{\boldsymbol{\theta}^{(d)}}(Y_n = l, X((n+1)\Delta) = j | X(n\Delta) = i) \\ &= \mathbb{P}_{\boldsymbol{\theta}^{(d)}}(Y_1 = l, X(\Delta) = j | X(0) = i) \quad i, j \in \mathcal{S}_X, l \in \mathcal{S}_Y, \end{aligned} \quad (1.5)$$

$$b_{22,\Delta}^{(l)} = \mathbb{1}_{\{0\}}(l).$$

These transmission matrices combine the transition and emission probabilities, thereby allowing us to account for a coupled system. The full mathematical formulation for deriving their forms is discussed in Section 1.3 and involves conditioning on the number of jumps from all  $d+1$  dark states within the interval  $[0, \Delta)$ . From this, we use Laplace transforms and the distributions of state holding times to iteratively compute matrices that converge to our set of transmission matrices. It should be noted here, that the forms of these matrices are not easily tractable with respect to the parameter vector  $\boldsymbol{\theta}^{(d)}$ , therefore requiring numerical approximations. However, a detailed algorithm that can be implemented to recover these matrices is also presented in Section 1.3.

## Accounting for false positive observations

Occasionally, random peaks in the background noise may exceed the threshold value used to determine a fluorophore in the On state, resulting in a false positive identification of the fluorophore. For experiments conducted over a large enough number of frames, this false positive rate may become significant in the observed process  $\{Y_n\}$ .

Specifically, if  $\omega \in [0, 1]$  denotes the probability of falsely observing a fluorophore, assumed independent of the general observation process, then we may use the updated transmission matrices

$$B_{\Delta}^{*(0)} = (1 - \omega)B_{\Delta}^{(0)} \quad (1.6)$$

$$B_{\Delta}^{*(1)} = B_{\Delta}^{(1)} + \omega B_{\Delta}^{(0)}, \quad (1.7)$$

in the PSHMM characterisation. The new parameter vector to estimate then becomes  $\boldsymbol{\theta}_{\omega}^{(d)} = \left( (\boldsymbol{\theta}^{(d)})^{\top} \quad \omega \right)^{\top} \in \Theta_{\omega}^{(d)} := \Theta^{(d)} \times [0, 1]$ .

### 1.3 Derivation of transmission probabilities

In this section, we will derive the forms of the transmission probabilities that are used as entries of the emission matrices  $\{B_{\Delta}^{(l)}\}_{l=0}^1$  in (1.4) needed in the computation of the PSHMM likelihood. We will firstly provide an extensive overview outlining all the necessary mathematical tools, namely the use of Laplace transforms and the distributions of state holding times, that are needed to compute these matrices for all  $d \in \mathbb{Z}_{\geq 0}$ .

In particular, when  $\{Y_n\}$  is recorded over exposure times  $\Delta$  for each frame, we will consider deriving the transmission probability function in (1.5), holding for any  $i, j \in \mathcal{S}_X = \{0, 0_1, \dots, 0_d, 1, 2\}$  and  $l \in \mathcal{S}_Y = \{0, 1\}$ . By further describing how the state of  $X(\Delta)$  will change the structure of these computations, we will then delve into its technicalities on a case by case basis. In doing so, a derivation to Algorithm 2 (provided in Section 1.B of Appendix 1) that computes the elements of the updated transmission matrices in (1.6

- 1.7) will be provided. Finally, we will discuss that the form of transmission probabilities when  $d = 0$  can be derived exactly. This is an experimentally useful and mathematically complete exercise, and is presented in Section 1.C of Appendix 1.

On the notation we will use in this section, we will make extensive use of the following: Firstly,  $\mathcal{S}_X = \{0, 0_1, \dots, 0_d, 1, 2\}$  is maintained to be the state space of the process  $\{X(t)\}$  and  $\bar{\mathcal{S}}_X := \mathcal{S}_X \setminus \{2\}$  the state space of the process without the photo-bleaching state 2. We let  $\mathbf{0}_n$  and  $\mathbf{1}_n$  denote the  $(n \times 1)$  column vectors of zeros and ones, respectively, and  $I_n$  to be the  $n \times n$  identity matrix, holding for any  $n \in \mathbb{N}$ . To introduce sub-matrices of matrices, 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  $A$ .

Secondly, we will continually refer to  $i \in \bar{\mathcal{S}}_X$  as the state of  $X(0)$ ,  $j \in \mathcal{S}_X$  as the state of  $X(\Delta)$  and  $l \in \mathcal{S}_Y$  as the state of  $Y_1$ , unless stated otherwise.

### 1.3.1 Overview

In the following, we will describe key mathematical concepts, with particular reference to multivariate counting processes and the idea of *labelling sets* that are needed to compute (1.5).

In the most general setting when  $d \in \mathbb{Z}_{\geq 0}$ , we compute (1.5) by conditioning on the number of transitions made by  $\{X(t)\}$  between states in a *labelling set*  $\mathcal{R}$ , as stated in Definition 1, before time  $\Delta$ .

**Definition 1** (Labelling set (Minin and Suchard, 2007)). *Let  $\{X(t)\}$  be a continuous time Markov process taking values in a discrete state space  $\mathcal{S}_X$ . A **labelling set**  $\mathcal{R}$  is a set which contains tuples of ordered index pairs of the form  $(p, q) \in \mathcal{R}$ , that label transitions from state  $p$  to state  $q$ , with  $p, q \in \mathcal{S}_X$ .*

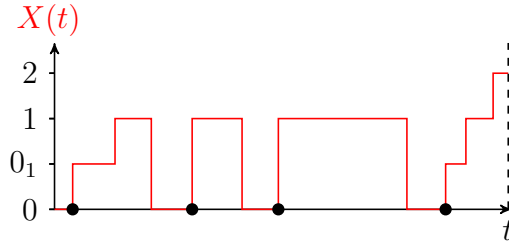
Using an arbitrary labelling set, we now define  $\{N_{\mathcal{R}}(t) : t \in \mathbb{R}_{\geq 0}\}$  to be the (univariate) counting process, which counts the number of transitions in the labelling set  $\mathcal{R}$  that have occurred by  $\{X(t)\}$  before time  $t$ . This process

has state space  $\mathbb{Z}_{\geq 0}$ . A concrete example detailing a labelling set and its resulting counting process is described in Example 1.

**Example 1.** Consider the photo-switching process  $\{X(t)\}$  under model  $\mathcal{M}_{S_X}^1$ . We wish to define the labelling set that considers all transitions from the first dark state 0. Using the representation in (1.3) for  $d = 1$ , when  $\{X(t)\}$  is in 0, it can subsequently transition to any one of the states  $0_1, 1, 2$ . Therefore, labelling all transitions from state 0 would imply (using Definition 1) that

$$\mathcal{R} = \{(0, 0_1), (0, 1), (0, 2)\}.$$

Now consider the following path of  $\{X(t)\}$  over some arbitrary time  $t > 0$ :



During this time,  $\{X(t)\}$  transitions from state 0 to state  $0_1$  twice and to state 1 twice, with transitions indicated in black. It does not transition to state 2 from 0 within this interval. Counting all of the transitions made within the labelling set  $\mathcal{R}$  before time  $t$  therefore enables  $N_{\mathcal{R}}(t) = 4$ .

We further define  $\{\mathbf{N}_{\mathcal{R}_n}(t) : t \in \mathbb{R}_{\geq 0}\}$  to be the random vector comprised of the  $n \in \mathbb{Z}_{>0}$  univariate counting processes  $\mathbf{N}_{\mathcal{R}_n}(t) = [N_{\mathcal{R}_1}(t) \ N_{\mathcal{R}_2}(t) \ \dots \ N_{\mathcal{R}_n}(t)]^\top$ , so that  $\mathbf{N}_{\mathcal{R}_n}(t)$  is a multivariate counting process. Here  $\mathcal{R}_n := \{\mathcal{R}_1, \mathcal{R}_2, \dots, \mathcal{R}_n\}$  is the set of  $n$  labelling sets  $\mathcal{R}_1, \mathcal{R}_2, \dots, \mathcal{R}_n$ . This process has state space  $\mathbb{Z}_{\geq 0}^n$ .

For some  $n \in \mathbb{N}$ , let  $\mathcal{R}_n^{ij}$  denote the set of labelling sets needed to compute  $b_{ij}^{(l)}$  (discussion on choosing  $\mathcal{R}_n^{ij}$  will follow). We define the probabilities

$$q_{ij}(\mathbf{k}, \Delta) = \mathbb{P}_{\theta^{(d)}}(\mathbf{N}_{\mathcal{R}_n^{ij}}(\Delta) = \mathbf{k}, X(\Delta) = j | X(0) = i)$$

$$\xi_{ij}(l, \mathbf{k}, \Delta) = \mathbb{P}_{\theta^{(d)}}(Y_1 = l | \mathbf{N}_{\mathcal{R}_n^{ij}}(\Delta) = \mathbf{k}, X(0) = i, X(\Delta) = j),$$

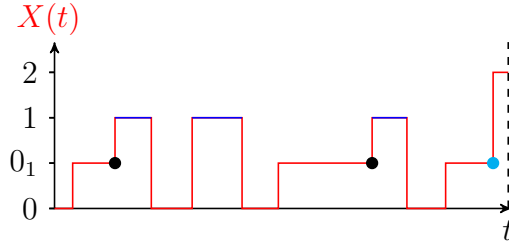
and write (by conditioning  $n$  times) the transmission probabilities in the form

$$b_{ij,\Delta}^{(l)} = \sum_{\mathbf{k} \in \mathbb{Z}_{\geq 0}^n} q_{ij}(\mathbf{k}, \Delta) \xi_{ij}(l, \mathbf{k}, \Delta). \quad (1.8)$$

Although we are free to choose any  $\mathcal{R}_n^{ij}$ , poor choices may lead to intractability of the above probabilities. An example of a poor choice is described in Example 2.

**Example 2.** Consider  $\{X(t)\}$  under model  $\mathcal{M}_{S_X}^1$ . Let  $\mathcal{R}_1 = \{(0_1, 1)\}$  be the set labelling transitions from state  $0_1$  to 1, and  $\mathcal{R}_2 = \{(0_1, 2)\}$  be the set labelling transitions from state  $0_1$  to 2. During a path of which  $X(0) = 0$  and  $X(t) = 2$ , now define the set  $\mathcal{R}_2^{0_1 2}$  to be comprised of these two labelling sets so that  $\mathcal{R}_2^{0_1 2} = \{\mathcal{R}_1, \mathcal{R}_2\}$ .

Consider the following path of  $\{X(t)\}$  over some arbitrary time  $t > 0$ :

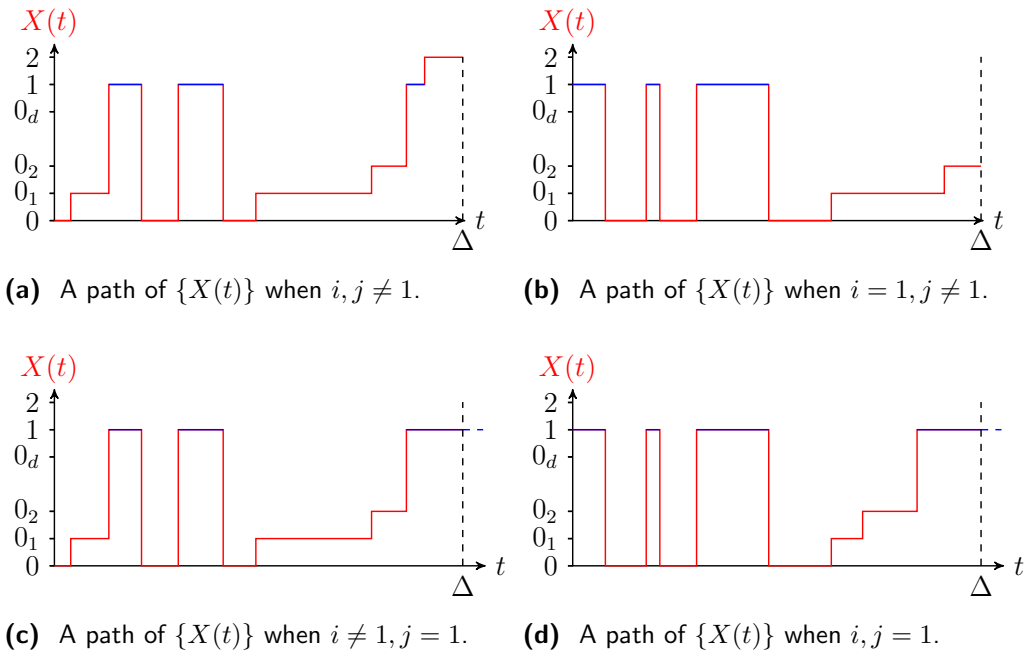


During this path, there are two transitions from  $0_1$  to 1 enabling  $N_{\mathcal{R}_1}(t) = 2$  (indicated in black) and there is one transition from  $0_1$  to 2 enabling  $N_{\mathcal{R}_2}(t) = 1$  (indicated in cyan). Hence,  $\mathbf{N}_{\mathcal{R}_2^{0_1 2}}(t) = (2 \ 1)^\top$ .

In the path above, we also see that there is a transition from state 0 to 1. By definition, an observation of a fluorophore within this period is dependent on **all** time pieces occupying the photon-emitting state 1 (indicated in blue). By using  $\mathcal{R}_2^{0_1 2}$ , we observe that there are two transitions from state  $0_1$  to 1 and so there are at least two time pieces in the On state. However, the transition from state 0 to 1 is ignored using this choice of labelling set. This means that we cannot fully characterise an observation of a molecule using  $\mathcal{R}_2^{0_1 2}$ .

We will thus now describe a method as to how one can choose  $\mathcal{R}_n^{ij}$  for effective computation of (1.8).

The first term  $q_{ij}(\mathbf{k}, \Delta)$  over any  $\mathcal{R}_n^{ij}$  can be computed using Laplace transforms. However, identifying which  $\mathcal{R}_n^{ij}$  is needed to compute (1.8) requires attention to  $\xi_{ij}(l, \mathbf{k}, \Delta)$ . When  $l = 1$ , this term describes the probability of observing a fluorophore given the number of transitions made in  $\mathcal{R}_n^{ij}$ . An observation of a fluorophore (as defined in (1.2)) is dependent on the total time spent in the On state 1 and as such we endeavour to characterise this time using the transitions that have occurred within the interval  $[0, \Delta)$ .



**Figure 1.6:** Figures 1.6a and 1.6b consider two possible paths of  $\{X(t)\}$  when  $j \neq 1$  and show that all individual time pieces in state 1 (blue) are distinctly exponentially distributed. Figures 1.6c and 1.6d consider two paths of  $\{X(t)\}$  when  $j = 1$  and show that all but the last individual time pieces in state 1 (blue) are exponential. The final time piece suffers from right-censoring.

Figures 1.6a - 1.6d show four possible paths of  $\{X(t)\}$  within this time interval. Firstly, Figures 1.6a and 1.6b highlight two paths when  $X(\Delta) \neq 1$ . In Figure 1.6a when  $X(0) \neq 1$ , each of the three time pieces in state 1 are exponentially independently and identically distributed (iid) (with scale parameter  $\sigma_1$ ), and thus the total time spent in this state is characterised by an Erlang( $3, \sigma_1$ ) density (truncated over the interval  $[0, \Delta)$ ). This is also true for the path shown in Figure 1.6b when  $X(0) = 1$ , since the memory-

less property ensures that the first time piece still remains exponential. It is easy to see that this property will hold for any  $i, j \in \mathcal{S}_X$  with  $j \neq 1$ . Computing  $b_{ij,\Delta}^{(l)}$  in this case would thus require knowledge of the number of time pieces in the On state, or equivalently, of the number of transitions made to state 1. This can be done by considering the sole ( $n = 1$ ) labelling set  $\mathcal{R}^0 = \{(0, 1), (0_1, 1), \dots, (0_d, 1)\}$  and setting  $\mathcal{R}_1^{ij} = \mathcal{R}^0$ , for all  $i \in \bar{\mathcal{S}}_X, j \in \mathcal{S}_X \setminus \{1\}$ . In this case, which from herein we will refer to as *Case*  $j \neq 1$ , we condition on the univariate counting process  $\{N_{\mathcal{R}^0}(t)\}$  and write (1.8) as

$$b_{ij,\Delta}^{(l)} = \sum_{k \in \mathbb{Z}_{\geq 0}} q_{ij}(k, \Delta) \xi_{ij}(l, k, \Delta). \quad (1.9)$$

Figures 1.6c and 1.6d highlight two paths when  $X(\Delta) = 1$ . In Figure 1.6c when  $X(0) \neq 1$ , there are three time pieces in the On state, the first two of which are iid exponentially distributed. However, since  $X(\Delta) = 1$ , the final time piece, suffers from *right-censoring*. This is owed to the fact that  $\{X(t)\}$  will still remain in the On state for an unknown time after the observation has ceased at time  $t = \Delta$ . This is also true for the path shown in Figure 1.6d when  $X(0) = 1$ , whereby exploiting the lack of memory property ensures that although the first three time pieces are iid exponential, the final piece is not. In both cases, the total time spent in the On state cannot be determined using its holding times. Nevertheless, since  $X(\Delta) = 1$  (photo-bleaching has almost surely not occurred), we can consider the holding times in each of the dark states  $0_0, \dots, 0_d$ . Using similar arguments to before, the holding time in each dark state  $0_p$  ( $p = 0, \dots, d$ ) is exponentially distributed (with scale parameter  $\sigma_{0_p}$ ), and the sum of all times spent in state  $0_p$  is characterised by a truncated Erlang( $k_p, \sigma_{0_p}$ ) density, with  $k_p \in \mathbb{Z}_{\geq 0}$  denoting the number of transitions made from state  $0_p$ , i.e. the number of  $0_p \rightarrow 1$  and  $0_p \rightarrow 0_{p+1}$  (for all  $p \neq d$ ) transitions over the interval  $[0, \Delta)$ . Subtracting the sum of all times spent in any dark state (characterised by the truncated sum of  $d + 1$  independent but non-identical Erlang( $k_p, \sigma_{0_p}$ ) densities) from  $\Delta$ , recovers the total time spent in the On state, as depicted in Figure 1.7.

The utilisation of  $d + 1$  Erlang densities in this setting thus invokes conditioning on a multivariate ( $n = d + 1$ ) counting process  $\mathcal{R}_{d+1}^{i1}$ . We can





### Computation of $q_{ij}(k, \Delta)$

In this section, we will consider using Laplace transforms to derive the form of

$$q_{ij}(k, \Delta) = \mathbb{P}_{\theta^{(d)}}(N_{\mathcal{R}^0}(\Delta) = k, X(\Delta) = j | X(0) = i),$$

utilising the labelling set  $\mathcal{R}^0 = \{(0, 1), (0_1, 1), \dots, (0_d, 1)\}$ . Using the infinitesimal definition of a Markov Process, for small  $h > 0$  so that  $o(h)$  is a function such that  $\lim_{h \downarrow 0} \frac{o(h)}{h} = 0$ , we have for any  $t \geq 0$  and  $i \in \bar{\mathcal{S}}_X$

$$\begin{aligned} q_{i0}(k, t+h) &= (1 - \sigma_0 h) q_{i0}(k, t) + \lambda_{10} q_{i1}(k, t) h + o(h) \\ q_{i0_p}(k, t+h) &= (1 - \sigma_{0_p} h) q_{i0_p}(k, t) + \lambda_{0_{p-1}0_p} q_{i0_{p-1}}(k, t) h + o(h) \quad p = 1 \dots d \\ q_{i1}(k, t+h) &= (1 - \sigma_1 h) q_{i1}(k, t) + \sum_{p=0}^d \lambda_{0_p1} q_{i0_p}(k-1, t) h + o(h) \\ q_{i2}(k, t+h) &= q_{i2}(k, t) + \sum_{p=0}^d \mu_{0_p} q_{i0_p}(k, t) h + \mu_1 q_{i1}(k, t) h + o(h). \end{aligned} \quad (1.12)$$

The above equations can be succinctly written in matrix form. Specifically, we consider  $Q_t(k)$  to denote the  $(d+2) \times (d+2)$  matrix such that for any  $t \geq 0$ ,

$$Q_t(k) = \begin{pmatrix} q_{00}(k, t) & q_{00_1}(k, t) & \dots & q_{01}(k, t) \\ q_{0_10}(k, t) & q_{0_10_1}(k, t) & \dots & q_{0_11}(k, t) \\ \dots & \vdots & \ddots & \dots \\ q_{10}(k, t) & q_{10_1}(k, t) & \dots & q_{11}(k, t) \end{pmatrix}. \quad (1.13)$$

Moreover, we define the *end-absorbed state* vector as

$$\bar{\mathbf{Q}}_t(k) = \left( q_{02}(k, t) \quad q_{0_12}(k, t) \quad \dots \quad q_{0_d2}(k, t) \quad q_{12}(k, t) \right)^\top.$$

From (1.12), we have that

$$\bar{\mathbf{Q}}_t(k) = \left( \int_0^t Q_s(k) ds \right) \boldsymbol{\mu}, \quad (1.14)$$

with  $\boldsymbol{\mu} = \left( \mu_0 \ \dots \ \mu_{0_d} \ \mu_1 \right)^\top \equiv (G)_{(1:d+2),d+3}$  as the  $(d+2)$  vector of photo-bleaching rates. Recovering  $\bar{\mathbf{Q}}_t(k)$  will therefore require an expression for  $Q_t(k)$ .

**Lemma 2.** *Let  $\{X(t) : t \in \mathbb{R}_{\geq 0}\}$  be an irreducible Markov Chain over a state space  $\mathcal{S}_X$  with cardinality  $n$  and let its generator  $G$  be such that its  $(i, j)$ th entry is  $\lambda_{ij}$ . Letting  $Q_t(k)$  denote the matrix with  $(i, j)$ th entry  $q_{ij}(k, t) = \mathbb{P}_{\boldsymbol{\theta}^{(d)}}(N_{\mathcal{R}}(t) = k, X(t) = j | X(0) = i)$ , which counts transitions in the labelling set  $\mathcal{R}$  in the time interval  $[0, t)$ . Then the Laplace transformed matrix (for all  $k \in \mathbb{Z}_{\geq 0}$ )  $F_s(k) = \int_0^\infty e^{-st} Q_t(k) dt$  takes the form*

$$F_s(k) = (sI_n - G_{\bar{\mathcal{R}}})^{-1} (G_{\mathcal{R}}(sI_n - G_{\bar{\mathcal{R}}})^{-1})^k. \quad (1.15)$$

Here  $G_{\mathcal{R}} \in \mathbb{R}^{n \times n}$  is the matrix with  $(i, j)$ th entry  $\lambda_{ij} \mathbb{1}_{\mathcal{R}}((i, j))$  and  $G_{\bar{\mathcal{R}}} = G - G_{\mathcal{R}}$ .

*Proof.* See [Minin and Suchard \(2007\)](#). □

A matrix differential Equation for  $Q_t(k)$  and its Laplace transformed matrix  $F_s(k) = \int_0^\infty e^{-st} Q_t(k) dt$  can be obtained by leveraging the result from [Minin and Suchard \(2007\)](#) given in Lemma 2. In particular, the result on  $F_s(k)$  in (1.15) requires  $\{X(t)\}$  to be irreducible on its state space. Although this is not true for our process, it is not difficult to see that  $\{X(t)\}$  is irreducible on  $\bar{\mathcal{S}}_X$ . Therefore, in order to use (1.15), we can use the sub-Markovian generator of  $\{X(t)\}$ :  $G_S \equiv (G)_{(1:d+2), (1:d+2)}$  which is a  $(d+2) \times (d+2)$  matrix gained by deleting the  $(d+3)$ th row and column from the generator  $G$  in (1.1)\*. We define  $G_{S, \mathcal{R}^0}$  to be the  $(d+2) \times (d+2)$  matrix filled with the transition rates *only* in the labelling set of interest  $\mathcal{R}^0$ .  $G_{S, \mathcal{R}^0}$  thus has  $(i, j)$ th entry

$$(G_{S, \mathcal{R}^0})_{i,j} := \begin{cases} \lambda_{0_{i-1}1} & \text{for } i = 1, \dots, d+1, j = d+2 \\ 0 & \text{otherwise.} \end{cases}$$

---

\*To avoid division by zero, in the case where  $\sigma_p = \sigma_q$  for some or all  $p \neq q \in \bar{\mathcal{S}}_X$ , we must replace all such  $\sigma_p$  with  $\sigma_q$  in the diagonal entries of  $G$ .

Letting  $G_{S,\bar{\mathcal{R}}^0} = G_S - G_{S,\mathcal{R}^0}$ , we use Lemma 2 to obtain that for any  $k \in \mathbb{Z}_{\geq 0}$ ,

$$F_s(k) = (sI_{d+2} - G_{S,\bar{\mathcal{R}}^0})^{-1} (G_{S,\mathcal{R}^0}(sI_{d+2} - G_{S,\bar{\mathcal{R}}^0})^{-1})^k. \quad (1.16)$$

Recovering  $Q_t(k)$  now follows from the inverse Laplace transform  $Q_t(k) = \mathcal{L}_s^{-1}[F_s(k)](t)$ . [Minin and Suchard \(2007\)](#) explains that a sufficient condition for obtaining a closed form expression is that  $G_{S,\bar{\mathcal{R}}^0}$  and  $G_S$  commute. Nonetheless, the non-commutative properties of  $G_{S,\bar{\mathcal{R}}^0}$  and  $G_S$ , coupled with the difficulties in attempting to gain  $Q_t(k)$  by brute-force, leave its form to be obtained computationally. Obtaining  $q_{ij}(k, \Delta)$  for all  $k \geq 1$  thus requires evaluating a numerical approximation of  $Q_\Delta(k)$ .

**Remark 4.** When  $k = 0$ ,  $Q_\Delta(0) = e^{G_{S,\bar{\mathcal{R}}^0}\Delta}$ . To compute,  $\bar{\mathbf{Q}}_\Delta(0)$ , we note from [Van Loan \(1978\)](#) that  $\bar{\mathbf{Q}}_\Delta(0) = \left( \int_0^\Delta e^{G_{S,\bar{\mathcal{R}}^0}s} ds \right) \boldsymbol{\mu} = (e^{A\Delta})_{(i_1:i_2),(i_2+1:i_3)} \boldsymbol{\mu}$

with  $i_1 = 2d+5, i_2 = 3(d+2), i_3 = 4(d+2)$ ;  $A = \begin{bmatrix} A_1 & \mathbf{0}_{2(d+2)} \mathbf{0}_{2(d+2)}^\top \\ \mathbf{0}_{2(d+2)} \mathbf{0}_{2(d+2)}^\top & A_2 \end{bmatrix}$ ,

$$A_1 = \begin{bmatrix} -G_{S,\bar{\mathcal{R}}}^\top & I_{d+2} \\ \mathbf{0}_{d+2} \mathbf{0}_{d+2}^\top & -G_{S,\bar{\mathcal{R}}}^\top \end{bmatrix} \text{ and } A_2 = \begin{bmatrix} G_{S,\bar{\mathcal{R}}} & I_{d+2} \\ \mathbf{0}_{d+2} \mathbf{0}_{d+2}^\top & \mathbf{0}_{d+2} \mathbf{0}_{d+2}^\top \end{bmatrix}.$$

### Computation of $\xi_{ij}(l, k, \Delta)$

In this section, we will relate  $\{X(t)\}$  to a renewal sequence. This allows us to characterise the distribution of waiting times for computation of  $\xi_{ij}(l, k, \Delta)$ . This construction, which we define as the photo-switching alternating renewal process (PSARP) is described in Definition 2.

**Definition 2.** Let  $\mathbf{U} = (U_0, U_1, \dots)$  denote the successive lengths of time  $\{X(t)\}$  is in the On state 1 and let  $\mathbf{D} = (D_0, D_1, \dots)$  denote the successive lengths of time that the process is not in state 1 before photo-bleaching. We define the **photo-switching alternating renewal process (PSARP)** as being characterised by the sequence of iid random vectors  $\Gamma$ , where

$$\Gamma = \begin{cases} ((U_0, D_0), (U_1, D_1), \dots) & \text{if } X(0) = 1 \\ ((D_0, U_0), (D_1, U_1), \dots) & \text{if } X(0) \notin \{1, 2\}. \end{cases}$$

A “renewal” can thus be thought of as returns to the On state if  $X(0) = 1$  or to a dark state if  $X(0) \notin \{1, 2\}$ . In particular, for  $n \in \mathbb{Z}_{\geq 0}$ , defining  $R_n = U_n + D_n$ , describes a renewal process with inter-arrival times  $R_0, R_1, \dots$

For full characterisation, we need to consider the distributions of  $\{U_n\}$  and  $\{D_n\}$ . Clearly each  $U_n \stackrel{iid}{\sim} \exp(\sigma_1)$ . To deal with  $\{D_n\}$ , we let  $\{J_n : n \in \mathbb{Z}_{\geq 0}\}$  be a discrete valued stochastic process on the state space  $\mathcal{S}_J = \{0, 1, \dots, d\}$ , which counts the number of jumps between dark states ( $0_p \rightarrow 0_{p+1}$   $p = 0, \dots, d-1$ ) during the  $n$ th renewal. We have that

$$D_0 | J_0 \stackrel{d}{=} \begin{cases} \sum_{s=0}^{J_0} D_0^s & \text{if } X(0) = 1 \\ \sum_{s=p}^{J_0} D_0^s & \text{if } X(0) = 0_p \quad p = 0, \dots, d, \end{cases}$$

and for all  $n \geq 1$  that  $D_n | J_n \stackrel{d}{=} \sum_{s=0}^{J_n} D_n^s$ . Here, each  $D_n^s \stackrel{iid}{\sim} \exp(\sigma_{0_s})$  and  $\stackrel{d}{=}$  denotes equivalence in distribution.

Once  $Y_1$  has been recorded,  $\{X(t)\}$  continues after time  $\Delta$  without observation. Section 1.3.1 explains that while the time taken between the final jump made by  $\{X(t) : t \in [0, \Delta]\}$  (or from 0 if no transitions have occurred) and its next transition is not necessarily exponentially distributed, the *first* time piece  $U_0$  or  $D_0^p$  (as defined in Definition 2) for some  $p = 0, \dots, d$ , remains exponential. If  $X(0) \notin \{1, 2\}$ ,  $X(\Delta) \neq 1$  and  $N_{\mathcal{R}^0}(\Delta) = k$ , then there are exactly  $k$  time pieces in this state ( $U_0, U_1, \dots, U_{k-1}$ ). By construction, these  $k$  pieces are iid exponentially distributed and their *sum* has an Erlang distribution with the shape and rate parametrisation  $\Upsilon(k) = \sum_{i=0}^{k-1} U_i \sim \text{Erlang}(k, \sigma_1)$ . Similarly, if  $X(0) = 1$ ,  $X(\Delta) \neq 1$  and  $N^0(\Delta) = k$ , then there are  $k + 1$  exponential time pieces in the On state ( $U_0, U_1, \dots, U_{k-1}, U_k$ ) and the total time spent in state 1 is governed by  $\Upsilon(k + 1)$ . Since  $Y_1 = 0$  if and only if the total time spent in the On state within the interval  $[0, \Delta]$  is less

than or equal to  $\delta$ , for  $k \in \mathbb{N}$ ,  $i \in \bar{\mathcal{S}}_X$  and  $j \neq 1$  we have

$$\begin{aligned}\xi_{ij}(0, k, \Delta) &= \mathbb{P}_{\theta^{(d)}}(\Upsilon(k + \mathbb{1}_{\{1\}}(i)) \leq \delta | \Upsilon(k + \mathbb{1}_{\{1\}}(i)) \leq \Delta) \\ &= \frac{1 - \sum_{m=0}^{k + \mathbb{1}_{\{1\}}(i) - 1} \frac{(\sigma_1 \delta)^m}{m!} e^{-\sigma_1 \delta}}{1 - \sum_{m=0}^{k + \mathbb{1}_{\{1\}}(i) - 1} \frac{(\sigma_1 \Delta)^m}{m!} e^{-\sigma_1 \Delta}}\end{aligned}\quad (1.17)$$

$$\xi_{1j}(0, 0, \Delta) = \frac{1 - e^{-\sigma_1 \delta}}{1 - e^{-\sigma_1 \Delta}} \quad (1.18)$$

$$\xi_{ij}(1, k, \Delta) = 1 - \xi_{ij}(0, k, \Delta). \quad (1.19)$$

For computational purposes<sup>†</sup>, we form (analogously to  $Q_\Delta(k)$ ) the  $(d+2) \times (d+1)$  matrix  $\Xi_\Delta^l(k)$  with elements gained from (1.17)-(1.19). For  $i, j \in \bar{\mathcal{S}}_X, j \neq 1$  and  $l \in \mathcal{S}_Y$  we define

$$\Xi_\Delta^l(k) = \begin{pmatrix} \xi_{00}(l, k, \Delta) & \xi_{00_1}(l, k, \Delta) & \dots & \xi_{00_d}(l, k, \Delta) \\ \xi_{01_0}(l, k, \Delta) & \xi_{01_0_1}(l, k, \Delta) & \dots & \xi_{01_0_d}(l, k, \Delta) \\ \dots & \vdots & \ddots & \dots \\ \xi_{10}(l, k, \Delta) & \xi_{10_1}(l, k, \Delta) & \dots & \xi_{10_d}(l, k, \Delta) \end{pmatrix}, \quad (1.20)$$

and  $\bar{\Xi}_\Delta^l(k)$  to be the  $(d+2)$  end state vector

$$\bar{\Xi}_\Delta^l(k) = \left( \xi_{02}(l, k, \Delta) \quad \dots \quad \xi_{12}(l, k, \Delta) \right)^\top. \quad (1.21)$$

### 1.3.3 Case $j = 1$

Using the same mathematical tools as in the case  $j \neq 1$  (Section 1.3.2), we derive the forms of  $q_{i1}(\mathbf{k}, \Delta)$  using Laplace transforms and  $\xi_{i1}(l, \mathbf{k}, \Delta)$  using state holding time distributions. Before doing so, we will describe an alternative method needed to compute  $b_{i1, \Delta}^{(l)}$ .

---

<sup>†</sup>The matrix representations  $Q_\Delta(k)$  and  $\Xi_\Delta^l(k)$  are used in Algorithm 2 (see Section 1.B of Appendix 1) for computing transmission matrices in the form (1.4).

### Equivalent definition

For computational feasibility, we would like to reduce the number of counts that can be made in the labelling set. Here, we will therefore seek an alternative method for computing the sum in (1.11).

Using the form of  $\mathbf{k}$  in (1.10), we let  $k \in \mathbb{Z}_{>0}$  be equal to  $\mathbf{k}^\top \mathbf{1}_{d+1}$ : the total number of transitions made from the dark states within the interval  $[0, \Delta)$ . When  $X(0) = i$ , we consider the set of feasible transitions made by  $\{X(t)\}$ . In particular, on the state space  $\bar{\mathcal{S}}_X$ , an On-Dark cycle of the form  $1 \rightarrow 0 \rightarrow 0_1 \rightarrow \dots \rightarrow 0_p$  for some  $p = 0, \dots, d$  is *regenerated* every time a dark state transitions back to the On state. When  $i = 0_p$  for some  $p$ , this cyclic structure implies that  $k_p \geq k_{p+1} \geq \dots \geq k_d$  since any subsequent dark state to  $0_p$  can make at most the number of transitions made by its preceding state. Using similar arguments, we also have that  $k_0 \geq k_1 \geq \dots \geq k_{p-1} \geq k_p - 1$ . When  $i = 0_p$ , the set of feasible transitions, which we denote as  $\mathcal{C}_k^i$ , is of the form

$$\mathcal{C}_k^i = \{\mathbf{k} \in \mathbb{Z}_{\geq 0}^{d+1} : \mathbf{k}^\top \mathbf{1}_{d+1} = k, k_p > 0, k_0 \geq \dots \geq k_{p-1} \geq k_p - 1 \geq \dots \geq k_d - 1\},$$

and when  $i = 1$ , it is easy to see that  $\mathcal{C}_k^1 \equiv \mathcal{C}_k^0$ .

When  $X(\Delta) = 1$ , we thus endeavour to compute the following form of the transmission probabilities

$$b_{i1, \Delta}^{(l)} = \sum_{k=0}^{\infty} \sum_{\mathbf{k} \in \mathcal{C}_k^i} q_{i1}(\mathbf{k}, \Delta) \xi_{i1}(l, \mathbf{k}, \Delta). \quad (1.22)$$

### Computation of $q_{i1}(\mathbf{k}, \Delta)$

In this section, in an analogous fashion to the computation of  $q_{ij}(k, \Delta)$ , we will invoke the use of Laplace transforms to derive  $q_{i1}(\mathbf{k}, \Delta)$ , whereby for  $\mathbf{k} \in \mathbb{Z}_{\geq 0}^{d+1}$

$$q_{i1}(\mathbf{k}, \Delta) = \mathbb{P}_{\boldsymbol{\theta}^{(d)}}(\mathbf{N}_{\mathcal{R}_{d+1}^1}(\Delta) = \mathbf{k}, X(\Delta) = 1 | X(0) = i),$$

and transitions are counted in  $\mathcal{R}_{d+1}^1 = \{\mathcal{R}_0^1, \mathcal{R}_1^1, \dots, \mathcal{R}_d^1\}$  where for  $p = 0, \dots, d-1$ ,  $\mathcal{R}_p^1 = \{(0_p, 1), (0_p, 0_{p+1})\}$  and  $\mathcal{R}_d^1 = \{(0_d, 1)\}$ .

Using the infinitesimal definition of a Markov process, we have for any  $t \geq 0$  and  $i \in \bar{S}_X$

$$\begin{aligned} q_{i0}(\mathbf{k}, t+h) &= (1 - \sigma_0 h) q_{i0}(\mathbf{k}, t) + \lambda_{10} q_{i1}(\mathbf{k}, t) h + o(h) \\ q_{i0_p}(\mathbf{k}, t+h) &= (1 - \sigma_{0_p} h) q_{i0_p}(\mathbf{k}, t) + \lambda_{0_{p-1}0_p} q_{i0_{p-1}}(\mathbf{k} - \mathbf{e}_{d+1}^p, t) h + o(h) \quad p = 1 \dots d. \\ q_{i1}(\mathbf{k}, t+h) &= (1 - \sigma_1 h) q_{i1}(\mathbf{k}, t) + \sum_{p=0}^d \lambda_{0_p1} q_{i0_p}(\mathbf{k} - \mathbf{e}_{d+1}^{p+1}, t) h + o(h), \end{aligned}$$

whereby  $\mathbf{e}_n^p$  denotes the  $p$ th canonical (standard) basis vector of  $\mathbb{R}^n$ . For  $s > 0$ , taking the limit  $h \downarrow 0$  and taking Laplace transforms reveals that

$$(s + \sigma_0) f_{i0}(\mathbf{k}, s) = \lambda_{10} f_{i1}(\mathbf{k}, s) \tag{1.23}$$

$$(s + \sigma_{0_p}) f_{i0_p}(\mathbf{k}, s) = \lambda_{0_{p-1}0_p} f_{i0_{p-1}}(\mathbf{k} - \mathbf{e}_{d+1}^p, s) \quad p = 1 \dots d, \tag{1.24}$$

$$(s + \sigma_1) f_{i1}(\mathbf{k}, s) = \sum_{p=0}^d \lambda_{0_p1} f_{i0_p}(\mathbf{k} - \mathbf{e}_{d+1}^{p+1}, s), \tag{1.25}$$

with  $\mathcal{L}_t[q_{ij}(\mathbf{k}, t)](s) =: f_{ij}(\mathbf{k}, s) = \int_0^\infty e^{-st} q_{ij}(\mathbf{k}, t) dt$ .

For  $\mathbf{k} \in \mathbb{Z}_{\geq 0}^{d+1}$ , this yields the recursion<sup>‡</sup>

$$f_{i1}(\mathbf{k}, s) = \frac{\lambda_{10}}{s + \sigma_1} \sum_{p=0}^d \frac{\lambda_{0_p1} \prod_{q=0}^{p-1} \lambda_{0_q0_{q+1}}}{\prod_{q=0}^p (s + \sigma_{0_q})} f_{i1} \left( \mathbf{k} - \sum_{r=0}^p \mathbf{e}_{d+1}^{r+1}, s \right), \tag{1.26}$$

where  $f_{i1}(\mathbf{k}, s) = 0$  for any  $\mathbf{k} \notin C_k^i$ .

**Remark 5.** Using Equations (1.23) - (1.25), we can obtain the initialisations:  $f_{i1}(\mathbf{0}_{d+1}, s) = \frac{\mathbb{1}_{\{1\}}(i)}{s + \sigma_1}$ ,  $f_{0_p1}(\mathbf{e}_{d+1}^{p+1}, s) = \frac{\lambda_{0_p1}}{(s + \sigma_{0_p})(s + \sigma_1)}$  for  $p = 1, \dots, d$  and  $f_{11}(\mathbf{e}_{d+1}^1, s) = \frac{\lambda_{10}\lambda_{01}}{(s + \sigma_0)(s + \sigma_1)^2}$ .

The inverse Laplace transform of (1.26), and its initialisations to recover  $q_{i1}(\mathbf{k}, t)$  and thus  $q_{i1}(\mathbf{k}, \Delta)$  are left as a computational exercise.

<sup>‡</sup>To avoid division by zero, when  $\sigma_p = \sigma_q$  for any  $p \neq q$ ,  $\sigma_p$  must be replaced with  $\sigma_q$  in the Laplace transforms.

**Remark 6.** When  $k = 0$ ,  $\mathbf{k}$  can only be the vector  $\mathbf{0}_{d+1}$  and hence  $q_{i1}(\mathbf{0}_{d+1}, \Delta) = 0$  for all  $i \neq 1$ ; it is however easily seen that  $q_{11}(\mathbf{0}_{d+1}, \Delta) = e^{-\sigma_1 \Delta}$ .

### Computation of $\xi_{i1}(l, \mathbf{k}, \Delta)$

Following on from the analysis presented in sections 1.3.1 and 1.3.2, we will in this section endeavour to compute  $\xi_{i1}(l, \mathbf{k}, \Delta)$ . In particular, by using the exponential time pieces in the dark states  $0 \dots, 0_d$ , we can recover the total time spent in state 1 by subtracting the total time spent in the dark states from  $\Delta$ .

If  $X(0) = i$  for some  $i \in \bar{\mathcal{S}}_X$ ,  $X(\Delta) = 1$  and  $\mathbf{N}_{\mathcal{R}_{d+1}^{i1}}(\Delta) = \mathbf{k}$  as defined in (1.10), we study the number of transitions  $k_0, \dots, k_d$  from each dark state  $0_0, \dots, 0_d$ . There are exactly  $k_p$  exponential time pieces in each dark state  $0_p$  for  $p = 0, \dots, d$  with a total time  $v_p^i(k_p)$  characterised by an Erlang distribution. For any  $p$ , we therefore have (using the PSARP construction in Definition 2) that  $v_p^i(k_p) = \sum_{s=0}^{k_p} D_s^p \sim \text{Erlang}(k_p, \sigma_{0_p})$ .

By defining the event  $A_i := \{Y_1 = 0 | \mathbf{N}_{\mathcal{R}_{d+1}^{i1}}(\Delta) = \mathbf{k}, X(0) = i, X(\Delta) = 1\}$ , we equivalently have

$$A_i = \left\{ \sum_{p=0}^d v_p^i(k_p) \geq \Delta - \delta \mid \sum_{p=0}^d v_p^i(k_p) \leq \Delta \right\},$$

so that  $\xi_{i1}(0, \mathbf{k}, \Delta) = \mathbb{P}_{\boldsymbol{\theta}^{(d)}}(A_i)$ .

By further defining  $\boldsymbol{\sigma} = \left( \sigma_0 \quad \sigma_{0_1} \quad \dots \quad \sigma_{0_d} \right)^\top$ , we let  $\Phi(\mathbf{k}, \boldsymbol{\sigma})$  denote the sum of  $d+1$  independent Erlang distributions, each with shape parameter  $k_p$  and rate parameter  $\sigma_{0_p}$  for  $p = 0, \dots, d$ . For each  $i$ , we have  $\sum_{p=0}^d v_p^i(k_p) \sim \Phi(\mathbf{k}, \boldsymbol{\sigma})$ . If  $F_\Phi(\phi | \mathbf{k}, \boldsymbol{\sigma})$  denotes its cumulative distribution function<sup>§</sup>, we have

<sup>§</sup> $F_\Phi(\cdot | \cdot, \cdot)$  can be computed in a recursive fashion using the algorithm presented in Moschopoulos (1985).



for  $k \in \mathbb{Z}_{>0}$ ,  $i \in \bar{\mathcal{S}}_X$  and  $\mathbf{k} \in \mathcal{C}_k^i$

$$\xi_{i1}(1, \mathbf{k}, \Delta) = \frac{F_\Phi(\Delta - \delta|\mathbf{k}, \boldsymbol{\sigma})}{F_\Phi(\Delta|\mathbf{k}, \boldsymbol{\sigma})} \quad (1.27)$$

$$\xi_{i1}(0, \mathbf{k}, \Delta) = 1 - \xi_{i1}(1, \mathbf{k}, \Delta) \quad (1.28)$$

$$\xi_{11}(1, \mathbf{0}, \Delta) = 1. \quad (1.29)$$

### 1.3.4 Transmission probabilities

In this section, we will use all derived forms of  $q_{ij}(k, \Delta)$ ,  $q_{i1}(\mathbf{k}, \Delta)$ ,  $\xi_{ij}(l, k, \Delta)$  and  $\xi_{i1}(l, \mathbf{k}, \Delta)$  to obtain  $b_{ij,\Delta}^{(l)}$ . We will describe how these probabilities in their matrix representations can be used to gain transmissions in the form of (1.4).

When  $i \in \bar{\mathcal{S}}_X$ ,  $j \neq 1$  and  $l \in \mathcal{S}_Y$  we can compute the inverse Laplace transform of (1.16) to obtain  $Q_\Delta^0(k)$ , and therefore  $\bar{Q}_\Delta^0(k)$  from (1.14). Furthermore, using (1.17)-(1.19) as entries in matrices  $\Xi_\Delta^l(k)$  and  $\bar{\Xi}_\Delta^l(k)$  (for  $l \in \mathcal{S}_Y$ ) as defined in (1.20) and (1.21) enables the computation of transmission probabilities (1.9). Moreover, we use (1.27)-(1.29) with inverse Laplace transforms  $q_{i1}(\mathbf{k}, \Delta)$  from (1.26) in the computation of (1.11).

For full matrix representation, we define for  $k \in \mathbb{Z}_{\geq 0}$

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

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

with  $\odot$  denoting the Hadamard (element wise) product and where  $\mathbf{v}_\Delta^l(\mathbf{k})$  for  $l \in \mathcal{S}_Y$  are the two  $(d+2)$  vectors such that

$$\mathbf{v}_\Delta^l(\mathbf{k}) = \left( \sum_{\mathbf{k} \in \mathcal{C}_k^0} q_{01}(\mathbf{k}, \Delta) \xi_{01}(l, \mathbf{k}, \Delta) \dots \sum_{\mathbf{k} \in \mathcal{C}_k^{0,d}} q_{0,d+1}(\mathbf{k}, \Delta) \xi_{0,d+1}(l, \mathbf{k}, \Delta) \sum_{\mathbf{k} \in \mathcal{C}_k^0} q_{11}(\mathbf{k}, \Delta) \xi_{01}(l, \mathbf{k}, \Delta) \right)^\top.$$

Then considering the transmission matrix representation in (1.4), we have

that each

$$B_{\Delta}^{(l)} = \sum_{k=0}^{\infty} B_{\Delta}^{(l)}(k).$$

In practice, the above sum, for small rates and  $\Delta$ , will converge extremely quickly. At least in this practical application, computations for  $k \geq 4$  are unlikely to be needed, enabling the required computational complexity to be  $O(4(d+3)^2)$ . Nevertheless, in the case  $d = 0$ , in Section 1.C of Appendix 1, we provide the forms of the transmission probabilities in closed form. Here, it is shown that they can be deduced without the need of computational approximations to the corresponding inverse Laplace transforms. An algorithm detailing all computational steps to evaluate these matrices suitable for any  $d \in \mathbb{Z}_{\geq 0}$  (any number of multiple dark states) is presented in Algorithm 2 found in Section 1.B of Appendix 1.

## 1.4 Inference

In this section, we derive the log-likelihood function of imaging observations using the PSHMM constructed in Section 1.2.3 and discuss a computational implementation that can avoid numerical underflow. Under any photo-switching model, we then use this log-likelihood to discuss how the static parameters of the PSHMM can be inferred using maximum likelihood estimation. Furthermore, we define the Bayesian Information Criterion (BIC) as a model selection tool to pick the most likely number of multiple dark states  $d$ . When only one experiment is conducted, we additionally describe a bootstrapping method to gain approximate 95% confidence intervals of the resulting parameter estimates.

### 1.4.1 Likelihood function

We now provide an algorithm for estimating the unknown parameters  $\theta_{\omega}^{(d)}$  of the PSHMM, which utilises a suitable adaptation of the forward-backward dynamic programming algorithm (Rabiner, 1989), making use of the transmission matrices in (1.4).

Let  $\mathbf{y} = (y_1 \ y_2 \ \dots \ y_{N_F})^\top$  be the sequence of observations across  $N_F$  frames for a single photo-switching fluorophore. We define the forward-backward probabilities as

$$\begin{aligned}\alpha_{n,i} &= \mathbb{P}_{\boldsymbol{\theta}_\omega^{(d)}}(Y_1 = y_1, \dots, Y_n = y_n, X(n\Delta) = i) & n = 1, \dots, N_F, \\ \beta_{n,i} &= \mathbb{P}_{\boldsymbol{\theta}_\omega^{(d)}}(Y_{n+1} = y_{n+1}, \dots, Y_{N_F} = y_{N_F} | X(n\Delta) = i) & n = 0, \dots, N_F - 1.\end{aligned}$$

For each such  $n$ , we define the forward-backward vectors as

$$\begin{aligned}\boldsymbol{\alpha}_n &= (\alpha_{n,0} \ \dots \ \alpha_{n,0_d} \ \alpha_{n,1} \ \alpha_{n,2})^\top \\ \boldsymbol{\beta}_n &= (\beta_{n,0} \ \dots \ \beta_{n,0_d} \ \beta_{n,1} \ \beta_{n,2})^\top.\end{aligned}$$

Now observe that for any  $n > 1$  and  $j \in \mathcal{S}_X$

$$\begin{aligned}\alpha_{n,i} &= \sum_{j \in \mathcal{S}_X} \mathbb{P}_{\boldsymbol{\theta}_\omega^{(d)}}(Y_1 = y_1, \dots, Y_n = y_n, X(n\Delta) = i, X((n-1)\Delta) = j) \\ &= \sum_{j \in \mathcal{S}_X} \mathbb{P}_{\boldsymbol{\theta}_\omega^{(d)}}(Y_1 = y_1, \dots, Y_{n-1} = y_{n-1}, X((n-1)\Delta) = j) \times \\ &\quad \mathbb{P}_{\boldsymbol{\theta}_\omega^{(d)}}(Y_n = y_n, X(n\Delta) = i | X((n-1)\Delta) = j, Y_1 = y_1, \dots, Y_{n-1} = y_{n-1}) \\ &= \sum_{j \in \mathcal{S}_X} \alpha_{n-1,j} \mathbb{P}_{\boldsymbol{\theta}_\omega^{(d)}}(Y_n = y_n, X(n\Delta) = i | X((n-1)\Delta) = j) \\ &\equiv \sum_{j \in \mathcal{S}_X} \alpha_{n-1,j} \mathbf{b}_{ji,\Delta}^{*(y_n)},\end{aligned}$$

which implies that  $\boldsymbol{\alpha}_n^\top = \boldsymbol{\alpha}_{n-1}^\top B_\Delta^{*(y_n)}$  for  $n = 2, \dots, N_F$ . Similar arguments show that  $\boldsymbol{\alpha}_1^\top = \boldsymbol{\nu}_X^\top B_\Delta^{*(y_1)}$  when  $n = 1$ . Additionally, we have for  $n < N_F$

and  $j \in \mathcal{S}_X$

$$\begin{aligned}
\beta_{n-1,i} &= \sum_{j \in \mathcal{S}_X} \mathbb{P}_{\boldsymbol{\theta}_\omega^{(d)}}(Y_n = y_n, \dots, Y_{N_F} = y_{N_F}, X(n\Delta) = j | X((n-1)\Delta) = i) \\
&= \sum_{j \in \mathcal{S}_X} \mathbb{P}_{\boldsymbol{\theta}_\omega^{(d)}}(Y_n = y_n, X(n\Delta) = j | X((n-1)\Delta) = i) \times \\
&\quad \mathbb{P}_{\boldsymbol{\theta}_\omega^{(d)}}(Y_{n+1} = y_{n+1}, \dots, Y_{N_F} = y_{N_F} | X(n\Delta) = j) \\
&= \sum_{j \in \mathcal{S}_X} \beta_{n,j} \mathbb{P}_{\boldsymbol{\theta}_\omega^{(d)}}(Y_n = y_n, X(n\Delta) = j | X((n-1)\Delta) = i) \\
&\equiv \sum_{j \in \mathcal{S}_X} \beta_{n,j} b_{ij,\Delta}^{*(y_n)},
\end{aligned}$$

which implies that  $\boldsymbol{\beta}_{n-1} = B_\Delta^{*(y_n)} \boldsymbol{\beta}_n$  for  $n = 1, \dots, N_F$ , with  $\boldsymbol{\beta}_{N_F} = \mathbf{1}_{d+3}$ , the  $(d+3) \times 1$  vector of ones. This yields the following recursion formula

$$\begin{aligned}
\boldsymbol{\alpha}_0 &= \boldsymbol{\nu}_X & \boldsymbol{\alpha}_n^\top &= \boldsymbol{\alpha}_{n-1}^\top B_\Delta^{*(y_n)} & n &= 1, \dots, N_F, \\
\boldsymbol{\beta}_{N_F} &= \mathbf{1}_{d+3} & \boldsymbol{\beta}_n &= B_\Delta^{*(y_{n+1})} \boldsymbol{\beta}_{n+1} & n &= 0, \dots, N_F - 1.
\end{aligned} \tag{1.30}$$

It now follows that the likelihood of observation vector  $\mathbf{y}$  given parameter vector  $\boldsymbol{\theta}^{(d)}$  is  $p_{\boldsymbol{\theta}_\omega^{(d)}}(\mathbf{y}) =: L_{\boldsymbol{\theta}_\omega^{(d)}}(\mathbf{y}) = \boldsymbol{\alpha}_n^\top \boldsymbol{\beta}_n$  for all  $n = 0, \dots, N_F$ . In particular, we have  $L_{\boldsymbol{\theta}_\omega^{(d)}}(\mathbf{y}) = \boldsymbol{\alpha}_{N_F}^\top \mathbf{1}_{d+3}$ , which can be computed using the transmission matrices together with recursive computation for  $\boldsymbol{\alpha}_n^\top$  as indicated in (1.30). In the situation where we have  $M \geq 1$  independent photo-switching fluorophores, the log-likelihood is given by

$$\begin{aligned}
\log p_{\boldsymbol{\theta}_\omega^{(d)}}(\mathcal{Y}) &=: \ell_{\boldsymbol{\theta}_\omega^{(d)}}(\mathcal{Y}) = \sum_{m=1}^M \log(\boldsymbol{\alpha}_{N_F,m}^\top \mathbf{1}_{d+3}) \\
&= \sum_{m=1}^M \log\left(\boldsymbol{\nu}_X^\top B_\Delta^{*(y_{1,m})} B_\Delta^{*(y_{2,m})} \dots B_\Delta^{*(y_{N_F,m})} \mathbf{1}_{d+3}\right),
\end{aligned} \tag{1.31}$$

where  $\mathcal{Y} = (\mathbf{y}^1 \ \mathbf{y}^2 \ \dots \ \mathbf{y}^M)$ ,  $y_{n,m}$  is the  $n$ th observation and  $\boldsymbol{\alpha}_{N_F,m}$  is the forward probability vector for molecule  $m = 1, \dots, M$ .

## Avoiding numerical underflow

Computing (1.31) over a large number of frames  $N_F$  is likely to lead to numerical underflow. In order to circumvent this issue, we can, for each molecule  $m = 1, \dots, M$  and for  $n = 2, \dots, N_F$  define

$$\begin{aligned}\bar{\alpha}_{1,m}^\top &= \boldsymbol{\nu}_X^\top B_\Delta^{*(y_{1,m})} / c_{1,m} & c_{1,m} &= \boldsymbol{\nu}_X^\top B_\Delta^{*(y_{1,m})} \mathbf{1}_{d+3} \\ \bar{\alpha}_{n,m}^\top &= \bar{\alpha}_{n-1,m}^\top B_\Delta^{*(y_{n,m})} / c_{n,m} & c_{n,m} &= \bar{\alpha}_{n-1,m}^\top B_\Delta^{*(y_{n,m})} \mathbf{1}_{d+3},\end{aligned}$$

so that  $\prod_{m=1}^M \prod_{n=1}^{N_F} c_{n,m}$

$$\begin{aligned}&= \prod_{m=1}^M \boldsymbol{\nu}_X^\top B_\Delta^{*(y_{1,m})} \mathbf{1}_{d+3} \frac{\boldsymbol{\nu}_X^\top B_\Delta^{*(y_{1,m})} B_\Delta^{*(y_{2,m})} \mathbf{1}_{d+3}}{\boldsymbol{\nu}_X^\top B_\Delta^{*(y_{1,m})} \mathbf{1}_{d+3}} \dots \frac{\boldsymbol{\nu}_X^\top B_\Delta^{*(y_{1,m})} \dots B_\Delta^{*(y_{N_F,m})} \mathbf{1}_{d+3}}{\boldsymbol{\nu}_X^\top B_\Delta^{*(y_{1,m})} \dots B_\Delta^{*(y_{N_F-1,m})} \mathbf{1}_{d+3}} \\ &= p_{\boldsymbol{\theta}_\omega^{(d)}}(\mathcal{Y}).\end{aligned}$$

Normalising the above forward-backward vectors at each time step therefore allows us to sequentially recover the more numerically stable  $\ell_{\boldsymbol{\theta}_\omega^{(d)}}(\mathcal{Y}) = \sum_{m=1}^M \sum_{n=1}^{N_F} \log(c_{n,m})$ . This function has a computational cost of  $O(N_F(d+3)^3)$ , and therefore increases linearly in time for a fixed photo-switching model. An algorithm to compute the log-likelihood of a dataset  $\mathcal{Y}$  for given  $\boldsymbol{\theta}_\omega^{(d)}$  is presented in Algorithm 1.

Inference of  $\boldsymbol{\theta}_\omega^{(d)}$  is now possible through standard maximum likelihood estimation. Although analytic differentiation of (1.31) is not easily feasible, maximising this function with respect to  $\boldsymbol{\theta}_\omega^{(d)}$  can be done either through numerically approximating derivatives (e.g. via quasi-Newton methods) or by using derivative-free optimisation, for example with the Nelder-Mead algorithm. A discussion on implementing the PSHMM algorithm, with particular attention given to the convergence of transmission matrices and choosing a starting point for optimisation is detailed in Section 1.D of Appendix 1. A discussion on the statistical properties of the PSHMM estimator, with special reference to identifiability and multi-modality of photo-switching parameters, can be found in the next chapter.

### 1.4.2 Model selection

To determine the *unknown* number of multiple  $d$  dark states, a model selection criterion can be utilised. Commonly used criteria include the Akaike information criterion (AIC) (Akaike, 1974), which measures a model’s goodness of fit, and the Bayesian information criterion (BIC) (Schwarz, 1978), which measures a model’s posterior probability. For this problem, we choose the BIC to determine the most likely model given data  $\mathcal{Y}$ . Although similar to the AIC, we found that the BIC penalises model complexity more heavily and therefore offers greater protection against over-fitting when the number of data-points is large, as is the case when dealing with experimental imaging data.

Specifically under model  $\mathcal{M}_A^d$ , the BIC uses an approximation to the posterior model probability  $\mathbb{P}(\mathcal{M}_A^d|\mathcal{Y}) \propto \mathbb{P}(\mathcal{Y}|\mathcal{M}_A^d)\mathbb{P}(\mathcal{M}_A^d)$ . When all candidate models of this form are (apriori) equally likely, maximising the posterior model probability given data  $\mathcal{Y}$  is equivalent to maximising the marginal likelihood

$$\mathbb{P}(\mathcal{Y}|\mathcal{M}_A^d) = \int_{\Theta_\omega^{(d)}} \ell_{\theta_\omega^{(d)}}(\mathcal{Y})\pi(\theta_\omega^{(d)}), \quad (1.32)$$

where  $\pi(\theta_\omega^{(d)})$  denotes the (apriori) distribution of the photo-switching parameter vector  $\theta_\omega^{(d)}$  under model  $\mathcal{M}_A^d$ . Utilising the Weak Law of Large Numbers (WLLN) and invoking a flat prior on  $\theta_\omega^{(d)}$ , i.e. that  $\pi(\theta_\omega^{(d)}) \propto 1$  on Laplace’s approximation to (1.32) yields (Schwarz, 1978)

$$-\log \mathbb{P}(\mathcal{Y}|\mathcal{M}_A^d) \approx (3(d+2) + |A|) \log(MN_F) - 2\ell_{\hat{\theta}_\omega^{(d)}}(\mathcal{Y}), \quad (1.33)$$

where  $\ell_{\hat{\theta}_\omega^{(d)}}(\mathcal{Y})$  denotes the maximised log-likelihood using the maximum likelihood estimates  $\hat{\theta}_\omega^{(d)}$ . The BIC defined in this context is exactly the right hand side of (1.33) and can be computed among all suitable models, with the most preferred model  $\mathcal{M}_{\hat{A}}^{\hat{d}}$  being chosen as that with the smallest BIC value.

### 1.4.3 Bootstrapping

When only one experiment is conducted to produce an  $N_F \times M$  dataset  $\mathcal{Y}$ , a single prediction  $\hat{\theta}_\omega^{(\hat{d})}$  is obtained. In this circumstance, a bootstrapping scheme can be used to gain approximate confidence intervals for each component of  $\theta_\omega^{(\hat{d})}$ .

In the same manner as is presented in Efron and Tibshirani (1993), we generate  $R$  (typically large) bootstrap datasets  $\mathcal{Y}^{*1}, \mathcal{Y}^{*2}, \dots, \mathcal{Y}^{*R}$  each consisting of re-sampled (with replacement) columns of  $\mathcal{Y}$ . From each dataset, we acquire bootstrap replicated parameter estimates  $\hat{\theta}_\omega^{(\hat{d})*1}, \hat{\theta}_\omega^{(\hat{d})*2}, \dots, \hat{\theta}_\omega^{(\hat{d})*R}$  using the same PSHMM maximum likelihood procedure used to obtain  $\hat{\theta}_\omega^{(\hat{d})}$ . For  $0.5 < p < 1$ , letting  $(\hat{\theta}_\omega^{(\hat{d})*})_{i,(p)}$  and  $(\hat{\theta}_\omega^{(\hat{d})*})_{i,(1-p)}$  be the  $100 \cdot p$ th and  $100 \cdot (1 - p)$ th empirical percentiles of the  $i$ th component of  $\theta_\omega^{(\hat{d})}$  obtained from  $(\hat{\theta}_\omega^{(\hat{d})*1})_i, (\hat{\theta}_\omega^{(\hat{d})*2})_i, \dots, (\hat{\theta}_\omega^{(\hat{d})*R})_i$ , a percentile bootstrap interval of length  $1 - 2p$  is given by (see Efron and Tibshirani, 1993)

$$[(\hat{\theta}_\omega^{(\hat{d})})_{i,\%lo}, (\hat{\theta}_\omega^{(\hat{d})})_{i,\%up}] \approx [(\hat{\theta}_\omega^{(\hat{d})*})_{i,(p)}, (\hat{\theta}_\omega^{(\hat{d})*})_{i,(1-p)}].$$

## 1.5 Simulations

In this section, we provide the results of simulation studies that have been conducted to assess and analyse the performance of the PSHMM method as detailed in Section 1.2.3. To make the results applicable, we restrict ourselves to realistic parameter values that typically occur in an experimental setting<sup>¶</sup>. We begin by presenting rate estimates from the PSHMM estimator, and compare these results to the exponential fitting estimator Lin et al. (2015). Secondly, we apply the Bayesian Information Criterion to a range of different datasets in order to assess its ability in selecting the correct number of multiple dark states  $d$ .

---

<sup>¶</sup>Further simulation studies addressing the statistical properties of the PSHMM estimator over a broader range of parameter values are presented in the next chapter.

---

**Algorithm 1** Compute log-likelihood  $\ell_{\theta_\omega^{(d)}}(\mathcal{Y})$

---

```

function PSHMM_LOG_LIKELIHOOD( $\mathcal{Y}, \theta_\omega^{(d)}, M, N_F, \Delta$ )
     $B_\Delta^{*(0)}, B_\Delta^{*(1)} \leftarrow \text{COMPUTE\_TRANSMISSIONS}(\theta_\omega^{(d)}, \Delta)$  ▷ Using
    Algorithm 2
     $l \leftarrow 0$ 
    for  $m = 1$  to  $M$  do
         $\alpha \leftarrow \mathbf{0}_{N_F} \mathbf{0}_{d+3}^\top$ 
         $\alpha[1, :] \leftarrow \nu_X^\top B_\Delta^{*(y_{1,m})}$ 
         $C \leftarrow \alpha[1, :] \mathbf{1}_{d+3}$ 
         $\alpha[1, :] \leftarrow \alpha[1, :] / C$ 
         $l \leftarrow l + \log(C)$ 
        for  $n = 2$  to  $N_F$  do
             $\alpha[n, :] \leftarrow \alpha[n-1, :] B_\Delta^{*(y_{n,m})}$ 
             $C \leftarrow \alpha[n, :] \mathbf{1}_{d+3}$ 
             $\alpha[n, :] \leftarrow \alpha[n, :] / C$ 
             $l \leftarrow l + \log(C)$ 
    return  $l$  ▷ Log-likelihood  $\ell_{\theta_\omega^{(d)}}(\mathcal{Y})$ 

```

---

### 1.5.1 Estimating rate parameters

Firstly, to test the performance of parameter estimation against the exponential fitting method of Lin et al. (2015), synthetic imaging data of photo-switching fluorophores was simulated. We begin our focus on the model  $\mathcal{M}_{\{1\}}^0$ , since for many practical applications the life-times of further dark (in particular the triplet ( $T_1$ )) states is short relative to  $\Delta$ . As such, this dark state has been considered as part of the meta-stable On state (Ha and Tinnefeld, 2012, Vogelsang et al., 2010). Since the predominant pathway to photo-bleaching is via the triplet state, a simplified model can be used in which the photo-bleaching state 2 is only accessible from state 1.

Details on the image simulation method and how the discretised state sequences were extracted can be found in Section 1.F of Appendix 1. Global parameter values are also noted. The extracted state sequences were analysed using an implementation of Algorithms 1 and 2. The resulting parameter estimates were compared to estimates derived from the exponential fitting method, which was extended in this study to allow for the calculation of

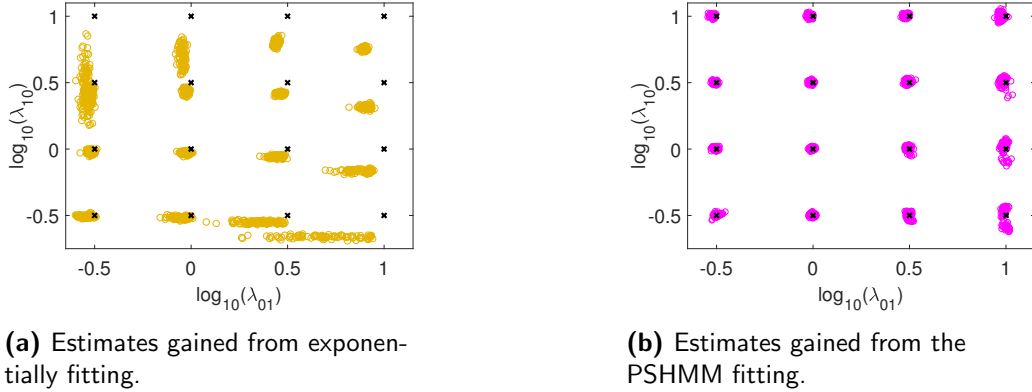


photo-bleaching rates (see Section 1.E of Appendix 1 for further details).

Table 1.3 (see Appendix 1.G) shows estimated parameter statistics over 16 image simulation studies with 100 replicates (datasets) per study under the  $\mathcal{M}_{\{1\}}^0$  model. Rate parameters  $\boldsymbol{\theta}^{(0)}$ , were chosen to cover a range of observed behaviours of organic fluorophores and fluorescent proteins (Dempsey et al., 2011) with  $M = 100$  fluorophores per study. The number of frames  $N_F$  in each study was adjusted to standardise the average number of transitions predicted from  $\boldsymbol{\theta}^{(0)}$ . Scatter plots of these rate estimates are presented in Figure 1.8. It is evident that the PSHMM yields estimates with much lower bias and root mean squared errors (RMSE) when compared to the exponential fitting method, although they have a tendency to increase as transition and photo-bleaching rates are increased. The reported empirical (2.5, 97.5) percentile intervals contain the true parameter values across all studies for the PSHMM method and further highlights the bias in estimates obtained from exponential fitting.

For experimenters, the effect of imaging parameters on the performance of the estimators is of particular interest and importance. Further simulation studies carried out under model  $\mathcal{M}_{\{1\}}^0$  highlight the consistency in both accuracy and precision of the PSHMM estimator across a range of different experimental conditions. Figure 1.9 compares the PSHMM with exponential fitting rate estimates when we vary the emission intensity of the fluorophores (measured in the mean number of photons each emits when in the On state for time  $\Delta$ ). Further investigation of other parameters, including the frame length ( $\Delta$ ), the number of frames ( $N_F$ ) and the detection threshold (proportional to  $\delta$ ) under this model, are provided in Section 1.H of Appendix 1. Across the full range of relevant parameters tested, the PSHMM estimator performs significantly better than exponential fitting.

To assess the accuracy of parameter estimates for the extended models  $d = 1$  and  $d = 2$  over fast, medium and slow switching scenarios, additional simulations were performed by directly sampling the continuous time processes  $\{X(t)\}$  and extracting the observation sequences  $\mathcal{Y}$  as in (1.2), using fixed values of  $\boldsymbol{\theta}^{(d)}$ . Results from the analyses of these simulations are shown in Tables 1.4 and 1.5 in Section 1.G of Appendix 1. While it is evident that



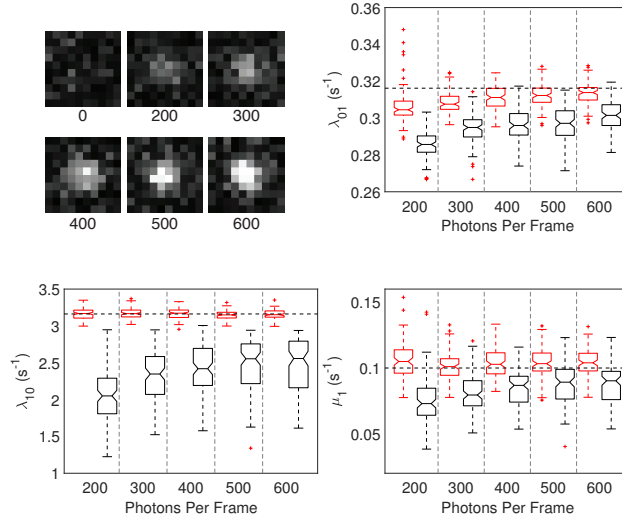
**Figure 1.8:** Estimates of  $\log_{10}(\lambda_{01})$  and  $\log_{10}(\lambda_{10})$  simulated from model  $\mathcal{M}_{\{1\}}^0$  using both exponential fitting (1.8a) and PSHMM fitting (1.8b) are plotted in dark yellow and pink, respectively. True rates are plotted as black crosses. Measurement unit is per second. Estimates for the photo-bleaching rate  $\mu_1$ , along with means, RMSEs and 2.5 and 97.5 empirical percentiles are given in Table 1.3 (see Appendix 1.G).

the estimates for  $\lambda_{0_d 0_{d+1}}$  and  $\lambda_{0_{d+1} 1}$  incur greater bias as  $d$  increases, the empirical (2.5, 97.5) percentile intervals predominantly cover true parameter values, albeit over a larger area due to the increase in the RMSEs. As is seen when  $d = 0$ , the exponential fitting method performs less well, yielding much higher bias and RMSEs for particular parameter values.

### Varying imaging parameters

Additionally, we analysed simulated image traces based on a range of different scenarios for the  $\mathcal{M}_{\{1\}}^0$  model. In particular, we considered varying the frame length  $\Delta$  (Figure 1.11), threshold (proportional to  $\delta$ ) (Figure 1.12), number of frames  $N_F$  (Figure 1.13) and photo-bleaching parameter  $\mu_1$  (Figure 1.14) that is presented in Section 1.H of Appendix 1. All datasets have  $M = 100$  with  $T^* = 0$ , known initial probability  $\nu_1^* = 1$  and unknown false positive rate  $\omega > 0$ .

Our PSHMM method is seen to perform extremely well across all different switching and sampling scenarios, especially in comparison to the exponential fitting method which incurs a consistent bias. It is worth noting that we see an increase in bias of the PSHMM estimator for low frame rates and high threshold values. This is expected as its ability to detect multiple tran-



**Figure 1.9:** Top Left: Examples of single simulated frames at the indicated number of photons per frame (Appendix 1). Box-plots showing quantiles from estimates of  $\lambda_{01}$ ,  $\lambda_{10}$  and  $\mu_1$  from both exponential fitting (black) and PSHMM fitting (red) are plotted against increasing photons per frame.  $N_F = 9872$  for all simulations. True rates given by the blue line.

sitions within a frame is diminished. Furthermore, it is indicated that the PSHMM estimator also exhibits a much lower variance than that from the exponential fitting, a property which also decreases with larger  $M$  (data not shown). Figure 1.14 finally highlights a reduction in bias with  $\mu_1$ , due to a greater number of transitions between hidden states. While this bias gradually decreases with the relative number of fluorophore blinks, our method of prediction also increases in its precision and accuracy. The low bias exhibited by the PSHMM estimator is supported by the fact the true switching rates predominantly lie within the first and third quantiles from the box plots in Figures 1.11 - 1.14 of Section 1.H (Appendix 1), a property not wholly matched by the exponential fitting estimates. These quantiles can be seen to diverge, representing an increase in the variance of the estimator when  $\Delta$  and  $\delta$  (Figures 1.11 and 1.12) are increased, owed to poor identification of the true switching times within a frame. Plots from varying  $N_F$  as shown in Figure 1.13 further highlight how the bias of the PSHMM estimates is unaffected in comparison to the exponential fitting. Furthermore, it is indicated that the PSHMM estimator also exhibits a much lower variance than that

from the exponential fitting, a property which also decreases with larger  $M$  (data not shown). Figure 1.14 finally highlights a reduction in bias with  $\mu_1$ , due to a greater number of transitions between hidden states.

### 1.5.2 Model selection

Using three different simulation studies, the BIC was also used to conduct model selection from the set of proposals  $\{\mathcal{M}_{\{1\}}^0, \mathcal{M}_{\{1\}}^1, \mathcal{M}_{\{1\}}^2\}$  (i.e. under the assumption that the photo-bleaching state was known to only be accessible by the On state). Applying model selection to the  $\mathcal{M}_{\{1\}}^0$  dataset used to estimate parameters in Table 1.3 of Section 1.G (Appendix 1) results in the true state model being chosen in all (100%) cases. 100 datasets each for  $d = 0, 1, 2$  were generated for studies 2, 17 and 20 with  $\Delta = \frac{1}{50}s$  and  $M = 300$ . These results presented in Table 1.1 demonstrate the accuracy of selecting the correct model using this criterion.

Predicted $\rightarrow$ True $\downarrow$	$\mathcal{M}_{\{1\}}^0$	$\mathcal{M}_{\{1\}}^1$	$\mathcal{M}_{\{1\}}^2$
$\mathcal{M}_{\{1\}}^0$	100	0	0
$\mathcal{M}_{\{1\}}^1$	0	98	2
$\mathcal{M}_{\{1\}}^2$	0	1	99

**Table 1.1:** Confusion table showing the empirical percentage of models predicted from three candidates:  $\mathcal{M}_{\{1\}}^0$ ,  $\mathcal{M}_{\{1\}}^1$  and  $\mathcal{M}_{\{1\}}^2$  under simulation studies 16, 19 and 20 (see Tables 1.3, 1.4 and 1.5 in Appendix 1.G), with  $M = 300$ ,  $\delta = \frac{1}{100}s$  and  $\Delta = \frac{1}{50}s$ . 100 datasets from each study were generated and the BIC was used to select the best fitted model.

## 1.6 Application to Alexa Fluor 647 data

In this section we apply the PSHMM fitting method presented in this chapter to Alexa Fluor 647 data analysed with the exponential fitting method in [Lin et al. \(2015\)](#). The BIC will firstly be used to select the most likely photo-kinetic model of the fluorophores imaged in this data, from which its resulting rate estimates, together with bootstrapped confidence intervals, will be presented under both methods.

In this experiment, antibodies labelled with Alexa Fluor 647 at a ratio of 0.13-0.3 dye molecules per antibody were sparsely absorbed to a cover slip and imaged by Total Internal Fluorescence microscopy to investigate the effect of eight different laser intensities on the photo-switching behaviour of the Alexa Fluor 647 dye. The study contains 27 experiments with differing combinations of laser intensities and frame rates<sup>||</sup>. These values, together with the number of molecules detected and the number of frames over which they were imaged are summarised in Table 1.2. The full details of this experiment can be found in [Lin et al. \(2015\)](#).

For each photo-switchable molecule detected, the discrete observation trace indicating if the molecule was observed in each frame, was extracted see (Section 1.F of Appendix 1). In all experiments, the true model and its associated parameters were unknown. Subsequently, we will show comparisons between estimates from both the PSHMM and modified exponential fitting methods

\*\*.

Cell	Laser intensity	$\Delta^{-1}$	$M$	$N_F$	Cell	Laser intensity	$\Delta^{-1}$	$M$	$N_F$	Cell	Laser intensity	$\Delta^{-1}$	$M$	$N_F$
1	1.3519	200	275	49796	10	10.8149	800	244	29418	19	43.2597	800	617	29059
2	43.2597	800	515	29179	11	43.2597	800	493	29400	20	135.1867	800	414	29218
3	86.5195	800	425	29551	12	43.2597	800	456	29071	21	10.8149	200	340	39721
4	5.4075	200	335	49815	13	86.5195	800	425	29426	22	43.2597	800	534	29778
5	21.6299	800	437	29467	14	86.5195	800	398	28989	23	135.1867	800	454	29191
6	21.6299	800	305	29074	15	135.1867	800	422	29295	24	21.6299	200	292	39703
7	21.6299	800	290	29145	16	135.1867	800	436	29270	25	86.5195	800	443	29107
8	10.8149	800	230	29438	17	86.5195	800	554	29327	26	2.7037	200	259	49533
9	10.8149	800	230	29257	18	5.4075	200	393	39758	27	135.1867	800	440	29198

**Table 1.2:** A description of the Alexa Fluor 647 datasets with reference to the laser intensities in  $\text{kW}/\text{cm}^2$  and frames sampled per second (or  $\Delta^{-1}$ ) measured in  $\text{s}^{-1}$  used to characterise each of the 27 cell experiments. The  $N_F \times M$  size of each cell's dataset is also included.

Initially, the BIC model selection criterion as outlined in Section 1.4.2 was used to select the most suitable model for the data from the range of models  $\mathcal{M}_\emptyset^0$ ,  $\mathcal{M}_{\{0\}}^0$ ,  $\mathcal{M}_{\{1\}}^0$ ,  $\mathcal{M}_\emptyset^1$ ,  $\mathcal{M}_{\{0\}}^1$ ,  $\mathcal{M}_{\{0_1\}}^1$ ,  $\mathcal{M}_{\{1\}}^1$ ,  $\mathcal{M}_\emptyset^2$ ,  $\mathcal{M}_{\{0\}}^2$ ,  $\mathcal{M}_{\{0_1\}}^2$ ,  $\mathcal{M}_{\{0_2\}}^2$  and  $\mathcal{M}_{\{1\}}^2$ , with the model  $\mathcal{M}_{\{1\}}^2$  being selected on all (100%) occasions. This supports the analysis of [Lin et al. \(2015\)](#), who hypothesise this, albeit assuming

<sup>||</sup>This dataset was acquired and made available for our use by Dr. Yu Lin and Prof. Joerg Brewersdorf at Yale University.

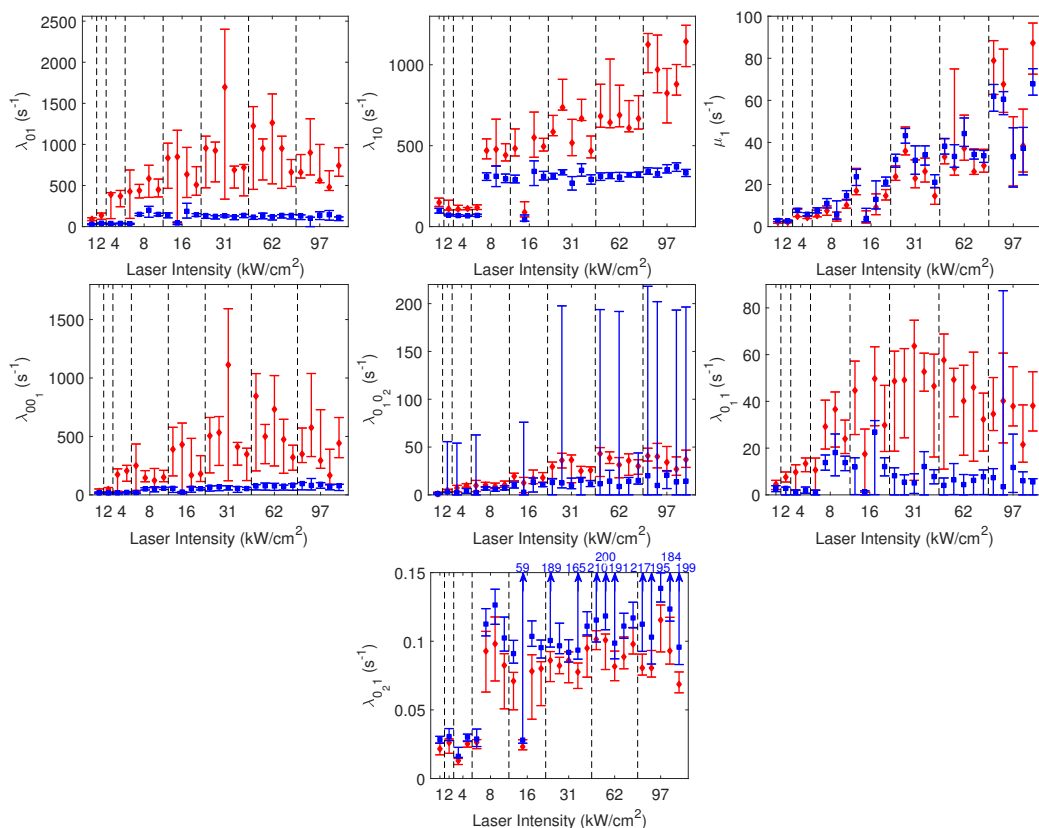
\*\*We modified the exponential fitting algorithm used by [Lin et al. \(2015\)](#) to allow for the absorption parameter (see Section 1.E of Appendix 1 for more details).

the  $\mathcal{M}_\emptyset^2$  (without bleaching) model for rate estimates gained from exponential fitting. PSHMM maximum likelihood estimates were then computed for the estimation of  $\boldsymbol{\theta}_\omega^{(2)} = \left( \lambda_{00_1} \quad \lambda_{01} \quad \lambda_{0_1 0_2} \quad \lambda_{0_1 1} \quad \lambda_{0_2 1} \quad \lambda_{10} \quad \mu_1 \quad \boldsymbol{\nu}_X \quad \delta \quad \omega \right)^\top$  for each of the 27 datasets. Associated with these, 95% bootstrapped intervals were computed using the method in Section 1.4.3 ( $R = 100$  due to computational intensity). The results are shown in Figure 1.10. Comparisons with exponential fitting bootstrapped re-estimates (where  $\boldsymbol{\nu}_X$ ,  $\delta$  and  $\omega$  are not estimable in this setting) are also shown.

The results indicate that the exponential fitting in general, predicts a much slower switching scenario for the Alexa Fluor 647 antibodies, with many estimates shown to be several orders of magnitude below those predicted by the PSHMM. This resembles the conclusions reached from the results of the simulation studies as described in Section 1.5.1 and are thought to occur as a result of the exponential fitting method missing events within frames. Incidentally, the higher variance of predictions from both methods are shown to be reported at higher laser intensities, where faster switching of fluorophores is promoted. This is especially pronounced in some particularly large simulated confidence sets for the exponential fitting estimates of  $\lambda_{0_1 0_2}$  and  $\lambda_{0_2 1}$  (see Figure 1.10).

## 1.7 Conclusions

In this chapter, we have formulated the most general continuous time photo-switching model and from it, carefully defined the observation process. We then linked it to the hidden continuous time photo-switching behaviour that we wish to infer upon. From this, we have formulated a hidden Markov model to link the observations to the continuous time photo-switching model. Importantly, images being formed by exposing the camera over a non-zero time interval violates the traditional assumption placed on HMMs that the emission and transition probabilities are decoupled. To tackle this, we introduced transmission matrices that capture all the dependencies present in the model and provided a detailed scheme for computing them for any continuous time photo-switching model. A modification of the forward-backward algorithm



**Figure 1.10:** Rate predictions and associated 95% bootstrap confidence sets are shown for  $\lambda_{01}$ ,  $\lambda_{10}$ ,  $\mu_1$ ,  $\lambda_{011}$ ,  $\lambda_{010_2}$ ,  $\lambda_{011}$  and  $\lambda_{021}$ , for 8 different values of laser intensity (see Table 1.2). Intervals in blue correspond to those from exponential fitting and those in red correspond to those gained from the PSHMM. Point estimates from each of the 27 datasets are given by diamond (PSHMM) or square (exponential).

tailored for these coupled HMMs was presented and numerical maximisation of the computed likelihood was performed to generate accurate estimates of the true photo-switching rates. Through a detailed simulation study, these were compared to estimates from an existing exponential fitting method. We found that our proposed method of parameter estimation is highly robust to a range of simulated experimental parameters, including low signal-to-noise ratios and fast frame rates, frequently outperforming estimates from exponential fitting. We further found that by using the BIC, it is possible to perform accurate model selection from a range of model proposals, thus providing a powerful new tool for chemists wishing to infer the number of quantum states a particular fluorophore can exist in. The model selection

and estimation methods presented in this chapter were then applied to real data collected from the study of [Lin et al. \(2015\)](#). This provided strong evidence of a relationship between laser intensity and photo-switching rates and the results support the hypothesis that Alexa Fluor 647 fluorophores have three off-states in addition to a photo-bleached state.



## Appendix 1

### 1.A Proof of Theorem 1

In this section, we will prove that the observation process  $\{Y_n\}$  as defined by Equation (1.2) does not exhibit the Markov property (of any order). We will show that this is true for all observations generated by the set of processes  $\{X(t)\}$  defined by the number of multiple off states  $d \in \mathbb{Z}_{\geq 0}$  and paths to the photo-bleaching state 2, as is depicted in Figure 1.3. This property provides the basis of the PSHMM inference we have presented and conducted in this chapter.

*Proof.* For all  $d \in \mathbb{Z}_{\geq 0}$ , any  $\boldsymbol{\lambda} = \left( \lambda_{01} \ \lambda_{00_1} \ \lambda_{0_1 1} \ \lambda_{0_1 0_2} \ \dots \ \lambda_{0_d 1} \ \lambda_{10} \right)^\top \in \mathbb{R}_{>0}^{2d+2}$  and any  $\boldsymbol{\mu} = \left( \mu_0 \ \dots \ \mu_{0_d} \ \mu_1 \right)^\top \in \mathbb{R}_{\geq 0}^{d+2}$  (as characterised by the model  $\mathcal{M}_{\mathcal{S}_X}^d$ ), we define for  $i, j \in \mathcal{S}_X$  and  $n \in \mathbb{N}$

$$\begin{aligned} b_{ij,\Delta}^{(1)} &= \mathbb{P}_{\boldsymbol{\theta}^{(d)}}(X(n\Delta) = j, Y_n = 1 | X((n-1)\Delta) = i) \\ \bar{b}_{i,\Delta}^{(1)} &= \mathbb{P}_{\boldsymbol{\theta}^{(d)}}(Y_n = 1 | X((n-1)\Delta) = i). \end{aligned} \quad (1.34)$$

We consider  $E_n^l$  to be the event that  $l \in \mathcal{S}_Y = \{0, 1\}$  is observed in the  $n$ th frame, i.e. that  $E_n^l = \{Y_n = l\}$ , and that  $F_n^j$  is the event that  $X$  takes the value  $j$  at time  $n\Delta$ , i.e. that  $F_n^j = \{X(n\Delta) = j\}$ .

Fixing  $n \in \mathbb{Z}_{>0}$ , we will show that the quantity  $S(n)$ , describing the probability of observing a 1 in the  $n$ th frame given observations of 1s in all previous  $n-1$  frames is dependent on the full history of the process  $\{Y_n\}$  from time  $n=1$ . Using the notation defined above, we have that  $S(n) = \mathbb{P}_{\boldsymbol{\theta}^{(d)}}(E_n^1 | \cap_{i=1}^{n-1} E_i^1)$ , with  $S(1) = \mathbb{P}_{\boldsymbol{\theta}^{(d)}}(E_1^1)$ . To obtain  $S(n)$ , we condition on the events  $F_n^k$  where  $k \in \bar{\mathcal{S}}_X$ , since starting a frame in the photo-bleaching state 2 would result in no observation of the fluorophore. Using the Markov property of  $\{X(t)\}$

and Bayes' theorem, we obtain that

$$\begin{aligned}
S(n) &= \sum_{k \in \bar{\mathcal{S}}_X} \mathbb{P}_{\boldsymbol{\theta}^{(d)}}(E_n^1 | F_n^k) \mathbb{P}_{\boldsymbol{\theta}^{(d)}}(F_n^k | \cap_{i=1}^{n-1} E_i^1) \\
&= \left( \frac{1}{\prod_{i=1}^{n-1} S(i)} \right) \left[ \sum_{k \in \bar{\mathcal{S}}_X} \bar{b}_{k,\Delta}^{(1)} \left( \sum_{j \in \bar{\mathcal{S}}_X} b_{jk,\Delta}^{(1)} \mathbb{P}_{\boldsymbol{\theta}^{(d)}}(F_{n-1}^j | \cap_{i=0}^{n-2} E_i^1) \right) \right].
\end{aligned}$$

In the above, we can further compute that for all  $k \in \bar{\mathcal{S}}_X$

$$\begin{aligned}
\mathbb{P}_{\boldsymbol{\theta}^{(d)}}(F_n^k | \cap_{i=0}^{n-1} E_i^1) &= \mathbb{P}_{\boldsymbol{\theta}^{(d)}}(F_n^k | E_{n-1}^1 | \cap_{i=0}^{n-2} E_i^1) \mathbb{P}_{\boldsymbol{\theta}^{(d)}}(\cap_{i=0}^{n-2} E_i^1) \\
&= \left( \sum_{j \in \bar{\mathcal{S}}_X} b_{jk,\Delta}^{(1)} \mathbb{P}_{\boldsymbol{\theta}^{(d)}}(F_{n-1}^j | \cap_{i=0}^{n-2} E_i^1) \right) \mathbb{P}_{\boldsymbol{\theta}^{(d)}}(\cap_{i=0}^{n-2} E_i^1).
\end{aligned} \tag{1.35}$$

By letting  $\boldsymbol{\nu}_X$  be the initial probability mass function for  $\{X(t)\}$  whereby  $(\boldsymbol{\nu}_X)_i = \mathbb{P}_{\boldsymbol{\theta}^{(d)}}(X(0) = i)$  ( $i \in \mathcal{S}_X$ ), and considering iterating the conditional probabilities in (1.35) backwards in time, we obtain the relationship

$$S(n) = \frac{\sum_j \bar{b}_{j,\Delta}^{(1)} \sum_{k_{n-1}} b_{k_{n-1}j,\Delta}^{(1)} \sum_{k_{n-2}} \cdots \sum_{k_2} \left( \prod_{i=2}^{n-2} b_{k_{n-i}k_{n-i+1},\Delta}^{(1)} \right) \sum_{k_1} b_{k_1k_2,\Delta}^{(1)} (\boldsymbol{\nu}_X)_{k_1}}{\sum_j \bar{b}_{j,\Delta}^{(1)} \sum_{k_{n-2}} b_{k_{n-2}j,\Delta}^{(1)} \sum_{k_{n-3}} \cdots \sum_{k_2} \left( \prod_{i=3}^{n-2} b_{k_{n-i}k_{n-i+1},\Delta}^{(1)} \right) \sum_{k_1} b_{k_1k_2,\Delta}^{(1)} (\boldsymbol{\nu}_X)_{k_1}},$$

which depends on the full history of the process  $\{Y_n\}$  from  $t = 0$ , since for all  $d$  and  $0 \leq \delta < \Delta$ , we clearly have that  $0 < \bar{b}_{0p,\Delta}^{(1)} < 1$  for all  $p = 0, \dots, d$  and  $0 < \bar{b}_{ij,\Delta}^{(1)} < 1$ , for all  $i, j \in \bar{\mathcal{S}}_X$ .  $\square$

## 1.B Algorithm to compute transmission matrices

Algorithm 2 presents the method for computing the updated transmission matrices  $B_{\Delta}^{*(0)}$  and  $B_{\Delta}^{*(1)}$  as detailed in Section 1.3, suitable for any  $d \geq 0$ . To align notation to that used computationally, 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  $A$ .

---

**Algorithm 2** Compute transmission matrices  $B_{\Delta}^{*(0)}$  and  $B_{\Delta}^{*(1)}$

---

```

1: function COMPUTE_TRANSMISSIONS( $\theta_{\omega}^{(d)}, \Delta$ )
2:   Compute  $G$  from (1.1) using  $\theta^{(d)}$ 
3:    $G_{S, \mathcal{R}^0} \leftarrow \mathbf{0}_{d+2} \mathbf{0}_{d+2}^{\top}$ 
4:    $G_S \leftarrow G[1 : d+2, 1 : d+2]^a$ 
5:    $\boldsymbol{\mu} \leftarrow G[1 : d+2, d+3]$ 
6:    $\sigma_1 \leftarrow -G[d+2, d+2]$ 
7:    $\boldsymbol{\sigma} \leftarrow -\text{diag}(G[1 : d+1, 1 : d+1])$ 
8:   for  $i = 1$  to  $d+1$  do
9:      $G_{S, \mathcal{R}^0}[i, d+2] \leftarrow G_S[i, d+2]$ 
10:   $G_{S, \bar{\mathcal{R}}^0} \leftarrow G_S - G_{S, \mathcal{R}^0}$ 
11:   $A_1 \leftarrow \begin{bmatrix} -G_{S, \bar{\mathcal{R}}^0}^{\top} & I_{d+2} \\ \mathbf{0}_{d+2} \mathbf{0}_{d+2}^{\top} & -G_{S, \bar{\mathcal{R}}^0}^{\top} \end{bmatrix}$ 
12:   $A_2 \leftarrow \begin{bmatrix} G_{S, \bar{\mathcal{R}}^0} & I_{d+2} \\ \mathbf{0}_{d+2} \mathbf{0}_{d+2}^{\top} & \mathbf{0}_{d+2} \mathbf{0}_{d+2}^{\top} \end{bmatrix}$ 
13:   $A \leftarrow \begin{bmatrix} A_1 & \mathbf{0}_{2(d+2)} \mathbf{0}_{2(d+2)}^{\top} \\ \mathbf{0}_{2(d+2)} \mathbf{0}_{2(d+2)}^{\top} & A_2 \end{bmatrix}$ 
14:   $Q_{\Delta}^0(0) \leftarrow e^{G_{S, \bar{\mathcal{R}}^0} \Delta}$ 
15:   $\bar{Q}_{\Delta}^0(0) \leftarrow e^{A \Delta} [i_1 : i_2, i_2 + 1 : i_3] \boldsymbol{\mu} \quad \triangleright i_1 = 2d+5, i_2 = 3(d+2) \text{ and}$ 
     $i_3 = 4(d+2)$ 
16:   $c \leftarrow \frac{1 - e^{-\sigma_1 \Delta}}{1 - e^{-\sigma_1 \Delta}}$ 
17:   $\Xi_{\Delta}^0(0) \leftarrow [\mathbf{1}_{d+1} \mathbf{1}_{d+1}^{\top} \quad c \mathbf{1}_{d+1}]^{\top}$ 
18:   $\Xi_{\Delta}^1(0) \leftarrow \mathbf{1}_{d+2} \mathbf{1}_{d+1}^{\top} - \Xi_{\Delta}^0(0)$ 
19:   $\bar{\Xi}_{\Delta}^0(0) \leftarrow [\mathbf{1}_{d+1}^{\top} \quad c]^{\top}$ 
20:   $\bar{\Xi}_{\Delta}^1(0) \leftarrow \mathbf{1}_{d+2} - \bar{\Xi}_{\Delta}^0(0)$ 
21:   $B_{\Delta}^{(0)} \leftarrow \begin{bmatrix} (Q_{\Delta}^0(0))_{(1:d+2), (1:d+1)} \odot \Xi_{\Delta}^0(0) & \mathbf{0}_{d+2} & \bar{Q}_{\Delta}^0(0) \odot \bar{\Xi}_{\Delta}^0(0) \\ & \mathbf{0}_{d+1}^{\top} & 0 & 1 \end{bmatrix}$ 
22:   $B_{\Delta}^{(1)} \leftarrow \begin{bmatrix} (Q_{\Delta}^0(0))_{(1:d+2), (1:d+1)} \odot \Xi_{\Delta}^1(0) & [\mathbf{0}_{d+1}^{\top} \quad e^{-\sigma_1 \Delta}]^{\top} & \bar{Q}_{\Delta}^0(0) \odot \bar{\Xi}_{\Delta}^1(0) \\ & \mathbf{0}_{d+1}^{\top} & 0 & 0 \end{bmatrix}$ 
23:   $k \leftarrow 1 \quad \triangleright // \text{Start convergence of transmission matrices}$ 
24:  while  $B_{\Delta}^{(0)}$  and  $B_{\Delta}^{(1)}$  have not converged do
25:     $Q_{\Delta}^0(k) \leftarrow \mathcal{L}_s^{-1}[(sI_{d+2} - G_{S, \bar{\mathcal{R}}^0})^{-1} (G_{S, \mathcal{R}^0}(sI_{d+2} - G_{S, \bar{\mathcal{R}}^0})^{-1})^k](\Delta) \triangleright$ 
    Compute inverse Laplace transform matrix
26:     $\bar{Q}_{\Delta}^0(k) \leftarrow \left( \int_0^{\Delta} Q_s^0(k) ds \right) \boldsymbol{\mu}$ 
27:    for  $i = 1$  to  $d+1$  do

```

---

<sup>a</sup>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  $|(G)_{p,p} - (G)_{q,q}| < \epsilon$ ;  $p \neq q = 1, \dots, d+2$ .

---

---

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

29:                      $\triangleright \Upsilon \sim \text{Erlang}(k, \sigma_1)$  and  $F_\Upsilon(u, k, \sigma_1) = \mathbb{P}_{\theta^{(d)}}(\Upsilon \leq u)$

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

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

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

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

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

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

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

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

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

39:                     Find all vectors  $\mathbf{k} = (k_0 \ k_1 \ \dots \ k_d)^\top \in \mathcal{C}_k^{0_{i-1}}$             $\triangleright$

$\mathcal{C}_k^{0_{i-1}} := \{\mathbf{k} : \mathbf{k}^\top \mathbf{1}_{d+1} = k, k_{i-1} > 0, k_0 \geq \dots \geq k_{i-1} - 1 \geq \dots \geq k_d - 1\}$

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

41:                     **For** each  $\mathbf{k}$ ,                      $f_{0_{i-1}}(\mathbf{k}, s) \leftarrow$

$\frac{\lambda_{10}}{s + \sigma_1} \sum_{p=0}^d \frac{\lambda_{0p1} \prod_{q=0}^{p-1} \lambda_{0q0q+1}}{\prod_{q=0}^p (s + \sigma_{0q})} f_{0_{i-1}}(\mathbf{k} - \sum_{r=0}^p \mathbf{e}_{d+1}^{r+1}, s)$             $\triangleright$  **Compute**

$f_{0_{i-1}}(\mathbf{k}, s)$  recursively via the initialisations  $f_{0_{i-1}}(\mathbf{0}_{d+1}, s) = \frac{1_{\{d+2\}}(i)}{s + \sigma_1}$ ,

$f_{0_p1}(\mathbf{e}_{d+1}^{p+1}, s) = \frac{\lambda_{0p1}}{(s + \sigma_{0p})(s + \sigma_1)}$  for  $p = 0, \dots, d$ , and  $f_{0_{d+1}1}(\mathbf{e}_{d+1}^1, s) =$

$\frac{\lambda_{10} \lambda_{01}}{(s + \sigma_0)(s + \sigma_1)^2}$ .

42:                     **For** each  $\mathbf{k}$ , compute  $q_{0_{i-1}1}^1(\mathbf{k}, \Delta) = \mathcal{L}_s^{-1}(f_{0_{i-1}}(\mathbf{k}, s))(\Delta)$

43:                      $\xi_{0_{i-1}1}^1(0, \mathbf{k}, \Delta) \leftarrow \frac{F_\Phi(\Delta|\mathbf{k}, \boldsymbol{\sigma}) - F_\Phi(\Delta - \delta|\mathbf{k}, \boldsymbol{\sigma})}{F_\Phi(\Delta|\mathbf{k}, \boldsymbol{\sigma})} \triangleright F_\Phi(\phi|\mathbf{k}, \boldsymbol{\sigma}) =$

$\mathbb{P}_{\theta^{(d)}}(\Phi \leq \phi)$ , where  $\Phi = \sum_{p=0}^m W_p$ ,  $W_p \stackrel{\text{indep}}{\sim} \text{Erlang}(k_p, \sigma_{0p})$

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

45:

46:                      $B_\Delta^{(0)}[i, d + 2] \leftarrow B_\Delta^{(0)}[i, d + 2] +$

$\sum_{\mathbf{k} \in \mathcal{C}_k^{0_{i-1}}} q_{0_{i-1}1}^1(\mathbf{k}, \Delta) \xi_{0_{i-1}1}^1(0, \mathbf{k}, \Delta)$

47:

48:                      $B_\Delta^{(1)}[i, d + 2] \leftarrow B_\Delta^{(1)}[i, d + 2] + \sum_{\mathbf{k} \in \mathcal{C}_k^{0_{i-1}}} q_{0_{i-1}1}^1(\mathbf{k}, \Delta) (1 -$

$\xi_{0_{i-1}1}^1(0, \mathbf{k}, \Delta))$

49:                      $k \leftarrow k + 1$

50:                      $B_\Delta^{*(0)} \leftarrow (1 - \omega) B_\Delta^{(0)}$

51:                      $B_\Delta^{*(1)} \leftarrow B_\Delta^{(1)} + \omega B_\Delta^{(0)}$

52:           **return**  $B_\Delta^{*(0)}, B_\Delta^{*(1)}$                       $\triangleright$  **Output** transmission matrices

---

## 1.C Exact solution of transmission probabilities when there is a single dark state

In this section, we look to find an exact solution of the transmission probabilities in the event of there being a single dark state 0; in this case  $d = 0$ .

With no multiple dark states, the counting process which we must consider conditioning on uses the sole labelling set  $\{(0, 1)\}$ , and is therefore univariate. Using the same notation as that from Section 1.3.1, this labelling set is  $\mathcal{R} := \mathcal{R}_1^1 = \mathcal{R}^0 = \{(0, 1)\}$ . We therefore consider  $\{N(t) : t \in \mathbb{R}_{\geq 0}\} := \{N_{\mathcal{R}_1^1}(t)\} = \{N_{\mathcal{R}^0}(t)\}$  to denote the number of jumps made by the process  $\{X(t)\}$ , counting the number of transitions in  $\mathcal{R}$ . We can thus calculate the transmission probabilities by conditioning on  $N(\Delta)$ , so that

$$b_{ij,\Delta}^{(l)} = \sum_{k=0}^{\infty} \xi_{ij}(l, k, \Delta) q_{ij}(l, \Delta), \quad (1.36)$$

with  $q_{ij}(k, \Delta) = \mathbb{P}_{\theta^{(a)}}(N(\Delta) = k, X(\Delta) = j | X(0) = i)$  and  $\xi_{ij}(l, k, \Delta) = \mathbb{P}_{\theta^{(a)}}(Y_1 = l | N(\Delta) = k, X(0) = i, X(\Delta) = j)$ .

The matrix representation of the transmission matrices  $B_{\Delta}^{(0)}$  and  $B_{\Delta}^{(1)}$  as in (1.4) holds by setting  $B_{\Delta}^{(l)} = \sum_{k=0}^{\infty} Q_{\Delta}(k) \odot \Xi_{\Delta}^l(k)$  for each  $l \in \mathcal{S}_Y$ , where for all  $k \in \mathbb{Z}_{\geq 0}$ , we define the  $3 \times 3$  matrices  $Q_{\Delta}(k)$  and  $\Xi_{\Delta}^l(k)$  as

$$Q_{\Delta}(k) = \begin{pmatrix} q_{00}(k, \Delta) & q_{01}(k, \Delta) & q_{02}(k, \Delta) \\ q_{10}(k, \Delta) & q_{11}(k, \Delta) & q_{12}(k, \Delta) \\ 0 & 0 & \mathbb{1}_{\{0\}}(k) \end{pmatrix} \quad (1.37)$$

$$\Xi_{\Delta}^l(k) = \begin{pmatrix} \xi_{00}(l, k, \Delta) & \xi_{01}(l, k, \Delta) & \xi_{02}(l, k, \Delta) \\ \xi_{10}(l, k, \Delta) & \xi_{11}(l, k, \Delta) & \xi_{12}(l, k, \Delta) \\ 0 & 0 & \mathbb{1}_{\{0\}}(k + j) \end{pmatrix}. \quad (1.38)$$

### Computation of $Q_{\Delta}(k)$

Similarly to the general case, we use Laplace transforms to calculate  $Q_{\Delta}(k)$  for  $k \in \mathbb{Z}_{\geq 0}$ . Firstly, using the same methodology as in Sections 1.3.2 and

1.3.3, for any  $t \geq 0$ , it is easy to obtain the following system of differential Equations

$$\begin{aligned} q'_{i0}(k, t) &= -\sigma_0 q_{i0}(k, t) + \lambda_{10} q_{i1}(k, t) \\ q'_{i1}(k, t) &= -\sigma_1 q_{i1}(k, t) + \lambda_{01} q_{i0}(k-1, t) \\ q'_{i2}(k, t) &= \mu_0 q_{i0}(k, t) + \mu_1 q_{i1}(k, t). \end{aligned}$$

Taking Laplace transforms of both sides of the above Equations, yields for  $i, j \in \mathcal{S}_X$  and  $k \in \mathbb{N}$

$$\begin{aligned} f_{ij}(k, s) &= \frac{\lambda_{01}^k \lambda_{10}^{k+i-j}}{(s + \sigma_1)^{k+i} (s + \sigma_0)^{k+1-j}} \\ f_{ij}(0, s) &= \frac{(1-j)\lambda_{10}^{i-j}}{(\sigma_1 - \sigma_0)^{i-j} (s + \sigma_0)} + \frac{i\lambda_{10}^{i-j}}{(\sigma_0 - \sigma_1)^{i-j} (s + \sigma_1)} \quad \sigma_0 \neq \sigma_1 \\ f_{ij}(0, s) &= \frac{\lambda_{10}^{i-j}}{(s + \sigma_1)^{(i-j)+1}} \mathbb{1}_{\geq 0}(i-j) \quad \sigma_0 = \sigma_1. \end{aligned}$$

Making use of the fact that for  $n \in \mathbb{Z}_{\geq 0}$  and some  $a \in \mathbb{R}$ ,  $\mathcal{L}_s^{-1} \left[ \frac{1}{(s+a)^{n+1}} \right] (t) = \frac{t^n e^{-at}}{n!}$ , we obtain that

$$q_{ij}(0, t) = \frac{(1-j)\lambda_{10}^{i-j}}{(\sigma_1 - \sigma_0)^{i-j}} e^{-\sigma_0 t} + \frac{i\lambda_{10}^{i-j}}{(\sigma_0 - \sigma_1)^{i-j}} e^{-\sigma_1 t} \quad \sigma_0 \neq \sigma_1 \quad (1.39)$$

$$q_{ij}(0, t) = \frac{\lambda_{10}^{i-j} t^{i-j} e^{-\sigma_1 t}}{(i-j)!} \mathbb{1}_{\geq 0}(i-j) \quad \sigma_0 = \sigma_1. \quad (1.40)$$

When  $k \in \mathbb{N}$  and  $\sigma_0 = \sigma_1$  it is easily seen that  $q_{ij}(k, t) = \mathcal{L}_s^{-1}[f_{ij}(k, s)](t)$  takes the form

$$q_{ij}(k, t) = \frac{\lambda_{01}^k \lambda_{10}^{k+i-j} t^{2k+i-j} e^{-\sigma_1 t}}{(2k+i-j)!}. \quad (1.41)$$

When  $\sigma_0 \neq \sigma_1$ , to find  $q_{ij}(k, t)$  we split the expressions for  $f_{ij}(k, s)$  into partial fractions using the method of derivatives in the same manner as in [Minin and Suchard \(2007\)](#). In particular, we write  $f_{ij}(k, s)$  as

$$f_{ij}(k, s) = \sum_{p=1}^{k+1-j} \frac{{}^{(ij)}A_{k:(k-p+1-j)}^0}{(s + \sigma_0)^p} + \sum_{p=1}^{k+i} \frac{{}^{(ij)}A_{k:(k-p+i)}^1}{(s + \sigma_1)^p},$$

where

$$\begin{aligned}
{}_{(ij)}A_{k:p}^0 &= \frac{1}{p!} \frac{d^p}{ds^p} \left\{ (s + \sigma_0)^{k+1-j} f_{ij}(k, s) \right\} \Big|_{s=-\sigma_0} \\
&= \binom{k+p+i-1}{p} \frac{(-1)^p \lambda_{01}^k \lambda_{10}^{k+i-j}}{(\sigma_1 - \sigma_0)^{k+p+i}} \\
{}_{(ij)}A_{k:p}^1 &= \frac{1}{p!} \frac{d^p}{ds^p} \left\{ (s + \sigma_1)^k f_{ij}(k, s) \right\} \Big|_{s=-\sigma_1} \\
&= \binom{k+p-j}{p} \frac{(-1)^p \lambda_{01}^k \lambda_{10}^{k+i-j}}{(\sigma_0 - \sigma_1)^{k+p+1-j}}.
\end{aligned}$$

We are thus able to obtain the following closed form expressions

$$q_{ij}(k, t) = \sum_{p=1}^{k+1-j} \frac{{}_{(ij)}A_{k:(k-p+1-j)}^0}{(p-1)!} t^{p-1} e^{-\sigma_0 t} + \sum_{p=1}^{k+i} \frac{{}_{(ij)}A_{k:(k-p+i)}^1}{(p-1)!} t^{p-1} e^{-\sigma_1 t}. \quad (1.42)$$

We now consider two cases to compute  $q_{i2}(k, t) = \sum_{p=0}^1 \mu_p \int_0^t q_{ip}(k, s) ds$ .

Firstly, when  $\sigma_0 = \sigma_1$ , we have that

$$\begin{aligned}
q_{i2}(0, t) &= \sum_{p_1=0}^1 \frac{\mu_{p_1} \lambda_{10}^{i-p_1} \mathbb{1}_{\geq 0}(i-p_1)}{\sigma_1^{i-p_1+1}} \sum_{p_2=i-p_1+1}^{\infty} \frac{(\sigma_1 t)^{p_2} e^{-\sigma_1 t}}{p_2!} \\
q_{i2}(k, t) &= \sum_{p_1=0}^1 \frac{\mu_{p_1} \lambda_{01}^k \lambda_{10}^{k+i-p_1}}{\sigma_1^{2k+i-p_1+1}} \sum_{p_2=2k+i-p_1+1}^{\infty} \frac{(\sigma_1 t)^{p_2} e^{-\sigma_1 t}}{p_2!}. \quad (1.43)
\end{aligned}$$

Secondly, when  $\sigma_0 \neq \sigma_1$ , we have

$$\begin{aligned}
q_{i2}(0, t) &= \frac{\mu_0 \lambda_{10}^i}{(\mu_0 + \lambda_{01})(\sigma_1 - \sigma_0)^i} (1 - e^{-\sigma_0 t}) + \frac{i}{\sigma_1} \left[ \frac{\mu_0 \lambda_{10}}{(\sigma_0 - \sigma_1)} + \mu_1 \right] (1 - e^{-\sigma_1 t}) \\
q_{i2}(k, t) &= \sum_{p_1=0}^1 \mu_{p_1} \sum_{p_2=0}^1 \sum_{p_3=1}^{k+1-p_2(1-p_1)} \frac{{}_{(ip_2)}A_{k:(k-p_3+ip_1+(1-p_2)(1-p_1))}^{p_1}}{\sigma_{p_1}^{p_3}} \sum_{d=p_3}^{\infty} \frac{(\sigma_{p_1} t)^d}{d!} e^{-\sigma_{p_1} t}. \quad (1.44)
\end{aligned}$$

Setting  $t = \Delta$  in all Equations (1.39)-(1.44) yields all entries of  $Q_{\Delta}(k)$  as in (1.37), for  $k \in \mathbb{Z}_{\geq 0}$ .



### Computation of $\Xi_{\Delta}^l(k)$

Using the same methodology as presented in sections 1.3.2 and 1.3.3, we will demonstrate computation of  $\Xi_{\Delta}^l(k)$ . In particular, since  $d = 0$  we have for  $s \in \mathbb{N}$  that  $R_s = U_s + D_s$  is a PSARP (see Definition 2) with inter-arrival times  $\mathbf{R} = (R_1, R_2, \dots)$ . Here,  $U_s \stackrel{iid}{\sim} \exp(\sigma_1)$  and  $D_s \stackrel{iid}{\sim} \exp(\sigma_0)$ .

Similarly to the general case, if  $X(0) = 0$ ,  $X(\Delta) \in \{0, 2\}$  and  $N(\Delta) = k$ , then there are no censored observations from the On state, so that for  $k \in \mathbb{N}$ ,  $i, j \in \mathcal{S}_X$  with  $i \neq 2$  and  $j \neq 1$

$$\xi_{ij}(0, k, \Delta) = \frac{1 - \sum_{m=0}^{k+i-1} \frac{(\sigma_1 \delta)^m}{m!} e^{-\sigma_1 \delta}}{1 - \sum_{m=0}^{k+i-1} \frac{(\sigma_1 \Delta)^m}{m!} e^{-\sigma_1 \Delta}} \quad (1.45)$$

$$\xi_{00}(0, 0, \Delta) = 1 \quad (1.46)$$

$$\xi_{10}(0, 0, \Delta) = \frac{1 - e^{-\sigma_1 \delta}}{1 - e^{-\sigma_1 \Delta}}. \quad (1.47)$$

If  $X(0) \in \{0, 1\}$ ,  $X(\Delta) = 1$  and  $N(\Delta) = k$  with  $k \in \mathbb{N}$ , there are exactly  $k$  time pieces in the dark state 0 ( $D_1, D_2, \dots, D_k$ ). These  $k$  pieces form iid exponential random variables and their sum  $v(k)$  also has an Erlang distribution  $v(k) = \sum_{i=1}^k D_i \sim \text{Erlang}(k, \sigma_0)$ . Now since the random event

$$\{Y_1 = 0 | N(\Delta) = k, X(0) \in \{0, 1\}, X(\Delta) = 1\} = \{v(k) \geq \Delta - \delta | v(k) \leq \Delta\},$$

we have for  $k \in \mathbb{N}$

$$\xi_{i1}(1, k, \Delta) = \frac{1 - \sum_{m=0}^{k-1} \frac{(\sigma_0(\Delta-\delta))^m}{m!} e^{-\sigma_0(\Delta-\delta)}}{1 - \sum_{m=0}^{k-1} \frac{(\sigma_0 \Delta)^m}{m!} e^{-\sigma_0 \Delta}} \quad (1.48)$$

$$\xi_{11}(1, 0, \Delta) = 1. \quad (1.49)$$

We may fill the matrices  $\Xi_{\Delta}^0(k)$  and  $\Xi_{\Delta}^1(k)$  in (1.38) by using (1.45)-(1.49) and the fact that for  $i, j \in \mathcal{S}_X$ ,  $\xi_{ij}(1, k, \Delta) = 1 - \xi_{ij}(0, k, \Delta)$ .

## 1.D Discussion on implementing the PSHMM algorithm

In this section, we discuss the implementation of PSHMM algorithm for practical parameter values. We will firstly discuss convergence of the transmission matrices defined in (1.3.4) that are needed to compute the log-likelihood of observations in (1.31). We then highlight key mathematical ideas that suggest appropriate initial parameters needed to optimise this likelihood, including why potential multi-modality effectuates the search for the maximum likelihood estimates from different starting values of the parameter space.

### 1.D.1 Convergence of transmission matrices

In most practical aspects,  $\Delta = \mathcal{O}(10^{-2})$  and  $\sigma_i \Delta < 1$  for all  $i \in \bar{S}_X$ . With these parameter values, convergence of the transmission matrices from Algorithm 2 (from Section 1.B) usually transpires up to and including  $k = 4$ , with  $k \geq 6$  probabilities seldom needed in practice. For the simulated parameters in this chapter, our Matlab implementation of this Algorithm thus approximates the matrices  $B_\Delta^{(0)}$  and  $B_\Delta^{(1)}$  using only  $k \leq 5$ .

### 1.D.2 Likelihood optimisation

The log-likelihood function in equation (1.31) is optimised with respect to  $\theta^{(d)}$  via the Nelder-Mead simplex to obtain maximum likelihood estimates. When  $d = 0$ , we are able to gain starting parameters for this search using the method described below. However when  $d > 0$ , we deploy a stochastic search with many starting values to mitigate for multiple modes and find the global maximum.

#### Multi-modality

It is seen that as  $d$  increases (specifically with higher switching rates), the likelihood function can become multi-modal along the direction of  $\lambda_{0_i 0_{i+1}}$  where

$i = 0, \dots, d$ , from both the PSHMM and exponential fitting methods. In these situations, an approach which searches the parameter space from many different starting points is useful to implement for a global maximum to be found. For this reason, we also implemented a function which utilises MATLAB's MultiStart property, with different start values as those gained from exponential fitting.

### Estimating initial rate parameters

To numerically maximise the log-likelihood function in (1.31), a starting value  $\hat{\boldsymbol{\theta}}_{\omega}^{(d)}$  needs to be determined primarily due to the potential multi-modality of the likelihood function and further to reduce the computational time given the size of the parameter space  $\Theta_{\omega}^{(d)}$ .

To do so, we consider a crude approximate estimate to start a maximum likelihood search. In the case  $d = 0$ , we have found that optimising the likelihood function of a first order Markov chain  $\{\bar{Y}_n : n \in \mathbb{Z}^*\}$  (used as a partial likelihood function for  $\{Y_n\}$ ) has yielded appropriate initial estimates for  $\lambda_{01}$  and  $\lambda_{10}$  using a *truncated* dataset  $\tilde{\mathcal{Y}}$ . Here, we consider *each* observation sequence before absorption:  $\tilde{\mathbf{y}}^j = \{y_i^j\}_{i=0}^{\sigma^j-1}$ ; where  $\sigma^j \in \mathbb{Z}_{\geq 0}$  denotes the last frame a fluorophore is seen in observation sequence  $j = 1, \dots, M$ . The truncated dataset is then defined as  $\tilde{\mathcal{Y}} = (\tilde{\mathbf{y}}^1 \quad \tilde{\mathbf{y}}^2 \quad \dots \quad \tilde{\mathbf{y}}^M)$ . The chain  $\{\bar{Y}_n\}$  is now constructed by using one-step transition probabilities from  $\{Y_n\}$  through setting  $\mu_0 = \mu_1 = \delta = 0$  and  $\boldsymbol{\nu}_X = \left( \frac{\lambda_{10}}{(\lambda_{01} + \lambda_{10})} \quad \frac{\lambda_{01}}{(\lambda_{01} + \lambda_{10})} \right)^{\top}$ , which is also the stationary distribution for  $\{X(t)\}$ .

**Remark 7.** In the case  $d = 0$ , we can calculate that the transition probability matrix  $\mathbf{P}^{\bar{Y}}$  of  $\{\bar{Y}_n\}$ , whereby  $(\mathbf{P}^{\bar{Y}})_{ij} = \mathbb{P}(Y_n = j | Y_{n-1} = i)$  takes the form

$$\mathbf{P}^{\bar{Y}} = \begin{pmatrix} e^{-\lambda_{01}\Delta} & 1 - e^{-\lambda_{01}\Delta} \\ \frac{\pi_X^0 e^{-\lambda_{01}\Delta} (1 - e^{-\lambda_{01}\Delta})}{1 - \pi_X^0 e^{-\lambda_{01}\Delta}} & \frac{\pi_X^1 + \pi_X^0 (1 - e^{-\lambda_{01}\Delta})^2}{1 - \pi_X^0 e^{-\lambda_{01}\Delta}} \end{pmatrix}, \quad (1.50)$$

where  $\pi_X^0 = \frac{\lambda_{10}}{(\lambda_{01} + \lambda_{10})}$  and  $\pi_X^1 = \frac{\lambda_{01}}{(\lambda_{01} + \lambda_{10})}$ .

Crude estimates for  $\lambda_{01}$  and  $\lambda_{10}$  can be obtained by optimising the restricted

likelihood function for  $\{\bar{Y}_n\}$  using the truncated dataset  $\tilde{\mathcal{Y}}$ ; this takes the form  $L_{\bar{Y}}(\tilde{\mathcal{Y}}; \lambda_{01}, \lambda_{10}) = \prod_{j=1}^M \prod_{i=0}^{o^j-1} p_{y_i^j y_{i+1}^j}^{\bar{Y}}$ . If  $n_{s_1 s_2}^j$  denotes the number of observed transitions from state  $s_1$  to  $s_2$  ( $s_1, s_2 \in \mathcal{S}_Y$ ) over  $o^j$  observations in sequence  $j = 1, \dots, M$ , then it is shown in [Rajarshi \(2013\)](#) that the maximum likelihood estimators of the transition probabilities in  $\mathbf{P}^{\bar{Y}}$  are given by  $\hat{p}_{00}^{\bar{Y}} = \frac{\sum_{j=1}^M n_{00}^j}{\sum_{j=1}^M n_{00}^j + n_{01}^j}$  and  $\hat{p}_{10}^{\bar{Y}} = \frac{\sum_{j=1}^M n_{10}^j}{\sum_{j=1}^M n_{10}^j + n_{11}^j}$ . By rearranging the expressions for  $\lambda_{01}$  and  $\lambda_{10}$  given in (1.50), we obtain the crude estimates

$$\hat{\lambda}_{01} = -\frac{\log(\hat{p}_{00}^{\bar{Y}})}{\Delta} \quad \hat{\lambda}_{10} = \frac{\hat{p}_{10}^{\bar{Y}} \hat{\lambda}_{01}}{(1 - e^{-\hat{\lambda}_{01} \Delta})(e^{-\hat{\lambda}_{01} \Delta} - \hat{p}_{10}^{\bar{Y}})}. \quad (1.51)$$

Moreover, gaining initial estimates of the photo-bleaching rates  $\hat{\mu}_0$  and  $\hat{\mu}_1$  (when at least one rate is non-zero) can be done by considering approximate absorption times:  $t^j = o^j \Delta$  for  $j = 1 \dots M$ . These times can be used to fit the photo-bleaching time distribution (see [Buchholz et al. \(2014\)](#))

$$f_{\tau}(\tau | \lambda_{01}, \lambda_{10}, \mu_0, \mu_1) = \boldsymbol{\nu}_X^{\top} e^{\mathbf{T}\tau} \mathbf{t},$$

where

$$\mathbf{T} = \begin{pmatrix} -(\lambda_{01} + \mu_0) & \lambda_{01} \\ \lambda_{10} & -(\lambda_{10} + \mu_1) \end{pmatrix}, \quad \mathbf{t} = \begin{pmatrix} \mu_0 \\ \mu_1 \end{pmatrix}.$$

Estimates are gained from maximising the log-likelihood

$$\ell(\mu_0, \mu_1 | \tau, \hat{\lambda}_{01}, \hat{\lambda}_{10}) = \sum_{j=1}^M \log(f_{\tau}(t^j | \hat{\lambda}_{01}, \hat{\lambda}_{10}, \mu_0, \mu_1)),$$

where  $\hat{\lambda}_{01}$  and  $\hat{\lambda}_{10}$  are the crude estimates gained from (1.51). In the case that  $d > 0$ , we may choose to set  $\hat{\mu}_{0_i} = \hat{\mu}_0$  for all  $i = 1, \dots, m$  where the model permits photo-bleaching from other such dark states.

The noise parameters  $\delta$  and  $\omega$  are started close to zero, as in general these will be small. Furthermore, if the initial mass  $\boldsymbol{\nu}_X$  is unknown then one can initialise with  $\nu_1 = \frac{1}{M} \sum_{j=1}^M y_0^j$  and  $\nu_0 = 1 - \nu_1$ . When  $d > 0$ , we set  $(\boldsymbol{\nu}_X)_i = 0$  for all  $i \notin \{0, 1\}$ .

We have found that initial values  $\hat{\lambda}_{10}$ ,  $\hat{\mu}_0$  and  $\hat{\mu}_1$  gained from the above analysis are generally superior to estimates gained from exponential fitting for all  $d \geq 0$ . However, in the presence of multiple dark states, exponential fitting is used to obtain initial estimates for all other rate parameters as is required for the PSHMM likelihood optimisation.

## 1.E Exponential fitting estimator

In this section, we provide details of the exponential fitting estimator of Lin et al. (2015) that is used for comparison with the PSHMM estimator in this chapter. We begin with an outline of the original method, demonstrating how maximum likelihood rate estimates of the switching rates under models of type  $\mathcal{M}_\emptyset^d$  ( $d \in \mathbb{Z}_{\geq 0}$ ), are obtained from the photo-switching data. We then demonstrate how it can be extended to models of type  $\mathcal{M}_{\{1\}}^d$  ( $d \in \mathbb{Z}_{\geq 0}$ ) that include the photo-bleached state 2.

### 1.E.1 Original method

Consider the irreducible Markov process  $\{X(t) : t \in [0, \infty)\}$  on the state space  $\mathcal{S}_X = \{0, 0_1, \dots, 0_d, 1\}$  equipped with the generator  $G$  and initial probability mass  $\nu_X$  as shown in Equation (1.1). We note here that the inclusion of the photo-bleaching state 2 is not accounted for in this method.

If this Markov chain was completely observable, one would be able to note the dwell times (times spent) in each of the  $d + 2$  states. In particular, if  $T_s$  denotes a random dwell time in state  $s \in \mathcal{S}_X$ , then  $T_s \sim \exp(\sigma_s)$ , where  $\sigma_{0_d} = \lambda_{0_d 1}$ ,  $\sigma_1 = \lambda_{10}$  and when  $d > 0$ ,  $\sigma_{0_i} = \lambda_{0_i 0_{i+1}} + \lambda_{0_i 1}$ , for  $i = 0, \dots, d + 1$ . Maximum likelihood estimation of each  $\sigma_s$  is subsequently straightforward. Specifically, if  $N$  realisations  $t_s^1, t_s^2, \dots, t_s^N$  from  $T_s$  are obtained, then the maximum likelihood estimator  $\hat{\sigma}_s$  of  $\sigma_s$  is given by

$$\hat{\sigma}_s = \frac{N}{\sum_{i=1}^N t_s^i}. \quad (1.52)$$

Suppose now that  $\{X(t)\}$  is not directly observable but instead the times spent in the Dark-On cycle  $\phi_0 = 0 \rightarrow 0_1 \rightarrow \dots 0_d \rightarrow 1$  and the On-Dark cycle  $\phi_1 = 1 \rightarrow 0$  are observed. Firstly, we note that  $T_1$  is the dwell time in an On-Dark cycle  $\phi_1$ . Now let random variable  $\bar{T}_0$  be a dwell time in the Dark-On cycle  $\phi_0$ ; that is,  $\bar{T}_0$  is the time taken to reach state 1 from state 0 along the path  $\phi_0$ . One can express  $\bar{T}_0 = \sum_{i=0}^d T_{0_i}$ , i.e.  $\bar{T}_0$  is equal

in distribution to a sum of exponentially distributed random variables, with the key example that  $\bar{T}_0 = T_0$  when  $d = 0$ .

Under the  $d = 2$  case, [Lin et al. \(2015\)](#) use an ODE method to derive the probability density function of  $\bar{T}_0$ , parameterised by the unknown rates  $\boldsymbol{\lambda}_G$  in  $G$  (1.1). As required, this method can easily be extended to account for different values of  $d \in \mathbb{Z}_{\geq 0}$ . Specifically, the density function of  $\bar{T}_0$  takes the form

$$f_{\bar{T}_0}(t) = \sum_{j=0}^d k_j \sigma_{0_j} e^{-\sigma_{0_j} t} \quad t > 0. \quad (1.53)$$

When  $d = 1$ , it can be shown  $k_0 = 1 + \frac{\lambda_{00_1}}{\sigma_{0_1} - \sigma_0}$  and  $k_1 = 1 - k_0$ . When  $d = 2$ , the mixture coefficients are given by

$$\begin{aligned} k_0 &= 1 + \frac{\lambda_{00_1}}{\sigma_{0_1} - \sigma_0} + \frac{\lambda_{00_1} \lambda_{0_1 0_2} (\sigma_{0_1} - \sigma_{0_2})}{A} \\ k_1 &= \frac{\lambda_{00_1}}{\sigma_0 - \sigma_{0_1}} + \frac{\lambda_{00_1} \lambda_{0_1 0_2} (\sigma_{0_2} - \sigma_0)}{A} \\ k_2 &= \frac{\lambda_{00_1} \lambda_{0_1 0_2} (\sigma_0 - \sigma_{0_1})}{A} \\ A &= \sigma_{0_1} \sigma_{0_2} (\sigma_{0_1} - \sigma_{0_2}) + \sigma_0 \sigma_{0_2} (\sigma_{0_2} - \sigma_0) + \sigma_0 \sigma_{0_1} (\sigma_0 - \sigma_{0_1}). \end{aligned}$$

Readers are directed to [Lin et al. \(2015\)](#) for a formal derivation of this result.

Recognising that the each row of  $G$  must sum to zero, the density in (1.53) enables the unknown photo-switching parameters

$$\boldsymbol{\lambda} = \left( \lambda_{00_1} \quad \dots \quad \lambda_{0_{d-1} 0_d} \quad \lambda_{01} \quad \dots \quad \lambda_{0_d 1} \right)^\top$$

to be estimated via maximisation of the log-likelihood function

$$\ell(\bar{t}_0^1, \bar{t}_0^2, \dots, \bar{t}_0^N | \sigma_{0_0}, \dots, \sigma_{0_d}) = \sum_{i=0}^N \log \left( \sum_{j=0}^d k_j \sigma_{0_j} e^{-\sigma_{0_j} \bar{t}_0^i} \right), \quad (1.54)$$

where  $\bar{t}_0^1, \bar{t}_0^2, \dots, \bar{t}_0^N$  are  $N$  realisations from  $\bar{T}_0$ . Numerical optimisation of (1.54), together with the maximum likelihood estimator  $\hat{\sigma}_1 = N / (\sum_{i=1}^N t_1^i)$

from Equation (1.52) can be used to obtain the maximum likelihood estimate  $\hat{\boldsymbol{\lambda}}$  of  $\boldsymbol{\lambda}$ .

In the context of the model that we have formulated in this chapter and that of Lin et al. (2015), the data that is received does not produce observations of random variables  $T_1$  or  $\bar{T}_0$ . Instead, one observes a *discrete* sequence of zeros and ones indicating whether or not a fluorophore is detected in each time frame. Specifically, an  $N_F \times 1$  observation sequence  $\mathbf{y}$  (from a single molecule), can be written in block vector form

$$\mathbf{y} = \left[ \mathbf{0}_{n_0^1}^\top \quad \mathbf{1}_{n_1^1}^\top \quad \mathbf{0}_{n_0^2}^\top \quad \mathbf{1}_{n_1^2}^\top \quad \dots \quad \mathbf{0}_{n_0^N}^\top \quad \mathbf{1}_{n_1^N}^\top \right]^\top, \quad (1.55)$$

where  $\mathbf{0}_n$  and  $\mathbf{1}_n$  are the  $n \times 1$  vectors of zeros and ones respectively,  $n_0^i, n_1^i \in \mathbb{Z}_{\geq 0}$  and  $N$  is such that  $\sum_{i=1}^N n_0^i + n_1^i = N_F$ .

Using the form in (1.55) for each observation vector, we consider the sequences of times  $\tilde{t}_j^i = n_j^i \Delta$  for  $j = 0, 1$  and  $i = 1, \dots, N$ . These sequences of times are assumed in Lin et al. (2015) to be the dwell times to obtain the maximum likelihood rate estimates;  $\{\tilde{t}_1^i\}_{i=1}^N$  is used to obtain  $\hat{\sigma}_1$  in Equation (1.52) and  $\{\tilde{t}_0^i\}_{i=1}^N$  is used to obtain  $\boldsymbol{\lambda}$  by maximising the log-likelihood in Equation (1.54). The likelihood function for dwell times  $\mathcal{T} = \{\tilde{t}_{0,m}^i, i = 1, \dots, N_m, m = 1, \dots, M\}$  recorded from  $M \geq 1$  independent molecules becomes

$$\ell(\mathcal{T} | \sigma_{0_0}, \dots, \sigma_{0_d}) = \sum_{m=1}^M \sum_{i=0}^{N_m} \log \left( \sum_{j=0}^d k_j \sigma_{0_j} e^{-\sigma_{0_j} \tilde{t}_{0,m}^i} \right), \quad (1.56)$$

which again can be numerically optimised, e.g. via quasi-Newton methods or the Nelder-Mead simplex.

## Drawbacks

When applied to fluorescence microscopy data, there are a number of problems associated with this method which effectuate poorer rate estimation when compared with the PSHMM method.



Firstly, the observed dwell times can only take a discrete set of values, even though the true dwell times are continuous random variables. Secondly, observed dwell times for the On-Dark cycle  $\phi_1$  may be an overestimate of the true dwell times as short transitions to a dark state may not be detected. Thirdly, observed dwell times for the Dark-On cycle  $\phi_0$  may be an overestimate as short transitions to the On state may not be detected. Lastly, the assumed distribution of the dwell time for the Dark-On cycle is incorrect if not all dark states on the path are reached. In general, this method therefore incorrectly estimates a much slower photo-switching model for a fluorophore via the consistent underestimation of the parameters  $\lambda_{01}, \lambda_{10}$  and poorer estimation of the remaining rates in  $\lambda_G$ .

### 1.E.2 Extension to handling the photo-bleached state

To gain estimates of these switching rates in the presence of the photo-bleached state 2, we consider *each* observation sequence before absorption:  $\tilde{\mathbf{y}}^j = \{y_i^j\}_{i=0}^{o^j-1}$ ; where  $o^j \in \mathbb{Z}_{\geq 0}$  denotes the last frame a fluorophore is seen in observation sequence  $j = 1, \dots, M$ . Photo-switching estimates are gained by fitting to the truncated dataset  $\tilde{\mathcal{Y}} = (\tilde{\mathbf{y}}^1 \quad \tilde{\mathbf{y}}^2 \quad \dots \quad \tilde{\mathbf{y}}^M)$ .

We are able to infer upon the photo-bleaching parameter  $\mu_1$  (when  $\mu_j = 0$  for all  $j \neq 1$ ) by using the fact that the random variable  $N_{10}$  denoting the number of  $1 \rightarrow 0$  transitions observed in a single sequence is geometrically distributed with success probability  $\frac{\mu_1}{\sigma_1}$ . Maximum likelihood estimation from the entire dataset  $\mathcal{Y}$  yields that  $\hat{\lambda}_{10} = \bar{n}_{10}\hat{\mu}_1$ , where  $\bar{n}_{10} = \frac{1}{M} \sum_{j=1}^M n_{10}^j$  and  $n_{10}^j$  denotes the number of observed  $1 \rightarrow 0$  transitions in sequence  $j$ . Since the exponentially fitting method, by considering dwell time sequences from the On state, always yields a maximum likelihood estimate for  $\hat{\sigma}_1 = \hat{\lambda}_{10} + \hat{\mu}_1$ , we obtain that  $\hat{\lambda}_{10} = \frac{\bar{n}_{10}\hat{\sigma}_1}{1+\bar{n}_{10}}$  and  $\hat{\mu}_1 = \frac{\hat{\sigma}_1}{1+\bar{n}_{10}}$ .

## 1.F Image simulation methods

In this section we provide the simulation methods for the simulation studies presented in Section 1.5, specifically Figure 7 and Table 2.

### 1.F.1 Imaging simulation

Simulated images of a fluorophore are produced from a sequence of continuous time state transitions realised from model  $\mathcal{M}_{\{1\}}^0$ , discretised to give a sequence of fractional On times  $q_0, \dots, q_{N_F-1}$  for each frame. All other state information (dark-state / bleached-state identity) is discarded. Let  $\lambda_p$  be the expected number of photons emitted per second by a fluorophore in the On state, then the expected number of photons emitted by the fluorophore in frame  $i$  is  $q_i \Delta \lambda_p$ .

To replicate the microscope point spread function, photon positions are assumed to be distributed according to a 2D Gaussian distribution with standard deviation  $\sigma = 135$  nm centred at the stationary position of the fluorophore  $\mathbf{s} \in \mathbb{R}^2$ . The photon positions are binned into a grid of  $N_{pix}$   $100 \times 100$  nm pixels  $\{C_k \subset \mathbb{R}^2; k = 1, \dots, N_{pix}\}$  representing the EMCCD camera. The expected number of photons from the fluorophore in pixel  $k$  for frame  $i$  is

$$\mu_{i,k} = q_i \Delta \lambda_p \int_{C_k} \mathcal{N}(\mathbf{x}; \mathbf{s}, \sigma^2 I_2) d\mathbf{x}$$

where  $\mathcal{N}(\cdot; \boldsymbol{\mu}, \Sigma)$  denotes the probability density function of the Gaussian distribution with mean  $\boldsymbol{\mu}$  and covariance matrix  $\Sigma$ . A constant mean background photon count of 5 is added to the expected photon count in every pixel.

A Poisson-Gamma-Normal noise model is used to simulate the EMCCD readouts. This model is adapted from [Hirsch et al. \(2013\)](#) and constant parameter values used are typical of commercial EMCCD cameras. The expected photon count per pixel ( $\mu_{i,k} + 5$ ) is converted to an expected electron count per pixel  $\mu_{i,k}^e$  by multiplying it by the EMCCD quantum efficiency, 0.9, and adding a spurious electron dark current of  $0.005 e^- s^{-1}$ , i.e.

$\mu_{i,k}^e = 0.9(\mu_{i,k} + 5) + 0.005$ . The electron count  $e_{i,k}$  for pixel  $k$  in frame  $i$  is sampled from a  $\text{Poisson}(\mu_{i,k}^e)$  distribution. The electron count after EMCCD gain  $\epsilon_{i,k}$ , is then sampled from a  $\text{Gamma}(e_{i,k}, \beta)$  distribution, with  $\beta$  equal to the EMCCD gain, set to 250 in this simulation (typical experimental ranges are 100–300). Gaussian distributed EMCCD read noise  $r_{i,k} \sim N(0, \sigma_{rn}^2)$ ,  $\sigma_{rn} = 6$ , is added to  $\epsilon_{i,k}$  to give the final electron count for pixel  $k$  in frame  $i$ . This is divided by an analogue to digital conversion sensitivity of 3.2 to give the digital camera count. As in an EMCCD camera, a base offset of 100 digital camera counts is added to prevent clipping of negative numbers on digitisation and the final count is discretised and truncated to the range  $[0, 65535]$ .

### 1.F.2 Image analysis and trace idealisation

To generate idealised traces, photon count vs time traces were extracted from image sequences and thresholded as follows. The position of molecules in the Alexa Fluor 647 image data were as previously determined by [Lin et al. \(2015\)](#) and the positions of molecules used in simulated data were known. The position of each molecule was used to extract  $5 \times 5$  pixel regions from the raw data centred (without interpolation) on the molecule to produce a sequence of ‘trace images’. The background intensity of each frame of the trace images was calculated as the mean of the 16 boundary pixels and subtracted from the trace images. The photon number for each frame of the trace images was calculated by correcting for the photon conversion factor of the camera and subsequently integrating the convolution of the trace image with a  $5 \times 5$  pixel Gaussian kernel. A hard threshold was applied to the photon number per frame trace at multiples of the standard deviation of the background,  $\sigma_{BG}$ .

### 1.F.3 Global parameter set for simulations

The global parameter set used in all simulation studies, unless being varied as part of the study, are as follows:

Parameter	Value
Threshold	$5\sigma_{BG}$
Expected number of photons per frame length ( $\Delta\lambda_p$ )	500
$\Delta$	1/30 s
$\lambda_{01}$	$0.3162 \text{ s}^{-1}$
$\lambda_{10}$	$3.1623 \text{ s}^{-1}$
$\mu_{10}$	$0.1 \text{ s}^{-1}$
$T^*$	0 s
$\nu_1^*$	1
$N_F$	9872 frames

For the Alexa Fluor 647 data, a fixed threshold of  $2\sigma_{BG}$  was applied.

## 1.G Rate estimates

In this section, we provide tables to show the photo-switching rates estimated by the PSHMM and exponential fitting method for the simulation studies conducted in Section 1.5.1. 95% simulation percentile intervals are also shown.

Study	$N_F$	$\theta^{(d)}$	PSHMM		PSHMM RMSE ( $\times 10^{-2}$ )	PSHMM (2.5%, 97.5%) percentiles	Exp		Exp RMSE ( $\times 10^{-2}$ )	Exp (2.5%, 97.5%) percentiles
			Mean	Bias			Mean	Bias		
1	16800	0.32	0.32	0.00	0.97	(0.30, 0.34)	0.29	-0.029	3.29	(0.26, 0.32)
		0.32	0.32	-0.001	0.76	(0.30, 0.33)	0.31	-0.007	0.89	(0.30, 0.32)
		0.01	0.01	0.001	0.21	(0.01, 0.02)	0.01	0.001	0.11	(0.01, 0.01)
2	11151	0.32	0.32	-0.001	0.66	(0.30, 0.33)	0.30	-0.012	1.47	(0.29, 0.32)
		1	1.00	0.003	1.91	(0.96, 1.04)	0.95	-0.053	5.95	(0.89, 0.99)
		0.03	0.03	0.001	0.35	(0.03, 0.04)	0.03	0.001	0.33	(0.03, 0.04)
3	9364	0.32	0.31	-0.004	0.68	(0.30, 0.32)	0.30	-0.017	1.90	(0.28, 0.32)
		3.16	3.16	0.002	6.56	(3.05, 3.28)	2.45	-0.712	77.85	(1.78, 3.92)
		0.11	0.11	0.001	1.02	(0.09, 0.13)	0.09	-0.017	2.12	(0.06, 0.12)
4	8799	0.32	0.30	-0.013	1.40	(0.29, 0.31)	0.28	-0.032	3.35	(0.27, 0.30)
		10	9.96	-0.042	23.03	(9.52, 10.42)	3.19	-6.809	690.87	(1.52, 5.91)
		0.33	0.35	0.014	3.79	(0.29, 0.42)	0.12	-0.210	21.49	(0.06, 0.25)
5	10962	1	1.00	-0.002	1.86	(0.96, 1.04)	0.90	-0.104	12.43	(0.74, 1.01)
		0.32	0.32	0.000	0.72	(0.30, 0.33)	0.30	-0.013	1.47	(0.29, 0.32)
		0.01	0.01	0.000	0.10	(0.01, 0.01)	0.01	0.001	0.11	(0.01, 0.01)
6	5312	1	1.00	-0.004	1.81	(0.96, 1.03)	0.95	-0.054	6.44	(0.87, 1.01)
		1	1.00	0.001	1.76	(0.96, 1.04)	0.93	-0.066	6.88	(0.89, 0.97)
		0.03	0.03	0.001	0.29	(0.03, 0.04)	0.03	0.001	0.28	(0.03, 0.04)
7	3526	1	0.99	-0.015	2.32	(0.95, 1.02)	0.95	-0.053	5.78	(0.91, 0.99)
		3.16	3.17	0.003	7.33	(3.01, 3.30)	2.71	-0.451	46.16	(2.50, 3.89)
		0.11	0.11	0.001	1.06	(0.08, 0.13)	0.10	-0.004	0.99	(0.08, 0.12)
8	2961	1	0.97	-0.033	3.91	(0.93, 1.00)	0.91	-0.095	9.75	(0.85, 0.95)
		10	9.94	0.003	27.21	(9.46, 10.47)	5.02	0.003	504.34	(3.66, 6.52)
		0.33	0.35	0.017	3.94	(0.28, 0.42)	0.20	-0.133	13.77	(0.14, 0.27)
9	9116	3.16	3.15	-0.008	6.88	(3.04, 3.29)	2.31	-0.855	95.19	(1.64, 3.04)
		0.32	0.31	-0.002	1.53	(0.28, 0.34)	0.28	-0.037	3.74	(0.27, 0.29)
		0.01	0.01	0.000	0.11	(0.01, 0.01)	0.01	0.001	0.13	(0.01, 0.01)
10	3466	3.16	3.13	-0.035	7.47	(3.01, 3.28)	2.83	-0.335	37.76	(2.49, 3.06)
		1	1.00	0.004	4.04	(0.90, 1.07)	0.87	-0.129	13.00	(0.84, 0.90)
		0.03	0.03	0.001	0.37	(0.03, 0.04)	0.04	0.002	0.36	(0.03, 0.04)
11	1680	3.16	3.11	-0.052	9.49	(2.98, 3.31)	2.92	-0.245	25.63	(2.75, 3.08)
		3.16	3.18	0.015	9.12	(2.99, 3.37)	2.60	-0.567	56.96	(2.48, 3.70)
		0.11	0.11	0.002	1.21	(0.09, 0.13)	0.11	-0.000	0.98	(0.09, 0.13)
12	1115	3.16	3.03	-0.135	14.73	(2.92, 3.15)	2.79	-0.377	38.19	(2.66, 3.92)
		10	9.99	-0.008	24.86	(9.54, 10.48)	6.35	-3.648	365.97	(5.72, 6.93)
		0.33	0.35	0.015	3.92	(0.29, 0.44)	0.27	-0.061	6.79	(0.22, 0.33)
13	8532	10	9.93	-0.069	25.64	(9.47, 10.47)	5.33	-4.666	506.70	(1.98, 8.60)
		0.32	0.32	0.000	4.42	(0.24, 0.37)	0.22	-0.099	9.95	(0.21, 0.23)
		0.01	0.01	0.000	0.09	(0.01, 0.01)	0.01	0.001	0.10	(0.01, 0.01)
14	2882	10	9.86	-0.142	29.52	(9.44, 10.39)	7.75	-2.246	241.85	(5.53, 8.71)
		1	1.03	0.026	10.53	(0.78, 1.16)	0.68	-0.323	32.37	(0.64, 0.71)
		0.03	0.03	0.001	0.36	(0.03, 0.04)	0.04	0.001	0.37	(0.03, 0.04)
15	1096	10	9.73	-0.266	38.40	(9.22, 10.34)	8.19	-1.814	184.84	(7.38, 8.66)
		3.16	3.21	0.049	20.73	(2.45, 3.47)	2.05	-1.108	110.97	(1.97, 3.16)
		0.11	0.11	0.004	1.22	(0.09, 0.13)	0.11	0.004	1.04	(0.09, 0.13)
16	531	10	9.50	-0.501	55.72	(9.10, 9.96)	7.93	-2.072	207.90	(7.55, 8.22)
		10	9.91	-0.095	54.47	(9.02, 10.88)	5.63	-4.368	436.96	(5.40, 5.89)
		0.33	0.34	0.007	4.51	(0.26, 0.43)	0.30	-0.029	4.06	(0.26, 0.36)

Table 1.3: Caption next page.

**Table 1.3:** Simulation results showing mean, bias, root mean squared error (RMSE) and the 2.5 and 97.5 empirical percentiles of the estimates of  $\theta^{(0)} = (\lambda_{01} \ \lambda_{10} \ \mu_1)^\top$  under model  $\mathcal{M}_{\{1\}}^0$  for both the PSHMM and exponential fitting (Exp) methods across 100 repeat experiments.  $\Delta = \frac{1}{30}s$ ,  $\delta, \omega > 0$  (unknown), and  $M = 100$ . For both methods, log – log scatterplots of  $\lambda_{01}$  and  $\lambda_{10}$  are shown in Figure 1.8.

Study	$N_F$	$\theta^{(d)}$	PSHMM	PSHMM	PSHMM	PSHMM	Exp	Exp	Exp	Exp
			Mean	Bias	RMSE ( $\times 10^{-2}$ )	(2.5%, 97.5%) percentiles	Mean	Bias	RMSE ( $\times 10^{-2}$ )	(2.5%, 97.5%) percentiles
17	11151	0.15	0.15	0.002	1.69	(0.12, 0.19)	0.15	-0.004	1.66	(0.11, 0.19)
		0.3	0.30	0.001	0.85	(0.28, 0.32)	0.30	-0.002	0.84	(0.28, 0.31)
		0.1	0.10	0.000	0.43	(0.09, 0.11)	0.10	0.002	0.45	(0.10, 0.11)
		0.80	0.80	-0.001	1.28	(0.78, 0.82)	0.76	-0.039	4.12	(0.74, 0.79)
		0.01	0.01	0.000	0.15	(0.01, 0.01)	0.02	0.010	0.97	(0.02, 0.02)
18	9364	0.35	0.36	0.005	5.44	(0.24, 0.43)	0.33	-0.022	5.32	(0.24, 0.43)
		1	1.00	0.003	3.68	(0.94, 1.07)	0.95	-0.049	5.83	(0.90, 1.01)
		0.3	0.30	-0.002	2.01	(0.26, 0.34)	0.29	-0.008	2.12	(0.25, 0.33)
		2.30	2.30	-0.003	5.01	(2.21, 2.39)	2.04	-0.262	26.48	(1.95, 2.11)
		0.10	0.10	0.002	0.98	(0.09, 0.12)	0.10	-0.005	1.03	(0.08, 0.11)
19	7000	2	2.03	0.033	18.14	(1.75, 2.45)	2.16	0.156	21.25	(1.89, 3.50)
		10	9.78	-0.218	54.49	(8.55, 10.53)	6.94	-3.061	306.69	(6.59, 7.34)
		0.7	0.71	0.011	4.88	(0.64, 0.83)	0.67	-0.031	4.85	(0.60, 0.76)
		10	10.00	0.002	63.62	(9.22, 11.65)	4.97	-5.030	503.14	(4.75, 5.17)
		0.33	0.34	0.005	7.29	(0.20, 0.56)	0.27	-0.068	7.30	(0.22, 0.32)

**Table 1.4:** Simulation results showing mean, bias, root mean squared error (RMSE) and the 2.5 and 97.5 empirical percentiles of the estimates of  $\theta^{(1)} = (\lambda_{00_1} \ \lambda_{01} \ \lambda_{0_1} \ \lambda_{10} \ \mu_1)^\top$  under model  $\mathcal{M}_{\{1\}}^1$  for both the PSHMM and exponential fitting (Exp) methods across 100 repeat experiments.  $\Delta = \frac{1}{30}s$ ,  $\delta = 0.01s$ ,  $\omega > 0$  (unknown), and  $M = 100$ .

Study	$N_F$	$\theta^{(d)}$	PSHMM	PSHMM	PSHMM	PSHMM	Exp	Exp	Exp	Exp
			Mean	Bias	RMSE ( $\times 10^{-2}$ )	(2.5%, 97.5%) percentiles	Mean	Bias	RMSE ( $\times 10^{-2}$ )	(2.5%, 97.5%) percentiles
20	7000	2	2.03	0.032	3.14	(1.75, 2.37)	2.05	0.054	2.12	(1.79, 3.31)
		10	9.85	-0.153	13.42	(9.20, 10.49)	7.04	-2.958	878.82	(6.68, 7.48)
		0.2	0.21	0.009	0.10	(0.16, 0.27)	0.18	-0.024	0.11	(0.13, 0.21)
		0.7	0.69	-0.012	0.37	(0.59, 0.83)	0.66	-0.037	0.35	(0.57, 0.75)
		0.01	0.01	-0.001	0.00	(0.01, 0.01)	0.01	0.005	0.02	(0.01, 0.02)
		10	9.63	-0.368	35.57	(8.73, 10.53)	4.91	-5.087	2588.85	(4.67, 5.16)
		0.33	0.32	-0.009	0.32	(0.24, 0.45)	0.32	-0.013	0.12	(0.26, 0.38)

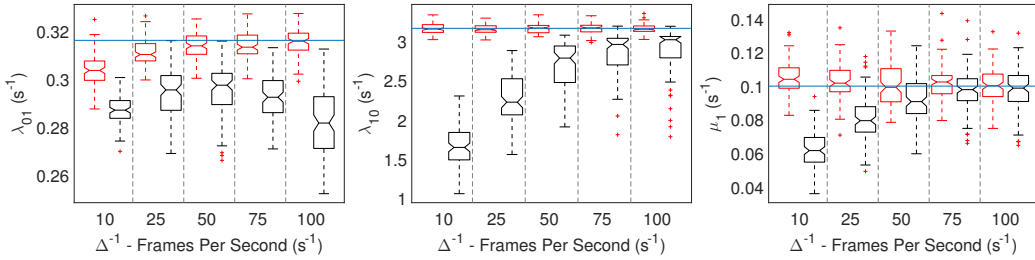
**Table 1.5:** Simulation results showing mean, bias, root mean squared error (RMSE) and the 2.5 and 97.5 empirical percentiles of the estimates of  $\theta^{(2)} = (\lambda_{00_1} \ \lambda_{01} \ \lambda_{0_1} \ \lambda_{0_1} \ \lambda_{0_1} \ \lambda_{0_1} \ \lambda_{10} \ \mu_1)^\top$  under model  $\mathcal{M}_{\{1\}}^2$  for both the PSHMM and exponential fitting (Exp) methods across 100 repeat experiments.  $\Delta = \frac{1}{30}s$ ,  $\delta = 0.01s$ ,  $\omega > 0$  (unknown), and  $M = 100$ .

## 1.H Further results

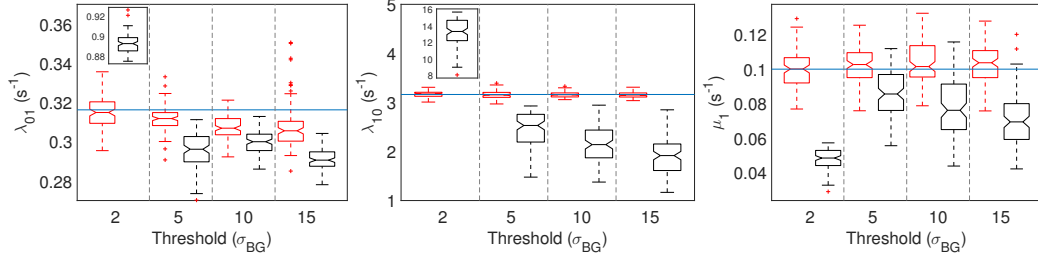
In this section, we present further simulation studies that explore the effects of varying the frame length  $\Delta$ , threshold (proportional to  $\delta$ ), number of frames  $N_F$  and absorption parameter  $\mu_1$  under the model  $\mathcal{M}_{\{1\}}^0$ .

We analysed simulated image traces (see Section 1.F of Appendix A) based on a range of different scenarios. All datasets were created with  $M = 100$  with  $T^* = 0$ , known initial probability  $\boldsymbol{\nu}_X^* = \begin{pmatrix} 0 & 1 & 0 \end{pmatrix}^\top$  and unknown false positive rate  $\omega > 0$ . In Section 1.5.1, we explored the effect of different photo-switching rates  $(\lambda_{01}, \lambda_{10})$ . Here, we consider varying the frame length  $\Delta$  (Figure 1.11), threshold (proportional to  $\delta$ ) (Figure 1.12), number of frames  $N_F$  (Figure 1.13) and absorption parameter  $\mu_1$  (Figure 1.14).

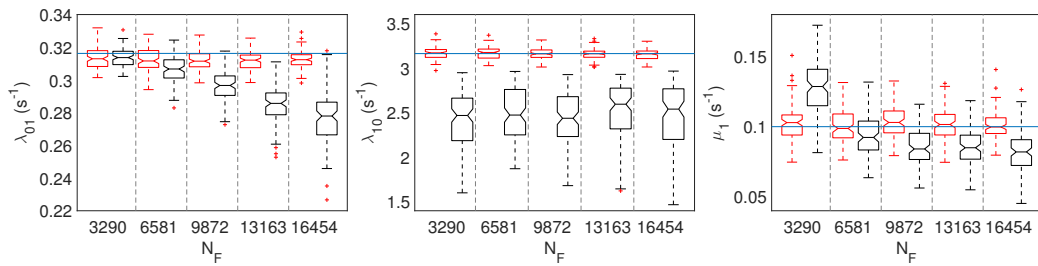
Our PSHMM method is seen to perform extremely well across all different switching and sampling scenarios, especially in comparison to the exponential fitting method which incurs a consistent bias. It is worth noting that we see an increase in bias of the PSHMM estimator for low frame rates and high threshold values. This is expected as its ability to detect multiple transitions within a frame is diminished. Furthermore, it is indicated that the PSHMM estimator also exhibits a much lower variance than that from the exponential fitting, a property which also decreases with larger  $M$  (data not shown). Figure 1.14 finally highlights a reduction in bias with  $\mu_1$ , due to a greater number of transitions between hidden states.



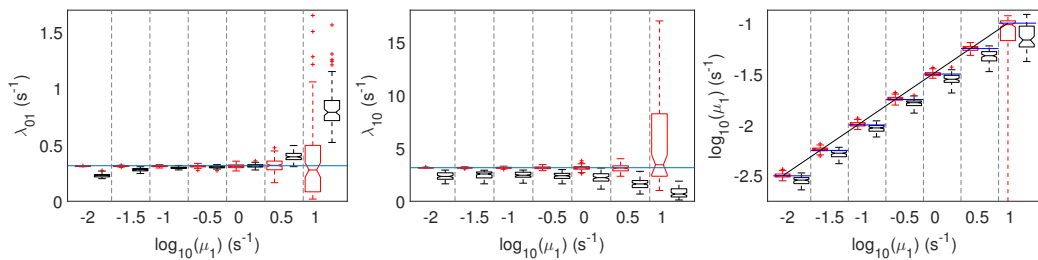
**Figure 1.11:** Boxplots showing quantiles from estimates of  $\lambda_{01}$ ,  $\lambda_{10}$  and  $\mu_1$  from both exponential fitting (black) and PSHMM fitting (red) are plotted against  $\log(\Delta)$ . True rates given by the blue line.



**Figure 1.12:** Boxplots showing quantiles from estimates of  $\lambda_{01}$ ,  $\lambda_{10}$  and  $\mu_1$  from both exponential fitting (black) and PSHMM fitting (red) are plotted against  $\log(\text{threshold})$ .  $N_F = 9872$  for all simulations. True rates given by the blue line.



**Figure 1.13:** Boxplots showing quantiles from estimates of  $\lambda_{01}$ ,  $\lambda_{10}$  and  $\mu_1$  from both exponential fitting (black) and PSHMM fitting (red) are plotted against  $N_F$ . True rates given by the blue line.



**Figure 1.14:** Boxplots showing quantiles from estimates of  $\lambda_{01}$ ,  $\lambda_{10}$  and varying  $\mu_1$  from both exponential fitting (black) and PSHMM fitting (red) are plotted against  $\log_{10}(\mu_1)$ . True rates given by the blue lines.



# 2

## PROPERTIES OF THE PHOTO-SWITCHING HIDDEN MARKOV MODEL ESTIMATOR

### 2.1 Introduction

In Chapter 1, we proposed a kinetic model to describe the photo-switching behaviour of fluorophores, and derived the PSHMM to estimate the unknown parameters of the model via maximum likelihood estimation. Although we showed through a set of simulation studies that this estimator can accurately determine the unknown photo-switching rates, small biases and variances can be observed when dealing with data that come from wider classes of models that stem from different imaging parameters. Therefore, to investigate the versatility of the PSHMM estimator under a range of different experimental conditions, it is crucial that its inherent properties are well studied. This is especially highlighted for the main application of the PSHMM estimator, since datasets that are likely to be corrupted by noise may cause certain unknown parameters to be unidentifiable in the corresponding analysis. For

example, situations whereby fluorophores are observed in high background noise may enable a limited number of observations to be recorded, resulting in data containing too few positive detections to properly identify the switching rates between the On and dark states.

It is well understood that highly noisy data can make parameter estimation and identification more difficult (Godfrey and DiStefano, 1985, Little et al., 2010). In the PSHMM, the noise comes in two forms. Firstly, the inclusion of  $\delta \in [0, \Delta)$ , the maximum time whereby a fluorophore in an On state would not be detected, results from a noise floor that the fluorophore signal needs to penetrate. Secondly, the inclusion of  $\omega \in [0, 1]$ , the false positive probability of observing a fluorophore when it is in reality not detected, results from random spikes of photon intensities due to background noise. On the other hand, the unknown number of multiple dark states  $d$  prevents transitions between them to be observed, and therefore results in the lack of information in the data since longer times spent in dark states result in null detections of a fluorophore.

The inclusion of  $\omega$  in the PSHMM is generally not seen to be problematic since advanced imaging abilities and segmentation algorithms (Henriques et al., 2010, Ovesný et al., 2014, Boyd et al., 2018) ensure that this probability is extremely low in practice. Nevertheless, as  $\delta \rightarrow \Delta$  and as  $d$  increases, the sparsity of the observation vector  $\mathbf{y}$  increases. In this scenario, we would assume parameter estimation/identification problems due to lack of high quality informative data on the hidden process  $\{X(t)\}$ , whereby the accretion of data may not necessarily result in better parameter estimation. Furthermore, while a set of parameters may be strictly identifiable, properties such as correlations between their estimates and multi-modality of the likelihood function may be such as to make numerical estimation troublesome (Jacquez and Greif, 1985).

Identifiability issues may further have an adverse effect on the large sample properties of the PSHMM estimator, namely that of *consistency*. The estimator being consistent would imply that it is possible to (within an arbitrary precision) recover the true parameters (being estimated) with a sufficiently large enough dataset. When designing potential imaging experiments to

recover photo-switching rates, it is therefore also important to understand what the effect of attaining more data has on the accuracy and validity of the resulting parameter estimates.

### 2.1.1 Chapter summary

This chapter will discuss and analyse the quality of the PSHMM estimates. The fact that it is not easily feasible to analytically tract the PSHMM log-likelihood function (with respect to the parameter vector  $\boldsymbol{\theta}^{(d)}$ ), poses difficulties in theoretically analysing the inherent properties of this estimator; therefore, the analyses made in this chapter will be wholly supported by simulation studies.

In Section 2.2, we begin by giving the formal definition of model identifiability and deduce that theoretical justification of parameter identification in the PSHMM is not feasible. To this end, in Section 2.2.1 we define the notion of *local identifiability*, which is characterised by the Hessian or *observed Fisher Information* matrix evaluated (locally) on the log-likelihood surface. Using simulation studies, we then evaluate potential identification issues through the analyses of these matrices evaluated at maximum likelihood estimates, including a short study verifying that the Hessian matrices generally appear non-singular. Since identification issues may arise due to ridges appearing on the log-likelihood surface, in Section 2.2.2 we first compare theoretical correlation structures between parameters with empirical correlations and identify pairs of parameters which appear to become unidentifiable/correlated as the noise parameter  $\delta \rightarrow \Delta$ . We then explore the log-likelihood surface between these parameters through contour plots and deduce that this characteristic does not greatly hinder estimation. In Section 2.2.3, we then discuss the issue of *multi-modality*, which although appears problematic when  $d = 2$ , is shown to diminish as the number of data points increases. Finally in Section 2.2.4, we explore the effect of the frame length  $\Delta$  and demonstrate that parameter identification issues effectively disappear (even with low signal to noise ratios) as the sampling rate  $\Delta \rightarrow 0$ .

In Section 2.3, we study consistency of the PSHMM estimator by examining

the convergence of mean-square errors in further simulation studies which vary the number of imaged molecules  $M$ . Here, we also find that the pairs of parameters that were previously deemed difficult to identify are also shown to be inconsistent when the signal to noise ratio decreases.

The results presented have been collected from nine further simulation studies that were conducted exclusively for this chapter. In particular, we investigate the effects of slow, medium and fast switching parameters on the three models  $\mathcal{M}_{\{1\}}^0$ ,  $\mathcal{M}_{\{1\}}^1$  and  $\mathcal{M}_{\{1\}}^2$ . The nine studies, which will be referenced to throughout this chapter, are driven by their parameter set-ups as is presented in Table 2.1 and were simulated by directly sampling the continuous time processes  $\{X(t)\}$  and extracting the observation sequences  $\mathcal{Y}$  as in (1.2). All studies were executed with  $\omega$  (false positive rate) equal to zero, as this was found to be a good approximation for experimental performance. Unless stated otherwise, the values shown of  $\boldsymbol{\theta}^{(d)}$ ,  $\Delta^{-1}$ ,  $N_F$  and  $M$  in this table are the simulation parameters used for each study. Any change made in one or more of these values in the subsequent analyses, will be clearly highlighted.

## 2.2 Identification of model parameters

Formally, identifiability in the PSHMM means that the log-likelihood (1.31) is an injective function from the parameter space  $\boldsymbol{\theta}^{(d)}$  to the space of distributions  $\Theta^{(d)}$  for the data; this is stated more generally in Definition 3. It should be clear that since it is not feasible to analytically tract the log-likelihood with respect to  $\boldsymbol{\theta}^{(d)}$ , obtaining such a theoretical result for the PSHMM is highly non-trivial. This motivates our exploration of the issue through empirical studies.

**Definition 3** (Little et al. (2010)). *Let  $\boldsymbol{\theta}^{(d)} \in \Theta^{(d)}$  be the parameter vector of interest in the PSHMM as defined in (1.3). If  $\mathcal{Y}$  denotes data matrix, then  $\boldsymbol{\theta}^{(d)}$  is said to be identifiable if there exists no other  $\boldsymbol{\theta}^{*(d)} \in \Theta^{(d)}$  such that  $\ell_{\boldsymbol{\theta}^{(d)}}(\mathcal{Y}) = \ell_{\boldsymbol{\theta}^{*(d)}}(\mathcal{Y})$  almost everywhere.*

We will begin by exploring the notion of *local identifiability*, which can be quantified by analyses of Fisher information matrices. Through simulation

studies, we will first show that the correlation between the parameters  $\lambda_{01}$  and  $\delta$  increases as  $\delta/\Delta \rightarrow 1$ . Here, there also becomes a clear mismatch between the sample correlation matrix and that derived from the Fisher information matrix. This property is seen to be heightened as the photo-switching becomes faster, i.e. as the transitions between all states in the state space  $\mathcal{S}_X$  become more rapid, from which the curvature of the log-likelihood surface is also seen to decrease. When  $\delta/\Delta$  is low (or when the signal to noise ratio is high), the correlation between  $\lambda_{01}$  and  $\delta$  is low and there is strong agreement between the empirical correlation matrices and those estimated by the Hessian matrices, indicating the unknown parameters of the model are in practically identifiable. Contour plots will be shown to support this statement.

We will also explore the effects of changing the sampling time  $\Delta$ . As  $\Delta \rightarrow 0$ , we would expect greater identification of the photo-switching parameters in  $\{X(t)\}$  as the observed process  $\{Y_n\}$  more closely aligns with the hidden process. For the same photo-switching parameters, therefore, we should see better estimation of unknown parameters as  $\Delta$  decreases. We will first show that the previously highlighted correlation structures start to diminish, and second that estimates from selected simulation studies (especially those with faster switching parameters) become unbiased as  $\Delta \rightarrow 0$ , for all values of  $\delta/\Delta$  studied.

### 2.2.1 Local identifiability

**Definition 4** (Rothenberg (1971)). *Parameter vector  $\boldsymbol{\theta}^{(d)}$  of the PSHMM is said to be locally identifiable if and only if the Fisher information matrix  $\mathcal{I}(\boldsymbol{\theta}^{(d)}) = \mathbb{E}_{\boldsymbol{\theta}^{(d)}} \left( \frac{\partial}{\partial \boldsymbol{\theta}^{(d)}} \ell_{\boldsymbol{\theta}^{(d)}}(\mathcal{Y}) \right)$  is non-singular.*

The definition of local identifiability as stated in Definition 4 implies that that  $\boldsymbol{\theta}^{(d)}$  is locally identifiable when all of the eigenvalues of its Fisher information matrix  $\mathcal{I}(\boldsymbol{\theta}^{(d)})$  are non-zero. A singular Fisher information matrix (with at least one eigenvalue equal to zero) implies there does not exist a unique maximum to the likelihood function (typically due to a flat ridge) and the model is locally unidentifiable (Little et al., 2010). Due to the complexity

Parameter	$d$	$\lambda_{001}$ ( $s^{-1}$ )	$\lambda_{01}$ ( $s^{-1}$ )	$\lambda_{0102}$ ( $s^{-1}$ )	$\lambda_{011}$ ( $s^{-1}$ )	$\lambda_{021}$ ( $s^{-1}$ )	$\lambda_{10}$ ( $s^{-1}$ )	$\mu_1$ ( $s^{-1}$ )	$\Delta^{-1}$ ( $s^{-1}$ )	$\nu_1^*$	$M$	$N_F$
Study												
1 (SLOW)	0		0.3162				1	0.0333	30	1	100	$10^4$
2 (MEDIUM)	0		1				3.162	0.1054	30	1	100	$10^4$
3 (FAST)	0		3.162				10	0.333	30	1	100	$10^4$
4 (SLOW)	1	0.15	0.3		0.1		0.8	0.01	30	1	100	$10^4$
5 (MEDIUM)	1	0.35	1		0.3		2.3	0.1	30	1	100	$10^4$
6 (FAST)	1	2	10		0.7		10	0.333	30	1	100	$10^4$
7 (SLOW)	2	0.15	0.3	0.05	0.1	0.001	0.8	0.05	30	1	100	$10^4$
8 (MEDIUM)	2	0.8	4	0.1	0.4	0.005	8	0.1	30	1	100	$10^4$
9 (FAST)	2	2	10	0.2	0.7	0.01	10	0.333	30	1	100	$10^4$

**Table 2.1:** Global parameter values for the stimulation studies conducted in this section. All studies have been conducted with the model format of  $M_{\{1\}}^d$  for  $d = 0, 1, 2$ , and with  $T^* = 0$ .

of the model, the Fisher information matrix can not be computed, however we can study local identifiability via the observed Fisher information matrix (the Hessian matrix of the log-likelihood function)

$$\mathcal{J}(\hat{\boldsymbol{\theta}}^{(d)}) = -\nabla\nabla^\top \ell_{\boldsymbol{\theta}^{(d)}}(\mathcal{Y})|_{\boldsymbol{\theta}^{(d)}=\hat{\boldsymbol{\theta}}^{(d)}},$$

evaluated at the maximum likelihood estimate  $\hat{\boldsymbol{\theta}}^{(d)}$  of  $\boldsymbol{\theta}^{(d)}$  (Colquhoun et al., 2003), which can be computed using the method of finite differences. This is averaged over several repeated simulations of data set  $\mathcal{Y}$ .

### Singularity analysis

The threshold of Viallefont et al. (1998) states that if the smallest eigenvalue  $\xi_{\min}$  of  $\mathcal{J}(\hat{\boldsymbol{\theta}}^{(d)})$  is larger than  $q \cdot \xi_{\max} \cdot 1 \times 10^{-9}$ , where  $q$  is the dimension of  $\boldsymbol{\theta}^{(d)}$  and  $\xi_{\max}$  is the largest eigenvalue, then it can be considered to be non-zero and hence  $\mathcal{J}(\hat{\boldsymbol{\theta}}^{(d)})$  non-singular. Adopting this threshold, 500 independent datasets were generated for various values of  $\delta$  for each study 1 – 9 as described in Table 2.1. In all cases the Hessian matrices were determined to be non-singular, and hence consider the model identifiable, with the exception of 0.33% of the Hessian matrices for Study 6 (when  $\delta \geq 0.3$ ), 0.073%

of the Hessian matrices for Study 7, and 0.018% of the Hessian matrices computed for Study 9. In all cases the average smallest eigenvalue was considerably larger than the threshold. This provides compelling evidence that the PSHMM model is locally identifiable for all parameter values studied.

### 2.2.2 Correlations

In order to test whether parameters are independently identifiable, we analyse the correlation structure between estimates of the model parameters. In particular, high correlations between particular sets of parameters highlight dependencies not constructed by the model, indicating that these parameters are individually unidentifiable and hence troublesome for numerical optimisation procedures. As previously noted, when  $\delta/\Delta \rightarrow 1$ , we would expect an increase in the sparsity of the datasets, which may encourage correlation structures to transpire.

Correlation between parameter estimates can be analysed in two ways. The first is through the correlation matrix as derived from the observed Fisher information matrix, namely

$$R(\hat{\boldsymbol{\theta}}^{(d)}) = \text{diag}(\mathcal{J}(\hat{\boldsymbol{\theta}}^{(d)})^{-1})^{-\frac{1}{2}} \mathcal{J}(\hat{\boldsymbol{\theta}}^{(d)})^{-1} \text{diag}(\mathcal{J}(\hat{\boldsymbol{\theta}}^{(d)})^{-1})^{-\frac{1}{2}}.$$

This matrix estimates the theoretical correlation structure between parameter estimates based on curvature of the likelihood surface. The second way is to look at the sample correlation matrix based on the estimates themselves. Not only do each individually provide evidence of correlation, or lack thereof, between parameter estimates, but if the model parameters are identifiable, these two matrices should approximately align. Therefore, when the two mismatch this is evidence that the numerical optimisation procedure we deploy is struggling to properly identify the parameters (Colquhoun et al., 2003).

For each study listed in Table 2.1, 500 datasets were simulated. The sample correlation matrix was computed from these 500 sets of estimates and the values reported for the correlation derived from the Hessian matrix were averaged over these 500 datasets.

Tables 2.2, 2.3 and 2.4 in Section 2.A of Appendix 2 show the correlation estimates for studies 1,2 and 3 (parameters provided in Table 2.1) as  $\delta/\Delta$  increases to 1. The correlations obtained from the Hessians and those obtained empirically agree quite well even when  $\delta/\Delta$  is large. The parameters, however which are most affected are  $\lambda_{01}$  (the transition rate between the first dark and On states) and  $\delta$ , whose correlation increases as  $\delta/\Delta$  increases. While this pattern is seen for all three switching scenarios, the correlation coefficient  $\rho(\lambda_{01}, \delta)$  becomes significantly larger as the switching moves faster; for example, the empirical correlation reaches 0.9603 when  $\delta/\Delta = 0.7, 0.8$  under the fast switching scenario. With this example, larger differences between the correlation estimates can also be seen. A change in correlation can also be observed between  $\lambda_{10}$  and  $\delta$ , and  $\mu_1$  and  $\delta$ , whose correlations becomes increasingly negative as  $\delta/\Delta$  increases. While these correlations do not become close to  $-1$ , the correlation pattern observed with all photo-switching parameters and  $\delta$  suggest issues with model identification with large values of  $\delta/\Delta$ .

Tables 2.5-2.7 in Section 2.A of Appendix 2 show the correlation estimates for studies 4, 5 and 6, as  $\delta/\Delta$  increases to 1. Again, it is observed that the values obtained from the Hessians and those obtained empirically agree well and the parameters, which are most affected are  $\lambda_{01}$  and  $\delta$ , whose correlation is highest, especially under fast switching, when  $\delta/\Delta = 0.5, 0.6$ . This is mimicked by the poorer agreement between the two sets of values. On the other hand, one may observe the positive correlations  $\rho(\lambda_{00_1}, \lambda_{01})$  and  $\rho(\lambda_{00_1}, \lambda_{0_1,1})$  which are, under all values of  $\delta/\Delta$ , consistently around the 0.55 and 0.65 mark respectively for the slow switching, reducing to around 0.3 – 0.4 under the fast switching. These correlations are inherent to the photo-switching model, since estimates for photo-switching parameters  $\lambda_{00_1}, \lambda_{01}, \lambda_{0_1,1}$  are drawn from the same sequence of zeros in  $\{Y_n\}$ . These correlations inherent to the PSHMM are not affected by the noise parameter  $\delta$  and do not appear to pose issues to identifiability.

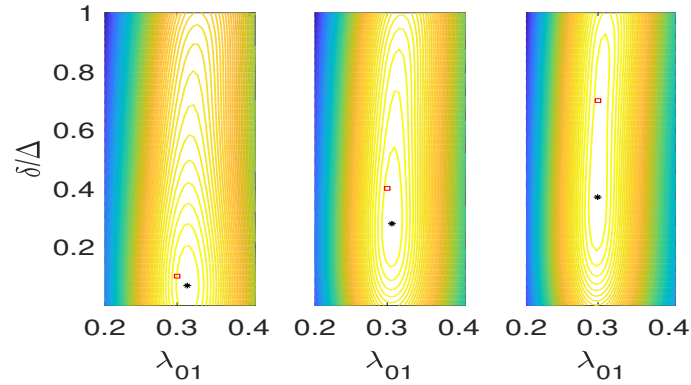
Tables 2.8-2.13 in Section 2.A of Appendix 2 show the correlation estimates for studies 7, 8 and 9. Under the medium and fast switching scenarios, it is noted that the correlations have a much better agreement than under the



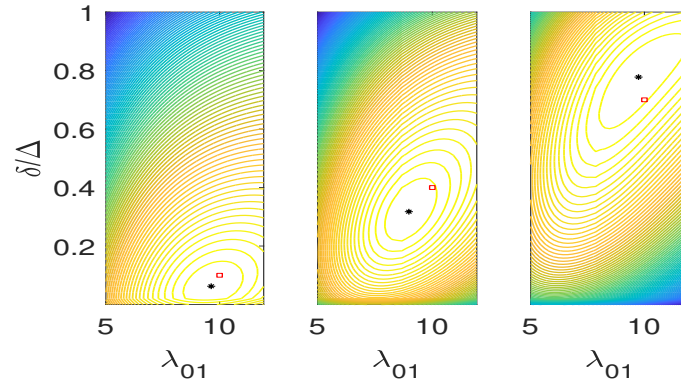
slow switching scenario. Most parameter pairs have seemingly low correlations with again the greatest affected pair being between  $\lambda_{01}$  and  $\delta$ , whose correlation is highest when  $\delta/\Delta = 0.6, 0.7$  for the medium switching and between  $0.4 - 0.7$  under fast switching. Similar to the  $d = 1$  case, the increased correlations between photo-switching rates from the multiple dark states ( $\lambda_{001}, \lambda_{01}, \lambda_{0102}, \lambda_{011}, \lambda_{021}$ ) is owed to the increased number of zeros in the datasets.

To look more closely at how the change in  $\delta$  affects the likelihood surface between  $\delta$  and its highest correlated parameters, Figures 2.1 and 2.2 show the likelihood contour surfaces for 3 different values of  $\delta$ ,  $\delta/\Delta = 0.1, 0.4, 0.7$ , under studies 4 and 6 respectively. For clearer comparisons, the  $\{X(t)\}$  data used to generate these surfaces are the same for all three values of  $\delta/\Delta$ . These contours show the change between  $\lambda_{01}$  and  $\delta$  about their maximum likelihood estimates with all other parameters of the model fixed at their maximum likelihood estimates. Figure 2.1 highlights the smaller correlation between  $\lambda_{01}$  and  $\delta$  with the contours appearing to extend vertically. Moreover, the maximum likelihood estimate for  $\lambda_{01}$  appears to be computed around the 0.3 (true value) mark for all ranges of  $\delta/\Delta$ .

This is in contrast to the faster switching scenario in Figure 2.2. The greater correlation and dependence between parameters can be observed through the contours lying in diagonal direction, which becomes more extreme as  $\delta/\Delta$  increases. While these closed contours show definitive maxima exist on the surfaces (it is structurally identifiable), the confidence area as depicted by the elliptical contours around the maximum likelihood estimates (plotted in asterisks) is seen to increase as  $\delta/\Delta$  increases. The larger area and flatness highlight the difficulty in correctly estimating/identifying these parameters. As previously stated, in the case of slow switching (Figure 2.1), the estimate of  $\lambda_{01}$  is not affected too badly by an increase in noise, however, the greater correlation does indeed affect this estimate in a faster switching scenario (Figure 2.2), as is depicted by a greater bias in the estimate.



**Figure 2.1:** Contour plots showing the log-likelihood surface of  $\lambda_{01}$  against  $\delta$  when  $d = 1$  under study 4 (of Table 2.1) with other rate parameters  $\lambda_{00_1}$ ,  $\lambda_{10}$  and  $\mu_1$  fixed at their maximum likelihood values. The left most plot shows the surface when  $\frac{\delta}{\Delta} = 0.1$ , the middle shows the surface when  $\frac{\delta}{\Delta} = 0.4$  and the rightmost when  $\frac{\delta}{\Delta} = 0.7$ . Maximum likelihood estimates are shown as asterisks (black) with the true values shown as squares (red). The process  $\{X(t)\}$  generating the data is the same for all three plots.

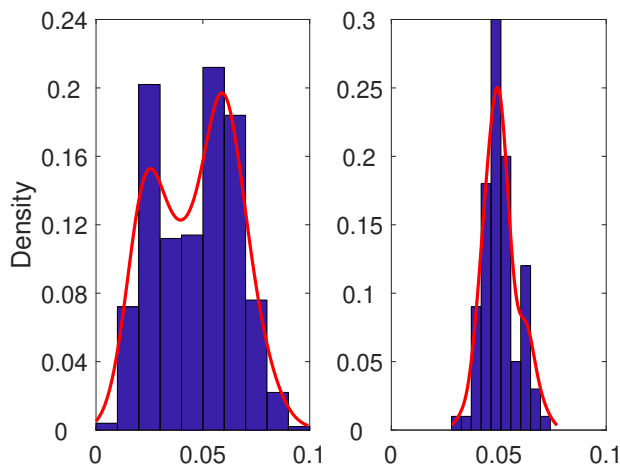


**Figure 2.2:** Contour plots showing the log-likelihood surface of  $\lambda_{01}$  against  $\delta$  when  $d = 1$  under study 6 (of Table 2.1) with other rate parameters  $\lambda_{00_1}$ ,  $\lambda_{10}$  and  $\mu_1$  fixed at their maximum likelihood values. The left most plot shows the surface when  $\frac{\delta}{\Delta} = 0.1$ , the middle shows the surface when  $\frac{\delta}{\Delta} = 0.4$  and the rightmost when  $\frac{\delta}{\Delta} = 0.7$ . Maximum likelihood estimates are shown as asterisks (black) with the true values shown as squares (red). The process  $\{X(t)\}$  generating the data is the same for all three plots.

### 2.2.3 Multi-modality

While multi-modality on a likelihood surface does not always necessarily relate to local identifiability, it is important to assess whether the existence of multiple modes can cause the optimisation method to wrongly determine the global maxima. We refer to this as a *global identifiability* issue. In Chapter

1, specifically Section 1.D of Appendix 1, we discussed how a computational implementation of the PSHMM optimiser should be initialised. For the 3 state case ( $d = 0$ ), an approximation scheme is used to find a suitable starting point, and for the simulations conducted in this thesis has always located the correct mode. In the 4 and 5 state cases ( $d = 1$  and  $d = 2$ , respectively), a stochastic search method is deployed that trials multiple starting points. Unimodal histograms for the parameter estimates would indicate a single global maximum, whereas a multi-modal histogram would indicate further dominant modes being located instead. We found a unimodal distribution for the estimates of all parameters under all simulation studies (figures omitted) with one exception. Figure 2.3 shows histograms of estimates of  $\lambda_{0_1 0_2}$  in Study 7 (slow switching) the 5 state ( $d = 2$ ) model. When the number of frames  $N_F = 10,000$ , it is clear that the optimisation procedure is finding two dominant modes. It should be noted however, that this global identifiability issue diminishes as  $N_F$  increases to 50,000, and the surface becomes unimodal.



**Figure 2.3:** Rate estimates for  $\lambda_{0_1 0_2} = 0.05$  under the slow switching scenario when  $d = 2$ , estimates show multi-modality when  $N_F = 10^4$  (left) and uni-modality when  $N_F = 5 \times 10^4$ .

## 2.2.4 The effect of frame length

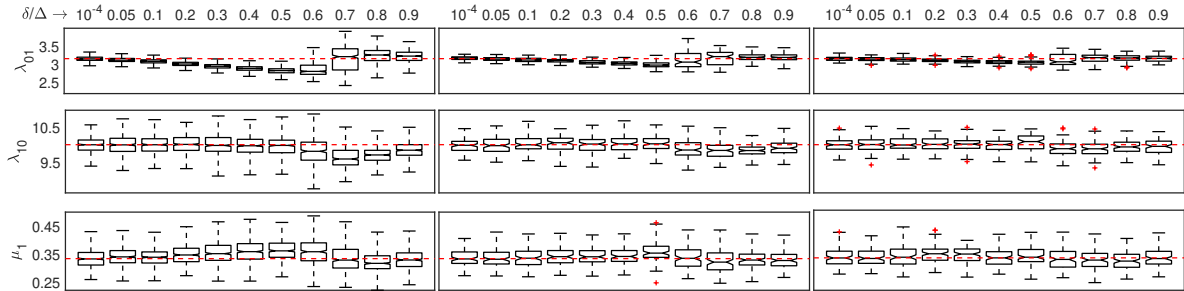
In order to test the effect of decreasing the frame length  $\Delta$  (or increasing the frame rate) on estimates, we simulated 500 datasets for the fast switching

scenarios (studies 3, 6 and 9 in Table 2.1) under the models  $d = 0, 1, 2$ . The fast switching parameters were used in this analysis due to the fact that the correlation structures between the noise parameter  $\delta$  and photo-switching parameters were the highest, as is explained in Section 2.2.2. In particular, small values of  $\Delta$  should yield less bias within the estimates and any identifiability issues in parameters should diminish as the hidden process  $\{X(t)\}$  is more closely seen. To address this issue, we show rate estimates for three sets of imaging parameters  $\Delta^{-1} = 30, 65, 100\text{s}^{-1}$  from datasets imaging  $M = 100$  molecules over a period of  $3 \times 10^5$  seconds.

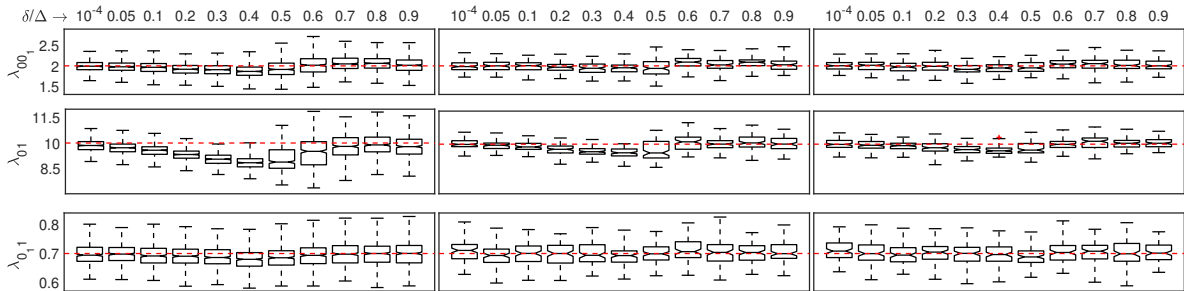
Figures 2.4 - 2.8 show box-plots of rate results for all simulations. Under all models, it is evident that rate estimates become unbiased as  $\Delta \rightarrow 0$ , under most values of the noise fraction  $\delta/\Delta$ . For instance, the bias in estimates that is highlighted most strongly is when  $\delta/\Delta \approx 0.4 - 0.5$ , for most cases. Although the same pattern in bias is seen across all three values of  $\Delta$ , this bias is significantly seen to decrease. This intuitively reflects the fact that  $\Delta \rightarrow 0$  implies that the observed process  $\{Y_n\}$  provides greater information of the process  $\{X(t)\}$ , with the  $\Delta, \delta \rightarrow 0$  case becoming completely informative of the hidden process. When  $\Delta^{-1} = 30\text{s}^{-1}$  for the  $d = 2$  case for example, bias in the parameters  $\lambda_{01} = \lambda_{10} = 10$  even at low noise floors is observed. This bias dissipates as the sampling time decreases, implying that there are identifiability issues for faster switching parameters if the sampling rate  $\Delta^{-1}$  is not large enough to capture the transitions between On and dark states.

One may also observe that while the bias in estimates decrease as  $\delta/\Delta$  ranges from  $10^{-4} - 0.6$ , that in some cases the estimator begins to improve and recover the true rates towards  $\delta/\Delta = 0.9$ . This is most notably seen for the  $\lambda_{01}$  and  $\lambda_{10}$  parameters in the  $d = 0, 1$  cases (Figures 2.4-2.6) and  $\lambda_{10}$  in the  $d = 2$  case (Figure 2.8). Intuitively, this effect may be due to the fact that in this situation,  $\{X(t)\}$  is likely to *only* be occupying the On state within each frame in which an observation is collected. Here, the PSHMM may first, be able to learn the distribution of holding times in the On state, and second, inherently be able to identify the frames in which transitions from a dark to an On state occur (before a subsequent observation of a fluorophore). Depending on the amount of data collected, this may therefore enable better

estimation of the parameters  $\lambda_{01}, \lambda_{10}$  for larger values of  $\delta/\Delta$ . However, it should be noted that in these situations, other parameter estimates (as seen in Figures 2.4 - 2.8) do not improve in the same manner. When considering the bias for all parameters estimated, we therefore observe and expect better rate estimation for all parameters for low values of  $\delta/\Delta$ .

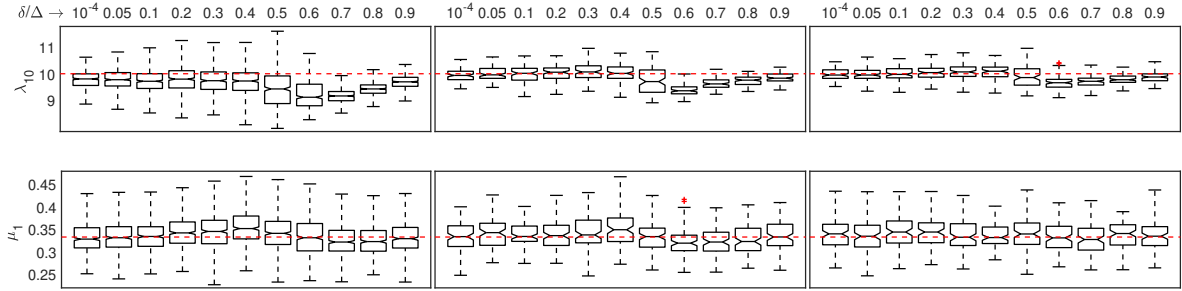


**Figure 2.4:** Box-plots showing rate estimates when  $d = 0$  under study 3 of Table 2.1 when  $\Delta^{-1} = 30$  (left),  $\Delta^{-1} = 65$  (middle) and  $\Delta^{-1} = 100$  (right). True rates indicated by the red dashed line.

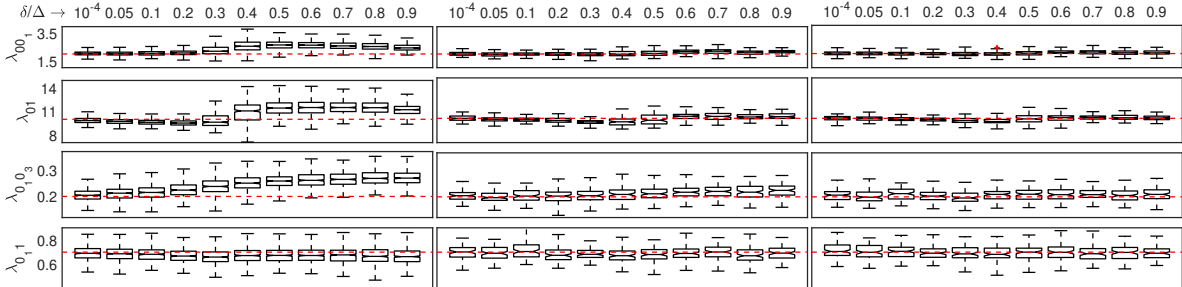


**Figure 2.5:** Box-plots showing rate estimates when  $d = 1$  under study 4 of Table 2.1 when  $\Delta^{-1} = 30$  (left),  $\Delta^{-1} = 65$  (middle) and  $\Delta^{-1} = 100$  (right). True rates indicated by the red dashed line.

To look more closely at how the change in  $\Delta$  affects the likelihood surface between  $\delta$  and its highest correlated parameter  $\lambda_{01}$ , Figure 2.9 shows the likelihood contour surfaces for 3 different values of  $\Delta$ ,  $\Delta^{-1} = 30, 65, 100s^{-1}$ , under study 6, for when  $\delta/\Delta = 0.7$ . Again, for clearer comparisons, the  $\{X(t)\}$  data used to generate these surfaces are the same for all three values of  $\delta/\Delta$ . These contours show the change between  $\lambda_{01}$  and  $\delta$  about their maximum likelihood estimates with all other parameters of the model fixed at their maximum likelihood estimates. One can observe that the correlation between  $\lambda_{01}$  and  $\delta$  reduces as  $\Delta$  decreases.



**Figure 2.6:** Box-plots showing rate estimates when  $d = 1$  under study 4 of Table 2.1 when  $\Delta^{-1} = 30$  (left),  $\Delta^{-1} = 65$  (middle) and  $\Delta^{-1} = 100$  (right). True rates indicated by the red dashed line.

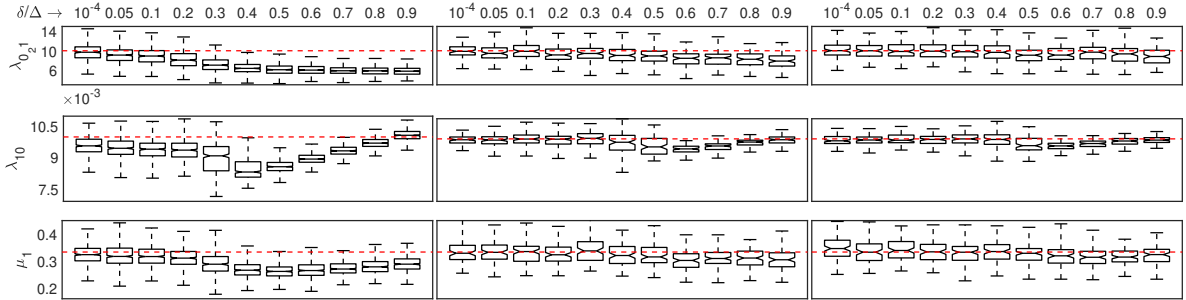


**Figure 2.7:** Box-plots showing rate estimates when  $d = 2$  under study 9 of Table 2.1 when  $\Delta^{-1} = 30$  (left),  $\Delta^{-1} = 65$  (middle) and  $\Delta^{-1} = 100$  (right). True rates indicated by the red dashed line.

### 2.2.5 Experimental impact

The analysis presented in this section indicates potential identification issues under three scenarios: when  $\delta/\Delta$  is large, when fluorophores are imaged in faster switching conditions (relative to the frame acquisition time  $\Delta$ ) and (specifically in the case of multiple dark states), when the data obtained are insufficiently large for rate estimation.

When designing fluorescence microscopy experiments, these results suggest that experimenters wishing to utilise a faster switching scenario during imaging should do so at shorter frame times, so as to improve subsequent rate estimation. On the other hand, while slower switching scenarios may be conducted across a larger range of frame times, experimenters should remain cautious as to the effect of high background noise  $\delta$ , and endeavour to conduct imaging under conditions that promote minimal background noise.



**Figure 2.8:** Box-plots showing rate estimates when  $d = 2$  under study 9 of Table 2.1 when  $\Delta^{-1} = 30$  (left),  $\Delta^{-1} = 65$  (middle) and  $\Delta^{-1} = 100$  (right). True rates indicated by the red dashed line.

In general, experiments that involve fluorophores that photo-switch between multiple dark states (larger values of  $d$ ) should additionally be imaged for a longer time than would be necessary for simpler models (i.e. when  $d = 0$ ).

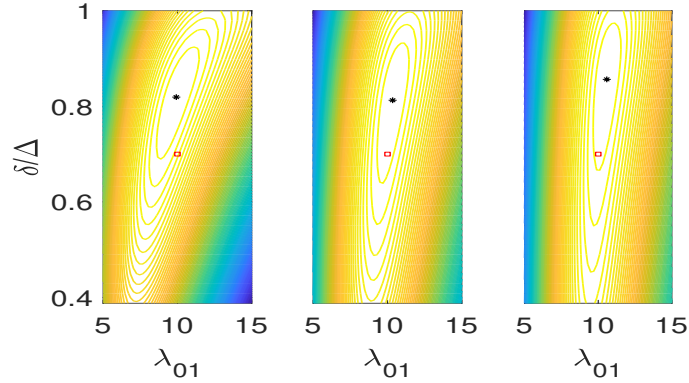
## 2.3 Consistency

In this section, we will use the identifiability analysis from the previous section to discuss when the PSHMM maximum likelihood estimator is consistent. Under studies 1-9, rate estimates will be computed as  $M$  is changed. In this case, the variance of the PSHMM estimator is shown to decrease as  $M$  increases, whilst consistency is shown through the convergence of mean-square errors only when  $\delta/\Delta$  is small and unknown transition rates are identifiable given the sampling time  $\Delta$ .

**Definition 5.** Let  $\mathcal{Y} = (\mathbf{y}^1 \ \mathbf{y}^2 \ \dots \ \mathbf{y}^M)$  be the data matrix, with each column being an  $N_F \times 1$  data vector from each independent imaging experiment. The PSHMM estimator  $\hat{\boldsymbol{\theta}}^{(d)} := \arg \max_{\boldsymbol{\theta}^{(d)} \in \Theta^{(d)}} \ell_{\boldsymbol{\theta}^{(d)}}(\mathcal{Y})$  of the true parameter vector  $\boldsymbol{\theta}^{*(d)}$  of the model is consistent if it converges in probability to  $\boldsymbol{\theta}^{*(d)}$ , i.e. that for any  $\epsilon > 0$ :

$$\lim_{N_F, M \rightarrow \infty} \mathbb{P}(|\hat{\boldsymbol{\theta}}^{(d)} - \boldsymbol{\theta}^{*(d)}| > \epsilon) = 0.$$

Specifically, the PSHMM estimator is consistent if it converges in probabil-



**Figure 2.9:** Contour plots showing the log-likelihood surface of  $\lambda_{01}$  against  $\delta/\Delta$  when  $d = 1$  under study 6 (of Table 2.1) with other rate parameters  $\lambda_{00_1}, \lambda_{10}$  and  $\mu_1$  fixed at their maximum likelihood values. The left most plot shows the surface when  $\Delta^{-1} = 30\text{s}^{-1}$ , the middle shows the surface when  $\Delta^{-1} = 65\text{s}^{-1}$  and the rightmost when  $\Delta^{-1} = 100\text{s}^{-1}$ . The true value for  $\delta/\Delta$  is 0.7 for each study. Maximum likelihood estimates are shown as asterisks (black) with the true values shown as squares (red). The process  $\{X(t)\}$  generating the data is the same for all three plots.

ity to the true parameter (see Definition 5). Again, consistency is difficult to show theoretically for the PSHMM estimator given the form of the log-likelihood function in (1.31). However, since convergence in mean square implies convergence in probability\*, we can loosely show that convergence in probability is attained by analysing if the mean-squared errors of parameter estimates tends to zero as  $M$  increases. From other studies, we have found that modifying the parameter  $N_F$  does not show any significant changes in bias/variance of the rate estimates due to the inclusion of the photo-bleaching state. In this section, we have therefore solely considered the effects of changing  $M$ .

Using the simulation studies 1–9 from the previous section (considered in Table 2.1), we were also able to compute the means and mean squared errors of the estimates, the results of which are shown for different  $\delta/\Delta$  in Tables 2.14 - 2.22 presented in Section 2.B of Appendix 2. To compare the mean squared

---

\*Let  $T_n$  be a sequence of random variables defined on a sample space  $\Omega$ , then  $T_n$  is convergent in mean square to  $T$  (also defined on  $\Omega$ ) if and only if  $\lim_{n \rightarrow \infty} \mathbb{E}((T_n - T)^2) = 0$ . By Markov's inequality, for any  $\epsilon > 0$  :  $\mathbb{P}((T_n - T)^2 \geq \epsilon^2) \leq \frac{\mathbb{E}((T_n - T)^2)}{\epsilon^2} \implies \lim_{n \rightarrow \infty} \mathbb{P}((T_n - T) \geq \epsilon) \leq \lim_{n \rightarrow \infty} \frac{\mathbb{E}((T_n - T)^2)}{\epsilon^2} = 0$ , which shows that convergence in mean square implies convergence in probability.

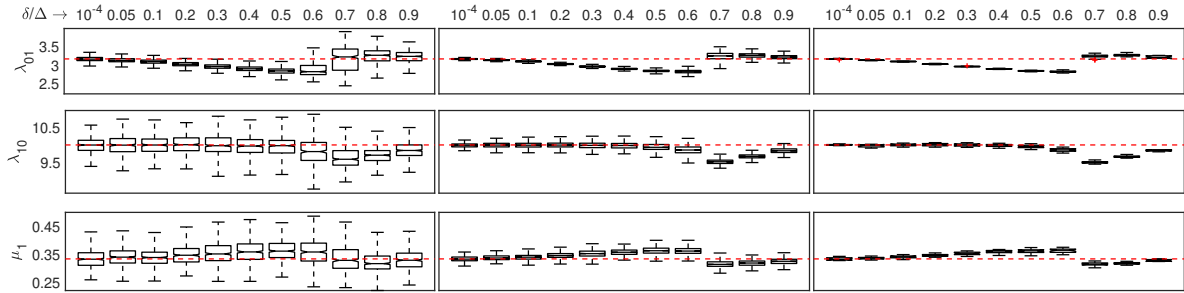


errors as the sample size  $M$  grows, we additionally show these errors under the same simulation studies executed with  $M = 1000$ . The mean squared errors, on the whole, are seen to decrease as the sample size  $M$  increases, for all switching parameters and values of  $\delta/\Delta$ . This reduction is highlighted more when  $\delta/\Delta$  is smaller, especially for faster switching rates. For example, in Table 2.16 (in Section 2.B of Appendix 2), while the mean squared error for parameter  $\lambda_{10}$  reduces for larger  $\delta/\Delta$ , it is not seen to reduce at the same rate as other parameters which would indicate that this bias would persist even if the sample size is increased more. This is particularly pronounced when  $\delta/\Delta$  is between 0.4 and 0.8. This pattern is also observed from other parameters, especially  $\lambda_{01}$  and  $\lambda_{00_1}$  in the faster switching scenarios. The fast reduction in MSE for parameters under slowing switching scenarios and with lower noise  $\delta/\Delta$  is reflected by better model identification (as shown in Section 2.2) and is therefore a good indication of consistency.

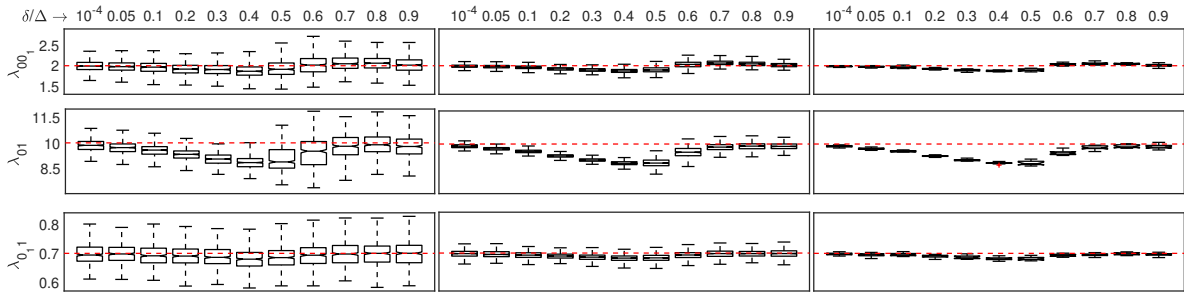
To more clearly visualise the effects of an increase in  $M$  on maximum likelihood estimation, we show rate estimates as  $\delta/\Delta$  increases when  $M = 100, 1000, 5000$  for the fast switching scenario (studies 3, 6 and 9).  $N_F$  is fixed at 10,000 under all simulations. Figures 2.10 - 2.14 show box-plots of rate results for all simulations. It is evident that while an asymptotic decrease in variance is pronounced, asymptotic bias persists for larger values of  $\delta/\Delta$ . Furthermore, stronger correlations between parameter estimates can be seen as  $\delta/\Delta$  increases, leading to inconsistency. For example, the estimates for  $\lambda_{01}$  under all  $d = 0, 1, 2$  models remain biased as  $M$  increase, albeit at a lower variance. This is also true for  $\lambda_{10}$  under studies 6 and 9.

From Figures 2.10 - 2.14, the pattern in bias that is captured by the majority of estimates under  $M = 100$  as  $\delta/\Delta$  increases persists as  $M$  increases. This is replicated by the mismatch in correlations that indicates identifiability issues. However, for low values of  $\delta/\Delta$ , the estimator appears to be asymptotically unbiased with mean squared errors converging to zero; in this case the estimator is seen to be consistent. It should be noted that the faster rates do affect these properties as is seen in Figure 2.14, whereby estimates for  $\lambda_{10}$ , although obtaining a smaller variance, remain biased for even the smallest values of  $\delta/\Delta$ . This is related to the identifiability analysis previously pre-

sented, as the sampling interval  $\Delta$  is too large to informatively detect and estimate faster switching rates. As  $\Delta \rightarrow 0$ , however, using the analysis from Section 2.2.4, these faster rates should become identifiable and consistency appears to hold.



**Figure 2.10:** Box-plots showing rate estimates when  $d = 0$  under study 3 of Table 2.1 when  $M = 100$  (left),  $M = 1000$  (middle) and  $M = 5000$  (right). True rates indicated by the red dashed line.

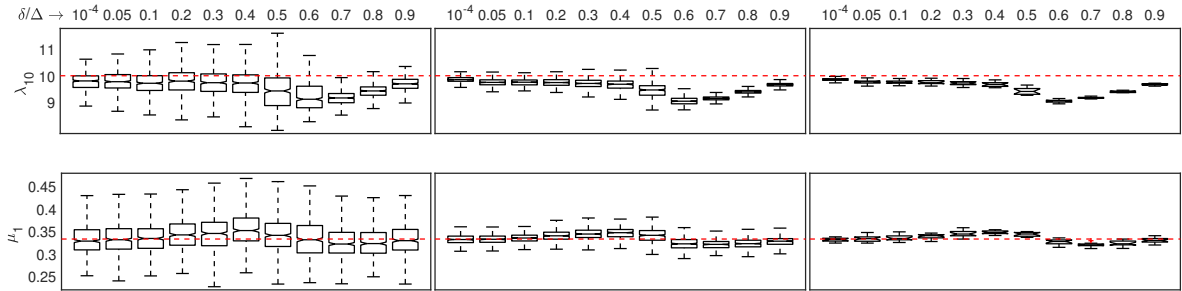


**Figure 2.11:** Box-plots showing rate estimates when  $d = 1$  under study 6 of Table 2.1 when  $M = 100$ ,  $M = 1000$  and  $M = 5000$  (right). True rates indicated by the red dashed line.

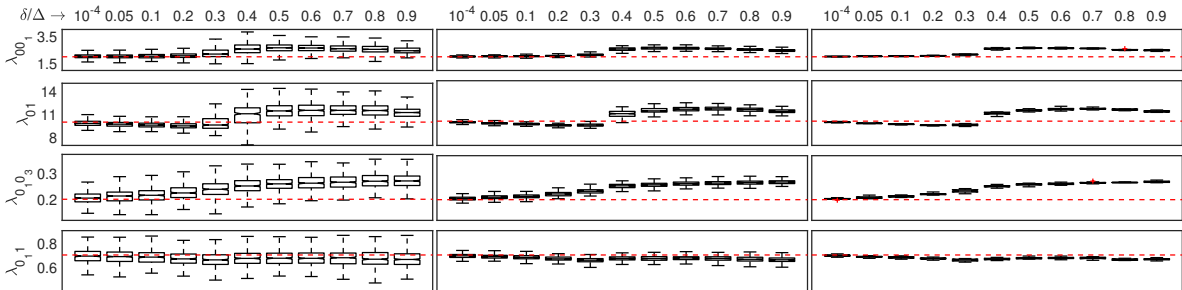
### 2.3.1 Experimental impact

The analysis presented in this section mimics the findings from the identifiability analysis presented in Section 2.2. Specifically, we found that the model fails to be consistent under two scenarios which also cause model identification problems: when  $\delta/\Delta$  is large and when fluorophores are imaged in faster switching conditions (relative to the frame acquisition time  $\Delta$ ).

Similarly to the experimental impact discussed in Section 2.2, these results suggest that first, experimenters wishing to utilise a faster switching scenario



**Figure 2.12:** Box-plots showing rate estimates when  $d = 1$  under study 6 of Table 2.1 when  $M = 100$ ,  $M = 1000$  and  $M = 5000$  (right). True rates indicated by the red dashed line.



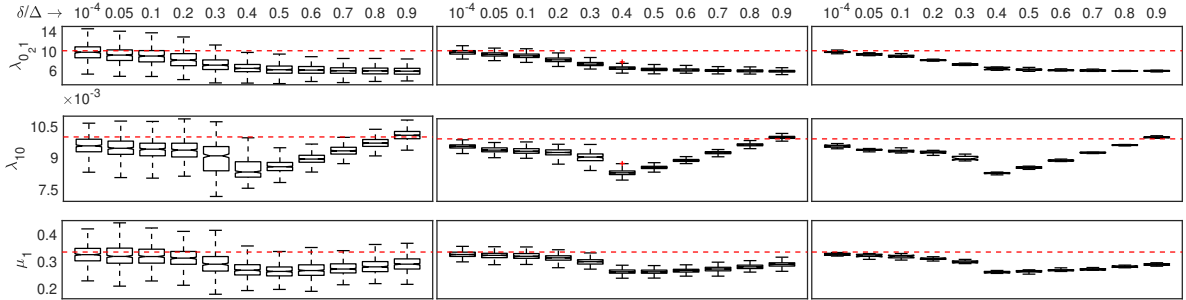
**Figure 2.13:** Box-plots showing rate estimates when  $d = 2$  under study 9 of Table 2.1 when  $M = 100$ ,  $M = 1000$  and  $M = 5000$  (right). True rates indicated by the red dashed line.

during imaging should do so at shorter frame times, and second, that imaging should be conducted in conditions that minimise the effects of background noise. In these situations, acquiring more data is likely to lead to lower variance and unbiased parameter estimates through the PSHMM.

## 2.4 Conclusions

In this chapter, we have analysed through extensive simulation studies the situations in which the PSHMM constructed in Chapter 1 can both identify unknown photo-switching parameters and is consistent.

We initially determined that potential identification issues may arise due to the noise in the model, which primarily comes in the form of  $\delta$ : the minimum time that the hidden signal  $\{X(t)\}$  must be active for within a frame for an observation of a molecule, and  $d$ : the number of multiple dark states. Us-



**Figure 2.14:** Box-plots showing rate estimates when  $d = 2$  under study 9 of Table 2.1 when  $M = 100$ ,  $M = 1000$  and  $M = 5000$  (right). True rates indicated by the red dashed line.

ing simulations studying slow, medium and fast photo-switching (from the models  $d = 0, 1, 2$ ), we found that identifiability issues somewhat arise as  $\delta/\Delta \rightarrow 1$ , i.e. the time a fluorophore needs to stay in the On state to be detected tends towards the frame rate. This was indicated by the observed correlation matrix, as derived from the observed covariance matrix  $\mathcal{J}(\hat{\theta}^{(d)})^{-1}$  showing strong correlation between  $\hat{\delta}$  and  $\hat{\lambda}_{01}$ , and hence the existence of a ridge on the likelihood surface, albeit still with curvature. While still technically identifiable, the correlation between these two parameter estimators indicated that when  $\delta/\Delta$  is large, they can be difficult to be independently identified and pose difficulties to numerical optimisation methods. This effect was more pronounced for faster transition rates due to an increased chance of transitions into the On state not being observed. However, for low values of  $\delta/\Delta$  as is typically encountered in practice, correlation between all elements of the estimator  $\hat{\theta}^{(d)}$  was low, providing clear empirical evidence of locally identifiability for all parameter values studied. Next, multi-modality of the parameter estimate  $\lambda_{01,02}$  in the  $d = 2$  model was shown to exist, although this effect disappeared as more data was used in the estimation procedure. Finally, studies on consistency of the PSHMM estimator corroborated our findings on identifiability. A break down in identifiability resulted in the estimator becoming inconsistent, with evidence suggesting that the mean squared error tends to zero when  $\delta/\Delta$  is within a suitable range (i.e.  $< 0.5$ ). However, as  $\delta/\Delta$  increased towards unity and under faster switching scenarios, consistency of the estimator disappeared with it becoming more biased (although with a reduction in variance).

With regards to experimental design, these results suggest that: first, experimenters wishing to utilise a faster switching scenario during imaging should do so at shorter frame times, second, that imaging should be conducted in conditions that minimise the effects of background noise  $\delta$ , and third, that experiments requiring a PSHMM model with multiple dark states  $d$  should involve acquiring the largest number of images possible.

## Appendix 2

### 2.A Correlation tables

In this section, we present tables to compare the empirical correlation structures between PSHMM parameter estimates (under studies 1-9) with those estimated from the Fisher information matrices, as analysed in Section 2.2.2.

$\frac{\delta}{\Delta} \rightarrow$	$10^{-4}$	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$\rho$ (Correlation) $\downarrow$											
$\rho(\lambda_{01}, \lambda_{10})$	0.0071 -0.0587	0.0054 0.0091	0.0030 -0.0624	-0.0055 -0.0043	-0.0208 0.0236	-0.0363 -0.0943	-0.0413 -0.1525	-0.0330 -0.1689	-0.0080 -0.0863	0.0152 -0.0972	0.0366 0.0306
$\rho(\lambda_{01}, \mu_1)$	0.0014 -0.0179	0.0011 0.0638	0.0005 0.0194	-0.0015 -0.0436	-0.0076 -0.0230	-0.0154 -0.0239	-0.0268 0.0021	-0.0355 -0.0847	-0.0401 -0.1393	-0.0373 -0.1372	-0.0289 -0.0354
$\rho(\lambda_{10}, \mu_1)$	0.0093 0.0406	0.0076 -0.0090	0.0067 -0.0096	0.0088 0.0892	0.0132 0.0388	0.0165 0.0368	0.0172 0.0742	0.0153 0.0861	0.0085 0.0569	0.0051 0.0635	0.0005 0.0411
$\rho(\lambda_{01}, \delta)$	0.0327 0.0472	0.0440 0.0357	0.0567 0.0202	0.0885 0.1463	0.1454 0.2580	0.2108 0.3753	0.3098 0.5584	0.3832 0.6469	0.4360 0.6479	0.4386 0.5965	0.4048 0.4922
$\rho(\lambda_{10}, \delta)$	-0.2216 -0.2008	-0.2081 -0.1640	-0.2034 -0.2127	-0.2182 -0.2624	-0.2361 -0.2431	-0.2480 -0.2636	-0.2187 -0.3307	-0.1714 -0.3591	-0.0954 -0.1750	-0.0380 -0.1876	0.0014 -0.0831
$\rho(\mu_1, \delta)$	-0.0382 0.0076	-0.0369 -0.0560	-0.0374 -0.0573	-0.0444 -0.0388	-0.0568 -0.0365	-0.0692 -0.0676	-0.0824 -0.1150	-0.08593 -0.1085	-0.0886 -0.1485	-0.0802 -0.2070	-0.0674 -0.0671

**Table 2.2:** Simulation results from  $d = 0$  slow switching (study 1 of Table 2.1): Mean approximate correlations obtained from the Hessian matrices compared with the values calculated directly from 500 fits.

$\frac{\delta}{\Delta} \rightarrow$	$10^{-4}$	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$\rho$ (Correlation) $\downarrow$											
$\rho(\lambda_{01}, \lambda_{10})$	0.0372 -0.0030	0.0258 0.0668	0.0149 -0.0379	-0.0150 -0.0657	-0.0541 -0.0721	-0.1049 -0.0936	-0.1763 -0.4365	-0.1362 -0.5085	-0.0367 -0.3351	0.0826 0.1128	0.1500 0.1405
$\rho(\lambda_{01}, \mu_1)$	-0.0017 0.0230	-0.0038 0.0594	-0.0060 -0.0698	-0.0133 0.0478	-0.0252 -0.0185	-0.0448 -0.0449	-0.0957 -0.2387	-0.1488 -0.3571	-0.1544 -0.2991	-0.1250 -0.1341	-0.0906 -0.0530
$\rho(\lambda_{10}, \mu_1)$	0.0180 0.0362	0.0203 0.0897	0.0238 -0.0147	0.0327 -0.0138	0.0430 -0.0369	0.0566 0.0824	0.0779 0.1797	0.0683 0.1535	0.0399 0.0563	0.0051 0.0003	-0.0105 -0.0259
$\rho(\lambda_{01}, \delta)$	0.0392 0.0147	0.0698 0.0403	0.0944 0.1103	0.1587 0.0717	0.2391 0.2736	0.3353 0.3533	0.5014 0.7927	0.6816 0.9073	0.7377 0.8907	0.7199 0.7393	0.6447 0.6183
$\rho(\lambda_{10}, \delta)$	-0.2886 -0.2183	-0.3132 -0.3704	-0.3375 -0.2748	-0.3779 -0.4224	-0.4082 -0.3994	-0.4366 -0.4762	-0.4456 -0.5830	-0.3017 -0.5811	-0.1455 -0.4164	-0.0024 -0.0146	0.0667 0.0001
$\rho(\mu_1, \delta)$	-0.0504 -0.0541	-0.0582 -0.0269	-0.0662 0.0410	-0.0834 -0.0409	-0.1021 -0.0845	-0.1255 -0.1696	-0.1722 -0.2951	-0.2071 -0.3826	-0.2009 -0.3302	-0.1661 -0.1768	-0.1290 -0.1421

**Table 2.3:** Simulation results from  $d = 0$  medium switching (study 2 of Table 2.1): Mean approximate correlations obtained from the Hessian matrices compared with the values calculated directly from 500 fits.

$\frac{\delta}{\Delta} \rightarrow$	$10^{-4}$	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$\rho$ (Correlation) $\downarrow$											
$\rho(\lambda_{01}, \lambda_{10})$	0.1114 0.1120	0.0740 0.0446	0.0213 0.0515	-0.0728 -0.0529	-0.1774 -0.3221	-0.2962 -0.3005	-0.4209 -0.4783	-0.5151 -0.7272	-0.2016 -0.5883	0.1049 0.0973	0.2962 0.2722
$\rho(\lambda_{01}, \mu_1)$	0.0004 0.0041	-0.0108 -0.0285	-0.0239 -0.0050	-0.0536 -0.1103	-0.0962 -0.1199	-0.1585 -0.1078	-0.2457 -0.2672	-0.4131 -0.6345	-0.4409 -0.6690	-0.3805 -0.3813	-0.3079 -0.2918
$\rho(\lambda_{10}, \mu_1)$	0.0374 -0.0232	0.0714 0.0787	0.0909 0.0687	0.1180 0.1526	0.1478 0.1492	0.1866 0.2669	0.2413 0.2172	0.3232 0.4944	0.1791 0.5065	0.0268 0.0598	-0.0521 -0.0299
$\rho(\lambda_{01}, \delta)$	0.0549 0.0606	0.1372 0.1019	0.1997 0.1897	0.3169 0.2986	0.4464 0.5680	0.5870 0.6007	0.7207 0.7456	0.8599 0.9531	0.9173 0.9603	0.9086 0.9186	0.8664 0.8544
$\rho(\lambda_{10}, \delta)$	-0.2615 -0.3324	-0.4980 -0.5800	-0.5683 -0.5959	-0.6140 -0.5643	-0.6454 -0.6925	-0.6736 -0.6900	-0.7026 -0.6598	-0.6837 -0.8279	-0.3339 -0.7023	-0.0428 -0.0835	0.1172 0.0202
$\rho(\mu_1, \delta)$	-0.0543 -0.0502	-0.1133 -0.1062	-0.1423 -0.1356	-0.1782 -0.2344	-0.2169 -0.2217	-0.2668 -0.2767	-0.3344 -0.2647	-0.4679 -0.6435	-0.4719 -0.7002	-0.4061 -0.4043	-0.3354 -0.3414

**Table 2.4:** Simulation results from  $d = 0$  fast switching (study 3 of Table 2.1): Mean approximate correlations obtained from the Hessian matrices compared with the values calculated directly from 500 fits.

$\frac{\delta}{\Delta} \rightarrow$	$10^{-4}$	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$\rho(\lambda_{00_1}, \lambda_{01})$	0.5926 0.5884	0.5915 0.5734	0.5916 0.5523	0.5896 0.5645	0.5901 0.5193	0.5898 0.5791	0.5857 0.5347	0.5832 0.5674	0.5823 0.5770	0.5828 0.5978	0.5855 0.5820
$\rho(\lambda_{00_1}, \lambda_{01_1})$	0.6754 0.6668	0.6761 0.6706	0.6752 0.7146	0.6764 0.6833	0.6769 0.6407	0.6798 0.6643	0.6807 0.5911	0.6807 0.6740	0.6834 0.6488	0.6828 0.6971	0.6844 0.6636
$\rho(\lambda_{01}, \lambda_{01_1})$	0.2854 0.2414	0.2856 0.2930	0.2845 0.2821	0.2841 0.2798	0.2868 0.2420	0.2910 0.3352	0.2928 0.2373	0.2953 0.3489	0.3004 0.2777	0.2973 0.3682	0.2958 0.3157
$\rho(\lambda_{00_1}, \lambda_{10})$	0.0158 0.0084	0.0152 -0.0103	0.0145 -0.0214	0.0118 0.0478	0.0078 0.0254	0.0033 -0.0038	-0.0013 0.0382	-0.0041 -0.0681	-0.0057 -0.1221	-0.0044 -0.0332	-0.0024 0.0114
$\rho(\lambda_{01}, \lambda_{10})$	0.0275 -0.0013	0.0259 0.0515	0.0238 -0.0358	0.0166 0.0359	0.0046 -0.0607	-0.0083 -0.0905	-0.0193 -0.0734	-0.0163 -0.1606	-0.0098 -0.1049	0.0028 0.0456	0.0164 -0.0530
$\rho(\lambda_{01_1}, \lambda_{10})$	0.0038 -0.0295	0.0032 -0.0067	0.0025 -0.0331	0.0001 0.0511	-0.0037 0.0147	-0.0079 -0.0397	-0.0111 0.0191	-0.0102 -0.0667	-0.0077 -0.0750	-0.0037 -0.0068	0.0012 -0.0099
$\rho(\lambda_{00_1}, \mu_1)$	0.0108 0.0321	0.0112 0.0828	0.0108 0.0191	0.0113 0.0008	0.0105 -0.0010	0.0098 0.0935	0.0090 0.0375	0.0082 -0.0906	0.0072 0.0402	0.0077 0.0407	0.0079 0.0007
$\rho(\lambda_{01}, \mu_1)$	0.0097 -0.0011	0.0099 0.0755	0.0096 0.0649	0.0093 -0.0289	0.0070 0.0220	0.0039 0.1154	-0.0000 -0.0762	-0.0032 -0.0721	-0.0064 -0.0061	-0.0036 0.0963	-0.0011 -0.0413
$\rho(\lambda_{01_1}, \mu_1)$	0.0326 0.0539	0.0336 0.1016	0.0328 0.0092	0.0338 -0.0256	0.0325 -0.0366	0.0313 0.0745	0.0301 0.0557	0.0297 0.0025	0.0283 0.0154	0.0292 0.0880	0.0293 0.0240
$\rho(\lambda_{10}, \mu_1)$	-0.0069 -0.0321	-0.0061 -0.0395	-0.0052 0.0123	-0.0037 -0.0677	-0.0014 -0.0500	0.0007 -0.0191	0.0016 -0.0013	-0.0005 -0.0007	-0.0035 0.0720	-0.0060 -0.0001	-0.0086 -0.0310
$\rho(\lambda_{00_1}, \delta)$	-0.0089 -0.0515	-0.0092 0.0053	-0.0084 0.0697	-0.0045 0.0345	0.0049 -0.0290	0.0177 0.0115	0.0365 0.0218	0.0568 0.1857	0.0773 0.0605	0.0748 0.1193	0.0707 0.0816
$\rho(\lambda_{01}, \delta)$	-0.0115 0.0189	-0.0076 0.0246	-0.0011 0.0555	0.0196 0.0912	0.0589 0.1491	0.1086 0.2846	0.1792 0.3455	0.2436 0.5210	0.3084 0.3632	0.2958 0.4294	0.2733 0.2540
$\rho(\lambda_{01_1}, \delta)$	-0.0013 -0.0417	0.0005 0.0392	0.0031 0.0722	0.0097 -0.0074	0.0225 0.0139	0.0389 0.0545	0.0609 0.1099	0.0811 0.1687	0.1017 0.0769	0.0952 0.0984	0.0874 0.0659
$\rho(\lambda_{10}, \delta)$	-0.1354 0.0256	-0.1663 -0.1309	-0.1876 -0.1557	-0.2165 -0.1972	-0.2343 -0.2832	-0.2425 -0.3085	-0.2265 -0.2948	-0.1626 -0.3686	-0.0920 -0.3085	-0.0420 -0.1145	0.0033 -0.0364
$\rho(\mu_1, \delta)$	-0.0124 0.0139	-0.0159 -0.0036	-0.0190 0.0458	-0.0241 -0.0462	-0.0307 0.0043	-0.0376 0.0336	-0.0439 -0.1000	-0.0456 -0.0441	-0.0462 -0.1469	-0.0393 0.0150	-0.0322 -0.0453

**Table 2.5:** Simulation results from  $d = 1$  slow switching (study 4 of Table 2.1): Mean approximate correlations obtained from the Hessian matrices compared with the values calculated directly from 500 fits.

$\frac{\delta}{\Delta} \rightarrow$	$10^{-4}$	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$\rho(\lambda_{00_1}, \lambda_{01})$	0.5729 0.5214	0.5758 0.5593	0.5748 0.5484	0.5717 0.5422	0.5663 0.5174	0.5603 0.5094	0.5570 0.4324	0.5502 0.4277	0.5493 0.5939	0.5529 0.5298	0.5555 0.4957
$\rho(\lambda_{00_1}, \lambda_{01_1})$	0.6802 0.7081	0.6824 0.6599	0.6823 0.6531	0.6850 0.6503	0.6839 0.7029	0.6856 0.7018	0.6899 0.6401	0.6918 0.6296	0.7003 0.6645	0.7023 0.7089	0.7008 0.6922
$\rho(\lambda_{01}, \lambda_{01_1})$	0.2924 0.2581	0.2947 0.2379	0.2943 0.2772	0.2937 0.2552	0.2903 0.2548	0.2917 0.2888	0.3038 0.2839	0.3105 0.2906	0.3204 0.3816	0.3199 0.2970	0.3094 0.2636
$\rho(\lambda_{00_1}, \lambda_{10})$	0.0489 0.0073	0.0475 0.0871	0.0439 0.0330	0.0360 0.0024	0.0253 -0.0025	0.0100 0.0689	-0.0072 -0.0672	-0.0157 -0.1155	-0.0152 -0.1029	-0.0069 -0.0185	0.0015 0.0112
$\rho(\lambda_{01}, \lambda_{10})$	0.0874 0.0537	0.0823 0.1334	0.0735 0.1303	0.0518 0.0098	0.0195 0.0352	-0.0318 -0.0818	-0.0778 -0.3192	-0.0651 -0.3792	-0.0341 -0.2396	0.0167 -0.0334	0.0638 0.0212
$\rho(\lambda_{01_1}, \lambda_{10})$	0.0240 -0.0204	0.0229 0.0454	0.0205 -0.0349	0.0151 0.0126	0.0065 -0.0303	-0.0071 0.0054	-0.0216 -0.1004	-0.0205 -0.1949	-0.0119 -0.0769	0.0019 -0.0142	0.0150 0.0592
$\rho(\lambda_{00_1}, \mu_1)$	0.0010 -0.0044	0.0011 -0.0558	0.0011 -0.0203	0.0006 0.0269	-0.0010 -0.0651	-0.0052 0.0270	-0.0140 -0.0735	-0.0208 -0.0301	-0.0238 -0.1018	-0.0214 -0.0145	-0.0158 0.0048
$\rho(\lambda_{01}, \mu_1)$	0.0015 0.0206	0.0013 -0.0976	0.0004 -0.0460	-0.0028 0.0222	-0.0109 -0.0486	-0.0300 -0.0313	-0.0602 -0.1469	-0.0828 -0.1258	-0.0918 -0.1143	-0.0807 -0.1881	-0.0581 -0.0141
$\rho(\lambda_{01_1}, \mu_1)$	0.0004 -0.0289	0.0004 -0.0396	0.0002 0.0082	-0.0006 0.0505	-0.0026 -0.0552	-0.0076 0.0147	-0.0173 -0.0762	-0.0242 -0.0785	-0.0263 -0.1116	-0.0235 -0.0469	-0.0168 0.0222
$\rho(\lambda_{10}, \mu_1)$	0.0110 0.0153	0.0167 -0.0353	0.0211 0.0207	0.0297 -0.0192	0.0402 0.0741	0.0572 -0.0107	0.0710 0.0750	0.0526 0.1038	0.0345 0.0899	0.0128 0.0152	-0.0032 0.1020
$\rho(\lambda_{00_1}, \delta)$	-0.0137 0.0136	-0.0149 -0.0547	-0.0129 -0.0208	-0.0060 -0.0296	0.0069 0.0109	0.0292 -0.0215	0.0633 0.0945	0.1106 0.1694	0.1342 0.2768	0.1364 0.0965	0.1191 0.0551
$\rho(\lambda_{01}, \delta)$	-0.0185 0.0276	-0.0117 -0.0156	0.0013 -0.0346	0.0371 0.0512	0.0956 0.1403	0.1903 0.3237	0.3050 0.6120	0.4493 0.6889	0.5195 0.6827	0.5192 0.5565	0.4580 0.4915
$\rho(\lambda_{01_1}, \delta)$	-0.0058 -0.0104	-0.0043 -0.0162	-0.0009 -0.0217	0.0078 -0.0425	0.0226 -0.0006	0.0474 0.0478	0.0828 0.1873	0.1281 0.3027	0.1464 0.2810	0.1470 0.1391	0.1254 0.0776
$\rho(\lambda_{10}, \delta)$	-0.2139 -0.0514	-0.2795 -0.2808	-0.3157 -0.2899	-0.3601 -0.3803	-0.3925 -0.3834	-0.4246 -0.5160	-0.4181 -0.5869	-0.2805 -0.5994	-0.1610 -0.4065	-0.0570 -0.1904	0.0222 -0.0170
$\rho(\mu_1, \delta)$	-0.0371 0.0300	-0.0518 -0.0303	-0.0618 -0.0464	-0.0793 -0.0858	-0.0991 -0.1312	-0.1296 -0.1113	-0.1616 -0.2092	-0.1713 -0.2037	-0.1673 -0.1956	-0.1464 -0.2347	-0.1137 -0.1104

**Table 2.6:** Simulation results from  $d = 1$  medium switching (study 5 of Table 2.1): Mean approximate correlations obtained from the Hessian matrices compared with the values calculated directly from 500 fits.



$\frac{\delta}{\Delta} \rightarrow$	$10^{-4}$	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$\rho(\lambda_{00_1}, \lambda_{01})$	0.3405 0.3628	0.3349 0.3507	0.3459 0.4301	0.3774 0.3108	0.4155 0.3824	0.4888 0.5229	0.6258 0.8100	0.6706 0.8409	0.6555 0.6629	0.6224 0.5735	0.5765 0.6091
$\rho(\lambda_{00_1}, \lambda_{01_1})$	0.3386 0.3038	0.3368 0.3429	0.3371 0.3030	0.3358 0.2940	0.3261 0.3251	0.3366 0.3622	0.3980 0.4578	0.4339 0.5063	0.4531 0.4640	0.4611 0.4623	0.4640 0.4496
$\rho(\lambda_{01}, \lambda_{01_1})$	0.1793 0.2272	0.1758 0.2038	0.1734 0.2124	0.1687 0.0852	0.1635 0.1339	0.1793 0.1729	0.2466 0.3222	0.2773 0.3821	0.2723 0.2721	0.2553 0.2359	0.2375 0.2230
$\rho(\lambda_{00_1}, \lambda_{10})$	0.0566 0.0602	-0.0493 -0.0620	-0.1066 -0.1243	-0.1763 -0.2787	-0.2428 -0.2494	-0.3332 -0.3709	-0.4572 -0.6973	-0.4535 -0.6761	-0.3202 -0.3627	-0.1847 -0.0915	-0.0797 -0.1371
$\rho(\lambda_{01}, \lambda_{10})$	0.5471 0.4775	0.4469 0.3501	0.3567 0.2732	0.1720 0.1298	-0.0583 -0.1502	-0.2968 -0.3976	-0.5318 -0.7826	-0.5674 -0.7792	-0.3987 -0.4798	-0.1932 -0.1459	0.0010 -0.0967
$\rho(\lambda_{01_1}, \lambda_{10})$	0.0783 0.0802	0.0583 0.0715	0.0426 0.0561	0.0158 -0.0467	-0.0177 0.0820	-0.0650 -0.0290	-0.1467 -0.2728	-0.1678 -0.2393	-0.1171 -0.1500	-0.0579 -0.0298	-0.0119 -0.0373
$\rho(\lambda_{00_1}, \mu_1)$	-0.0112 0.0160	-0.0310 -0.0869	-0.0462 0.0047	-0.0707 -0.1182	-0.1018 -0.0918	-0.1655 -0.1394	-0.2615 -0.4245	-0.2681 -0.4593	-0.2104 -0.2448	-0.1575 -0.0935	-0.1178 -0.1078
$\rho(\lambda_{01}, \mu_1)$	0.0406 0.0659	0.0408 -0.0173	0.0282 0.0020	-0.0127 -0.0232	-0.0801 -0.0253	-0.1876 -0.2111	-0.3309 -0.5190	-0.3646 -0.5624	-0.3174 -0.4151	-0.2661 -0.2480	-0.2224 -0.2524
$\rho(\lambda_{01_1}, \mu_1)$	0.0046 0.0045	0.0036 0.0090	0.0011 0.0858	-0.0049 -0.0460	-0.0152 0.0048	-0.0413 -0.1148	-0.0927 -0.1870	-0.1072 -0.1522	-0.0872 -0.1387	-0.0658 -0.0166	-0.0515 0.0270
$\rho(\lambda_{10}, \mu_1)$	0.0922 0.0813	0.1494 0.0900	0.1779 0.1089	0.2144 0.2483	0.2438 0.2469	0.3051 0.2569	0.3727 0.5041	0.3394 0.5874	0.2195 0.2500	0.1180 0.1503	0.0422 0.0054
$\rho(\lambda_{00_1}, \delta)$	0.1052 0.0726	0.1923 0.2080	0.2366 0.2657	0.2874 0.3300	0.3344 0.3623	0.4121 0.4415	0.5496 0.7632	0.5971 0.7753	0.5611 0.5765	0.5027 0.4441	0.4321 0.4436
$\rho(\lambda_{01}, \delta)$	-0.0966 -0.0648	-0.0979 -0.0544	-0.0471 0.0780	0.0956 0.1247	0.2808 0.3901	0.4847 0.5974	0.7125 0.9035	0.8244 0.9103	0.8510 0.8719	0.8399 0.8292	0.7972 0.8121
$\rho(\lambda_{01_1}, \delta)$	-0.0082 -0.0179	-0.0042 -0.0416	0.0050 -0.0140	0.0237 0.0748	0.0498 -0.0006	0.0939 0.0760	0.1867 0.2925	0.2345 0.3215	0.2294 0.2336	0.2082 0.1871	0.1870 0.1832
$\rho(\lambda_{10}, \delta)$	-0.4631 -0.5135	-0.7184 -0.7308	-0.7860 -0.7868	-0.8158 -0.8493	-0.7972 -0.8411	-0.8155 -0.8743	-0.8300 -0.9301	-0.7621 -0.9184	-0.5873 -0.6647	-0.3941 -0.3895	-0.1998 -0.2742
$\rho(\mu_1, \delta)$	-0.0964 -0.0937	-0.1701 -0.1357	-0.2031 -0.1218	-0.2426 -0.2575	-0.2726 -0.2991	-0.3368 -0.3191	-0.4184 -0.5532	-0.4186 -0.6135	-0.3587 -0.4383	-0.3059 -0.2904	-0.2603 -0.2848

**Table 2.7:** Simulation results from  $d = 1$  fast switching (study 6 of Table 2.1): Mean approximate correlations obtained from the Hessian matrices compared with the values calculated directly from 500 fits.

$\frac{\delta}{\Delta} \rightarrow$	$10^{-4}$	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$\rho$ (Correlation) $\downarrow$											
$\rho(\lambda_{00_1}, \lambda_{01})$	-0.1610 0.4636	-0.1158 0.3054	-0.1025 0.3245	-0.1637 0.4173	-0.0808 0.5135	-0.1255 0.3869	-0.0980 0.4449	-0.0860 0.4316	-0.1826 0.3913	-0.1768 0.4898	-0.0720 0.3780
$\rho(\lambda_{00_1}, \lambda_{0_1 0_2})$	0.3396 -0.1316	0.3015 -0.0657	0.2835 -0.0279	0.3388 -0.1000	0.2714 -0.0148	0.3261 -0.0698	0.2891 -0.0362	0.2844 -0.1160	0.3623 -0.1319	0.3685 -0.0965	0.2800 -0.0631
$\rho(\lambda_{01}, \lambda_{0_1 0_2})$	-0.7763 -0.6749	-0.7605 -0.5007	-0.7673 -0.6502	-0.7897 -0.6557	-0.7430 -0.5843	-0.7548 -0.6500	-0.7661 -0.6122	-0.7743 -0.6609	-0.8030 -0.6682	-0.8119 -0.6017	-0.7640 -0.6394
$\rho(\lambda_{00_1}, \lambda_{0_1 1})$	0.0027 0.5237	0.0574 0.3805	0.0735 0.5875	0.0146 0.5831	0.0939 0.2380	0.0663 0.5986	0.0995 0.5905	0.1137 0.6323	0.0163 0.5082	0.0198 0.6147	0.1254 0.6844
$\rho(\lambda_{01}, \lambda_{0_1 1})$	0.7683 0.7166	0.7494 0.8141	0.7518 0.6065	0.7716 0.6667	0.7377 0.6780	0.7506 0.6503	0.7473 0.6476	0.7553 0.6628	0.7828 0.6814	0.7884 0.6770	0.7483 0.6096
$\rho(\lambda_{0_1 0_2}, \lambda_{0_1 1})$	-0.7324 -0.5998	-0.7136 -0.4808	-0.7292 -0.5624	-0.7438 -0.5811	-0.6997 -0.6077	-0.6879 -0.5032	-0.7144 -0.5283	-0.7195 -0.5261	-0.7415 -0.5840	-0.7529 -0.5094	-0.7105 -0.5109
$\rho(\lambda_{00_1}, \lambda_{0_2 1})$	-0.4348 0.0908	-0.3985 -0.0216	-0.3920 -0.0453	-0.4337 -0.0207	-0.3760 -0.0235	-0.4084 -0.0908	-0.3922 -0.0088	-0.3824 0.0799	-0.4514 0.0372	-0.4481 0.0485	-0.3744 -0.0510
$\rho(\lambda_{01}, \lambda_{0_2 1})$	0.7245 0.5706	0.7024 0.5451	0.7040 0.4805	0.7385 0.4705	0.6843 0.4554	0.7144 0.4899	0.7134 0.5016	0.7226 0.5154	0.7601 0.5428	0.7738 0.4880	0.7155 0.4971
$\rho(\lambda_{0_1 0_2}, \lambda_{0_2 1})$	-0.8287 -0.7142	-0.8049 -0.7098	-0.8244 -0.7551	-0.8402 -0.7281	-0.7972 -0.7510	-0.8054 -0.6973	-0.8246 -0.7406	-0.8330 -0.7352	-0.8537 -0.7373	-0.8639 -0.6680	-0.8268 -0.7415
$\rho(\lambda_{0_1 1}, \lambda_{0_2 1})$	0.7449 0.5949	0.7287 0.5811	0.7287 0.5247	0.7503 0.5173	0.7140 0.5580	0.7299 0.4431	0.7273 0.5270	0.7299 0.5238	0.7556 0.5570	0.7669 0.5020	0.7302 0.4588
$\rho(\lambda_{00_1}, \lambda_{10})$	0.4647 -0.0491	0.4306 0.0298	0.4233 0.0261	0.4575 -0.0113	0.4072 0.0706	0.4354 0.0172	0.4158 0.0422	0.4101 0.0175	0.4715 -0.0017	0.4679 -0.0025	0.4006 0.1008
$\rho(\lambda_{01}, \lambda_{10})$	-0.6628 -0.5564	-0.6388 -0.2211	-0.6383 -0.5275	-0.6866 -0.5618	-0.6215 -0.4047	-0.6596 -0.5259	-0.6584 -0.5020	-0.6667 -0.4994	-0.7177 -0.5451	-0.7327 -0.4701	-0.6606 -0.4685
$\rho(\lambda_{0_1 0_2}, \lambda_{10})$	0.7532 0.7268	0.7308 0.7349	0.7380 0.7866	0.7723 0.7284	0.7111 0.7298	0.7427 0.6084	0.7500 0.7387	0.7631 0.7241	0.8001 0.6935	0.8221 0.7056	0.7652 0.7207
$\rho(\lambda_{0_1 1}, \lambda_{10})$	-0.6547 -0.4979	-0.6326 -0.2404	-0.6310 -0.5101	-0.6698 -0.5017	-0.6124 -0.4507	-0.6387 -0.4887	-0.6380 -0.4657	-0.6408 -0.3815	-0.6875 -0.4739	-0.7033 -0.4260	-0.6450 -0.3933
$\rho(\lambda_{0_2 1}, \lambda_{10})$	-0.7887 -0.6504	-0.7671 -0.5952	-0.7734 -0.6845	-0.8041 -0.6518	-0.7503 -0.6771	-0.7830 -0.6168	-0.7865 -0.6430	-0.7956 -0.6223	-0.8318 -0.6434	-0.8475 -0.5618	-0.8000 -0.6607
$\rho(\lambda_{00_1}, \mu_1)$	-0.4635	-0.4290	-0.4215	-0.4598	-0.4089	-0.4410	-0.4233	-0.4161	-0.4810	-0.4797	-0.4087

**Table 2.8:** Simulation results from  $d = 2$  slow switching (study 7 of Table 2.1): Mean approximate correlations obtained from the Hessian matrices compared with the values calculated directly from 500 fits.

$\frac{\delta}{\Delta} \rightarrow$	$10^{-4}$	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$\rho$ (Correlation) ↓											
	0.0628	-0.0465	-0.0558	-0.0237	-0.0410	-0.0343	-0.0751	0.0032	0.0164	-0.0128	-0.0537
$\rho(\lambda_{01}, \mu_1)$	0.7134 0.6626	0.6893 0.3925	0.6901 0.6148	0.7297 0.6245	0.6654 0.5208	0.7005 0.6370	0.6989 0.5776	0.7075 0.6132	0.7507 0.6428	0.7644 0.5733	0.7011 0.6022
$\rho(\lambda_{010_2}, \mu_1)$	-0.8275 -0.8870	-0.8067 -0.8837	-0.8203 -0.8959	-0.8412 -0.8767	-0.7915 -0.8818	-0.8145 -0.8643	-0.8246 -0.8927	-0.8362 -0.8835	-0.8593 -0.8770	-0.8756 -0.8559	-0.8328 -0.8819
$\rho(\lambda_{011}, \mu_1)$	0.7067 0.6151	0.6868 0.4014	0.6878 0.5766	0.7176 0.5828	0.6662 0.5771	0.6896 0.5499	0.6897 0.5319	0.6919 0.5133	0.7279 0.5949	0.7405 0.5246	0.6931 0.5236
$\rho(\lambda_{021}, \mu_1)$	0.8828 0.7657	0.8664 0.7285	0.8782 0.7994	0.8914 0.7742	0.8564 0.7882	0.8784 0.7764	0.8845 0.7867	0.8889 0.7827	0.9093 0.7849	0.9150 0.7263	0.8896 0.7742
$\rho(\lambda_{10}, \mu_1)$	-0.7789 -0.8129	-0.7556 -0.8276	-0.7605 -0.8237	-0.7950 -0.8102	-0.7335 -0.8129	-0.7729 -0.7683	-0.7761 -0.7842	-0.7898 -0.7749	-0.8301 -0.7782	-0.8498 -0.7852	-0.7951 -0.8033
$\rho(\lambda_{001}, \delta)$	-0.0117 -0.0587	-0.0110 -0.0580	-0.0108 0.0102	-0.0052 0.0161	0.0048 0.0344	0.0120 0.0484	0.0206 0.0241	0.0321 -0.0216	0.0477 0.0364	0.0621 0.0362	0.0642 0.0294
$\rho(\lambda_{01}, \delta)$	0.0003 0.0039	0.0041 0.0034	0.0120 0.0240	0.0306 0.1295	0.0564 0.1571	0.0752 0.1277	0.0946 0.1704	0.1178 0.1942	0.1359 0.2498	0.1516 0.2150	0.1630 0.1237
$\rho(\lambda_{010_2}, \delta)$	-0.0014 0.0140	-0.0019 -0.0551	-0.0036 -0.0075	-0.0071 0.0114	-0.0104 0.0020	-0.0072 0.0084	-0.0128 0.0394	-0.0134 -0.0730	-0.0134 -0.0802	-0.0118 -0.0697	-0.0130 0.0352
$\rho(\lambda_{011}, \delta)$	0.0011 -0.0186	0.0023 -0.0120	0.0059 0.0512	0.0134 -0.0097	0.0238 0.1336	0.0244 0.0405	0.0338 -0.0117	0.0401 0.0456	0.0469 0.1134	0.0525 0.0161	0.0551 0.0316
$\rho(\lambda_{021}, \delta)$	0.0047 0.0193	0.0057 0.0371	0.0078 0.0195	0.0113 -0.0824	0.0138 0.0192	0.0110 0.0113	0.0147 -0.0329	0.0142 0.0613	0.0121 0.1009	0.0108 0.0259	0.0110 -0.0146
$\rho(\lambda_{10}, \delta)$	-0.0488 -0.0760	-0.0597 -0.1348	-0.0679 -0.0990	-0.0723 -0.1065	-0.0840 -0.1229	-0.0763 -0.0849	-0.0713 -0.1186	-0.0559 -0.1887	-0.0316 -0.2027	-0.0210 -0.0679	-0.0162 -0.0344
$\rho(\mu_1, \delta)$	-0.0030 -0.0219	-0.0048 0.0949	-0.0043 -0.0145	-0.0022 -0.0152	-0.0096 0.0317	-0.0132 -0.0101	-0.0096 -0.0325	-0.0113 0.0822	-0.0109 0.0891	-0.0127 0.0197	-0.0153 -0.0336

**Table 2.9:** Simulation results from  $d = 2$  slow switching (study 7 of Table 2.1) continued: Mean approximate correlations obtained from the Hessian matrices compared with the values calculated directly from 500 fits.

$\frac{\delta}{\Delta} \rightarrow$	$10^{-4}$	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$\rho$ (Correlation) $\downarrow$											
$\rho(\lambda_{00_1}, \lambda_{01})$	0.3418 0.3578	0.3428 0.4383	0.3604 0.3207	0.3767 0.3759	0.3931 0.3966	0.4104 0.4616	0.4493 0.4572	0.5074 0.7093	0.4984 0.5343	0.4728 0.4565	0.4644 0.4863
$\rho(\lambda_{00_1}, \lambda_{0_1 0_2})$	0.2102 0.1448	0.2077 0.1084	0.1803 0.1646	0.1570 0.0317	0.1424 0.1551	0.1291 0.0747	0.1274 0.0551	0.1185 0.0984	0.1034 0.0732	0.0927 0.0845	0.0831 0.0274
$\rho(\lambda_{01}, \lambda_{0_1 0_2})$	-0.1329 -0.1305	-0.1320 -0.0815	-0.1142 -0.0471	-0.1002 -0.1104	-0.0877 -0.0044	-0.0703 0.0860	-0.0367 -0.0391	0.0025 0.0872	0.0044 0.0595	-0.0098 0.0264	-0.0167 -0.0565
$\rho(\lambda_{00_1}, \lambda_{0_1 1})$	0.3630 0.4174	0.3698 0.3864	0.4039 0.4574	0.4334 0.5145	0.4572 0.4500	0.4730 0.4464	0.4971 0.5019	0.5341 0.6346	0.5586 0.5959	0.5672 0.5705	0.5791 0.6507
$\rho(\lambda_{01}, \lambda_{0_1 1})$	0.3114 0.2614	0.3084 0.2775	0.2987 0.2635	0.2885 0.3035	0.2814 0.2272	0.2684 0.2364	0.2712 0.2422	0.2913 0.4806	0.2905 0.3408	0.2815 0.2312	0.2782 0.2962
$\rho(\lambda_{0_1 0_2}, \lambda_{0_1 1})$	-0.1423 -0.0358	-0.1352 -0.0399	-0.0936 -0.0072	-0.0594 -0.0452	-0.0387 0.1050	-0.0274 0.0116	-0.0020 -0.0167	0.0287 0.1275	0.0519 0.1244	0.0544 0.1939	0.0591 0.1078
$\rho(\lambda_{00_1}, \lambda_{0_2 1})$	-0.2231 -0.1357	-0.2208 -0.1891	-0.1900 -0.1442	-0.1644 -0.0181	-0.1446 -0.0645	-0.1310 -0.0120	-0.1145 -0.0687	-0.0855 0.0151	-0.0705 0.0627	-0.0660 0.0828	-0.0603 -0.0005
$\rho(\lambda_{01}, \lambda_{0_2 1})$	0.1673 0.1498	0.1672 0.0747	0.1515 0.0893	0.1382 0.1506	0.1277 0.0716	0.1116 0.1465	0.0955 0.0547	0.0741 0.0630	0.0647 0.1546	0.0657 0.0216	0.0657 0.0338
$\rho(\lambda_{0_1 0_2}, \lambda_{0_2 1})$	-0.5349 -0.4590	-0.5316 -0.4133	-0.4878 -0.3344	-0.4483 -0.3654	-0.4222 -0.2513	-0.4031 -0.1643	-0.3735 -0.1913	-0.3306 -0.1074	-0.2913 -0.0021	-0.2810 -0.0288	-0.2680 0.0257
$\rho(\lambda_{0_1 1}, \lambda_{0_2 1})$	0.3805 0.2976	0.3760 0.2930	0.3486 0.2831	0.3219 0.3103	0.3080 0.2394	0.2900 0.2386	0.2792 0.1643	0.2569 0.1985	0.2381 0.1807	0.2325 0.1182	0.2247 0.1320
$\rho(\lambda_{00_1}, \lambda_{10})$	0.1437 0.1258	0.1126 0.1137	0.0845 0.0199	0.0469 0.0010	0.0092 -0.0612	-0.0437 -0.0805	-0.1405 -0.2612	-0.1785 -0.5137	-0.0853 -0.1376	-0.0070 0.0150	0.0227 -0.0170
$\rho(\lambda_{01}, \lambda_{10})$	0.2225 0.2128	0.1748 0.2225	0.1430 0.1431	0.0592 0.0456	-0.0525 -0.1341	-0.2120 -0.2405	-0.4267 -0.5987	-0.4160 -0.7957	-0.1634 -0.2786	0.0438 0.0774	0.1643 0.1346
$\rho(\lambda_{0_1 0_2}, \lambda_{10})$	0.1310 0.0825	0.1186 0.1071	0.0940 0.0554	0.0741 0.0044	0.0598 0.0366	0.0450 0.0182	0.0238 -0.0310	0.0271 -0.0885	0.0516 -0.0364	0.0674 -0.0015	0.0713 0.0547
$\rho(\lambda_{0_1 1}, \lambda_{10})$	-0.0228 -0.0055	-0.0215 0.0790	-0.0105 0.0003	-0.0110 0.0121	-0.0192 -0.0290	-0.0373 0.0069	-0.0817 -0.1295	-0.1139 -0.3429	-0.0749 -0.1393	-0.0329 0.0212	-0.0130 0.0021
$\rho(\lambda_{0_2 1}, \lambda_{10})$	-0.1744 -0.1129	-0.1627 -0.1435	-0.1390 -0.1276	-0.1210 -0.0343	-0.1089 -0.0845	-0.0985 -0.1230	-0.0920 -0.0509	-0.0986 -0.0557	-0.1032 -0.0876	-0.1003 -0.0911	-0.0947 -0.0323
$\rho(\lambda_{00_1}, \mu_1)$	-0.2455	-0.2455	-0.2151	-0.1915	-0.1745	-0.1658	-0.1752	-0.1719	-0.1454	-0.1232	-0.1102

**Table 2.10:** Simulation results from  $d = 2$  medium switching (study 8 of Table 2.1): Mean approximate correlations obtained from the Hessian matrices compared with the values calculated directly from 500 fits.

$\frac{\delta}{\Delta} \rightarrow$	$10^{-4}$	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$\rho$ (Correlation) $\downarrow$	-0.1973	-0.1578	-0.1436	-0.0563	-0.1368	-0.1149	-0.0934	-0.2910	-0.0182	-0.1163	0.0363
$\rho(\lambda_{01}, \mu_1)$	0.1665 0.1581	0.1670 0.0993	0.1486 0.1173	0.1304 0.1754	0.1088 0.0089	0.0703 0.1282	0.0022 -0.0726	-0.0672 -0.3225	-0.0600 -0.0989	-0.0274 -0.1461	-0.0145 -0.0194
$\rho(\lambda_{010_2}, \mu_1)$	-0.5887 -0.5547	-0.5900 -0.4743	-0.5488 -0.4429	-0.5129 -0.4524	-0.4876 -0.3355	-0.4676 -0.2417	-0.4470 -0.2201	-0.4131 -0.1795	-0.3757 -0.1525	-0.3664 -0.0547	-0.3534 0.0107
$\rho(\lambda_{011}, \mu_1)$	0.3555 0.2864	0.3539 0.2858	0.3216 0.2606	0.2923 0.2965	0.2732 0.2075	0.2515 0.2217	0.2267 0.1741	0.1864 -0.0343	0.1695 0.0314	0.1726 -0.0467	0.1678 0.0264
$\rho(\lambda_{021}, \mu_1)$	0.7892 0.7308	0.7931 0.7359	0.7682 0.6911	0.7430 0.6600	0.7238 0.5947	0.7036 0.5151	0.6845 0.4457	0.6461 0.2904	0.6193 0.2393	0.6136 0.2314	0.6031 0.2140
$\rho(\lambda_{10}, \mu_1)$	-0.1696 -0.1419	-0.1516 -0.1207	-0.1213 -0.1143	-0.0932 0.0096	-0.0673 -0.0065	-0.0352 -0.0526	0.0128 0.1267	0.0072 0.2870	-0.0516 0.0659	-0.0883 0.0484	-0.0981 0.0283
$\rho(\lambda_{001}, \delta)$	0.0097 0.0344	0.0251 0.0906	0.0392 0.1098	0.0637 0.1030	0.0929 0.1259	0.1395 0.1979	0.2312 0.3136	0.3502 0.6346	0.3348 0.3826	0.2749 0.2479	0.2391 0.1893
$\rho(\lambda_{01}, \delta)$	-0.0247 0.0371	0.0104 0.0456	0.0460 0.0871	0.1360 0.1160	0.2580 0.3167	0.4269 0.4716	0.6343 0.7803	0.8224 0.9485	0.8432 0.8742	0.7971 0.7884	0.7568 0.7136
$\rho(\lambda_{010_2}, \delta)$	-0.0009 -0.0917	-0.0027 -0.0066	0.0045 -0.0172	0.0112 0.0377	0.0186 0.0262	0.0274 0.0610	0.0407 0.0370	0.0517 0.1253	0.0462 0.1161	0.0366 0.0567	0.0336 -0.0179
$\rho(\lambda_{011}, \delta)$	-0.0095 0.0148	-0.0073 0.0379	-0.0073 0.0230	0.0006 0.0587	0.0130 0.0403	0.0362 0.0344	0.0876 0.1317	0.1690 0.4143	0.1751 0.2533	0.1494 0.1249	0.1338 0.1302
$\rho(\lambda_{021}, \delta)$	0.0071 0.0087	0.0150 0.0584	0.0095 -0.0125	0.0072 0.0114	0.0046 0.0028	0.0043 0.0583	0.0111 0.0296	0.0204 0.0512	0.0195 0.1283	0.0168 -0.0333	0.0161 -0.0154
$\rho(\lambda_{10}, \delta)$	-0.3872 -0.2960	-0.5744 -0.6377	-0.6153 -0.6105	-0.6602 -0.6919	-0.6947 -0.7027	-0.7337 -0.7630	-0.7813 -0.8379	-0.5962 -0.8906	-0.3004 -0.4383	-0.0994 -0.0889	0.0205 -0.0137
$\rho(\mu_1, \delta)$	-0.0088 0.0353	-0.0105 0.0577	-0.0234 0.0100	-0.0378 -0.0561	-0.0562 -0.0918	-0.0808 -0.0357	-0.1196 -0.1748	-0.1437 -0.3398	-0.1233 -0.1752	-0.0956 -0.2138	-0.0854 -0.1197

**Table 2.11:** Simulation results from  $d = 2$  medium switching (study 8 of Table 2.1) continued: Mean approximate correlations obtained from the Hessian matrices compared with the values calculated directly from 500 fits.

$\frac{\delta}{\Delta} \rightarrow$	$10^{-4}$	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$\rho$ (Correlation) ↓											
$\rho(\lambda_{00_1}, \lambda_{01})$	0.3226 0.3530	0.3564 0.2565	0.3883 0.3992	0.4924 0.4931	0.6686 0.8714	0.7631 0.8827	0.7393 0.8089	0.6992 0.7678	0.6671 0.7209	0.6401 0.7226	0.6163 0.6324
$\rho(\lambda_{00_1}, \lambda_{0_1 0_2})$	0.1205 0.1615	0.1371 0.1546	0.1522 0.1761	0.1932 0.2190	0.2358 0.4098	0.2296 0.4319	0.1957 0.2675	0.1732 0.2230	0.1615 0.1780	0.1535 0.2474	0.1484 0.1447
$\rho(\lambda_{01}, \lambda_{0_1 0_2})$	-0.0160 0.0199	-0.0010 0.0016	0.0170 -0.0031	0.0689 0.1587	0.1544 0.4162	0.1760 0.3949	0.1431 0.3343	0.1192 0.2595	0.1055 0.2576	0.0960 0.2741	0.0907 0.2093
$\rho(\lambda_{00_1}, \lambda_{0_1 1})$	0.3840 0.3631	0.3753 0.3306	0.3716 0.3419	0.3504 0.3466	0.3733 0.3618	0.4291 0.5052	0.4677 0.4664	0.4892 0.5257	0.5101 0.5900	0.5303 0.5914	0.5467 0.5367
$\rho(\lambda_{01}, \lambda_{0_1 1})$	0.2566 0.2175	0.2469 0.1456	0.2396 0.1721	0.2196 0.2223	0.2317 0.2091	0.2671 0.3397	0.2786 0.2823	0.2735 0.2565	0.2705 0.2823	0.2684 0.3085	0.2630 0.2422
$\rho(\lambda_{0_1 0_2}, \lambda_{0_1 1})$	0.1555 0.1101	0.1397 0.1276	0.1423 0.1696	0.1310 0.0494	0.1647 0.2502	0.2001 0.4050	0.2222 0.2364	0.2270 0.3057	0.2381 0.2715	0.2495 0.3348	0.2555 0.3640
$\rho(\lambda_{00_1}, \lambda_{0_2 1})$	-0.1023 -0.0643	-0.1287 -0.2151	-0.1449 -0.2247	-0.1908 -0.2111	-0.1978 -0.4299	-0.1537 -0.3191	-0.1004 -0.2016	-0.0767 -0.1728	-0.0563 -0.0955	-0.0370 -0.1175	-0.0258 -0.0431
$\rho(\lambda_{01}, \lambda_{0_2 1})$	0.0598 0.0633	0.0412 -0.0582	0.0217 0.0131	-0.0344 -0.1040	-0.0945 -0.4517	-0.0891 -0.3535	-0.0446 -0.2757	-0.0249 -0.2656	-0.0089 -0.1532	0.0057 -0.1856	0.0166 -0.1403
$\rho(\lambda_{0_1 0_2}, \lambda_{0_2 1})$	-0.1144 -0.1028	-0.1613 -0.1426	-0.1743 -0.2482	-0.2246 -0.2027	-0.2222 -0.2314	-0.1757 -0.0931	-0.1218 -0.0403	-0.0939 0.0285	-0.0688 0.0814	-0.0434 0.0039	-0.0304 0.1128
$\rho(\lambda_{0_1 1}, \lambda_{0_2 1})$	0.1776 0.2054	0.1938 0.1622	0.2014 0.1693	0.2144 0.2174	0.1960 -0.0014	0.1692 -0.0196	0.1605 0.0495	0.1530 0.0043	0.1489 0.0344	0.1435 -0.0082	0.1411 0.0687
$\rho(\lambda_{00_1}, \lambda_{10})$	-0.1279 -0.1493	-0.1442 -0.1895	-0.1935 -0.2607	-0.3236 -0.3754	-0.4959 -0.7682	-0.4884 -0.7460	-0.3454 -0.4048	-0.2112 -0.1836	-0.1089 -0.0485	-0.0308 0.0547	0.0135 0.1022
$\rho(\lambda_{01}, \lambda_{10})$	0.4418 0.3562	0.3439 0.3696	0.2393 0.2138	-0.0537 -0.1255	-0.4460 -0.7916	-0.5453 -0.7995	-0.3998 -0.4417	-0.2325 -0.2177	-0.0835 -0.0264	0.0564 0.1237	0.1530 0.2645
$\rho(\lambda_{0_1 0_2}, \lambda_{10})$	-0.0801 -0.1580	-0.0752 -0.1416	-0.0813 -0.1723	-0.1086 -0.1874	-0.1412 -0.3643	-0.1003 -0.3863	-0.0339 -0.0279	0.0058 0.0105	0.0311 0.0135	0.0471 0.0831	0.0565 0.1246
$\rho(\lambda_{0_1 1}, \lambda_{10})$	0.1203 0.1284	0.1013 0.1092	0.0866 0.0637	0.0508 0.1215	-0.0503 -0.1283	-0.1218 -0.2744	-0.1120 -0.0791	-0.0779 -0.0893	-0.0506 -0.0258	-0.0276 -0.0098	-0.0139 0.0145
$\rho(\lambda_{0_2 1}, \lambda_{10})$	0.0939 0.1706	0.0820 0.2060	0.0865 0.2060	0.1021 0.1375	0.0967 0.4010	0.0301 0.2234	-0.0345 0.0684	-0.0573 -0.0768	-0.0689 0.0238	-0.0694 -0.0354	-0.0685 -0.1246
$\rho(\lambda_{00_1}, \mu_1)$	-0.1149	-0.1340	-0.1544	-0.2243	-0.3015	-0.2857	-0.2128	-0.1656	-0.1327	-0.1052	-0.0906

**Table 2.12:** Simulation results from  $d = 2$  fast switching (study 9 of Table 2.1): Mean approximate correlations obtained from the Hessian matrices compared with the values calculated directly from 500 fits.

$\frac{\delta}{\Delta} \rightarrow$	$10^{-4}$	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$\rho$ (Correlation) $\downarrow$	-0.1227	-0.1989	-0.1324	-0.1101	-0.4881	-0.3325	0.0572	0.0196	0.1341	0.2721	0.2246
$\rho(\lambda_{01}, \mu_1)$	0.0736 0.0433	0.0474 -0.0127	0.0205 0.0151	-0.0731 -0.0464	-0.2292 -0.5420	-0.2706 -0.4629	-0.2108 -0.0440	-0.1702 -0.0765	-0.1434 0.0686	-0.1197 0.1439	-0.1106 0.1185
$\rho(\lambda_{0_1 0_2}, \mu_1)$	-0.1442 -0.1792	-0.1719 -0.0526	-0.1791 -0.0951	-0.2215 0.0309	-0.2377 -0.1222	-0.2021 0.0194	-0.1534 0.2899	-0.1265 0.3101	-0.1071 0.3597	-0.0907 0.4460	-0.0828 0.3249
$\rho(\lambda_{0_1 1}, \mu_1)$	0.0990 0.0734	0.1122 0.1633	0.1148 0.0462	0.1252 0.1292	0.0841 -0.0900	0.0428 -0.0806	0.0375 -0.1047	0.0398 -0.1696	0.0402 -0.1339	0.0391 -0.0172	0.0384 -0.1325
$\rho(\lambda_{0_2 1}, \mu_1)$	0.3356 0.3900	0.3797 0.3971	0.3948 0.3593	0.4513 0.3572	0.4571 0.4253	0.4090 0.1470	0.3590 -0.1956	0.3293 -0.2335	0.3053 -0.3394	0.2799 -0.4054	0.2662 -0.3861
$\rho(\lambda_{10}, \mu_1)$	0.1818 0.2083	0.1730 0.2031	0.1932 0.1800	0.2468 0.2156	0.3011 0.5718	0.2370 0.5468	0.1326 0.2826	0.0710 0.3285	0.0281 0.2487	-0.0017 0.2693	-0.0196 0.3157
$\rho(\lambda_{00_1}, \delta)$	0.2862 0.3317	0.3072 0.3202	0.3486 0.4446	0.4606 0.5140	0.6247 0.8560	0.6987 0.8311	0.6506 0.6868	0.5851 0.6312	0.5300 0.5043	0.4784 0.5148	0.4323 0.3463
$\rho(\lambda_{0_1}, \delta)$	-0.1171 0.0380	-0.0230 0.0149	0.0611 0.0767	0.3069 0.3770	0.6540 0.9100	0.8472 0.9211	0.8578 0.8556	0.8361 0.8544	0.8164 0.8037	0.7897 0.7933	0.7638 0.6842
$\rho(\lambda_{0_1 0_2}, \delta)$	0.1187 0.1516	0.1268 0.2215	0.1365 0.2373	0.1683 0.2975	0.2037 0.4224	0.1889 0.4253	0.1459 0.2382	0.1197 0.2510	0.1041 0.2036	0.0933 0.1671	0.0893 0.1251
$\rho(\lambda_{0_1 1}, \delta)$	-0.0705 -0.0607	-0.0611 -0.0930	-0.0536 -0.0387	-0.0292 -0.1213	0.0721 0.1667	0.1701 0.2773	0.1878 0.1776	0.1755 0.1937	0.1664 0.1175	0.1568 0.1607	0.1447 0.0881
$\rho(\lambda_{0_2 1}, \delta)$	-0.1494 -0.1951	-0.1557 -0.2923	-0.1653 -0.3103	-0.1846 -0.2549	-0.1743 -0.4483	-0.1195 -0.3200	-0.0639 -0.2484	-0.0414 -0.2362	-0.0227 -0.1840	-0.0045 -0.1652	0.0117 -0.0590
$\rho(\lambda_{10}, \delta)$	-0.7786 -0.7288	-0.7828 -0.7674	-0.8115 -0.8005	-0.8588 -0.8653	-0.8723 -0.9448	-0.7580 -0.9320	-0.6020 -0.7030	-0.4491 -0.4571	-0.3053 -0.2589	-0.1681 -0.1178	-0.0732 0.0043
$\rho(\mu_1, \delta)$	-0.2283 -0.2079	-0.2331 -0.2474	-0.2570 -0.2130	-0.3149 -0.2348	-0.3713 -0.5936	-0.3353 -0.5430	-0.2602 -0.1652	-0.2192 -0.1521	-0.1924 0.0032	-0.1693 0.0379	-0.1622 -0.0300

**Table 2.13:** Simulation results from  $d = 2$  fast switching (study 9 of Table 2.1) continued: Mean approximate correlations obtained from the Hessian matrices compared with the values calculated directly from 500 fits.

## 2.B Consistency tables

In this section, we present tables to compare the convergence in mean squared errors of the PSHMM parameter estimates (under studies 1-9) when  $M = 100$  and  $M = 1000$ , as analysed in Section 2.3.

$\frac{\delta}{\Delta} \rightarrow$	$10^{-4}$	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$\hat{\lambda}_{01}$ Mean	1.0014	0.9966	0.9912	0.9850	0.9773	0.9705	0.9683	0.9936	1.0061	1.0127	1.0079
$\hat{\lambda}_{01}$ MSE (100) $\times 10^{-4}$	0.3356	0.3902	0.4012	0.4044	0.4503	0.4779	0.5474	0.5950	0.7025	0.5211	0.4466
$\hat{\lambda}_{01}$ MSE (1000) $\times 10^{-4}$	0.0376	0.0403	0.0343	0.0718	0.0977	0.1455	0.1746	0.1862	0.0927	0.0659	0.0531
$\hat{\lambda}_{10}$ Mean	3.1604	3.1634	3.1610	3.1572	3.1662	3.1674	3.1578	3.1250	3.1262	3.1296	3.1454
$\hat{\lambda}_{10}$ MSE (100) $\times 10^{-4}$	3.6676	3.5676	3.8007	3.8985	4.1441	3.9991	4.0541	4.2626	3.7268	4.0394	3.7124
$\hat{\lambda}_{10}$ MSE (1000) $\times 10^{-4}$	0.4015	0.4324	0.4274	0.3815	0.4071	0.4204	0.4033	0.5648	0.6505	0.4453	0.4247
$\hat{\mu}_1$ Mean	0.1055	0.1058	0.1073	0.1080	0.1084	0.1089	0.1092	0.1061	0.1053	0.1044	0.1042
$\hat{\mu}_1$ MSE (100) $\times 10^{-4}$	0.1325	0.1211	0.1200	0.1138	0.1229	0.1206	0.1342	0.1295	0.1232	0.1016	0.1155
$\hat{\mu}_1$ MSE (1000) $\times 10^{-4}$	0.0115	0.0108	0.0125	0.0105	0.0136	0.0113	0.0107	0.0130	0.0130	0.0162	0.0136

**Table 2.14:**  $d = 0$  slow switching (study 1 of Table 2.1): Means and Mean Squared Errors (MSE) from datasets with  $M = 100$  (100) and  $M = 1000$  (1000) are shown for all parameter estimates.

$\frac{\delta}{\Delta} \rightarrow$	$10^{-4}$	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$\hat{\lambda}_{01}$ Mean	1.0014	0.9966	0.9912	0.9850	0.9773	0.9705	0.9683	0.9936	1.0061	1.0127	1.0079
$\hat{\lambda}_{01}$ MSE (100) $\times 10^{-4}$	3.6014	3.1477	4.0102	6.3179	8.5862	12.5950	18.1615	24.8810	19.6845	10.8940	6.4757
$\hat{\lambda}_{01}$ MSE (1000) $\times 10^{-4}$	0.3129	0.4614	0.9002	2.8093	5.8887	9.4462	14.7635	16.2443	3.9079	2.1320	1.0608
$\hat{\lambda}_{10}$ Mean	3.1604	3.1634	3.1610	3.1572	3.1662	3.1674	3.1578	3.1250	3.1262	3.1296	3.1454
$\hat{\lambda}_{10}$ MSE (100) $\times 10^{-4}$	36.9219	42.3621	40.6597	38.6860	39.7586	44.7820	60.3105	71.4363	61.3570	50.1797	40.2596
$\hat{\lambda}_{10}$ MSE (1000) $\times 10^{-4}$	3.4391	3.6448	3.9765	3.8151	3.7223	3.9205	4.7250	7.5561	32.5732	16.5867	5.6184
$\hat{\mu}_1$ Mean	0.1055	0.1058	0.1073	0.1080	0.1084	0.1089	0.1092	0.1061	0.1053	0.1044	0.1042
$\hat{\mu}_1$ MSE (100) $\times 10^{-4}$	1.0219	0.9556	1.0623	1.3524	1.2432	1.2929	1.4148	1.2926	1.1855	1.2989	0.9957
$\hat{\mu}_1$ MSE (1000) $\times 10^{-4}$	0.1093	0.1167	0.1311	0.1158	0.1848	0.1678	0.2003	0.2111	0.1536	0.1536	0.1225

**Table 2.15:**  $d = 0$  medium switching (study 2 of Table 2.1): Means and Mean Squared Errors (MSE) from datasets with  $M = 100$  (100) and  $M = 1000$  (1000) are shown for all parameter estimates.



$\frac{\delta}{\Delta} \rightarrow$	$10^{-4}$	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$\hat{\lambda}_{01}$ Mean	3.1587	3.1236	3.0899	3.0167	2.9548	2.8934	2.8359	2.8888	3.1553	3.2415	3.2202
$\hat{\lambda}_{01}$ MSE (100) $\times 10^{-3}$	4.2580	5.2544	9.0954	25.0617	47.3891	78.0354	114.2236	125.1663	122.2452	53.9577	29.2550
$\hat{\lambda}_{01}$ MSE (1000) $\times 10^{-3}$	0.2806	1.4529	5.2052	20.2176	44.1040	75.7929	109.6569	122.9487	16.5624	14.1061	6.1912
$\hat{\lambda}_{10}$ Mean	9.9869	10.0025	9.9970	10.0068	10.0000	9.9712	9.9610	9.8077	9.6313	9.7021	9.8473
$\hat{\lambda}_{10}$ MSE (100) $\times 10^{-3}$	46.0991	67.4772	67.8874	67.2548	82.5422	85.0991	75.3934	189.8646	235.7081	133.0161	74.5628
$\hat{\lambda}_{10}$ MSE (1000) $\times 10^{-3}$	4.0829	6.5112	6.2349	6.8307	9.0871	8.9417	13.7226	36.7124	243.7806	114.5101	33.4812
$\hat{\mu}_1$ Mean	0.3336	0.3412	0.3387	0.3476	0.3549	0.3598	0.3642	0.3590	0.3337	0.3205	0.3308
$\hat{\mu}_1$ MSE (100) $\times 10^{-3}$	1.0400	1.1957	1.0078	1.3647	1.8971	2.1253	2.2552	2.8067	2.2045	1.3641	1.1986
$\hat{\mu}_1$ MSE (1000) $\times 10^{-3}$	0.1057	0.1187	0.1353	0.2655	0.5306	0.7058	1.0352	0.9740	0.4821	0.3477	0.1859

**Table 2.16:**  $d = 0$  fast switching (study 3 of Table 2.1): Means and Mean Squared Errors (MSE) from datasets with  $M = 100$  (100) and  $M = 1000$  (1000) are shown for all parameter estimates.

$\frac{\delta}{\Delta} \rightarrow$	$10^{-4}$	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$\hat{\lambda}_{001}$ Mean	0.1522	0.1499	0.1521	0.1491	0.1501	0.1500	0.1493	0.1505	0.1517	0.1508	0.1515
$\hat{\lambda}_{001}$ MSE (100) $\times 10^{-4}$	3.1832	3.3354	3.2151	3.1054	3.0920	3.3521	2.3554	3.1366	3.4032	3.2362	3.4182
$\hat{\lambda}_{001}$ MSE (1000) $\times 10^{-4}$	0.3359	0.3018	0.3093	0.2870	0.3105	0.3521	0.3355	0.3240	0.3449	0.3322	0.3361
$\hat{\lambda}_{01}$ Mean	0.3008	0.2993	0.3002	0.2985	0.2985	0.2976	0.2980	0.3000	0.3005	0.3007	0.3000
$\hat{\lambda}_{01}$ MSE (100) $\times 10^{-4}$	0.8397	0.7758	0.7117	0.8869	0.8883	0.9291	0.9691	1.0938	1.0441	1.0730	0.8439
$\hat{\lambda}_{01}$ MSE (1000) $\times 10^{-4}$	0.0960	0.0820	0.0921	0.1022	0.1356	0.2059	0.2391	0.2297	0.1271	0.1090	0.1040
$\hat{\lambda}_{011}$ Mean	0.1002	0.0995	0.1000	0.0994	0.0994	0.0996	0.0997	0.0999	0.1004	0.0999	0.1001
$\hat{\lambda}_{011}$ MSE (100) $\times 10^{-4}$	0.2087	0.2164	0.2058	0.2201	0.2054	0.2090	0.1745	0.1948	0.2117	0.2425	0.2098
$\hat{\lambda}_{011}$ MSE (1000) $\times 10^{-4}$	0.0197	0.0216	0.0242	0.0176	0.0221	0.0256	0.0233	0.0215	0.0191	0.0199	0.0240
$\hat{\lambda}_{10}$ Mean	0.7985	0.7993	0.8005	0.7996	0.8008	0.8010	0.7997	0.7991	0.7971	0.7986	0.7988
$\hat{\lambda}_{10}$ MSE (100) $\times 10^{-4}$	1.7727	1.7760	1.8119	1.7244	1.8805	1.8472	2.2378	2.0228	1.8457	1.9318	1.7723
$\hat{\lambda}_{10}$ MSE (1000) $\times 10^{-4}$	0.1852	0.1642	0.1766	0.1630	0.1801	0.1611	0.2042	0.3483	0.2588	0.2562	0.1874
$\hat{\mu}_1$ Mean	0.0102	0.0101	0.0102	0.0101	0.0100	0.0101	0.0099	0.0101	0.0100	0.0100	0.0101
$\hat{\mu}_1$ MSE (100) $\times 10^{-4}$	0.0218	0.0246	0.0201	0.0232	0.0276	0.0278	0.0224	0.0238	0.0224	0.0226	0.0237
$\hat{\mu}_1$ MSE (1000) $\times 10^{-4}$	0.0022	0.0024	0.0028	0.0021	0.0024	0.0025	0.0026	0.0023	0.0028	0.0025	0.0025

**Table 2.17:**  $d = 1$  slow switching (study 4 of Table 2.1): Means and Mean Squared Errors (MSE) from datasets with  $M = 100$  (100) and  $M = 1000$  (1000) are shown for all parameter estimates.

$\frac{\delta}{\Delta} \rightarrow$	$10^{-4}$	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$\hat{\lambda}_{00_1}$ Mean	0.3553	0.3584	0.3544	0.3469	0.3438	0.3438	0.3408	0.3538	0.3472	0.3534	0.3495
$\hat{\lambda}_{00_1}$ MSE (100) $\times 10^{-4}$	25.6851	27.5699	27.2189	23.9852	27.1779	26.6249	23.8982	28.9833	27.3760	31.2609	28.7500
$\hat{\lambda}_{00_1}$ MSE (1000) $\times 10^{-4}$	2.8673	2.5538	2.7678	3.0469	3.0582	3.3915	3.6247	3.4359	3.5640	2.7995	3.2578
$\hat{\lambda}_{01}$ Mean	1.0029	1.0013	0.9947	0.9861	0.9774	0.9745	0.9734	0.9907	0.9997	1.0004	0.9999
$\hat{\lambda}_{01}$ MSE (100) $\times 10^{-4}$	11.8527	11.8486	12.4051	13.3163	16.8455	20.6807	23.6586	31.5621	27.5247	21.7894	19.0594
$\hat{\lambda}_{01}$ MSE (1000) $\times 10^{-4}$	1.3297	1.4966	1.6546	3.5872	5.9372	9.6060	13.8229	11.2374	3.0040	2.1362	2.0034
$\hat{\lambda}_{0_{11}}$ Mean	0.3007	0.3028	0.2999	0.2989	0.2964	0.2966	0.2950	0.2996	0.2996	0.3009	0.2988
$\hat{\lambda}_{0_{11}}$ MSE (100) $\times 10^{-4}$	3.7738	3.5633	3.6592	3.3376	3.8073	3.7841	3.5575	3.8889	4.1254	3.4619	4.0927
$\hat{\lambda}_{0_{11}}$ MSE (1000) $\times 10^{-4}$	0.4245	0.3556	0.3786	0.4918	0.5113	0.4931	0.5018	0.5469	0.5059	0.3990	0.4637
$\hat{\lambda}_{10}$ Mean	2.2919	2.2960	2.2995	2.3010	2.3050	2.3006	2.2971	2.2874	2.2788	2.2849	2.2908
$\hat{\lambda}_{10}$ MSE (100) $\times 10^{-4}$	24.5487	28.2653	25.7279	23.7907	27.0397	31.0944	35.0341	41.3352	33.0594	28.5781	25.8814
$\hat{\lambda}_{10}$ MSE (1000) $\times 10^{-4}$	2.4204	2.4919	2.1457	2.8788	3.2393	3.0825	3.1199	11.6558	13.9423	7.3326	3.6860
$\hat{\mu}_1$ Mean	0.0996	0.1012	0.1007	0.1023	0.1015	0.1016	0.1020	0.1013	0.1001	0.0995	0.1001
$\hat{\mu}_1$ MSE (100) $\times 10^{-4}$	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
$\hat{\mu}_1$ MSE (1000) $\times 10^{-4}$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

**Table 2.18:**  $d = 1$  medium switching (study 5 of Table 2.1): Means and Mean Squared Errors (MSE) from datasets with  $M = 100$  (100) and  $M = 1000$  (1000) are shown for all parameter estimates.

$\frac{\delta}{\Delta} \rightarrow$	$10^{-4}$	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$\hat{\lambda}_{00_1}$ Mean	1.9923	1.9843	1.9661	1.9170	1.9068	1.8723	1.9401	2.0162	2.0622	2.0605	2.0236
$\hat{\lambda}_{00_1}$ MSE (100) $\times 10^{-2}$	1.6820	1.7604	2.1581	2.5568	2.8720	3.9995	4.7420	5.2883	4.0860	3.2989	3.6349
$\hat{\lambda}_{00_1}$ MSE (1000) $\times 10^{-2}$	0.1811	0.2328	0.3443	0.7816	1.2921	1.9940	1.5430	0.7706	0.8508	0.5470	0.3272
$\hat{\lambda}_{01}$ Mean	9.8520	9.7112	9.5725	9.3138	9.0396	8.8526	9.0706	9.4378	9.8018	9.9327	9.7973
$\hat{\lambda}_{01}$ MSE (100) $\times 10^{-2}$	13.9078	19.4594	29.9885	57.2287	103.9807	144.7740	143.2314	116.6074	57.4840	47.1627	48.6892
$\hat{\lambda}_{01}$ MSE (1000) $\times 10^{-2}$	3.0531	9.3754	20.7562	53.2153	94.8435	138.2305	136.8150	35.6913	8.4465	6.2657	6.4313
$\hat{\lambda}_{0_{11}}$ Mean	0.6980	0.6986	0.6929	0.6913	0.6888	0.6804	0.6880	0.6942	0.6986	0.6988	0.7000
$\hat{\lambda}_{0_{11}}$ MSE (100) $\times 10^{-2}$	0.1321	0.1175	0.1289	0.1280	0.1430	0.1735	0.1687	0.1759	0.1706	0.1695	0.1843
$\hat{\lambda}_{0_{11}}$ MSE (1000) $\times 10^{-2}$	0.0153	0.0154	0.0171	0.0207	0.0319	0.0440	0.0441	0.0201	0.0150	0.0140	0.0180
$\hat{\lambda}_{10}$ Mean	9.7762	9.7862	9.7305	9.7983	9.7404	9.7100	9.4153	9.2234	9.1590	9.4201	9.6822
$\hat{\lambda}_{10}$ MSE (100) $\times 10^{-2}$	15.4335	20.4386	27.7001	28.6734	30.5012	33.6122	79.2853	91.6354	77.9001	39.1059	15.6054
$\hat{\lambda}_{10}$ MSE (1000) $\times 10^{-2}$	3.1389	8.2192	7.9311	9.2028	11.4808	15.3201	38.0538	94.2513	75.0687	37.7118	12.0628
$\hat{\mu}_1$ Mean	0.3320	0.3342	0.3365	0.3448	0.3463	0.3546	0.3426	0.3342	0.3259	0.3257	0.3328
$\hat{\mu}_1$ MSE (100) $\times 10^{-2}$	0.1006	0.1155	0.1127	0.1387	0.1518	0.1884	0.1712	0.1917	0.1239	0.1180	0.1188
$\hat{\mu}_1$ MSE (1000) $\times 10^{-2}$	0.0104	0.0101	0.0106	0.0190	0.0255	0.0326	0.0304	0.0279	0.0250	0.0220	0.0129

**Table 2.19:**  $d = 1$  fast switching (study 6 of Table 2.1): Means and Mean Squared Errors (MSE) from datasets with  $M = 100$  (100) and  $M = 1000$  (1000) are shown for all parameter estimates.

$\frac{\delta}{\Delta} \rightarrow$	$10^{-4}$	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$\hat{\lambda}_{00_1}$ Mean	0.1655	0.1708	0.1679	0.1781	0.1684	0.1575	0.1709	0.1669	0.1657	0.1675	0.1721
$\hat{\lambda}_{00_1}$ MSE (100) $\times 10^{-4}$	44.0687	48.1666	40.4240	56.9849	46.9907	34.6941	48.7498	40.4209	37.4097	39.4570	54.9917
$\hat{\lambda}_{00_1}$ MSE (1000) $\times 10^{-4}$	6.5107	4.6626	3.1803	5.7639	5.2326	4.9994	4.3972	4.3537	3.4098	4.1759	3.9254
$\hat{\lambda}_{0_1}$ Mean	0.2935	0.2979	0.2974	0.2927	0.2952	0.2915	0.2952	0.2933	0.2905	0.2920	0.2938
$\hat{\lambda}_{0_1}$ MSE (100) $\times 10^{-4}$	7.9565	7.2124	7.2222	7.3178	7.2362	7.5725	8.0150	6.8996	6.4214	7.3674	6.8179
$\hat{\lambda}_{0_1}$ MSE (1000) $\times 10^{-4}$	1.6610	2.0638	1.9069	2.1283	1.4913	1.6825	1.6591	2.0276	1.2090	1.2242	1.2403
$\hat{\lambda}_{0_1 0_2}$ Mean	0.0541	0.0527	0.0523	0.0568	0.0525	0.0547	0.0540	0.0549	0.0577	0.0564	0.0537
$\hat{\lambda}_{0_1 0_2}$ MSE (100) $\times 10^{-4}$	2.5311	2.9952	2.9581	2.3374	2.7167	2.4805	2.4671	2.3162	1.7318	1.9756	2.1198
$\hat{\lambda}_{0_1 0_2}$ MSE (1000) $\times 10^{-4}$	1.1641	1.4721	1.4550	1.0681	0.8202	0.9414	0.8395	0.7101	0.6520	0.7776	0.5784
$\hat{\lambda}_{0_{11}}$ Mean	0.0918	0.0969	0.0966	0.0927	0.0946	0.0880	0.0948	0.0937	0.0873	0.0898	0.0941
$\hat{\lambda}_{0_{11}}$ MSE (100) $\times 10^{-4}$	11.2553	11.4354	10.3173	10.4587	10.2103	11.9950	9.3411	9.2835	9.7223	9.8362	11.2910
$\hat{\lambda}_{0_{11}}$ MSE (1000) $\times 10^{-4}$	2.4837	2.6489	3.0001	1.8920	1.5318	2.1322	2.0623	2.0265	1.5063	1.6543	1.2973
$\hat{\lambda}_{0_{21}}$ Mean	0.0010	0.0012	0.0012	0.0009	0.0011	0.0010	0.0010	0.0010	0.0008	0.0008	0.0010
$\hat{\lambda}_{0_{21}}$ MSE (100) $\times 10^{-4}$	0.0047	0.0079	0.0077	0.0026	0.0075	0.0041	0.0045	0.0036	0.0017	0.0017	0.0035
$\hat{\lambda}_{0_{21}}$ MSE (1000) $\times 10^{-4}$	0.0017	0.0025	0.0034	0.0014	0.0013	0.0012	0.0010	0.0012	0.0011	0.0013	0.0011
$\hat{\lambda}_{10}$ Mean	0.8198	0.8181	0.8135	0.8262	0.8172	0.8222	0.8161	0.8208	0.8250	0.8267	0.8200
$\hat{\lambda}_{10}$ MSE (100) $\times 10^{-4}$	28.5496	29.2364	28.3205	26.0696	30.5973	26.4407	25.8326	26.6544	20.9132	25.9590	24.5221
$\hat{\lambda}_{10}$ MSE (1000) $\times 10^{-4}$	11.0639	11.2100	11.5840	9.2298	6.7888	8.7008	6.2439	6.7714	6.0394	7.3362	5.4836
$\hat{\mu}_1$ Mean	0.0295	0.0355	0.0353	0.0241	0.0342	0.0311	0.0325	0.0313	0.0228	0.0255	0.0326
$\hat{\mu}_1$ MSE (100) $\times 10^{-4}$	18.0527	18.2972	18.0223	16.9144	17.8902	15.6854	15.8466	15.0540	13.7984	13.4112	14.6161
$\hat{\mu}_1$ MSE (1000) $\times 10^{-4}$	10.1650	10.9061	10.2482	9.1750	6.0234	7.8785	6.6199	5.1603	5.1085	5.9133	4.7159

**Table 2.20:**  $d = 2$  slow switching (study 7 of Table 2.1): Means and Mean Squared Errors (MSE) from datasets with  $M = 100$  (100) and  $M = 1000$  (1000) are shown for all parameter estimates.

$\frac{\delta}{\Delta} \rightarrow$	$10^{-4}$	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$\hat{\lambda}_{00_1}$ Mean	0.8036	0.7956	0.7886	0.7854	0.7771	0.7741	0.7632	0.8116	0.8310	0.8319	0.8279
$\hat{\lambda}_{00_1}$ MSE (100) $\times 10^{-3}$	3.0144	2.8559	2.9198	2.9900	3.5072	3.4486	4.3369	5.7624	5.6709	5.2478	5.6638
$\hat{\lambda}_{00_1}$ MSE (1000) $\times 10^{-3}$	0.3128	0.2841	0.3111	0.6009	0.9634	1.2946	1.4308	0.9600	1.3473	1.3295	0.7073
$\hat{\lambda}_{0_1}$ Mean	3.9984	3.9547	3.9089	3.8364	3.7578	3.6906	3.6473	3.9421	4.1152	4.1213	4.0996
$\hat{\lambda}_{0_1}$ MSE (100) $\times 10^{-3}$	9.9119	11.3783	16.1007	36.5443	68.6703	106.2609	147.0739	144.2045	71.3273	49.7364	39.7939
$\hat{\lambda}_{0_1}$ MSE (1000) $\times 10^{-3}$	1.2355	3.2177	9.2571	29.9110	61.3052	101.7778	132.2323	35.7616	24.8360	19.5067	9.6962
$\hat{\lambda}_{0_1 0_2}$ Mean	0.1002	0.1019	0.1001	0.1002	0.0998	0.0992	0.0995	0.1001	0.1029	0.1017	0.1017
$\hat{\lambda}_{0_1 0_2}$ MSE (100) $\times 10^{-3}$	0.1459	0.1478	0.1169	0.1333	0.1114	0.1104	0.0982	0.0984	0.1200	0.1025	0.1032
$\hat{\lambda}_{0_1 0_2}$ MSE (1000) $\times 10^{-3}$	0.0186	0.0128	0.0156	0.0168	0.0120	0.0106	0.0101	0.0163	0.0124	0.0180	0.0178
$\hat{\lambda}_{0_1 1}$ Mean	0.4013	0.3980	0.3966	0.3953	0.3971	0.3953	0.3893	0.3970	0.4045	0.4050	0.4049
$\hat{\lambda}_{0_1 1}$ MSE (100) $\times 10^{-3}$	0.8079	0.7637	0.7999	0.8251	0.8119	0.7423	0.8910	1.0031	1.0903	1.0056	1.0403
$\hat{\lambda}_{0_1 1}$ MSE (1000) $\times 10^{-3}$	0.0791	0.0683	0.0653	0.1070	0.1447	0.1444	0.1803	0.0983	0.1160	0.1195	0.0974
$\hat{\lambda}_{0_2 1}$ Mean	0.0051	0.0050	0.0052	0.0052	0.0053	0.0053	0.0052	0.0052	0.0053	0.0052	0.0053
$\hat{\lambda}_{0_2 1}$ MSE (100) $\times 10^{-3}$	0.0012	0.0011	0.0011	0.0010	0.0009	0.0009	0.0007	0.0006	0.0006	0.0005	0.0006
$\hat{\lambda}_{0_2 1}$ MSE (1000) $\times 10^{-3}$	0.0002	0.0001	0.0001	0.0002	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
$\hat{\lambda}_{10}$ Mean	7.9687	7.9803	7.9715	7.9827	7.9840	7.9804	7.9415	7.6940	7.6285	7.7437	7.8902
$\hat{\lambda}_{10}$ MSE (100) $\times 10^{-3}$	25.0985	36.8189	38.8441	39.8706	39.5881	46.9644	75.1187	202.7373	163.6210	86.1649	32.2554
$\hat{\lambda}_{10}$ MSE (1000) $\times 10^{-3}$	3.3643	4.0670	4.1188	4.4424	3.9562	5.9792	15.2332	195.7356	157.2629	65.2038	13.7654
$\hat{\mu}_1$ Mean	0.1013	0.0987	0.1026	0.1073	0.1128	0.1157	0.1152	0.1093	0.1073	0.1082	0.1109
$\hat{\mu}_1$ MSE (100) $\times 10^{-3}$	1.0780	0.9975	0.7788	0.7643	0.8492	0.8663	0.7543	0.4665	0.3883	0.3751	0.4219
$\hat{\mu}_1$ MSE (1000) $\times 10^{-3}$	0.1209	0.1109	0.1112	0.1647	0.1523	0.2716	0.2475	0.1378	0.0711	0.0851	0.1268

**Table 2.21:**  $d = 2$  medium switching (study 8 of Table 2.1): Means and Mean Squared Errors (MSE) from datasets with  $M = 100$  (100) and  $M = 1000$  (1000) are shown for all parameter estimates.

$\frac{\delta}{\Delta} \rightarrow$	$10^{-4}$	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$\hat{\lambda}_{00_1}$ Mean	2.0359	2.0293	2.0469	2.0815	2.2804	2.5864	2.6748	2.6492	2.6063	2.5731	2.4697
$\hat{\lambda}_{00_1}$ MSE (100) $\times 10^{-2}$	2.4447	2.3738	3.0898	4.2635	20.8307	50.8394	55.0500	51.6242	44.0403	41.0364	28.3079
$\hat{\lambda}_{00_1}$ MSE (1000) $\times 10^{-2}$	0.2685	0.3666	0.5155	0.7732	2.6830	33.2383	40.8341	40.2410	35.8971	28.6887	22.1064
$\hat{\lambda}_{0_1}$ Mean	9.8536	9.7290	9.6305	9.5301	9.9097	10.9416	11.4587	11.5278	11.5300	11.5069	11.1842
$\hat{\lambda}_{0_1}$ MSE (100) $\times 10^{-2}$	17.2377	21.3207	26.7591	37.4432	94.3525	239.7721	312.6215	330.2248	302.7605	295.8218	186.1543
$\hat{\lambda}_{0_1}$ MSE (1000) $\times 10^{-2}$	2.9915	8.3804	14.5863	27.9367	26.5326	97.9595	183.0150	229.2597	239.9449	205.5458	153.9862
$\hat{\lambda}_{0_1 0_2}$ Mean	0.2049	0.2116	0.2156	0.2253	0.2391	0.2524	0.2590	0.2631	0.2661	0.2703	0.2709
$\hat{\lambda}_{0_1 0_2}$ MSE (100) $\times 10^{-2}$	0.0531	0.0780	0.0884	0.1358	0.2402	0.3600	0.4144	0.4701	0.5090	0.5766	0.5755
$\hat{\lambda}_{0_1 0_2}$ MSE (1000) $\times 10^{-2}$	0.0071	0.0148	0.0225	0.0558	0.1152	0.2843	0.3347	0.3911	0.4220	0.4433	0.4663
$\hat{\lambda}_{0_1 1}$ Mean	0.6910	0.6854	0.6826	0.6704	0.6632	0.6737	0.6747	0.6752	0.6734	0.6721	0.6661
$\hat{\lambda}_{0_1 1}$ MSE (100) $\times 10^{-2}$	0.3199	0.3288	0.3566	0.3887	0.4659	0.4169	0.4301	0.4615	0.4717	0.5270	0.5524
$\hat{\lambda}_{0_1 1}$ MSE (1000) $\times 10^{-2}$	0.0380	0.0438	0.0698	0.1232	0.2166	0.1193	0.1187	0.1018	0.1261	0.1720	0.2028
$\hat{\lambda}_{0_2 1}$ Mean	0.0098	0.0091	0.0089	0.0082	0.0071	0.0064	0.0062	0.0061	0.0059	0.0059	0.0058
$\hat{\lambda}_{0_2 1}$ MSE (100) $\times 10^{-2}$	0.0003	0.0004	0.0004	0.0006	0.0010	0.0014	0.0016	0.0016	0.0017	0.0018	0.0018
$\hat{\lambda}_{0_2 1}$ MSE (1000) $\times 10^{-2}$	0.0000	0.0001	0.0001	0.0004	0.0008	0.0013	0.0015	0.0016	0.0017	0.0017	0.0018
$\hat{\lambda}_{10}$ Mean	9.5735	9.4538	9.4108	9.3594	8.9483	8.5021	8.5916	8.9477	9.3422	9.7067	10.0918
$\hat{\lambda}_{10}$ MSE (100) $\times 10^{-2}$	36.2423	52.0110	56.4493	65.8577	173.3694	258.4795	207.4896	116.7524	49.0383	14.6224	7.3112
$\hat{\lambda}_{10}$ MSE (1000) $\times 10^{-2}$	15.9769	31.8406	38.4982	48.2685	88.8552	281.1336	200.7809	114.7007	47.4346	9.0499	1.1104
$\hat{\mu}_1$ Mean	0.3233	0.3205	0.3193	0.3121	0.2905	0.2688	0.2614	0.2664	0.2727	0.2801	0.2898
$\hat{\mu}_1$ MSE (100) $\times 10^{-2}$	0.1398	0.1778	0.1546	0.1737	0.3280	0.5032	0.5754	0.5245	0.4283	0.3572	0.2602
$\hat{\mu}_1$ MSE (1000) $\times 10^{-2}$	0.0207	0.0297	0.0368	0.0639	0.1449	0.5308	0.5420	0.4790	0.4057	0.3094	0.2209

**Table 2.22:**  $d = 2$  fast switching (study 9 of Table 2.1) of Table 2.1): Means and Mean Squared Errors (MSE) from datasets with  $M = 100$  (100) and  $M = 1000$  (1000) are shown for all parameter estimates.

# 3

## MOLECULAR COUNTING

### 3.1 Introduction

Under different experimental conditions, the varying photo-switching rates between the photo-kinetic states of a fluorophore give rise to a collection of sparsely localised molecules at each time frame. The temporally varying fluorescence not only allows for inference upon photo-switching rates, using for example the PSHMM method detailed in Chapter 1, but can also provide quantitative information on molecular stoichiometries. In particular, being able to accurately count the number of fluorescently labelled molecules from the recorded localisations would allow much greater insight into the cellular structures and processes under observation. However, this is a notoriously difficult task as deriving the probability distribution for the number of localisations 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 photo-activated localisation microscopy (PALM) and are not wholly applicable or adaptable for counting molecules that are imaged via (d)STORM. For instance, the PALM methods of [Lee](#)

et al. (2012), Fricke et al. (2015), Nino et al. (2017), Rollins et al. (2014) assume a four state kinetic model (inactive, photon-emitting/On, dark and photo-bleached) for the photo-activatable fluorophore. Each fluorophore begins in a non-emissive dark (inactive\*) state before briefly moving into the photon-emitting On state. While there is then the small possibility of a small number of repeat transitions between this and a temporary dark state, to align with PALM, these methods assume that photo-bleaching is most likely to occur after the fluorophore has reached the On state for the first time. However, this kinetic model is inappropriate for (d)STORM, in which all fluorophores start in the On state, before stochastically moving back and forth between this and one or more dark states, and then permanently photo-bleaching.

The analysis of Nieuwenhuizen et al. (2015) is applicable for (d)STORM experiments, however, it assumes an underlying photo-kinetic model with  $d = 0$ , enabling fluorophores to only occupy three states (On, dark and photo-bleached). In this method, the switching between the On and dark state is modelled with a Poisson distribution and the bleaching is governed by a geometric distribution, therefore giving rise to a Poisson-geometric mixture distribution on the number of blinks. However, from Lin et al. (2015) and the analysis presented in Chapter 1, empirical evidence supports the existence of multiple dark states, rendering this method difficult to apply in practice.

Importantly, common to the methods in Lee et al. (2012), Fricke et al. (2015), Nino et al. (2017), Nieuwenhuizen et al. (2015), 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. Firstly, a fluorophore can briefly transition from the On state into a dark state and back again within a single camera frame of time  $\Delta$ ; this transition will not be detected as a separate blink. Secondly, a 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

---

\*An inactive state is a transient dark state which fluorophores imaged in PALM occupy before their photo-switching behaviour has been *activated*, and they begin to be observed.

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, [Rollins et al. \(2014\)](#) 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 (d)STORM, particularly in densely labelled environments, since the nuanced photo-switching behaviour means we cannot be certain of a specific fluorophore’s photo-kinetic state at any one time.

### 3.1.1 Chapter summary

In this chapter, we present a method of molecular counting that utilises the photo-switching and observation model developed in Chapter 1. In our method, both missed blinks and false positives are fully accounted for in the modelling, something which has been absent from molecular counting methods thus far. Furthermore, we perform counting using just the localisation count, making our method highly scalable and able to count thousands of molecules with computational ease.

In Section 3.2, we will use the observation process  $\{Y_n\}$  to carefully derive the *exact* form of the probability mass function of the number of localisations  $N_l$  a single fluorophore produces during an imaging experiment, and also prove the form of its probability generating function. Derivatives of this probability generating function will be subsequently used to calculate the expected number and variance of localisations per molecule. Importantly, we will show that these statistics are fully characterised by the underlying photo-switching parameters  $\theta_\omega^{(d)}$ . We then use the fast Fourier transform to extend this distribution to characterise the probability mass function of  $N_l$ , obtained from  $M$  (unknown) molecules during an experiment. In Section 3.3 we discuss how using training data to estimate the unknown photo-switching rates, leads us to being able to compute the posterior distribution of  $M$  given  $N_l$ , and therefore an estimate for  $M$  via its mode. In Section 3.4, we demonstrate the validity of our method through simulations conducted



under different experimental conditions, from three photo-switching models. Finally, in Section 3.5, we 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.

## 3.2 Cumulative localisations

In this section, we will describe how the probability mass function of the total number of localisations collected during an experiment can be computationally recovered given photo-switching parameters  $\boldsymbol{\theta}_\omega^{(d)}$ . Specifically, given an *unknown* number of  $M$  independently fluorescing molecules, with localisation processes  $\{Y_{n,1}\}, \{Y_{n,2}\}, \dots, \{Y_{n,M}\}$ , we will use the PSHMM derived in Chapter 1 to characterise the distribution of

$$N_l = \sum_{m=1}^M \sum_{n=1}^{N_F} Y_{n,m}, \quad (3.1)$$

the *cumulative* number of localisations obtained over an experiment of length  $N_F$  frames. In order to do so, we will firstly derive the probability mass function for  $N_l$  made by *a single fluorophore* (i.e. when  $M = 1$ ), and provide simulations to highlight key characteristics of this function, including its variability among a number of different photo-kinetic models. We then derive its probability generating function (pgf) and give expressions for its first two moments, from which both the expected value and variance of  $N_l$  can be deduced. We will subsequently explain how the Fast Fourier Transform can be used to efficiently perform  $M$  convolutions of this density, thereby recovering the probability distribution of  $N_l$ , when  $M > 1$ . This density, which will be seen as a function of  $M$  and the parameter vector  $\boldsymbol{\theta}_\omega^{(d)}$ , will then be used to derive the posterior mass function of  $M$  given  $N_l$  and  $\boldsymbol{\theta}_\omega^{(d)}$ .

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

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

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

where the sum is taken over the values  $Y_1, \dots, Y_n$  from the observed process  $\{Y_n\}$ . Since each  $Y_i$  is a function of the photo-switching parameters  $\boldsymbol{\theta}_\omega^{(d)}$ , we will show that the probability mass function of  $S_n$  given  $\boldsymbol{\theta}_\omega^{(d)}$  can be recovered computationally. In particular, we will be looking to find the probability mass function for  $S_{N_F}$  when a microscopy experiment is conducted over a known number of  $N_F$  frames.

### 3.2.1 Probability mass function

For any  $n \geq 1$ , Proposition 3 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 3.

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

$$\mathbf{R}(s, n) := \left( R(0, s, n) \quad R(0_1, s, n) \quad \dots \quad R(0_d, s, n) \quad R(1, s, n) \quad R(2, s, n) \right),$$

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

$$R(j, s, n) := \mathbb{P}_{\boldsymbol{\theta}_\omega^{(d)}}(X(n\Delta) = j, S_n = s). \quad (3.2)$$

By recursively computing

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

the probability mass of  $S_n$  follows

$$p_{\boldsymbol{\theta}_\omega^{(d)}}(S_n = s) := \mathbb{P}_{\boldsymbol{\theta}_\omega^{(d)}}(S_n = s) = \mathbf{R}(s, n) \mathbf{1}_{d+3} \quad s \in \mathcal{S}_{S_n}. \quad (3.3)$$

*Proof.* Let  $\mathbf{R}$  be as defined in (3.2).

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

$$\begin{aligned} R(j, s, 1) &= \sum_{i \in \mathcal{S}_X} \mathbb{P}_{\boldsymbol{\theta}_\omega^{(d)}}(X(\Delta) = j, Y_0 = s | X(0) = i) \mathbb{P}_{\boldsymbol{\theta}^{(d)}}(X(0) = i) \\ &= \sum_{i \in \mathcal{S}_X} b_{ij, \Delta}^{*(s)} \mathbb{P}_{\boldsymbol{\theta}^{(d)}}(X(0) = i). \end{aligned}$$

$$\text{Then, } \mathbf{R}(s, 1) = \boldsymbol{\nu}_X^\top B_\Delta^{*(s)}.$$

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

$$\begin{aligned} R(j, 0, n) &= \sum_{i \in \mathcal{S}_X} \mathbb{P}_{\boldsymbol{\theta}_\omega^{(d)}}(X(n\Delta) = j, S_n = 0 | X((n-1)\Delta) = i, S_{n-1} = 0) \\ &\quad \times R(i, 0, n-1) \\ &= \sum_{i \in \mathcal{S}_X} b_{ij, \Delta}^{*(0)} R(i, 0, n-1). \end{aligned}$$

$$\text{Then, } \mathbf{R}(0, n) = \mathbf{R}(0, n-1) B_\Delta^{*(0)}.$$

For  $1 \leq s < n$  we have

$$\begin{aligned} R(j, s, n) &= \sum_{x=s-1}^n \sum_{i \in \mathcal{S}_X} \mathbb{P}_{\boldsymbol{\theta}_\omega^{(d)}}(X(n\Delta) = j, S_n = s | X((n-1)\Delta) = i, S_{n-1} = x) \\ &\quad \times R(i, s-x, n-1) \\ &= \sum_{x=0}^1 \sum_{i \in \mathcal{S}_X} b_{ij, \Delta}^{*(x)} R(i, s-x, n-1). \end{aligned}$$

$$\text{Then, } \mathbf{R}(s, n) = \mathbf{R}(s, n-1) B_\Delta^{*(0)} + \mathbf{R}(s-1, n-1) B_\Delta^{*(1)}.$$

And finally for  $s = n$ , we have

$$\begin{aligned} R(j, n, n) &= \sum_{i \in \mathcal{S}_X} \mathbb{P}_{\boldsymbol{\theta}_\omega^{(d)}}(X(n\Delta) = j, S_n = n | X((n-1)\Delta) = i, S_{n-1} = n-1) \\ &\quad \times R(i, n-1, n-1) \\ &= \sum_{i \in \mathcal{S}_X} b_{ij, \Delta}^{*(1)} R(i, n-1, n-1). \end{aligned}$$

Then,  $\mathbf{R}(n, n) = \mathbf{R}(n-1, n-1) B_\Delta^{*(1)}$ .

Now since

$$\mathbb{P}_{\boldsymbol{\theta}_\omega^{(d)}}(S_n = s) = \sum_{j \in \mathcal{S}_X} \mathbb{P}_{\boldsymbol{\theta}_\omega^{(d)}}(X(n\Delta) = j, S_n = s),$$

we obtain

$$p_{\boldsymbol{\theta}_\omega^{(d)}}(S_n = s) := \mathbb{P}_{\boldsymbol{\theta}_\omega^{(d)}}(S_n = s) = \mathbf{R}(s, n) \mathbf{1}_{d+3} \quad s \in \mathcal{S}_{S_n}.$$

□

---

**Algorithm 3** Compute probability mass function (PMF) for  $S_{N_F}$

---

**function** PMF\_S( $\boldsymbol{\theta}_\omega^{(d)}, \Delta, N_F$ )

$B_\Delta^{*(0)}, B_\Delta^{*(1)} \leftarrow \text{COMPUTE\_TRANSMISSIONS}(\boldsymbol{\theta}_\omega^{(d)}, \Delta)$  ▷ Using

Algorithm 2

$R_0, R_1 \leftarrow \mathbf{0}_{N_F+1} \mathbf{0}_{d+3}^\top$

$R_0[1, :] \leftarrow \boldsymbol{\nu}_X^\top B_\Delta^{*(0)}$

$R_0[2, :] \leftarrow \boldsymbol{\nu}_X^\top B_\Delta^{*(1)}$

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

$R_1[1, :] \leftarrow R_0[1, :] B_\Delta^{*(0)}$

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

$R_1[s, :] \leftarrow R_0[s, :] B_\Delta^{*(0)} + R_0[s-1, :] B_\Delta^{*(1)}$

$R_1[n+1, :] \leftarrow R_0[n, :] B_\Delta^{*(1)}$

$R_0 \leftarrow R_1$

$\mathbf{p} \leftarrow R_0 \mathbf{1}_{d+3}$

▷  $\mathbf{p}[i] = \mathbb{P}_{\boldsymbol{\theta}_\omega^{(d)}}(S_{N_F} = i-1)$  for  $i = 1, \dots, N_F + 1$

**return**  $\mathbf{p}$

▷ Probability mass function for  $S_{N_F}$

---

Figure 3.1 presents the exact distributions  $p_{\theta_\omega^{(d)}}(S_{N_F} = s)$  for  $s \in \mathbb{Z}_{\geq 0}$  when compared with 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 states and the photo-bleaching rates. Moreover, as is to be expected, the average number of localisations 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 3.1b, which compares the mass functions under the  $d = 1$  model with  $\mu_0 = 0$  when  $\mu_1$  varies. When  $\mu_1$  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  $\mu_1$  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 localisations recorded per molecule.

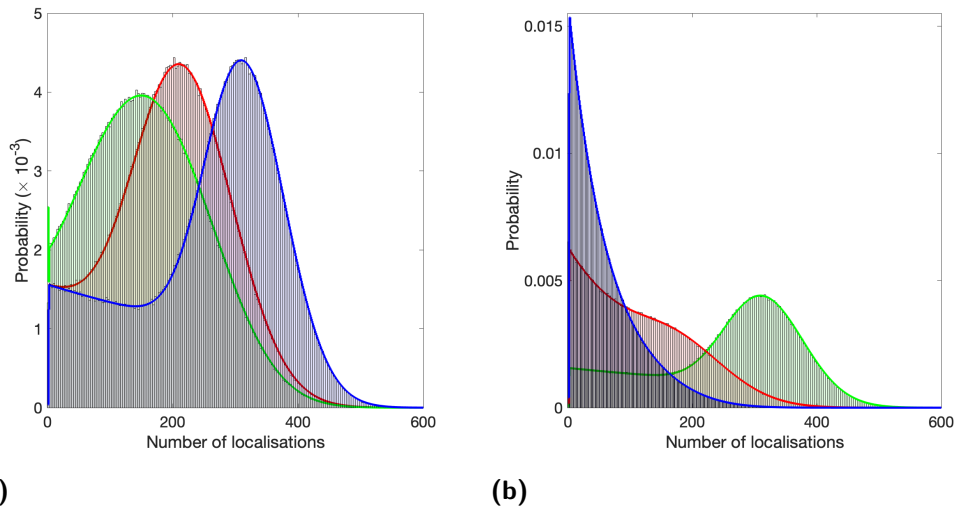
### 3.2.2 Probability generating function

Here, we will show that the probability generating function of  $S_n$  can be computed in a similar recursive fashion to its probability mass function. Proposition 4 gives its exact form.

**Proposition 4.** *For any  $n \in \mathbb{Z}_{>0}$ , the probability generating function (pgf) of  $S_n$ ,  $G_{S_n}(z) = \mathbb{E}_{\theta_\omega^{(d)}}(z^{S_n})$  is given by*

$$G_{S_n}(z) = \boldsymbol{\nu}_X^\top (B_\Delta^{*(0)} + zB_\Delta^{*(1)})^n \mathbf{1}_{d+3}. \quad (3.4)$$

*Proof.* By defining the vector quantity  $\mathbf{G}_{S_n}(z) := \sum_{i=0}^n \mathbf{R}(i, n)z^i$ , we have  $G_{S_n}(z) = \mathbf{G}_{S_n}(z)\mathbf{1}_{d+3}$ , similarly to (3.3). We therefore need to equivalently show that  $\mathbf{G}_{S_n}(z) = \boldsymbol{\nu}_X^\top (B_\Delta^{*(0)} + zB_\Delta^{*(1)})^n$ .



**Figure 3.1:** Figure 3.1a shows the theoretical and histogram estimate (from  $10^6$  simulations) of  $p_{\theta_\omega^{(d)}}(S_{N_F} = s)$  under 3 photo-switching models:  $d = 0$  (blue),  $d = 1$  (red) and  $d = 2$  (green). In all simulations,  $\mu_1 > 0, \mu_0 = \dots = \mu_{0_d} = 0, N_F = 1000, \nu_0 = \nu_1 = 0.5, \Delta = \frac{1}{30}s, \delta = 10^{-3}s$  and  $\omega = 10^{-6}$ ; rates chosen (where appropriately zero) are  $\lambda_{00_1} = 0.35s^{-1}, \lambda_{01} = 1s^{-1}, \lambda_{0_10_2} = 0.2s^{-1}, \lambda_{0_11} = 0.3s^{-1}, \lambda_{0_21} = 0.1s^{-1}, \lambda_{10} = 2.3s^{-1}, \mu_1 = 0.05s^{-1}$ . Figure 3.1b shows the theoretical and histogram estimate (from  $10^6$  simulations) of  $p_{\theta_\omega^{(d)}}(S_{N_F} = s)$  when  $d = 1$  with  $\mu_1 = 0.5s^{-1}$  (blue),  $\mu_1 = 0.2s^{-1}$  (red) and  $\mu_1 = 0.05s^{-1}$  (green).

The statement in (3.4) is true for  $n = 1$ , since

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

Assuming (3.4) is true for  $n = k$ , consider  $n = k + 1$ :

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

□

## Moments

In theory, the moments of the distribution  $p_{\boldsymbol{\theta}_\omega^{(d)}}(S_n = s)$  are fully characterised by the pgf given in (3.4). In particular, the expected value of  $S_n$ , denoted by  $\mathbb{E}_{\boldsymbol{\theta}_\omega^{(d)}}(S_n) = G'_{S_n}(1)$  and its second moment  $\mathbb{E}_{\boldsymbol{\theta}_\omega^{(d)}}(S_n^2) = G''_{S_n}(1) + \mathbb{E}_{\boldsymbol{\theta}_\omega^{(d)}}(S_n)$  can be determined by differentiating this pgf from first principles. Lemma 5 computes  $G'_{S_n}(1)$  and  $G''_{S_n}(1)$  to give explicit expressions for the expected value  $\mathbb{E}_{\boldsymbol{\theta}_\omega^{(d)}}(S_n)$  and variance  $\text{Var}_{\boldsymbol{\theta}_\omega^{(d)}}(S_n)$  of the random variable  $S_n$ .

**Lemma 5.** *The expected value  $\mathbb{E}_{\theta_\omega^{(d)}}(S_n)$  and variance  $\text{Var}_{\theta_\omega^{(d)}}(S_n)$  of  $S_n$  follows*

$$\mathbb{E}_{\theta_\omega^{(d)}}(S_n) = \boldsymbol{\nu}_X^T \left[ \sum_{i=1}^n e^{G\Delta(n-i)} B_\Delta^{*(1)} e^{G\Delta(i-1)} \right] \mathbf{1}_{d+3} \quad (3.5)$$

$$\text{Var}_{\theta_\omega^{(d)}}(S_n) = G''_{S_n}(1) + \mathbb{E}_{\theta_\omega^{(d)}}(S_n) - \mathbb{E}_{\theta_\omega^{(d)}}^2(S_n), \quad (3.6)$$

where

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

and  $e^G$  denotes the matrix exponential of the generator matrix  $G$  as defined in (1.1).

*Proof.* See Section 3.A of Appendix 3. □

### 3.2.3 Extension to $M$ molecules

When  $M \in \mathbb{Z}_{>0}$  independent molecules are imaged, the total number of localisations  $N_l$  (which can take a minimum value of 0 and a maximum value of  $MN_F$ ), can be written as

$$N_l = \sum_{m=1}^M S_{N_F, m} = \sum_{m=1}^M \sum_{i=1}^{N_F} Y_{i, m} \quad N_l \in [0, MN_F], \quad (3.7)$$

where  $S_{N_F, m}$  denotes the total number of localisations made by the  $m$ th fluorophore over an experiment consisting of imaging  $N_F$  frames. Using this expression for  $N_l$ , it now follows that the expected number and variance of *total* localisations are  $\mathbb{E}_{\theta, M}(N_l) = M\mathbb{E}_{\theta_\omega^{(d)}}(S_{N_F})$  and  $\text{Var}_{\theta, M}(N_l) = M\text{Var}_{\theta_\omega^{(d)}}(S_{N_F})$ , which can be computed using (3.5) and (3.6).



## Probability mass function

The probability mass function for  $N_l$  follows

$$p_{\boldsymbol{\theta}_\omega^{(d)},M}(N_l) = \sum_{\substack{s_1, \dots, s_M \\ :s_1 + \dots + s_M = N_l}} \prod_{i=1}^M p_{\boldsymbol{\theta}_\omega^{(d)}}(S_{N_F} = s_i),$$

which can be obtained by applying  $M$  convolutions of the mass function for  $S_{N_F}$ . This is most efficiently achieved via the Fast Fourier Transform.

Specifically, 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

$$\begin{aligned} \gamma_{S_{N_F}}(u) &:= \mathbb{E}_{\boldsymbol{\theta}_\omega^{(d)}}(e^{iuS_{N_F}}) = \sum_{s=0}^{\infty} \mathbb{P}_{\boldsymbol{\theta}_\omega^{(d)}}(S_{N_F} = s) e^{ius} \\ &= \sum_{s=0}^{N_F} p_{\boldsymbol{\theta}_\omega^{(d)}}(S_{N_F} = s) e^{ius}, \end{aligned}$$

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

$$\begin{aligned} \mathbb{E}_{\boldsymbol{\theta}_\omega^{(d)},M}(e^{iuN_l}) &= \mathbb{E}_{\boldsymbol{\theta}_\omega^{(d)},M}\left(e^{iu(S_{N_F,1} + \dots + S_{N_F,M})}\right) \\ &= \mathbb{E}_{\boldsymbol{\theta}_\omega^{(d)},M}\left(e^{iuS_{N_F,1}} \dots e^{iuS_{N_F,M}}\right) \\ &= \prod_{m=1}^M \mathbb{E}_{\boldsymbol{\theta}_\omega^{(d)}}\left(e^{iuS_{N_F,m}}\right) \quad (\text{since } S_{N_F,1}, \dots, S_{N_F,M} \text{ are independent}) \\ &= \gamma_{S_{N_F}}^M(u) \quad (\text{since } S_{N_F,1}, \dots, S_{N_F,M} \text{ are identically distributed}). \end{aligned} \tag{3.8}$$

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  $\{0, \dots, N\}$ . When  $N = N_F$ , this enables

$$\mathcal{F}_{s \rightarrow k}(p_{\boldsymbol{\theta}_\omega^{(d)}}(S_{N_F})) := \gamma_{S_{N_F}}(-u_{N_F}) = \sum_{s=0}^{N_F} p_{\boldsymbol{\theta}_\omega^{(d)}}(S_{N_F} = s) e^{-it_{N_F} ks},$$

to be seen as the Discrete Fourier Transform (DFT) of the probability mass

$p_{\boldsymbol{\theta}_\omega^{(d)}}(S_{N_F} = s)$ , where  $\mathcal{F}_{s \rightarrow k}(\cdot)$  denotes the discrete Fourier operator. The inverse DFT can then recover the probabilities via

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

Using the characteristic function of  $N_l$  from (3.8), it now follows that probability mass  $p_{\boldsymbol{\theta}_\omega^{(d)}, M}(N_l = s) := \mathbb{P}_{\boldsymbol{\theta}_\omega^{(d)}, M}(N_l = s)$  (where  $N_l$  takes values in the set  $\{0, \dots, MN_F\}$ ), can be recovered via

$$p_{\boldsymbol{\theta}_\omega^{(d)}, M}(N_l = s) = \frac{1}{MN_F + 1} \sum_{k=0}^{MN_F} \gamma_{S_{N_F}}^M(-t_{MN_F}k) e^{it_{MN_F}ks}, \quad (3.9)$$

so that  $p_{\boldsymbol{\theta}_\omega^{(d)}, M}(N_l = s) = \mathcal{F}_{k \rightarrow s}^{-1}(\gamma_{S_{N_F}}^M(-t_{MN_F}k)) = \mathcal{F}_{k \rightarrow s}^{-1}(\mathcal{F}_{s \rightarrow k}^M(p_{\boldsymbol{\theta}_\omega^{(d)}}(S_{N_F})))$ . 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  $p_{\boldsymbol{\theta}_\omega^{(d)}}(S_{N_F} = s)$ . The first  $N_F + 1$  elements of  $\mathbf{p}$  are therefore those outputted by Algorithm 3 and the remaining  $N_F(M - 1)$  elements are zeros. Algorithm 4 provides the scheme to compute the probability distribution of  $N_l$  using this reasoning.

---

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

---

- 1: **function** PMF\_NL( $\mathbf{p}_1, M$ ) ▷  $\mathbf{p}_1 \leftarrow \text{PMF\_S}(\boldsymbol{\theta}_\omega^{(d)}, \Delta, N_F)$  from Algorithm 3
  - 2:      $\mathbf{p}_2 \leftarrow [\mathbf{p}_1^\top \mathbf{0}_{N_F(M-1)}^\top]^\top$
  - 3:      $\mathbf{f} \leftarrow \mathcal{F}(\mathbf{p}_2)$  ▷ Apply Discrete Fourier Transform (DFT) to  $\mathbf{p}_2$  to get  $\mathbf{f}$
  - 4:      $\mathbf{f}_M \leftarrow \mathbf{f}^M$  ▷  $\mathbf{f}_M[i] = \mathbf{f}[i]^M$  for  $i = 1, \dots, MN_F + 1$
  - 5:      $\mathbf{p} \leftarrow \mathcal{F}^{-1}(\mathbf{f}_M)$  ▷ Apply inverse DFT to  $\mathbf{f}_M$  to get  $\mathbf{p}$ , where  $\mathbf{p}[i] = \mathbb{P}_{\boldsymbol{\theta}_\omega^{(d)}, M}(N_l = i - 1)$  for  $i = 1, \dots, MN_F + 1$
  - 6:     **return**  $\mathbf{p}$  ▷ Probability mass function for  $N_l$
-

### 3.3 Inference

In this section, we discuss how the (posterior) distribution of  $M$  can be obtained using the derived probability mass function of  $N_l$  in (3.9) given  $M$  and  $\boldsymbol{\theta}_\omega^{(d)}$ . Here, we also discuss how the mode of the resulting distribution can be used to estimate  $M$ , and provide a scheme to compute posterior 95% credible intervals.

The task of interest is to estimate  $M$ , the unknown number of molecules in a (d)STORM experiment, from  $N_l$ , the number of localisations recorded across  $N_F$  frames. Given a datum  $N_l$ , that takes the functional form in Equation (3.7), inferring both  $M$  and  $\boldsymbol{\theta}_\omega^{(d)}$  from their joint posterior distribution is difficult due to poor model specification. To deal with this, we choose to utilise a method of inference, formally termed as *modular inference* (Jacob et al., 2017), which can be implemented to determine parameters of interest in misspecified models, or models which produce poor quality data, as is the case here.

Specifically, since  $\boldsymbol{\theta}_\omega^{(d)}$  is well estimated through, for example, the PSHMM (maximum likelihood) fitting method outlined in Chapter 1, we may obtain an estimate  $\hat{\boldsymbol{\theta}}_\omega^{(\hat{d})}$  of the photo-switching parameter vector  $\boldsymbol{\theta}_\omega^{(d)}$  using *training* data  $\mathcal{D}_{tr}$  or through the *module*  $(\boldsymbol{\theta}_\omega^{(d)}, \mathcal{D}_{tr})$ . This training data consists of a set of observations of the localisation process  $\{Y_n\}$  from a known number of molecules, from which  $\hat{\boldsymbol{\theta}}_\omega^{(\hat{d})}$  can be obtained via the PSHMM. Subsequently,  $M$  can be inferred by utilising the determined estimate  $\hat{\boldsymbol{\theta}}_\omega^{(\hat{d})}$  to compute its posterior distribution under the module  $(M, N_l)$  as

$$p_{\hat{\boldsymbol{\theta}}_\omega^{(\hat{d})}, m}(M = m | N_l) \propto p_{\hat{\boldsymbol{\theta}}_\omega^{(\hat{d})}, m}(N_l) \pi_M(m), \quad (3.10)$$

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 the interval  $[M_{\min}, M_{\max}]$ . Furthermore, we choose  $M_{\min} = \max\left(\lceil \frac{N_l}{N_F} \rceil, 1\right)$  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  $\bar{m} = \left\lceil \frac{N_l}{\mathbb{E}_{\hat{\boldsymbol{\theta}}_\omega^{(\hat{d})}}(S_{N_F})} \right\rceil$

and  $M_{\max} = \bar{m} + \left\lceil 4\sqrt{\bar{m}\text{Var}_{\hat{\theta}_{\omega}^{(\hat{d})}}(S_{N_F})} \right\rceil$  and consider the range  $[M_{\min}, M_{\max}]$  suitable for inference. Here,  $\mathbb{E}_{\hat{\theta}_{\omega}^{(\hat{d})}}(S_{N_F})$  and  $\text{Var}_{\hat{\theta}_{\omega}^{(\hat{d})}}(S_{N_F})$  can be computed using (3.5) and (3.6).

For a given prior distribution  $\pi_M$ , Algorithm 5 computes  $p_{\hat{\theta}_{\omega}^{(\hat{d})},m}(M = m|N_l)$  using this described method.

---

**Algorithm 5** Compute posterior distribution  $p_{\hat{\theta}_{\omega}^{(\hat{d})},m}(M = m|N_l)$

---

**function** COMPUTE\_POSTERIOR( $\mathcal{D}_{tr}, N_l, \Delta, N_F, \pi_M$ )

Use  $\mathcal{D}_{tr}$  to obtain  $\hat{\theta}_{\omega}^{(\hat{d})}$

$\mathbf{p} \leftarrow \text{PMF\_S}(\hat{\theta}_{\omega}^{(\hat{d})}, \Delta, N_F)$

▷ From Algorithm 3

$e' \leftarrow \mathbb{E}_{\hat{\theta}_{\omega}^{(\hat{d})}}(S_{N_F})$

▷ Using (3.5)

$v' \leftarrow \text{Var}_{\hat{\theta}_{\omega}^{(\hat{d})}}(S_{N_F})$

▷ Using (3.6)

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

$\bar{m} \leftarrow \left\lceil \frac{N_l}{e'} \right\rceil$

$M_{\max} \leftarrow \bar{m} + \left\lceil 4\sqrt{\bar{m}v'} \right\rceil$

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

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

$\mathbf{p}_2 \leftarrow \text{PMF\_NL}(\mathbf{p}, i)$

▷ From Algorithm 4

$\mathbf{p}^*[i] \leftarrow \mathbf{p}_2[N_l + 1]\pi_M(i)$

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

▷ Normalise probabilities

**return**  $\mathbf{p}^*$

▷  $\mathbf{p}^*[m] = \mathbb{P}_{\hat{\theta}_{\omega}^{(\hat{d})},m}(M = m|N_l)$

---

Subsequently, the estimate  $\hat{M}$  of the number of molecules is found by locating the mode of the posterior  $p_{\hat{\theta}_{\omega}^{(\hat{d})},m}(M = m|N_l)$ , also known as the *maximum a posteriori* (MAP) estimate.

Under this inference mechanism, a 95% credible interval or *highest density region* (HDR) (Hyndman, 1996) 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 analysing the potential number of molecules that are truly imaged. Specifically, this region is chosen to be  $I = \left\{ m : p_{\hat{\theta}_{\omega}^{(\hat{d})},m}(m|N_l) \geq k_{0.05} \right\}$ , where

Parameter	$d$	$\lambda_{00_1}$ (s <sup>-1</sup> )	$\lambda_{0_1}$ (s <sup>-1</sup> )	$\lambda_{0_1 0_2}$ (s <sup>-1</sup> )	$\lambda_{0_1 1}$ (s <sup>-1</sup> )	$\lambda_{0_2 1}$ (s <sup>-1</sup> )	$\lambda_{10}$ (s <sup>-1</sup> )	$\mu_1$ (s <sup>-1</sup> )	$\Delta^{-1}$ (s <sup>-1</sup> )	$\delta$ (s)	$\omega$	$\nu_0$	$\nu_1$	$M$	$N_F$
Study															
1 (SLOW)	0		0.3162				1	0.0333	30	0.0033	10 <sup>-5</sup>	0.2	0.8	100	10 <sup>4</sup>
2 (MEDIUM)	0		1				3.162	0.1054	30	0.0033	10 <sup>-5</sup>	0.2	0.8	100	10 <sup>4</sup>
3 (FAST)	0		3.162				10	0.333	30	0.0033	10 <sup>-5</sup>	0.2	0.8	100	10 <sup>4</sup>
4 (SLOW)	1	0.15	0.3		0.1		0.8	0.01	30	0.0033	10 <sup>-5</sup>	0.2	0.8	100	10 <sup>4</sup>
5 (MEDIUM)	1	0.35	1		0.3		2.3	0.1	30	0.0033	10 <sup>-5</sup>	0.2	0.8	100	10 <sup>4</sup>
6 (FAST)	1	2	10		0.7		10	0.333	30	0.0033	10 <sup>-5</sup>	0.2	0.8	100	10 <sup>4</sup>
7 (SLOW)	2	0.15	0.3	0.05	0.1	0.001	0.8	0.05	30	0.0033	10 <sup>-5</sup>	0.2	0.8	100	10 <sup>4</sup>
8 (MEDIUM)	2	0.8	4	0.1	0.4	0.005	8	0.1	30	0.0033	10 <sup>-5</sup>	0.2	0.8	100	10 <sup>4</sup>
9 (FAST)	2	2	10	0.2	0.7	0.01	10	0.333	30	0.0033	10 <sup>-5</sup>	0.2	0.8	100	10 <sup>4</sup>

**Table 3.1:** Global parameter values for the stimulation studies conducted in this section.

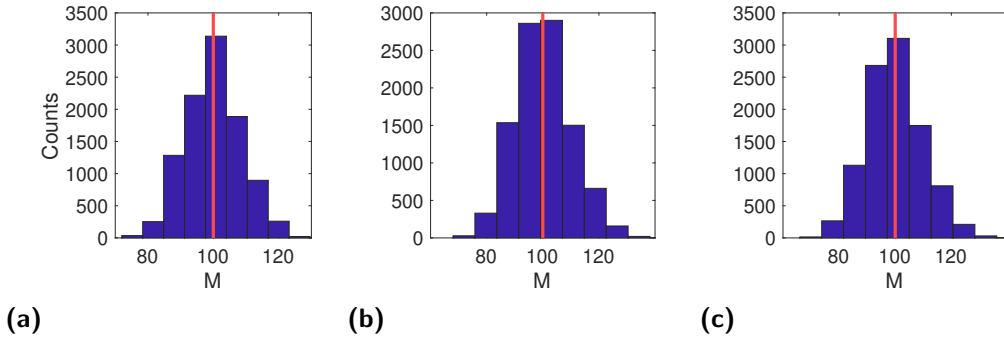
$k_{0.05}$  is the largest value such that

$$p_I := \sum_{m \in I} p_{\hat{\theta}_{\omega}^{(d)}, m}(M = m | N_I) \geq 0.95.$$

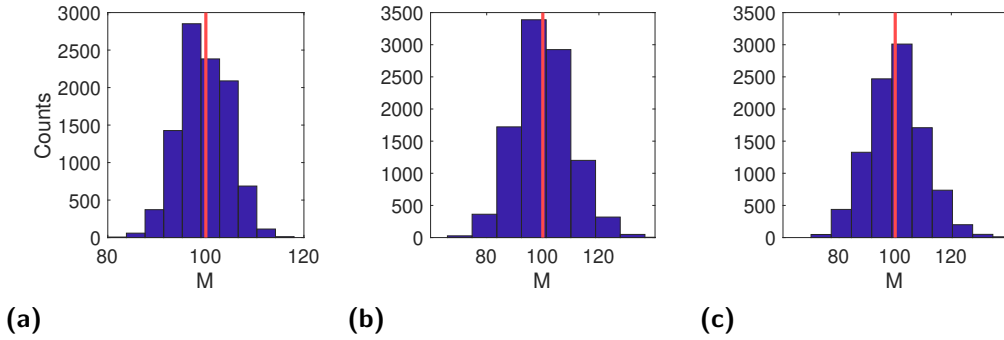
### 3.4 Simulations

In this section, we provide posterior estimates of  $M$  from nine simulation studies highlighting slow, medium and fast switching scenarios under the three photo-switching models  $\mathcal{M}_{\{1\}}^d$  with  $d$ , the number of multiple dark states, equalling 0, 1 and 2.

For each simulation study,  $10^4$  independent datasets, each containing 350 molecules were simulated. From this, the localisations from 250 molecules were used to estimate  $\theta_{\omega}^{(d)}$ . The number of localisations from the remaining 100 molecules were used to estimate  $M$  through the posterior mode of (3.10), under a uniform prior,  $\pi_M(m) \propto 1$ . The true parameter values for each study can be found in Table 3.1, where  $T^* = 0$  in all simulations. Figures 3.2a - 3.4c 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.



**Figure 3.2:** Simulation results from studies 1-3 in Table 3.1. Histograms represent counts of  $\hat{M}$  under the slow (Figure 3.2a), medium (Figure 3.2b) and fast (Figure 3.2c) scenarios when  $d = 0$ , from  $10^4$  independently generated datasets with  $M = 100$  and  $N_F = 10^4$ . For each estimate,  $\theta_\omega^{(0)}$  was determined using a training data set with  $M = 250$  and  $N_F = 10^4$ .

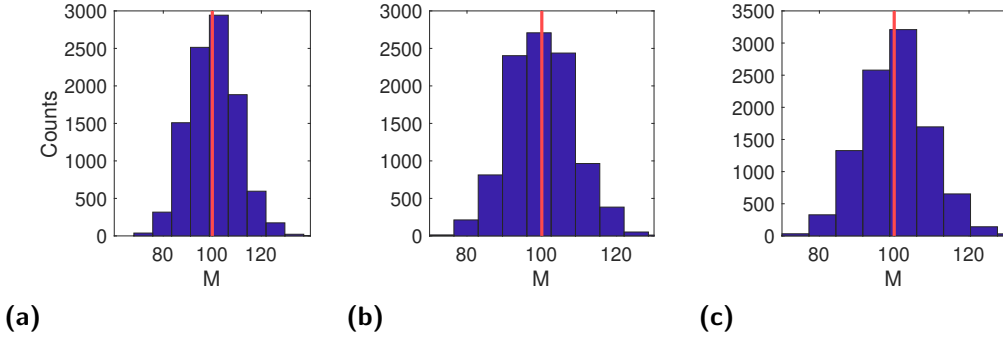


**Figure 3.3:** Simulation results from studies 4-6 in Table 3.1. Histograms represent counts of  $\hat{M}$  under the slow (Figure 3.3a), medium (Figure 3.3b) and fast (Figure 3.3c) switching scenarios when  $d = 1$ , from  $10^4$  independently generated datasets with  $M = 100$  and  $N_F = 10^4$ . For each estimate,  $\theta_\omega^{(1)}$  was determined using a training data set with  $M = 250$  and  $N_F = 10^4$ .

### 3.5 Application to Alexa Fluor 647 data

In this section, we use the Alexa Fluor 647 dataset as was previously analysed in Chapter 1, for the purpose of validating the theory and counting method presented in this chapter. For each of the experiments in this dataset, we apply our method to obtain an estimate for  $M$  and present the 95% credible intervals from the resulting posterior distributions.

Using this dataset, consisting of 27 independent experiments, the photo-



**Figure 3.4:** Simulation results from studies 1-3 in Table 3.1. Histograms represent counts of  $\hat{M}$  under the slow (Figure 3.4a), medium (Figure 3.4b) and fast (Figure 3.4c) switching scenarios when  $d = 2$ , from  $10^4$  independently generated datasets with  $M = 100$  and  $N_F = 10^4$ . For each estimate,  $\theta_\omega^{(2)}$  was determined using a training data set with  $M = 250$  and  $N_F = 10^4$ .

emission time trace of each photo-switchable molecule detected was extracted<sup>†</sup>. Therefore, in each experiment, the number of fluorophores present was known and therefore acts as a ground truth against which our estimate  $\hat{M}$  can be compared. For each dataset (labelled 1 - 27), each photo-switchable molecule detected had its discrete observation trace  $\{Y_n\}$  extracted from its photo-emission time trace. 70% of these traces were then used as the *training set*  $M_{tr}$ , used to identify model parameters  $\hat{\theta}_\omega^{(2)}$  via the PSHMM method described in Chapter 1. The remaining 30% (the test set) was then used to validate the inference method outlined in this chapter. Here,  $M$  (known) is the 30% of molecules that remain in the test set, and  $N_l$  is the (total) number of localisations recorded from these  $M$  molecules. Additionally, we fitted the photo-switching parameters using the  $d = 2$  photo-kinetic model  $\mathcal{M}_{\{1\}}^2$ , as was predicted by both the PSHMM in Chapter 1, and the method detailed in Lin et al. (2015) for this dataset.

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 3.5. With this are shown two examples of the posterior distribution of  $M$  given  $N_l$  (using (3.10)). The remaining figures can be found in Figure 3.6 of Section 3.B in Appendix 3. The values of the laser intensity, frame rate  $\Delta^{-1}$ , number of molecules in each dataset ( $M_{tr}, M$ ), the number of frames

<sup>†</sup>See Section 1.F.2 of Appendix 1 for more details.

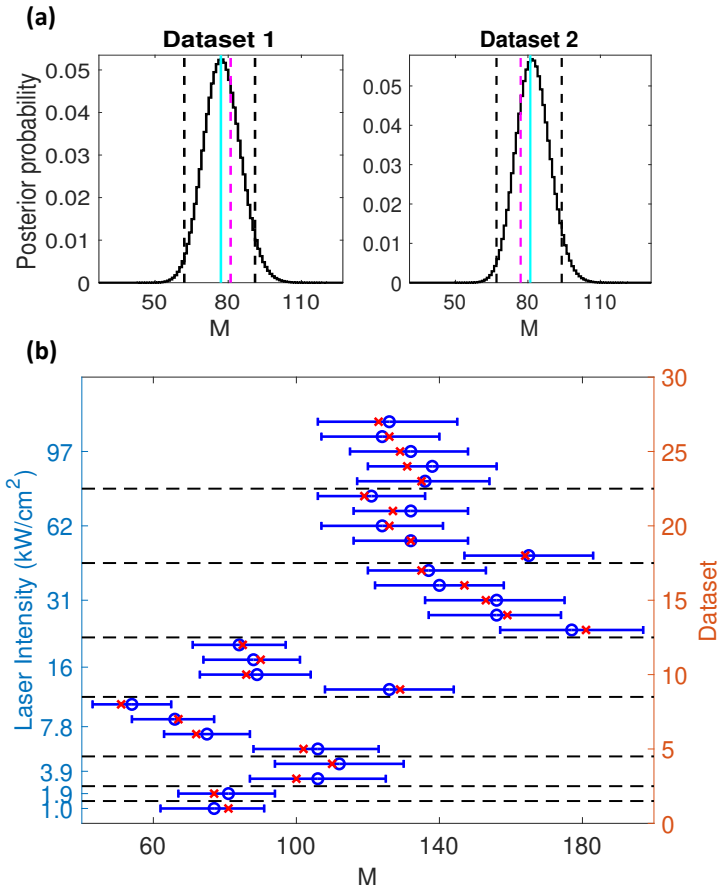
Dataset	Laser intensity	$\Delta^{-1}$	$M_{tr}$	$M$	$N_F$	$N_l$	$\hat{M}$	$I$	$p_I$
1	1.0	200	192	81	49796	4340	77	[62, 91]	0.951
2	1.9	200	180	77	49533	5300	81	[67, 94]	0.950
3	3.9	200	234	100	49815	2443	106	[87, 125]	0.955
4	3.9	200	295	110	39758	2834	112	[94, 130]	0.956
5	7.8	200	238	102	39721	2679	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 3.2:** A description of the Alexa Fluor 647 datasets, with reference to the laser intensities in  $\text{kW}/\text{cm}^2$  and frames  $\Delta^{-1}$  sampled per second used to characterise each of the 27 experiments. For each dataset, a training set of size  $M_{tr}$  was used to find the maximum likelihood estimate  $\hat{\theta}_\omega^{(2)}$ . A hold out test set of size  $M$  was used to validate the inference procedure.

over which they were imaged ( $N_F$ ), the total number of localisations ( $N_l$ ), the posterior mode  $\hat{M}$ , its 95% credible interval ( $I$ ) and its corresponding value  $p_I$  are summarised in Table 3.2. The maximum likelihood estimates  $\hat{\theta}_\omega^{(d)}$  used for each study is presented in Table 3.3 of Section 3.B of Appendix 3.

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.





**Figure 3.5:** (a) Posterior distributions of  $M$  given  $\hat{\theta}_\omega^{(2)}$  and  $N_l$  for the Alexa Fluor 647 datasets 1 and 2 (descriptions of which can be found in Table 3.3). For each study,  $\hat{M}$  is given by the corresponding posterior mode plotted in cyan, with the true values of  $M$  shown in magenta (dotted). 95% credible intervals for each  $\hat{M}$  are shown in black (dotted). (b)

Posterior estimates of  $M$  given  $\hat{\theta}_\omega^{(2)}$  and  $N_l$  for the 27 Alexa Fluor 647 datasets (descriptions of which can be found in Table 3.2) with varying laser intensities (kW/cm<sup>2</sup>). For each study,  $\hat{M}$  is given by the corresponding posterior mode plotted in blue (circle), with the true values of  $M$  shown in red (crosses) and 95% credible intervals for each  $\hat{M}$  are shown by blue error bars.

### 3.6 Conclusions

In this chapter, for an arbitrary number of fluorophores in a (d)STORM experiment, we have derived the distribution of the number of localisations per fluorophore and also provided expressions for its first two moments. This has allowed us to present an inference procedure for estimating the unknown

number of molecules, given an observed number of localisations. 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.

The method that we have constructed in this chapter is an example of modular inference, which separates the rate estimation (training) procedure from the counting procedure. While the training procedure requires a separate experiment to estimate fluorophore switching rates, it does mean that the counting process is computationally cheap and therefore highly scalable. In particular, this method can count several thousand molecules from tens of thousands of localisations with relative computational ease.

## Appendix 3

### 3.A Proof of Lemma 5

In this section, we provide a proof for Lemma 5.

*Proof.* In the following, we utilize the following expansion

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

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

From the definition of a derivative, we have

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

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

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

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

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

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

This gives

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

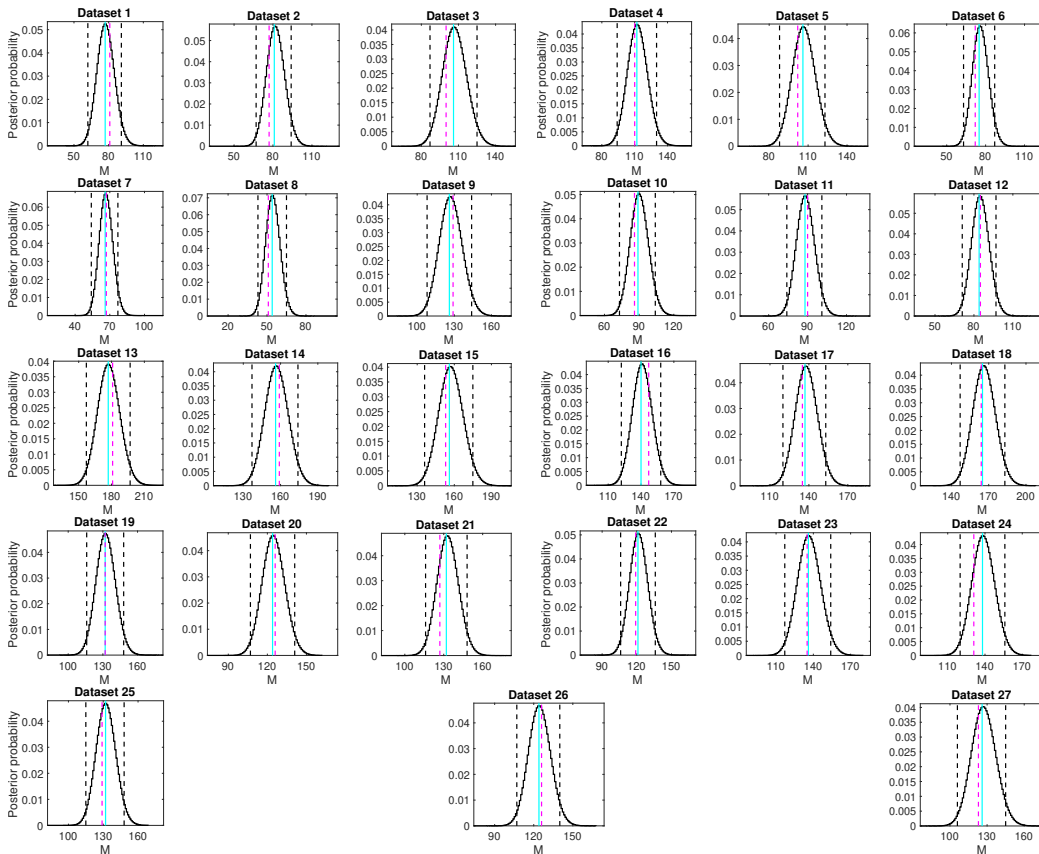
so that  $\mathbb{E}_{\boldsymbol{\theta}_\omega^{(d)}}(S_n^2) = G''_{S_n}(1) + \mathbb{E}_{\boldsymbol{\theta}_\omega^{(d)}}(S_n)$  and therefore  $\text{Var}_{\boldsymbol{\theta}_\omega^{(d)}}(S_n) = G''_{S_n}(1) + \mathbb{E}_{\boldsymbol{\theta}_\omega^{(d)}}(S_n) - \mathbb{E}_{\boldsymbol{\theta}_\omega^{(d)}}^2(S_n)$ .  $\square$

### 3.B Alexa Fluor 647 results

In this section, we provide the posterior distributions of  $M$  given  $N_l$  from the 27 Alexa Fluor 647 experiments studied. We additionally provide a table to detail the imaging parameters  $\hat{\theta}_\omega^{(d)}$  used when deriving the posterior distribution of  $M$  given  $\hat{\theta}_\omega^{(2)}$  for these experiments.

Figure 3.6 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.

Table 3.3 provides the number of each study, the laser intensity used,  $\Delta$ ,  $M$



**Figure 3.6:** Posterior distributions of  $M$  given  $\hat{\theta}_\omega^{(2)}$  and  $N_l$  for the 27 Alexa Fluor 647 datasets (descriptions of which can be found in Table 3.3). For each study,  $\hat{M}$  is given by the corresponding posterior mode plotted in cyan, with the true values of  $M$  shown in magenta (dotted). 95% credible intervals for each  $\hat{M}$  are shown in black (dotted).

(used for training and testing),  $N_F$  and the maximum likelihood parameter estimates in  $\hat{\theta}_\omega^{(2)}$ .

Dataset	Laser intensity	$\Delta^{-1}$	$M_{tr}$	$M$	$N_F$	$\hat{\lambda}_{00_1}$ $\times$ $\Delta$	$\hat{\lambda}_{0_1}$ $\times$ $\Delta$	$\hat{\lambda}_{0_1 0_2}$ $\times$ $\Delta$	$\hat{\lambda}_{0_1 1}$ $\times 10$ $\times \Delta$	$\hat{\lambda}_{0_2 1}$ $\times 10^4$ $\times \Delta$	$\hat{\lambda}_{10}$ $\times$ $\Delta$	$\hat{\mu}_1$ $\times 10^2$ $\times \Delta$	$\frac{\hat{\delta}}{\Delta}$	$\hat{\omega}$ $\times 10^5$	$\hat{\nu}_X$
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 3.3:** A description of the Alexa Fluor 647 datasets, with reference to the laser intensities in  $\text{kW}/\text{cm}^2$  and frames  $\Delta^{-1}$  sampled per second used to characterise 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  $\hat{\theta}_\omega^{(2)}$  via the PSHMM (estimated values shown). A hold out test set of size  $N_F \times M$  (test) was used in the posterior computations of  $M$ .

# 4

## SPATIO-TEMPORAL MODELLING

### 4.1 Introduction

Under special experimental conditions, molecules can be placed on a coverslip with sufficient spatial separation to neighbouring fluorophores as to enable observations per molecule to be extracted without difficulty in each frame (Lin et al., 2015). In this situation, their true spatial locations can be determined as the centroid of the observed localisations and their photo-switching rates can be inferred using, for example, the PSHMM estimator presented in Chapter 1.

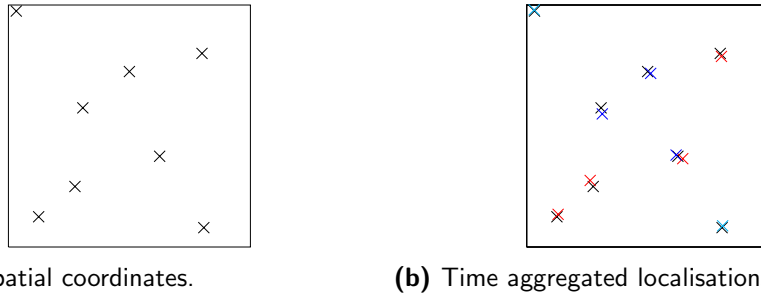
In most experiments, however, an *unknown* number of fluorophores is used in attaining the resulting super-resolution image of a structure of interest. Specifically, when a dense set of fluorophores (of unknown cardinality) is filmed over a number of frames, the photo-switching property enables accurate localisations to be made when molecules are detectable. However, their resulting non-stationary photo-switching behaviours invoke non constant detection rates per frame, which in turn generate a random number of *offspring* coordinates per molecule during an experiment. Such measurements are typically collected by first using a segmentation algorithm (Olivo-Marin, 2002,

Henriques et al., 2010, Ovesný et al., 2014) to locate high photon intensity regions of each image, and then by fitting point spread functions (Ober et al., 2015, Sage et al., 2015) to determine the precise coordinates of each localisation expected to be observed in the image. Due to the manner in which measurements are collected, the localisations will therefore differ between frames, as the same molecules are localised with measurement errors about their true spatial locations. Figures 4.1a and 4.1b provide a visualisation as to how these fluorophores are localised in space-time.

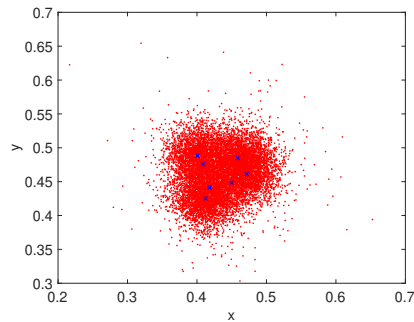
When multiple localisations are made, the *superimposed* image (see Figure 4.1b) does not in general allow for molecules to be individually resolved. These images are formed by the temporal aggregation of obtained (localisation) measurements that may invoke clusters of observations to develop in any one region of an image. Figure 4.2 shows an example of all localisations collected from seven spatially close molecules during an experiment. Without knowledge of the exact number of molecules present, it is observed here that identifying the true positions of the molecules (indicated in blue) may be challenging. These situations are in fact not uncommon in experimental set-ups and not only prove challenging in estimating the number of molecules present with their true spatial locations, but also render rate estimation procedures developed for this problem difficult to use. From just utilising the offspring measurements collected during an experiment, the inference problem therefore lies in identifying the *number* of fluorophores, their true spatial positions and the photo-switching rates which have generated the observations.

Spatial point processes or *random finite sets* are random variables that take values as sets (of points) with a random cardinality, and are therefore naturally suited to the spatial modelling of molecules that are imaged in fluorescence microscopy. The idea of modelling the clusters obtained in the superimposed image (Hsu and Baumgart, 2011, Wiemhöfer et al., 2012) or the spatial distributions of molecules themselves (Owen et al., 2010) via spatial point patterns is familiar. Modelling clusters of specific shapes and sampling locations of their centres through, for example, both maximum likelihood based and MCMC methods (Geyer and Møller, 1994, Møller and Waagepetersen,





**Figure 4.1:** Figure 4.1a shows the true spatial locations of a set of fluorophores from one fluorescence microscopy experiment. Figure 4.1b shows the aggregated measurements of localised fluorophores detected from the first 3 frames; objects detected from frame 1 (blue), frame 2 (red) and frame 3 (cyan) are shown with respect to the true locations (black).



**Figure 4.2:** Superimposed image of the localisations (red) attained from seven (spatially close) fluorescing molecules with true positions indicated in blue (crosses). Measurement unit is nanometre (nm)  $\times 10^4$ .

2003) are well studied in the literature. For example, in the case of Neymann-Scott point processes (that produce Gaussian distributed clusters), Mrkvička et al. (2014) develop a targeted MCMC algorithm able to jointly sample the cluster centres (with its unknown cardinality) and other parameters of the model. While this method could be applied to the superimposed image, ignoring the time domain is likely to result in poor estimation of the underlying photo-switching parameters. Furthermore, using the superimposed image will in general lead to spatial biases due to false positive localisations and cases where molecules are spatially close.

On the other hand, Mahler (2007b) introduced a Bayesian filtering approach aimed at sequentially tracking the spatial locations of *dynamic* objects or *targets* modelled as hidden finite point processes, that are observed in highly

noisy environments over time. However, the intractability of the derived Bayes filter for the tracking of a large number of targets has subsequently motivated numerous approximations to this original approach. Important developments include but are not limited to: the Probability Hypothesis Density (PHD) filter and its cardinalised version (Mahler, 2003, Vo and Ma, 2005, Vo et al., 2006), which propagates the first moment measure of the hidden point process to track both the positions and unknown number of targets, and their implementations through sequential Monte Carlo (SMC) type algorithms (Mahler, 2007a, Vo et al., 2003, 2005, Whiteley et al., 2010). All of these methods allow for the spatial tracking of hundreds of objects across time and have been used in many radar/sonar tracking, navigation and computer vision applications (Bar-Shalom et al., 2001, Maggio et al., 2008). Furthermore, estimation of the static parameters of the model in these approaches has been made possible through, for example, sequential Monte Carlo approximations of the marginal likelihood (Whiteley et al., 2009), an online Expectation-Maximisation (EM) algorithm (Yildirim et al., 2015), or a reversible-jump MCMC sampler (Jiang et al., 2015). For instance, the MCMC approach mentioned here is designed to jointly sample target numbers, locations and the static parameters of the model, and do so by designing a likelihood function which incorporates a data association variable between measurements and objects. Common to all these methods, however, is the assumption that targets are dynamic across time, an assumption that is violated in the case of fluorescence imaging of stationary fluorophores. While methods for estimating the number of molecules and their spatial locations could still be applied using the above methods, one would need to take extra precautions during implementation. For example, in the case of the sequential Monte Carlo filter of Vo et al. (2003), the static nature of the molecules is likely to lead to particle degeneracy, therefore requiring additional procedures to ensure the validity of the resulting estimates. Nevertheless, the methods described here are attractive due to their temporal incorporation and we thus look to formulate a model, specific to this application, which rests on similar principles. To the best of our knowledge, inference of detected points which so heavily rely on the temporal nature at which they occur has not been studied for this application.

### 4.1.1 Chapter summary

Motivated by the work of Mahler (2007b), we will in this chapter formally derive a suitable spatio-temporal model for fluorophores imaged in fluorescence microscopy, that is based upon point process theory. We therefore begin in Section 4.2 by mathematically defining the notion of (simple) finite point processes or *random finite sets* and define their probability density function with respect to a commonly used reference measure. In Section 4.3, we then formally characterise the time varying model that unifies the true spatial positions of fluorophores and the offspring observations that are made at each time frame. Here, we will show that a single photo-switching fluorophore can be regarded as an element of a pure birth spatial parent point process of molecular positions. Through this setting, its probability of detection and birth in a given frame can be derived, and are observed to be deterministic functions of the unknown photo-switching parameter  $\boldsymbol{\theta}^{(d)}$ . Furthermore, some asymptotic results regarding detections and births are also given. By placing a suitable prior distribution on the number of total fluorophores imaged in an experiment, a prior distribution of true molecular positions is then retrieved. From this, the full Bayes filter appropriate to this model is subsequently derived. In Section 4.4, we discuss how an MCMC algorithm, which we adapt to this problem, can be used to infer the parent point process and its associated photo-switching parameter vector  $\boldsymbol{\theta}^{(d)}$ . We will then define the Deviance Information Criterion (DIC) and illustrate how, using the outputs of MCMC simulations, it can be used as a model selection tool to pick the most likely number of molecules  $M$  and multiple dark states  $d$ . In Section 4.5, three simulation studies will be provided to demonstrate the validity of this method. Finally, in Section 4.6, we will apply this MCMC approach on a real dataset consisting of Alexa Fluor 488 molecules.

## 4.2 Random finite sets

In this section, we will begin by giving a mathematical definition of a random finite set (RFS) and describe the notion of a *probability density* of an

RFS with respect to a non-Lebesgue reference measure. Through an example which gives the probability density of an independent and identically distributed cluster process, we will study the well known spatial Poisson process. We then briefly describe a convolution theorem for random sets which is shown to be useful in analysing the superposition property of independent Poisson processes.

A *random finite set* (RFS) or a *simple-finite point process*  $C$  on  $\mathcal{C} \subset \mathbb{R}^n$  is a random variable that takes values in the space of finite subsets of  $\mathcal{C}$ , denoted  $\mathcal{F}(\mathcal{C})$ . In each of these subsets, the number of points in the set is random and the points themselves are distinct\*, random and unordered. An RFS  $C$  is therefore specified by a discrete random variable  $N_C$  on  $\mathbb{Z}_{\geq 0}$  with probability mass function  $p_{N_C}(\cdot)$  which determines the number of points in  $C$ , and a probability distribution  $p_{C|n_C}(\cdot)$  on the product space  $\mathcal{C}^{n_C} = \mathcal{C} \times \cdots \times \mathcal{C}$ , for  $n_C \in \mathbb{Z}_{>0}$ , which determines the joint spatial distribution of the points given that there are  $N_C = n_C$  points in the set  $C$  (Daley and Vere-Jones, 1998, Stoyan et al., 1987, Mahler, 2007b).

**Definition 6.** *A random finite set (RFS) or a simple-finite point process  $C$  on  $\mathcal{C} \subset \mathbb{R}^n$  is a measurable mapping*

$$C : \Omega \rightarrow \mathcal{F}(\mathcal{C}),$$

*from a sample space  $\Omega$  to the space of finite subsets  $\mathcal{F}(\mathcal{C})$ .*

Formally, an RFS as defined in Definition 6 is a measurable mapping from a sample space  $\Omega$  to  $\mathcal{F}(\mathcal{C})$ . Here,  $\Omega$  is equipped with a probability measure  $\mathbb{P}$  defined on a  $\sigma$ -algebra of events  $\sigma(\Omega)$ . Intuitively, this means that any probability distribution  $P$  of the RFS  $C$  can be written as the probability measure of its pre-image in the sample space  $\Omega$ , i.e that  $P = \mathbb{P}(C^{-1})$ . A more formal definition is stated in Definition 7.

**Definition 7.** *Let  $C$  be an RFS on  $\mathcal{C} \subset \mathbb{R}^n$ . The probability distribution of  $C$  on  $\mathcal{C}$  is the probability measure  $P$  on  $\mathcal{F}(\mathcal{C})$ , the space of finite subsets of*

---

\*This is equivalent to a finite point process  $C$  being simple.

$\mathcal{C}$ , defined as

$$P(\mathcal{B}) = \mathbb{P}(\{C \in \mathcal{B}\}) := \mathbb{P}(\{\omega \in \Omega | C(\omega) \in \mathcal{B}\}),$$

for any Borel measurable subset  $\mathcal{B} \subseteq \mathcal{F}(\mathcal{C})$ .

For a more comprehensive overview on point processes, including a detailed introduction to its measure theoretic formulation, the interested reader is directed to [Daley and Vere-Jones \(1998\)](#), [Stoyan et al. \(1987\)](#).

#### 4.2.1 Probability density function

Defining the probability density of an RFS is highly non-trivial since  $\mathcal{F}(\mathcal{C})$  does not inherit the standard (Euclidean) notion of a density defined on  $\mathbb{R}^n$ . However, by utilising a suitable choice of *reference* measure, we can firstly define the notion of a *set integral* and finally the *probability density function* of an RFS, which can be shown to be mathematically consistent with that of a usual probability density.

A measure  $\bar{\mu}$  on  $\mathcal{C}$  is defined as the relative *size* of subsets of  $\mathcal{C}$  and can be described as a countably additive function  $\bar{\mu} : A \rightarrow [0, \infty]$ , for any  $A \subseteq \mathcal{C}$ . In RFS and point process theory, the conventional choice of reference measure ([Vo et al., 2005](#), [Mahler, 2007b](#)) is given by

$$\bar{\mu}(\mathcal{T}) = \sum_{i=0}^{\infty} \frac{\lambda_L^i(\chi^{-1}(\mathcal{T}) \cap \mathcal{C}^i)}{i!}, \quad (4.1)$$

defined for any subset  $\mathcal{T} \subseteq \mathcal{F}(\mathcal{C})$ . Here,  $\mathcal{C}^i$  is the  $i$ th Cartesian product of  $\mathcal{C}$  such that  $\mathcal{C}^0 := \{\emptyset\}$ ,  $\lambda_L^i$  is the  $i$ th product unitless Lebesgue measure on  $\mathcal{C}^i$  and  $\chi : \uplus_{i=0}^{\infty} \mathcal{C}^i \rightarrow \mathcal{F}(\mathcal{C})$  is a mapping from vectors to sets such that  $\chi([\mathbf{c}_1, \dots, \mathbf{c}_i]^\top) = \{\mathbf{c}_j : j = 1, \dots, i\}$ , defined for any (distinct)  $\mathbf{c}_1, \dots, \mathbf{c}_i \in \mathcal{C}$ .

This choice of reference measure is analogous to the Lebesgue measure on  $\mathcal{C}$  and therefore allows the definition of a set integral of a function  $f : \mathcal{F}(\mathcal{C}) \rightarrow \mathbb{R}$ , as is provided in Definition 8.

**Definition 8** (Mahler (2007b, 2003)). *The integral of a non-negative function  $f : \mathcal{F}(\mathcal{C}) \rightarrow [0, \infty)$  over all  $\mathcal{T} \subseteq \mathcal{F}(\mathcal{C})$  with respect to the reference measure  $\bar{\mu}$  defined in (4.1) is given by*

$$\int_{\mathcal{T}} f(C) \bar{\mu}(dC) = \sum_{i=0}^{\infty} \frac{1}{i!} \int_{\chi^{-1}(\mathcal{T}) \cap \mathcal{C}^i} f(\{\mathbf{c}_1, \dots, \mathbf{c}_i\}) \lambda_L^i(d\mathbf{c}_1 \dots d\mathbf{c}_i).$$

*It now follows that the set integral of  $f$  over a closed subset  $A \subseteq \mathcal{C}$  is*

$$\int_A f(C) \delta C = \sum_{i=0}^{\infty} \frac{1}{i!} \int_{A^i} f(\{\mathbf{c}_1, \dots, \mathbf{c}_i\}) \lambda_L^i(d\mathbf{c}_1 \dots d\mathbf{c}_i),$$

*noting that the two integrals are related via*

$$\int_A f(C) \delta C = \int_{\mathcal{U}} f(C) \bar{\mu}(dC),$$

*with  $\mathcal{U} = \cup_{i=0}^{\infty} A^i$ .*

By defining the notion of a set integral with respect to the reference measure  $\bar{\mu}$ , we are able to now define the probability density function  $f_C(C)$  of an RFS  $C$  on  $\mathcal{C}$ . Technically,  $f_C(C)$  (if it exists) is defined as the *Radon-Nikodým derivative*<sup>†</sup> of the probability distribution  $P$  with respect to the reference measure  $\bar{\mu}$  and can be written as

$$f_C(C) = \frac{dP}{d\bar{\mu}} \iff P(\mathcal{T}) = \mathbb{P}(\{C \in \mathcal{T}\}) = \int_{\mathcal{T}} f_C(C') \bar{\mu}(dC'),$$

for any  $\mathcal{T} \subseteq \mathcal{F}(\mathcal{C})$ . To ensure the existence of  $f_C(C)$ , it is implicitly assumed here that  $P(\cdot)$  is absolutely continuous with respect to  $\bar{\mu}$ .

**Example 3** (Independently, identically distributed cluster processes (Mahler, 2007b)). *Let  $N_C$  be a discrete random variable on  $\mathbb{Z}_{\geq 0}$  with distribution  $p_{N_C}(n) := \mathbb{P}(N_C = n)$  and let  $f(\mathbf{c})$  be a probability density function on some*

<sup>†</sup>The Radon-Nikodým Theorem (Mahler, 2007b, Vo et al., 2005) states that given two  $\sigma$ -finite measures  $\bar{\mu}_1$  and  $\bar{\mu}_2$  on  $\mathcal{F}(\mathcal{C})$  whereby  $\bar{\mu}_2$  is absolutely continuous with respect to  $\bar{\mu}_1$ , there exists a function  $g : \mathcal{F}(\mathcal{C}) \rightarrow [0, \infty)$  such that for all  $\mathcal{T} \subseteq \mathcal{F}(\mathcal{C})$ ,  $\bar{\mu}_1(\mathcal{T}) = 0 \implies \bar{\mu}_2(\mathcal{T}) = 0$  if and only if  $\bar{\mu}_2(\mathcal{T}) = \int_{\mathcal{T}} g(C') \bar{\mu}_1(dC')$ . Here  $g = \frac{d\bar{\mu}_2}{d\bar{\mu}_1}$  is known as the Radon-Nikodým derivative or the density of  $\bar{\mu}_2$  with respect to  $\bar{\mu}_1$ .

bounded set  $\mathcal{C}$ . For any  $C = \{\mathbf{c}_1, \dots, \mathbf{c}_n\}$ , with  $|C| = n$ , the density function of  $C$  is defined to be

$$f_C(C) := n! p_{N_C}(n) \prod_{\mathbf{c} \in C} f(\mathbf{c}), \quad (4.2)$$

where it is easily seen that the set integral over  $\mathcal{C}$  is

$$\begin{aligned} \int_{\mathcal{C}} f_C(C) \delta C &= p_{N_C}(0) + \sum_{n=1}^{\infty} p_{N_C}(n) \int_{\mathcal{C}^n} f(\mathbf{c}_1) \dots f(\mathbf{c}_n) \, d\mathbf{c}_1 \dots d\mathbf{c}_n \\ &= \sum_{n=0}^{\infty} p_{N_C}(n) \\ &= 1. \end{aligned}$$

**Remark 8.** For some  $\gamma > 0$ , letting  $N_C \sim \text{Poi}(\gamma)$  enables  $p_{N_C}(n) = \frac{e^{-\gamma} \gamma^n}{n!}$ . In this case

$$f_C(C) = e^{-\gamma} \prod_{\mathbf{c} \in C} \underbrace{\gamma f(\mathbf{c})}_{v(\mathbf{c})}, \quad (4.3)$$

for some spatial density  $f(\mathbf{c})$  on  $\mathcal{C}$ . When  $f(\mathbf{c}) \propto 1$  (uniform),  $C$  is known as the (homogeneous) Poisson RFS. Moreover, the function  $v(\mathbf{c}) = \gamma f(\mathbf{c})$  is known as the **intensity function** of  $C$  and is such that

$$\mathbb{E}(|C \cap A|) = \int_A v(\mathbf{c}) \, d\mathbf{c},$$

gives the expected number of points in the RFS  $C$  that are in the region  $A \subseteq \mathcal{C}$ .

Taken directly from [Mahler \(2007b\)](#), Example 3 defines the probability density function of an independent and identically distributed (iid) cluster process  $C$  on  $\mathcal{C}$ , which (analogously to densities defined on the Euclidean space) integrates to unity over its domain. This kind of process is uniquely characterised by a random variable  $N_C$  which describes the cardinality of its points, and  $f(\mathbf{c})$  (a probability density on  $\mathcal{C}$ ) which defines their spatial distributions. Importantly, we see from Remark 8 that the well known Poisson RFS is an example of an iid cluster process with a Poisson cardinality distribution. Fur-

thermore, it is uniquely characterised by its intensity function  $v(\mathbf{c})$  (given in (4.3)), which describes the expected number of points per unit volume on  $\mathcal{C}$ .

An interesting property of the Poisson RFS, and one which we will utilise in the next section, is that the *superposition* of two independent Poisson processes defined on the same space  $\mathcal{C}$  is also a Poisson RFS<sup>‡</sup>. This is shown in Example 4, and relies on Theorem 6, the Fundamental Theorem of Convolution (Mahler, 2007b) for random finite sets. This theorem is useful in providing an explicit expression for the probability density of an RFS which is formed by superposition.

**Theorem 6** (Fundamental Theorem of Convolution (Mahler, 2007b)). *Let  $C = \cup_{i=1}^n C_i$ , where  $C_1, \dots, C_n$  are statistically independent random sets on  $\mathcal{C}$ . The probability density of  $C$  is related to the probability densities of  $C_1, \dots, C_n$  via*

$$f_C(C) = \sum_{W_1 \uplus \dots \uplus W_n = C} f_{C_1}(W_1) \dots f_{C_n}(W_n),$$

where the summation is taken over all mutually disjoint subsets  $W_1, \dots, W_n$  of  $C$  such that  $\cup_{i=1}^n W_i = C$ .

**Example 4.** *Let  $C_1$  and  $C_2$  be two independent Poisson random finite sets on  $\mathcal{C}$  with intensity functions  $v_1(\mathbf{c}) = \gamma_1 f(\mathbf{c})$  and  $v_2(\mathbf{c}) = \gamma_2 f(\mathbf{c})$ . By the fundamental theorem of convolution, the probability distribution of  $C = C_1 \cup$*

---

<sup>‡</sup>This result can be inductively extended to hold for the superposition of a countable number of Poisson processes.



$C_2$  follows

$$\begin{aligned}
f_C(C) &= \sum_{W \subseteq C_1 \cup C_2} f_{C_1}(W) f_{C_2}(C_1 \cup C_2 \setminus W) \\
&= \sum_{W \subseteq C_1 \cup C_2} e^{-(\gamma_1 + \gamma_2)} \left[ \gamma_1^{|W|} \prod_{\mathbf{c} \in W} f(\mathbf{c}) \right] \left[ \gamma_2^{|C_1 \cup C_2 \setminus W|} \prod_{\mathbf{c} \in C_1 \cup C_2 \setminus W} f(\mathbf{c}) \right] \\
&= e^{-(\gamma_1 + \gamma_2)} \prod_{\mathbf{c} \in C_1 \cup C_2} f(\mathbf{c}) \sum_{W \subseteq C_1 \cup C_2} \gamma_1^{|W|} \gamma_2^{|C_1 \cup C_2 \setminus W|} \\
&= e^{-(\gamma_1 + \gamma_2)} \prod_{\mathbf{c} \in C_1 \cup C_2} f(\mathbf{c}) \sum_{k=0}^{|C_1 \cup C_2|} \binom{|C_1 \cup C_2|}{k} \gamma_1^k \gamma_2^{|C_1 \cup C_2| - k} \\
&= e^{-(\gamma_1 + \gamma_2)} (\gamma_1 + \gamma_2)^{|C_1 \cup C_2|} \prod_{\mathbf{c} \in C_1 \cup C_2} f(\mathbf{c}) \\
&= e^{-(\gamma_1 + \gamma_2)} \prod_{\mathbf{c} \in C} (\gamma_1 + \gamma_2) f(\mathbf{c}),
\end{aligned}$$

which is recognised as the probability density of a Poisson RFS with intensity function  $v(\mathbf{c}) = (\gamma_1 + \gamma_2)f(\mathbf{c})$ .

### 4.3 Spatio-temporal model of a fluorophore

In this section, we define a spatio-temporal model suitable for fluorophores that are imaged under general experimental conditions. Here, we formulate the true positions of molecules as elements of an unobserved *parent* RFS which produces *offspring* observation sets collected at each time frame. From this, we derive the time variant birth and detection probabilities of fluorophores that are needed to characterise the parent and offspring RFS densities, and highlight its asymptotic properties. These densities are subsequently used in the derivation of the full Bayes filter under this model.

In the most general case, an experiment can give rise to an unknown initial probability mass  $\nu_X^*$ . Due to the stochastic nature of  $\{X(t)\}$ , fluorophores that are imaged under such experimental conditions may therefore take a non-zero amount of time to first reach the On state 1, and that this time will vary between molecules. Although the set-up of  $\{Y_n\}$  implies that reaching

the On state is not sufficient for the detection of a molecule, once a molecule has reached this state, it is *activated* as part of the underlying molecular configuration. This procedure therefore induces a pure spatio-temporal (hidden) birth process of true molecular positions, as new fluorophores are detected over time.

Specifically, suppose at time  $n$  there are  $K(n) := |C_n|$  fluorophores that have *already* reached their photon-emission states, where  $C_n = \{\mathbf{c}_{n,1}, \dots, \mathbf{c}_{n,K(n)}\}$  and each *parent*  $\mathbf{c} \in \mathcal{C}$  denotes the true position of a fluorophore that is present in the RFS  $C_n$ . Here, we assume that imaging occurs on some bounded region  $\mathcal{C} \subset \mathbb{R}^2$ . When  $N_F$  frames are imaged during an experiment, at every time  $n = 1, \dots, N_F$ , a set of *offspring* measurements  $Z_n = \{\mathbf{z}_{n,1}, \dots, \mathbf{z}_{n,M(n)}\}$  are recorded. This is said to mean that  $|Z_n| =: M(n) \in \mathbb{Z}_{\geq 0}$  distinct vectors  $\{\mathbf{z}_i\}_{i=1}^{M(n)}$ , with each  $\mathbf{z} \in \mathcal{C}$  are collected in frame  $n$ . We define  $Z_n$  to be the *multi-target* measurement formed by the  $K(n)$  present fluorophores and false positive observations (background noise).

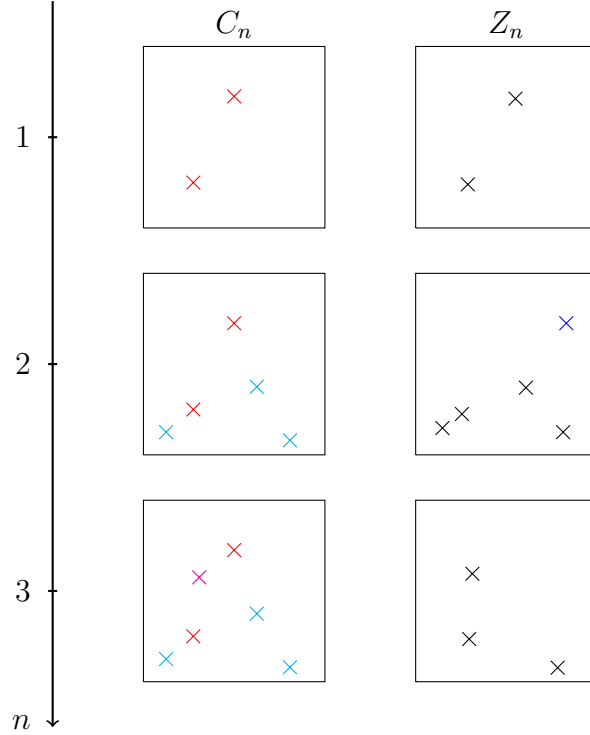
A suitable model we choose to place is therefore given by

$$\begin{aligned} C_1 &= B_1 \cup B_0 \\ C_n &= C_{n-1} \cup B_n \quad n > 1 \end{aligned} \tag{4.4}$$

$$Z_n = \Phi(C_n) \cup A_n \quad n \geq 1, \tag{4.5}$$

where  $C_n$  denotes the *hidden parent* RFS and  $Z_n$  denotes the *observed* RFS. In particular,  $B_n$  denotes the independent birth RFS of fluorophores “born” in frame  $n$ , with  $B_0$  denoting the RFS of fluorophores activated *before* imaging begins. Moreover,  $\Phi(C_n)$  denotes the RFS of primary target generated measurements and  $A_n$  independently denotes the RFS of false positive measurements. An illustration of this model over three frames is shown in Figure 4.3, which compares the hidden RFS  $C_1, C_2$  and  $C_3$  with their observation counterparts  $Z_1, Z_2$  and  $Z_3$ .

In order to statistically characterise the RFS densities of  $Z_n$  and  $C_n$ , we must first derive the detection and birth probabilities of a single fluorophore in each frame  $n$ . Specifically, the detection probabilities are needed to formulate the likelihood function for  $Z_n$ , and the birth probabilities determine



**Figure 4.3:** An illustration of the RFS  $C_n$  compared with the RFS  $Z_n$  against time  $n$ . Left:  $C_1 = B_1 \cup B_0$  (with  $B_0 = \emptyset$ ) is plotted in red at time  $n = 1$ ,  $C_2 = C_1 \cup B_2$  is plotted with  $B_2$  in cyan at time  $n = 2$  and  $C_3 = C_2 \cup B_3$  is plotted with  $B_3$  in magenta at time  $n = 3$ . Right: Observation sets  $Z_1, Z_2, Z_3$  are shown, with false positive observations plotted in blue.

the cardinality distribution of births occurring in each frame.

### 4.3.1 Probability of detection

For a single fluorophore, we define its detection probability in frame  $n$  as

$$p_{D,n}(\boldsymbol{\theta}^{(d)}) = \mathbb{P}_{\boldsymbol{\theta}^{(d)}}(Y_n = 1).$$

**Proposition 7.** *The probability of detection  $p_{D,n}(\boldsymbol{\theta}^{(d)})$  of a single fluorophore at frame  $n \in \mathbb{Z}_{>0}$  is given by*

$$p_{D,n}(\boldsymbol{\theta}^{(d)}) = \boldsymbol{\nu}_X^\top e^{G^{(n-1)\Delta}} B_\Delta^{(1)} \mathbf{1}_{d+3}, \quad (4.6)$$

where  $G$  denotes the Markovian generator of  $\{X(t)\}$  given in (1.1), and where  $e^{G\Delta} = B_{\Delta}^{(0)} + B_{\Delta}^{(1)}$  denotes its transition probability matrix over time instance  $\Delta$ .

*Proof.* For any  $n \in \mathbb{Z}_{>0}$ , we have

$$\begin{aligned} p_{D,n}(\boldsymbol{\theta}^{(d)}) &= \sum_{i_1, j \in \mathcal{S}_X} \mathbb{P}_{\boldsymbol{\theta}^{(d)}}(X(n\Delta) = j, Y_n = 1 | X((n-1)\Delta) = i_1) \\ &\quad \times \mathbb{P}_{\boldsymbol{\theta}^{(d)}}(X((n-1)\Delta) = i_1) \\ &= \sum_{\substack{i_1, \dots, i_n, j \\ \in \mathcal{S}_X}} b_{i_1 j, \Delta}^{(1)} \left( \prod_{l=1}^{n-1} \mathbb{P}_{\boldsymbol{\theta}^{(d)}}(X(\Delta) = i_l | X(0) = i_{l+1}) \right) \mathbb{P}_{\boldsymbol{\theta}^{(d)}}(X(0) = i_n), \end{aligned}$$

where the notational convention  $\prod_{\emptyset} := 1$  is used for the case  $n = 1$ .  $\square$

Proposition 7 gives the form of the detection probability for any  $n > 0$ , which is seen as a function of the photo-switching parameters  $\boldsymbol{\theta}^{(d)}$ . Furthermore, Figure 4.4a shows how  $p_{D,n}(\boldsymbol{\theta}^{(d)})$  decays over time and is in general non constant. In particular, Remark 9 proves that this probability under any photo-switching model  $\mathcal{M}_A^d$ ,  $A \subseteq \mathcal{S}_X$  tends to a deterministic constant.

**Remark 9.** Using (4.6), we see that asymptotically

$$\lim_{n \rightarrow \infty} p_{D,n}(\boldsymbol{\theta}^{(d)}) = \lim_{n \rightarrow \infty} \boldsymbol{\nu}_X^\top e^{G(n-1)\Delta} B_{\Delta}^{(1)} \mathbf{1}_{d+3} = \boldsymbol{\pi}_X^\top B_{\Delta}^{(1)} \mathbf{1}_{d+3},$$

where  $\boldsymbol{\pi}_X$  is the stationary distribution of  $\{X(t)\}$ . When the photo-bleaching state 2 is not included in the model, i.e. under model  $\mathcal{M}_{\emptyset}^d$ , the irreducibility of  $\{X(t)\}$  on  $\mathcal{S}_X$  implies that  $\boldsymbol{\pi}_X$  is the unique solution of  $\boldsymbol{\pi}_X^\top G = 0$  and  $\boldsymbol{\pi}_X^\top \mathbf{1}_{d+2} = 1$ . When the photo-bleaching state is included:  $\boldsymbol{\pi}_X = \mathbf{e}_{d+3}^{d+3}$ , the  $(d+3)$ th canonical (standard) basis vector of  $\mathbb{R}^{d+3}$ , which, given the form of  $B_{\Delta}^{(1)}$  in (1.4), implies that  $\lim_{n \rightarrow \infty} p_{D,n}(\boldsymbol{\theta}^{(d)}) = 0$ , for all  $\boldsymbol{\theta}^{(d)} \in \Theta^{(d)}$ . Under any model, we therefore see that  $p_{D,n}(\boldsymbol{\theta}^{(d)})$  tends to a deterministic constant.

### 4.3.2 Number of births

We will now introduce the notion of the *birth probability*. Intuitively, the probability that a fluorophore is “born” in frame  $n$  represents the fact that its hidden Markovian signal  $\{X(t)\}$  reaches the photon-emission On state 1 for the *first time*, and that this first passage time lies within the time interval  $[(n-1)\Delta, n\Delta)$ . Specifically, at every time instance  $n \in \mathbb{Z}_{>0}$ , we define the birth probability of a single molecule to be

$$\begin{aligned} p_{B,n}(\boldsymbol{\theta}^{(d)}) &= \mathbb{P}_{\boldsymbol{\theta}^{(d)}}(\text{There exists } t \in [(n-1)\Delta, n\Delta) : X(t) = 1, X(s) \neq 1 \\ &\quad \text{for all } s \in [-T^*, (n-1)\Delta)) \\ &= \mathbb{P}_{\boldsymbol{\theta}^{(d)}, \delta=0}(Y_n = 1, Y_{n-1} = 0, \dots, Y_1 = 0, X(s) \neq 1 \text{ for all } s \in [-T^*, 0)) \\ &\equiv \mathbb{P}_{\boldsymbol{\theta}^{(d)}, \delta=0}(Y_n = 1, Y_{1:n-1} = 0, X(s) \neq 1 \text{ for all } s \in [-T^*, 0)), \end{aligned}$$

where we use the notational convention that  $\{Y_{1:n-1} = 0\} := \{Y_1 = 0, Y_2 = 0, \dots, Y_{n-1} = 0\}$ . Here, we shall remind the reader that an experiment is defined to begin at time  $-T^*$ ; that is to say, the Markov process  $\{X(t)\}$  associated with each molecule is initialised at this time. While molecules are not observed in the time interval  $[-T^*, 0)$  (see Remark 1 of Chapter 1), the signals generated by  $\{X(t)\}$  may activate fluorophores into the On state within this period. Proposition 8 proves that the form of  $p_{B,n}(\boldsymbol{\theta}^{(d)})$  therefore relies now on the probability mass  $\boldsymbol{\nu}_X^*$  of  $\{X(-T^*)\}$  (see Remark 2 of Chapter 1) and the transmission matrices  $B_{T^*, \delta=0}^{(0)}$ ,  $B_{\Delta, \delta=0}^{(0)}$  and  $B_{\Delta, \delta=0}^{(1)}$ , which can be readily computed using the same methodology as is presented in Section 1.3.1, setting  $\delta = 0$ . We further note that the birth of a fluorophore does not necessarily mean it is detected in that frame, with its detection probability given in (4.6).

**Proposition 8.** *The birth probability  $p_{B,n}(\boldsymbol{\theta}^{(d)})$  of a single fluorophore at frame  $n \in \mathbb{Z}_{>0}$  is given by*

$$p_{B,n}(\boldsymbol{\theta}^{(d)}) = \boldsymbol{\nu}_X^{*\top} (B_{T^*, \delta=0}^{(0)}) (B_{\Delta, \delta=0}^{(0)})^{n-1} B_{\Delta, \delta=0}^{(1)} \mathbf{1}_{d+3}, \quad (4.7)$$

where  $B_{\Delta, \delta=0}^{(0)}$  and  $B_{\Delta, \delta=0}^{(1)}$  denote the two transmission matrices computed when  $\delta = 0$ . Moreover, the probability  $p_{B,0}(\boldsymbol{\theta}^{(d)})$  that a molecule is activated

before imaging begins is given by

$$p_{B,n}(\boldsymbol{\theta}^{(d)}) = \boldsymbol{\nu}_X^{*\top} B_{T^*,\delta=0}^{(1)} \mathbf{1}_{d+3}. \quad (4.8)$$

*Proof.* For each  $i, j \in \mathcal{S}_X$ ,  $l \in \{0, 1\}$  and  $t > 0$ , define the transmission probabilities under the noiseless system as  $b_{ij,t,\delta=0}^{(l)} = \mathbb{P}_{\boldsymbol{\theta}^{(d)}}(X(t) = j, Y_0 = l | X(0) = i, \delta = 0)$ . Then

$$\begin{aligned} p_{B,n}(\boldsymbol{\theta}^{(d)}) &= \sum_{j,i_1 \in \mathcal{S}_X} b_{i_1 j, \Delta, \delta=0}^{(1)} \mathbb{P}_{\boldsymbol{\theta}^{(d)}, \delta=0}(X((n-1)\Delta) = i_1 | Y_{1:n-1} = 0) \\ &\quad \times \mathbb{P}_{\boldsymbol{\theta}^{(d)}, \delta=0}(Y_{1:n-1} = 0, X(s) \neq 1 \forall s \in [-T^*, 0)) \\ &= \sum_{j, i_1, i_2 \in \mathcal{S}_X} b_{i_1 j, \Delta, \delta=0}^{(1)} b_{i_2 i_1, \Delta, \delta=0}^{(0)} \mathbb{P}_{\boldsymbol{\theta}^{(d)}, \delta=0}(X((n-2)\Delta) = i_2 | Y_{1:n-2} = 0) \\ &\quad \times \mathbb{P}_{\boldsymbol{\theta}^{(d)}, \delta=0}(Y_{1:n-2} = 0, X(s) \neq 1 \forall s \in [-T^*, 0)) \\ &= \sum_{\substack{i_1, \dots, i_{n+1}, j \\ \in \mathcal{S}_X}} b_{i_1 j, \Delta, \delta=0}^{(1)} \left( \prod_{l=1}^{n-1} b_{i_{l+1} i_l, \Delta, \delta=0}^{(0)} \right) b_{i_{n+1} i_n, T^*, \delta=0}^{(0)} \mathbb{P}_{\boldsymbol{\theta}^{(d)}}(X(-T^*) = i_{n+1}), \end{aligned}$$

where the notational convention  $\prod_{\emptyset} := 1$  is used for the case  $n = 1$ . By using similar arguments, the probability that  $\{X(t)\}$  with initial probability mass  $\boldsymbol{\nu}_X^*$ , first reaches the photon emission state within the time  $[-T^*, 0)$  is simply  $\boldsymbol{\nu}_X^\top B_{T^*, \delta=0}^{(1)} \mathbf{1}_{d+3}$ .  $\square$

Proposition 9 shows that for any model placed on the photo-switching behaviour of the fluorophores, the birth probability  $p_{B,n}(\boldsymbol{\theta}^{(d)}) \rightarrow 0$  as  $n \rightarrow \infty$ , and so we will see a decrease in the number of births. Figure 4.4b specifically shows how this birth probability decays over time.

**Proposition 9.** For any  $\boldsymbol{\theta}^{(d)} \in \Theta^{(d)}$  and  $\Delta > 0$ ,

$$\lim_{n \rightarrow \infty} p_{B,n}(\boldsymbol{\theta}^{(d)}) = 0.$$

*Proof.* See Section 4.A of Appendix 4.  $\square$

The birth probabilities are useful in determining the average number of births in any given frame. In particular, if  $M$  (total number of independently

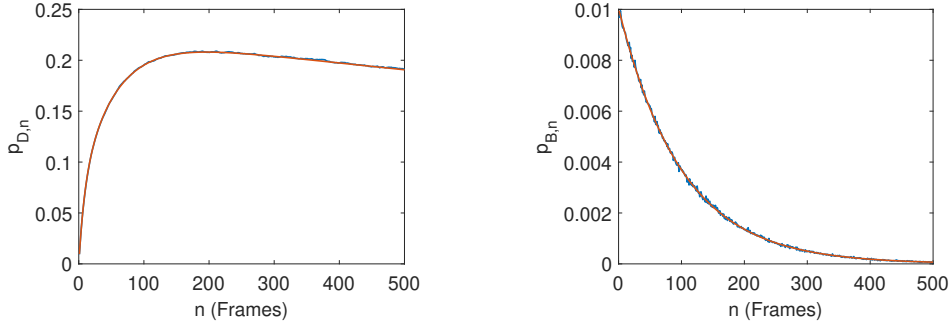
imaged molecules) is *known* prior to the experiment, then the number of births at frame  $n$ , which will herein be referred to as  $N_{B,n}$ , clearly satisfies  $N_{B,n} \sim \text{Bin}(M, p_{B,n}(\boldsymbol{\theta}^{(d)}))$  with expected value  $Mp_{B,n}(\boldsymbol{\theta}^{(d)})$ . In most microscopy experiments however,  $M$  is unknown. Since each fluorophore will generate an independent set of spatial measurements (which frequently overlap between molecules) until the end of the experiment is reached, there is only partial information available of  $M$  and as such a suitable prior distribution must be invoked for proper inference.

For this problem, we deem a Poisson prior on  $M$ , namely that  $M \sim \text{Poi}(\lambda_M)$ , suitable over the observation window. Here,  $\lambda_M$  is an appropriate parameter characterising the number of molecules expected to be imaged prior to the experiment. By Remark 10, this choice of prior enables  $N_{B,n}$  to be Poisson distributed with mean parameter  $\lambda_M p_{B,n}(\boldsymbol{\theta}^{(d)})$  and therefore motivates our choice of placing a Poisson prior distribution on the RFS  $B_n$ .

**Remark 10.** Using  $N_{B,n}|M \sim \text{Bin}(M, p_{B,n}(\boldsymbol{\theta}^{(d)}))$  and  $M \sim \text{Poi}(\lambda_M)$ , we have  $\mathbb{E}(s^{N_{B,n}}) = \mathbb{E}_M(\mathbb{E}_{N_{B,n}|M}(s^{N_{B,n}}|M)) = \mathbb{E}_M((1-p_{B,n}(\boldsymbol{\theta}^{(d)})+p_{B,n}(\boldsymbol{\theta}^{(d)})s)^M) = \exp(-\lambda_M p_{B,n}(\boldsymbol{\theta}^{(d)})(1-s))$ , which is recognised as the pgf of a Poisson random variable with rate parameter  $\lambda_M p_{B,n}(\boldsymbol{\theta}^{(d)})$ . Hence

$$N_{B,n} \sim \text{Poi}(\lambda_M p_{B,n}(\boldsymbol{\theta}^{(d)})).$$

**Remark 11.** Since  $N_{B,n} \sim \text{Poi}(\lambda_M p_{B,n}(\boldsymbol{\theta}^{(d)}))$  from Remark 10, we have for any  $\epsilon > 0$  that  $\mathbb{P}(N_{B,n} > \epsilon) \leq 1 - e^{-\lambda_M p_{B,n}(\boldsymbol{\theta}^{(d)})} \rightarrow 0$  as  $n \rightarrow \infty$  by Proposition 9, so the sequence of random variables  $\{N_{B,n}\} \rightarrow 0$  in probability (and hence in distribution). In particular,  $N_{B,n}$  converges to the degenerate distribution centred at 0.



(a) Probability of a detection over an experiment. (b) Probability of a birth over an experiment.

**Figure 4.4:** Plots showing how the probability of detection of a single fluorophore (left) and the the number of births in a single frame (right) changes over the course of the experiment. Blue shows the simulated (empirical) estimates and orange shows the true probabilities (calculated directly). Simulations carried out under model  $M_{\{1\}}^1$  whereby  $T^* = 0$ ,  $\nu_{0_1}^* = \nu_{0_1} = 1$ ,  $\Delta = \frac{1}{30}s$ ,  $\delta = 10^{-3}s$  and  $\alpha = 0$ ; rates chosen are  $\lambda_{00_1} = 0.35s^{-1}$ ,  $\lambda_{01} = 1s^{-1}$ ,  $\lambda_{0_1,1} = 0.3s^{-1}$ ,  $\lambda_{10} = 2.3s^{-1}$ ,  $\mu_1 = 0.05s^{-1}$ .

Specifically, both the birth RFS  $B_n$  and RFS for false positives are assumed to be a priori Poisson distributed, with respective spatial densities (see (4.3) in Remark 8)  $f_{B_n}(C')$  and  $f_{A_n}(C')$  defined for any  $C' \subseteq \mathcal{C}$ . By the above,  $B_n$  is modelled to have cardinality distribution  $N_{B_n} \sim \text{Poi}(\lambda_M p_{B,n}(\boldsymbol{\theta}^{(d)}))$  where  $p_{B,n}(\boldsymbol{\theta}^{(d)})$  as is given in (4.7) denotes the birth probability of a fluorophore (switching between  $d + 1$  dark states) at time step  $n$ . Moreover,  $A_n$  is modelled to have cardinality distribution  $N_{A_n} \sim \text{Poi}(\alpha)$ , where  $\alpha$  is the unknown average number of false positive observations produced in a *single* frame. In this manner, the spatial densities (with respect to the reference measure in (4.1)), take the form

$$f_{B_n}(C') = e^{-\lambda_M p_{B,n}(\boldsymbol{\theta}^{(d)})} \prod_{\mathbf{c} \in C'} \lambda_M p_{B,n}(\boldsymbol{\theta}^{(d)}) b(\mathbf{c})$$

$$f_{A_n}(C') = e^{-\alpha} \prod_{\mathbf{c} \in C'} \alpha a(\mathbf{c}) \quad n \geq 1,$$

where  $b(\mathbf{c})$ ,  $a(\mathbf{c})$  denote the spatial distributions of births and false positives, respectively. In particular, these densities must satisfy  $\int_{\mathcal{C}} b(\mathbf{c}) \, d\mathbf{c} = \int_{\mathcal{C}} a(\mathbf{c}) \, d\mathbf{c} = 1$ .

Since false positives are likely to occur uniformly over  $\mathcal{C}$ , we choose  $a(\mathbf{c}) =$



$\frac{1}{\lambda_L(\mathcal{C})}$ . Furthermore, since in most imaging applications, the spatial structures of cells are not known apriori, we also choose  $b(\mathbf{c}) = \frac{1}{\lambda_L(\mathcal{C})}$  to account for this uncertainty. It is important to note that this distribution can be suitably adjusted in experiments where one may have a better insight into  $b(\mathbf{c})$ . This choice of prior distribution on the birth point processes further informs us of the prior distribution of the RFS  $C_n$ . Specifically, by Equation (4.4) of the model,  $C_n$  can be written as  $C_n = \cup_{i=0}^n B_i$ . Since the birth processes  $\{B_i\}_{i=0}^n$  are statistically independent on  $\mathcal{C}$ , we observe (through an application of Example 4) that under any choice of  $b(\mathbf{c})$ ,  $C_n$  is apriori Poisson distributed with intensity function  $v(\mathbf{c}) = \lambda_M \sum_{i=0}^n p_{B_i}(\boldsymbol{\theta}^{(d)})b(\mathbf{c})$  for any  $\mathbf{c} \in \mathcal{C}$ .

In order to specify the function  $\Phi(C_n)$  of  $Z_n$ , we note that each fluorophore or parent  $\mathbf{c} \in C_n$  generates an offspring  $\mathbf{z} \in \Phi(C_n)$  such that  $\mathbf{z} = \emptyset$  with probability  $1 - p_{D,n}(\boldsymbol{\theta}^{(d)})$  or is a variate from the density  $f(\mathbf{z}^*|\mathbf{c})$  otherwise. Here,  $p_{D,n}(\boldsymbol{\theta}^{(d)})$  denotes the detection probability of a parent in frame  $n$  (as is given in (4.6)) and

$$f(\mathbf{z}^*|\mathbf{c}) \propto \exp\left(-\frac{1}{2\sigma_{\mathbf{z}^*}^2}(\mathbf{z}^* - \mathbf{c})^\top(\mathbf{z}^* - \mathbf{c})\right)$$

denotes the likelihood that offspring point  $\mathbf{z}^*$  is generated from parent  $\mathbf{c}$ . As is common in fluorescence microscopy experiments, the localisation standard deviation  $\sigma_{\mathbf{z}^*}$  can be measured from the localisation algorithm used, and is therefore assumed to be known for each observation  $\mathbf{z} \in Z_n$ .

### 4.3.3 Derivation of the Bayes filter

From Remark 11, one can see that  $B_n \rightarrow \{\emptyset\}$  in probability and thus that

$$f_{B_n}(C) \xrightarrow{D} \begin{cases} 1 & \text{if } C = \emptyset \\ 0 & \text{otherwise,} \end{cases}$$

where the limit denotes convergence in distribution. By defining the limiting RFS  $C^* := \lim_{n \rightarrow \infty} C_n$ , which is interpreted as the *implicit* point pattern configuration of the parents, the true configuration of  $C^*$  can thereby be

updated through time. Specifically, for any  $n \in \mathbb{Z}_{>0}$ , we wish to infer upon the parent RFS  $C_n$  (with its cardinality), and posterior estimates for

$$\boldsymbol{\theta}_\alpha^{(d)} := ((\boldsymbol{\theta}^{(d)})^\top \quad \alpha)^\top \in \Theta_\alpha^{(d)} := \Theta^{(d)} \times (0, \infty)$$

under a photo-switching model with  $d + 1$  dark states, given all observations  $Z^{(n)} := \cup_{i=1}^n Z_i$  through the Bayes-updated density

$$\begin{aligned} f_{C_n, \Theta_\alpha^{(d)} | Z^{(n)}}(C_n, \boldsymbol{\theta}_\alpha^{(d)} | Z^{(n)}) &\propto \\ f_{Z_n | C_n, \Theta_\alpha^{(d)}}(Z_n | C_n, \boldsymbol{\theta}_\alpha^{(d)}) f_{C_n | Z^{(n-1)}, \Theta_\alpha^{(d)}}(C_n | Z^{(n-1)}, \boldsymbol{\theta}_\alpha^{(d)}) \pi(\boldsymbol{\theta}_\alpha^{(d)}). \end{aligned} \quad (4.9)$$

Here,  $\pi(\boldsymbol{\theta}_\alpha^{(d)})$  denotes a suitable prior probability density for  $\boldsymbol{\theta}_\alpha^{(d)}$ .

Equation (4.9) can be computed by first using the *multi-target* likelihood as described in Mahler (2007b) by

$$\begin{aligned} f_{Z_n | C_n, \Theta_\alpha^{(d)}}(Z_n | C_n, \boldsymbol{\theta}_\alpha^{(d)}) &= \\ (1 - p_{D,n}(\boldsymbol{\theta}^{(d)}))^{|C_n|} f_{A_n}(Z_n) \sum_{\phi} \prod_{i: \phi(i) > 0} \frac{p_{D,n}(\boldsymbol{\theta}^{(d)}) f(\mathbf{z}_{\phi(i)} | \mathbf{c}_i) \mu(S)}{\alpha (1 - p_{D,n}(\boldsymbol{\theta}^{(d)}))}, \end{aligned} \quad (4.10)$$

where the sum is taken over all mappings  $\phi : \{1, \dots, |C_n|\} \rightarrow \{0, 1, \dots, |Z_n|\}$ , whereby  $\phi(i) = \phi(i') > 0 \implies i = i'$ . It should be noted here that when  $|Z_n| = 0$ , no observations are collected and the likelihood is therefore reduced to  $f_{Z_n | C_n, \Theta_\alpha^{(d)}}(\emptyset | C_n, \boldsymbol{\theta}_\alpha^{(d)}) = e^{-\alpha} (1 - p_{D,n}(\boldsymbol{\theta}^{(d)}))^{|C_n|}$ .

We can additionally show that up to a proportionality constant, the density  $f_{C_n | Z^{(n-1)}, \Theta_\alpha^{(d)}}(C_n | Z^{(n-1)}, \boldsymbol{\theta}_\alpha^{(d)})$  takes the form in (4.11) as is stated in Proposition 10.

**Proposition 10.** *For any  $n \in \mathbb{Z}_{>1}$ , the density  $f_{C_n | Z^{(n-1)}, \Theta_\alpha^{(d)}}(C_n | Z^{(n-1)}, \boldsymbol{\theta}_\alpha^{(d)})$  is proportional to the function*

$$\begin{aligned} \sum_{W_n \subseteq C_n} \sum_{W_{n-1} \subseteq W_n} \dots \sum_{W_1 \subseteq W_2} f_{B_n}(C_n \setminus W_n) f_{B_0}(W_1) \times \\ \left[ \prod_{i=1}^{n-1} f_{B_i}(W_{i+1} \setminus W_i) f_{Z_i | C_n, \Theta_\alpha^{(d)}}(Z_i | W_{i+1}, \boldsymbol{\theta}_\alpha^{(d)}) \right]. \end{aligned} \quad (4.11)$$

*Proof.* For every  $n \in \mathbb{Z}_{>1}$ , we can write  $\{C_n|Z^{(n-1)}, \boldsymbol{\theta}_\alpha^{(d)}\} = \{C_{n-1}|Z^{(n-1)}, \boldsymbol{\theta}_\alpha^{(d)}\} \cup \{B_n\}$ , where the point processes  $\{C_{n-1}|Z^{(n-1)}, \boldsymbol{\theta}_\alpha^{(d)}\}$  and  $\{B_n\}$  are independent by model construction.

By Theorem 6, we have for all  $n > 1$  that

$$f_{C_n|Z^{(n-1)}, \boldsymbol{\theta}_\alpha^{(d)}}(C_n|Z^{(n-1)}, \boldsymbol{\theta}_\alpha^{(d)}) \propto \sum_{W \subseteq C_n} f_{C_{n-1}|Z^{(n-1)}, \boldsymbol{\theta}_\alpha^{(d)}}(W|Z^{(n-1)}, \boldsymbol{\theta}_\alpha^{(d)}) f_{B_n}(C_n \setminus W), \quad (4.12)$$

where the sum is taken over all subsets  $W$  of  $C_n$ .

Since  $f_{B_0 \cup B_1}(C_1)$  characterises the distribution of  $C_1$ , we have

$$\begin{aligned} f_{C_1|Z^{(1)}, \boldsymbol{\theta}_\alpha^{(d)}}(C_1|Z_1, \boldsymbol{\theta}_\alpha^{(d)}) &\propto f_{Z_1|C_1, \boldsymbol{\theta}_\alpha^{(d)}}(Z_1|C_1, \boldsymbol{\theta}_\alpha^{(d)}) f_{B_0 \cup B_1}(C_1) \\ &\propto f_{Z_1|C_1, \boldsymbol{\theta}_\alpha^{(d)}}(Z_1|C_1, \boldsymbol{\theta}_\alpha^{(d)}) \sum_{W \subseteq C_1} f_{B_0}(W_1) f_{B_1}(C_1 \setminus W). \end{aligned}$$

Using this as an initialisation for (4.12), coupled with the Bayes update rule given in (4.9), yields the desired result. □

**Remark 12.** When  $\nu_1^* = 1$ , as is the case in (d)STORM experiments (see Remark 2), all fluorophores have been activated prior to imaging. In this situation, we therefore have that  $p_{B,0}(\boldsymbol{\theta}^{(d)}) = 1$  and  $p_{B,n}(\boldsymbol{\theta}^{(d)}) = 0$  for all  $n \geq 1$ , enabling  $C^* = C_1$ . Moreover, the birth densities in (4.11) are only non-zero when each  $W_i$  is chosen to be  $C_n$ , and  $f_{C_n|Z^{(n)}, \boldsymbol{\theta}_\alpha^{(d)}}(C_n|Z^{(n)}, \boldsymbol{\theta}_\alpha^{(d)})$  reduces to

$$f_{C_n|Z^{(n)}, \boldsymbol{\theta}_\alpha^{(d)}}(C_n|Z^{(n)}, \boldsymbol{\theta}_\alpha^{(d)}) \propto \left( e^{-\lambda_M} \prod_{\mathbf{c} \in C_n} \lambda_M b(\mathbf{c}) \right) \prod_{i=1}^n f_{Z_i|C_n, \boldsymbol{\theta}_\alpha^{(d)}}(Z_i|C_n, \boldsymbol{\theta}_\alpha^{(d)}). \quad (4.13)$$

#### 4.3.4 Limitations

While the full Bayes filter for this problem has been derived, we should highlight that the likelihood in (4.10) requires computations of all unique

combinations of the set  $\{1, \dots, M\}$  that have cardinality smaller than or equal to the size of the observation set. This is therefore only computationally feasible to compute when the number of molecules  $M$  is small, or when the cardinality of the observation set is small. Furthermore, while Remark 12 highlights a more computationally feasible density that can be implemented when all births of molecules have occurred in the first time step, the full density in (4.11) is observed to being even more restrictive on  $M$ , requiring multiple computations of power sets. In any case, this filter heavily limits the number of molecules that can be analysed at once.

## 4.4 Inference

In this section, we describe how a Metropolis-within-Gibbs MCMC sampler can be used to infer the unknown parameters  $C_{N_F}$  and  $\boldsymbol{\theta}_\alpha^{(d)}$  from the posterior density

$$f_{C_{N_F}, \boldsymbol{\theta}_\alpha^{(d)} | Z^{(N_F)}}(C_{N_F}, \boldsymbol{\theta}_\alpha^{(d)} | Z^{(N_F)}).$$

While there are many different variations of MCMC algorithms which can be implemented to sample from the intended target distribution, we construct a sampler that relies on a transformation of the parameter vector  $\boldsymbol{\theta}_\alpha^{(d)}$  and which is found to perform well for this problem. We subsequently describe the specific steps of this sampler and discuss how posterior samples of  $C_{N_F}$  and  $\boldsymbol{\theta}_\alpha^{(d)}$  obtained from this algorithm can be used for parameter estimation and model selection.

In the previous section, specifically in Equation (4.9), we derived the conditional density  $f_{C_n, \boldsymbol{\theta}_\alpha^{(d)} | Z^{(n)}}(C_n, \boldsymbol{\theta}_\alpha^{(d)} | Z^{(n)})$  for any  $n \geq 1$ . Using this, we now wish to perform inference on  $C_{N_F}$  and  $\boldsymbol{\theta}^{(d)}$  given all observation sets  $Z^{(N_F)}$  collected during an experiment consisting of  $N_F$  frames. Since inference of point processes and their parameters is typically done through Markov Chain Monte Carlo (MCMC) (Geyer and Møller, 1994, Møller and Waagepetersen, 2003), we choose to implement a suitable Metropolis-within-Gibbs (MwG) MCMC algorithm that targets  $C_{N_F}$  and  $\boldsymbol{\theta}^{(d)}$  from their joint posterior distribution via continual sampling from the conditional distri-

butions  $f_{C_n|\Theta_\alpha^{(d)}, Z^{(n)}}(C_n|\boldsymbol{\theta}_\alpha^{(d)}, Z^{(n)})$  and  $f_{\Theta_\alpha^{(d)}|C_n, Z^{(n)}}(\boldsymbol{\theta}_\alpha^{(d)}|C_n, Z^{(n)})$  through two Metropolis-Hastings kernels.

In Møller and Waagepetersen (2003), the authors describe two Metropolis-Hastings type algorithms which can be implemented to sample from unnormalised point processes with a known density, as is the case for  $C_{N_F}$  here. The first, called the *fixed number of points Metropolis-Hastings* sampler (Algorithm 7.1 of Møller and Waagepetersen (2003)) can sample locations of the point process of a *fixed* cardinality. The second, called the *birth-death-move Metropolis-Hastings* sampler (Algorithm 7.5 of Møller and Waagepetersen (2003)) is a special type of the well known reversible jump MCMC (Green, 1995) algorithm which can be implemented to jointly sample both the locations and the cardinality of the process. The latter therefore also produces a (posterior) distribution over the unknown number of points the process has. The ergodicity and irreducibility properties of the resulting chains are also proved in this reference.

For our problem, we have chosen to use the fixed number of points Metropolis-Hastings algorithm to update cluster centres  $C_{N_F}$  from the conditional density  $f_{C_n|\Theta_\alpha^{(d)}, Z^{(n)}}(C_n|\boldsymbol{\theta}_\alpha^{(d)}, Z^{(n)})$ . This sampler is chosen for two reasons. Firstly, the form of the posterior density inhibits computations for large  $M$ , which will affect the irreducibility of a chain constructed by the birth-death-move algorithm. Secondly, the posterior distribution of  $\boldsymbol{\theta}^{(d)}$  is seen to be *non-nested* between models with differing cardinalities  $M$  and  $d$ . This is said to mean that the resulting posterior distribution of  $\boldsymbol{\theta}^{(d)}$  is not invariant to  $M$ . This is not surprising however, since for a fixed dataset  $Z^{(N_F)}$ , the detection and birth probabilities are likely to vary between models of varying dimension. For this reason, constructing efficient jump proposals within the same MCMC chain is also difficult.

Sampling the unknown parameters of the model  $\boldsymbol{\theta}^{(d)}$  from the conditional density  $f_{\Theta_\alpha^{(d)}|C_n, Z^{(n)}}(\boldsymbol{\theta}_\alpha^{(d)}|C_n, Z^{(n)})$ , can be done through an application of the standard Metropolis-Hastings sampler. Although updating the parent process  $C_{N_F}$  through the fixed number of points MCMC sampler is not seen to be problematic, we found that implementing a Metropolis-Hastings algorithm which updates parameters in  $\boldsymbol{\theta}_\alpha^{(d)}$  independently resulted in poor

mixing of the chain. Poor mixing was observed due to two main reasons. Firstly, the target space  $\Theta_\alpha^{(d)}$  is *restricted*, in that parameters can only take values on some smaller subset of the real hyper-plane. This makes it difficult to construct efficient proposals from distributions living in the same space. Secondly, the parameters in  $\boldsymbol{\theta}^{(d)}$  appear in the model through the birth and detection probabilities, and are therefore subject to time induced *correlations*. The correlation analysis of photo-switching parameters presented in Chapter 2 also supports this observation. To this end, we decide to implement a sampler which utilises a bijective transformation that maps the parameter vector  $\boldsymbol{\theta}_\alpha^{(d)}$  to the real valued hyper-plane. The inherent correlation structure between transformed parameters can then be exploited by an application of the Adaptive Metropolis (AM) algorithm detailed in [Haario et al. \(2001\)](#). This algorithm constructs new proposals using the empirical covariance of previously accepted samples. A further description of this sampler, including a proof detailing its ergodicity, can also be found in this reference.

The MwG MCMC algorithm we use here will therefore use the fixed number of points Metropolis-Hastings algorithm to *shift* the locations of a given number of  $M$  centres in  $C_{N_F}$  from the density  $f_{C_{N_F}|Z^{(N_F)},\Theta_\alpha^{(d)}}(C_{N_F}|Z^{(N_F)},\boldsymbol{\theta}_\alpha^{(d)},M) \propto$

$$f_{Z_{N_F}|C_{N_F},\Theta_\alpha^{(d)}}(Z_{N_F}|C_{N_F},\boldsymbol{\theta}_\alpha^{(d)},M)f_{C_{N_F}|Z^{(N_F-1)},\Theta_\alpha^{(d)}}(C_{N_F}|Z^{(N_F-1)},\boldsymbol{\theta}_\alpha^{(d)},M), \quad (4.14)$$

which can be evaluated (up to a proportionality constant) from (4.10) and (4.11). Further, it will also *update*  $\boldsymbol{\theta}^{(d)}$  from (4.9) given a photo-switching model of the form  $\mathcal{M}_A^d$ , where  $A \subseteq \mathcal{S}_X$  denotes the set of states in  $\mathcal{S}_X$  from which the photo-bleaching state 2 is accessible. Multiple chains will therefore be created over different models, resulting in the application of a model selection criterion to pick the most suitable model from a range of given proposals. For this problem, we have chosen to use the Deviance Information Criterion (DIC) for our analysis, which can easily be evaluated using the output of the corresponding MCMC chain.

Before describing the specific *update* and *shift* moves that are needed for implementing this sampler, we first detail the transformation which is used to

update  $\boldsymbol{\theta}^{(d)}$  and define a suitable prior distribution for it under this mapping.

### Change of variables

Under a model  $\mathcal{M}_A^d$   $A \subseteq \mathcal{S}_X$ , the parameter vector of interest is  $\boldsymbol{\theta}_\alpha^{(d)}$  which takes values in

$$\Theta_\alpha^{(d)} = \mathbb{R}_{>0}^{2(d+1)+|A|+1} \times \mathbb{S}^{d+3} \times [0, \Delta),$$

where  $\mathbb{S}^{d+3}$  denotes the simplex of  $d + 3$  - dimensional probability vectors.

We now endeavour to find a bijective mapping  $g : \Theta_\alpha^{(d)} \rightarrow \mathbb{R}^{3(d+2)+|A|}$ . This not just enables Gaussian proposals to be used for the transformed parameters in the corresponding Metropolis-Hastings step, but is likely to improve mixing of the resulting chain. We define

$$\boldsymbol{\theta}'_\alpha^{(d)} := g(\boldsymbol{\theta}_\alpha^{(d)}) = \left[ g_1^\top([\boldsymbol{\lambda}_G \ \alpha]^\top) \quad g_2^\top(\boldsymbol{\nu}'_X) \quad g_3(\delta) \right]^\top,$$

and is such that its inverse mapping  $\boldsymbol{\theta}_\alpha^{(d)} := g^{-1}(\boldsymbol{\theta}'_\alpha^{(d)})$  is

$$g^{-1}(\cdot) = \left[ g_1^{-1\top}(\cdot) \quad g_2^{-1\top}(\cdot) \quad g_3^{-1}(\cdot) \right]^\top.$$

Firstly, we let  $[\boldsymbol{\lambda}'_G \ \alpha']^\top := g_1([\boldsymbol{\lambda}_G \ \alpha]^\top) = \log([\boldsymbol{\lambda}_G \ \alpha]^\top)$ . This implies that the inverse transformation is defined by  $[\boldsymbol{\lambda}_G \ \alpha]^\top = g_1^{-1}([\boldsymbol{\lambda}'_G \ \alpha']^\top) = \exp([\boldsymbol{\lambda}_G \ \alpha]^\top)$ . Moreover, we choose each photo-switching rate in  $\boldsymbol{\lambda}_G$  and  $\alpha$  to be a priori log-normally distributed, such that

$$\log([\boldsymbol{\lambda}_G \ \alpha]^\top) \sim \mathcal{N}(\mathbf{0}_{2(d+1)+|A|+1}, I_{2(d+1)+|A|+1}).$$

The prior distribution of  $[\boldsymbol{\lambda}'_G \ \alpha']^\top$  under this transformation is therefore given by  $\pi_{g_1}([\boldsymbol{\lambda}'_G \ \alpha']^\top) \propto \exp\left(-\frac{1}{2}[\boldsymbol{\lambda}'_G \ \alpha']^\top [\boldsymbol{\lambda}'_G \ \alpha']\right)$ .

Secondly, we let  $\boldsymbol{\nu}'_X := g_2(\boldsymbol{\nu}_X) = \left[ \log\left(\frac{\nu_0}{\nu_2}\right) \quad \dots \quad \log\left(\frac{\nu_{0d}}{\nu_2}\right) \quad \log\left(\frac{\nu_1}{\nu_2}\right) \right]^\top$  with  $\nu_2 = 1 - \sum_{i \in \bar{\mathcal{S}}_X} \nu_i$ . The inverse transformation is defined by  $\boldsymbol{\nu}_X = g_2^{-1}(\boldsymbol{\nu}'_X) = \left[ \frac{e^{-\nu'_0}}{1 + \sum_{i \in \bar{\mathcal{S}}_X} e^{-\nu'_i}} \quad \dots \quad \frac{e^{-\nu'_{0d}}}{1 + \sum_{i \in \bar{\mathcal{S}}_X} e^{-\nu'_i}} \quad \frac{e^{-\nu'_1}}{1 + \sum_{i \in \bar{\mathcal{S}}_X} e^{-\nu'_i}} \right]^\top$ . Moreover, we

let  $\boldsymbol{\nu}_X$  be prior distributed from a Logistic-Normal distribution with mean parameter  $\mathbf{0}_{d+2}$  and covariance matrix  $I_{d+2}$ . This distribution has probability density function

$$\pi(\boldsymbol{\nu}_X) = \frac{1}{(2\pi)^{\frac{d+3}{2}}} \frac{1}{\prod_{i \in \mathcal{S}_X} \nu_i} e^{-\frac{1}{2} \log\left(\frac{\nu_X^-}{\nu_2}\right)^\top \log\left(\frac{\nu_X^-}{\nu_2}\right)},$$

where  $\boldsymbol{\nu}_X^- = (\nu_0 \ \dots \ \nu_{0_d} \ \nu_1)^\top$ . It is known from [Atchinson and Shen \(1980\)](#) that the distribution of  $\boldsymbol{\nu}'_X$  resulting from the transformation  $g_2$  follows  $g_2(\boldsymbol{\nu}_X) \sim \mathcal{N}(\mathbf{0}_{d+2}, I_{d+2})$ , enabling  $\pi_{g_2}(\boldsymbol{\nu}'_X) \propto \exp\left(-\frac{1}{2} \boldsymbol{\nu}'_X{}^\top \boldsymbol{\nu}'_X\right)$ .

Finally, we let  $\delta' := g_3(\delta) = -\log\left(\frac{\Delta-\delta}{\delta}\right)$ , with inverse  $g_3^{-1}(\delta') = \frac{\Delta}{e^{-\delta'}+1}$ . Here we define the prior distribution of  $\delta'$  to be

$$\pi_{g_3}(\delta') = \frac{e^{-\delta'}}{(1 + e^{-\delta'})^2},$$

which is equivalent to the untransformed parameter  $\delta \sim \text{Unif}(0, \Delta)$ .

The components of  $\boldsymbol{\theta}'_\alpha^{(d)}$  are chosen to be apriori independent, that is

$$\pi_g(\boldsymbol{\theta}'_\alpha^{(d)}) = \pi_{g_1}([\boldsymbol{\lambda}'_G \ \alpha']^\top) \pi_{g_2}(\boldsymbol{\nu}'_X) \pi_{g_3}(\delta'), \quad (4.15)$$

the product of priors of its individual components.

#### 4.4.1 Update moves

We now describe the Metropolis-Hastings step suitable for updating the parameter vector  $\boldsymbol{\theta}'_\alpha^{(d)}$ . Specifically, let  $\bar{\boldsymbol{\theta}}_\alpha^{(d)}$  be the current value of the *transformed* parameter vector and let  $\bar{C}$  be the current set of parent locations from the RFS  $C_{N_F}$  with  $M$  centres. A *proposal*  $\boldsymbol{\theta}'_\alpha^{(d)}$  is then generated via

$$\boldsymbol{\theta}'_\alpha^{(d)} \sim \mathcal{N}(\bar{\boldsymbol{\theta}}_\alpha^{(d)}, \tilde{\Sigma}_u),$$

where  $\tilde{\Sigma}_u \in \mathbb{R}^{(3(d+2)+|A|) \times (3(d+2)+|A|)}$  is the proposal covariance matrix. For the first 100 iterations of the sampler,  $\tilde{\Sigma}_u = \frac{(0.1)^2}{(3(d+2)+|A|)} I_{(3(d+2)+|A|)}$ . For the



remaining iterations,  $\tilde{\Sigma}_u = \frac{(2.38)^2}{(3(d+2)+|A|)}\hat{\Sigma} + 10^{-3}\frac{(0.1)^2}{(3(d+2)+|A|)}I_{(3(d+2)+|A|)}$ , with  $\hat{\Sigma}$  being the current *empirical* estimate of the covariance structure of the (transformed) parameters  $\boldsymbol{\theta}'_\alpha^{(d)}$ .

This proposal is then accepted with probability  $\min(1, h_u)$ , where the *update hastings ratio*  $h_u$  follows

$$h_u = \frac{f_{C_{N_F}|Z^{(N_F)}, \Theta_\alpha^{(d)}}(\bar{C}|Z^{(N_F)}, g^{-1}(\boldsymbol{\theta}'_\alpha^{(d)}), M)\pi_g(\boldsymbol{\theta}'_\alpha^{(d)})}{f_{C_{N_F}|Z^{(N_F)}, \Theta_\alpha^{(d)}}(\bar{C}|Z^{(N_F)}, g^{-1}(\bar{\boldsymbol{\theta}}_\alpha^{(d)}), M)\pi_g(\bar{\boldsymbol{\theta}}_\alpha^{(d)})}, \quad (4.16)$$

and where  $f_{C_{N_F}|Z^{(N_F)}, \Theta_\alpha^{(d)}}(\bar{C}|Z^{(N_F)}, g^{-1}(\cdot), M)$  can be evaluated from (4.14).

It should be noted here that since we are generating proposals from the multivariate Normal distribution, the proposal densities cancel in the Hastings' ratio to reach (4.16).

#### 4.4.2 Shift moves

Here, we define the Metropolis-Hastings step which is taken from Møller and Waagepetersen (2003), suitable for updating cluster centres of the parent RFS  $C_{N_F}$ . Let  $\bar{C} = \{\bar{\mathbf{c}}_1, \bar{\mathbf{x}}_2, \dots, \bar{\mathbf{c}}_M\}$  be the current configuration of the parent RFS with  $M$  clusters, and let  $\bar{\boldsymbol{\theta}}_\alpha^{(d)}$  be the current value of the *transformed* parameter vector. The location of a single cluster in  $\bar{C}$  is now shifted randomly. Specifically, an index  $I$  is chosen uniformly from the set  $\{1, 2, \dots, M\}$  and

$$\mathbf{c}' \sim \mathcal{N}(\bar{\mathbf{c}}_I, \tilde{\Sigma}_s)$$

is sampled. Here,  $\tilde{\Sigma}_s \in \mathbb{R}^{2 \times 2}$  denotes a user defined proposal covariance matrix. For this problem, we let  $\tilde{\Sigma}_s = c_{\Sigma_s}I_2$ , where the constant  $c_{\Sigma_s} > 0$  is appropriately chosen to achieve an acceptance rate of around 23%.

The proposal  $C' = \{\bar{\mathbf{c}}_1, \dots, \bar{\mathbf{c}}_{I-1}, \mathbf{c}', \bar{\mathbf{c}}_{I+1}, \dots, \bar{\mathbf{c}}_M\}$  is then accepted with probability  $\min(1, h_s)$ , where the *shift hastings ratio*  $h_s$  follows

$$h_s = \frac{f_{Z_{N_F}|C_{N_F}, \Theta_\alpha^{(d)}}(Z_{N_F}|C', g^{-1}(\bar{\boldsymbol{\theta}}_\alpha^{(d)}), M)f_{C_{N_F}|Z^{(N_F-1)}, \Theta_\alpha^{(d)}}(C'|Z^{(N_F-1)}, g^{-1}(\bar{\boldsymbol{\theta}}_\alpha^{(d)}), M)}{f_{Z_{N_F}|C_{N_F}, \Theta_\alpha^{(d)}}(Z_{N_F}|\bar{C}, g^{-1}(\bar{\boldsymbol{\theta}}_\alpha^{(d)}), M)f_{C_{N_F}|Z^{(N_F-1)}, \Theta_\alpha^{(d)}}(\bar{C}|Z^{(N_F-1)}, g^{-1}(\bar{\boldsymbol{\theta}}_\alpha^{(d)}), M)},$$

which can be computed using (4.10) and (4.11).

An algorithm detailing our MwG MCMC method to output  $N_{MC}$  posterior samples from  $C_{N_F}$  and the (transformed) parameter vector  $\boldsymbol{\theta}'_{\alpha}^{(d)}$ , is given in Algorithm 6.

---

**Algorithm 6** Metropolis-within-Gibbs algorithm to obtain  $N_{MC}$  samples from the posterior  $f_{C_{N_F}, \Theta_\alpha^{(d)} | Z^{(N_F)}}(C_{N_F}, \boldsymbol{\theta}'^{(d)} | Z^{(N_F)}, M)$  under model  $\mathcal{M}_A^d$

---

**function** RUN\_MWG( $Z^{(N_F)}, \Delta, N_{MC}, M, A, d, \mathcal{C}, \tilde{\Sigma}_s$ )

$\boldsymbol{\theta}'^{(d)} \leftarrow \mathbf{0}_{N_{MC}} \mathbf{0}_{3(d+2)+|A|}^\top$

$C \leftarrow \mathbf{0}_{M \times 2 \times N_{MC}}$   $\triangleright$  Array of zeros in  $\mathbb{R}^{M \times 2 \times N_{MC}}$ , with  $i$ th slice  $\mathbf{0}_M \mathbf{0}_2^\top$

Sample  $\bar{C} \leftarrow \text{Unif}_M(\mathcal{C})$   $\triangleright$  Initialise  $M$  uniform points on  $\mathcal{C}$

$C[:, :, 1] \leftarrow \bar{C}$

Sample  $\boldsymbol{\theta}' \sim \pi_g(\boldsymbol{\theta}'^{(d)})$   $\triangleright$  Initialise from prior given in (4.15)

$\boldsymbol{\theta}'^{(d)}[:, 1] \leftarrow \boldsymbol{\theta}'$

$\tilde{\Sigma}_u \leftarrow \frac{(0.1)^2}{(3(d+2)+|A|)} I_{(3(d+2)+|A|)}$   $\triangleright$  Initialise update covariance

**for**  $i = 2$  to  $N_{MC}$  **do**

$C' \leftarrow C[:, :, i-1]$   $\triangleright$  //Shift pattern  $C$

$l \sim \text{Unif}(\{1, \dots, M\})$   $\triangleright$  Sample index from set  $\{1, \dots, M\}$

Sample  $\mathbf{c}' \sim \mathcal{N}(C[l, :, i-1], \tilde{\Sigma}_s)$   $\triangleright$  Propose shift

$C'[l, :] \leftarrow \mathbf{c}'$

$h_s \leftarrow \frac{f_{C_{N_F} | Z^{(N_F)}, \Theta_\alpha^{(d)}}(C' | Z^{(N_F)}, g^{-1}(\boldsymbol{\theta}'^{(d)}[:, i-1]), M)}{f_{C_{N_F} | Z^{(N_F)}, \Theta_\alpha^{(d)}}(C[:, :, i-1] | Z^{(N_F)}, g^{-1}(\boldsymbol{\theta}'^{(d)}[:, 1]), M)}$

Sample  $U \sim \text{Unif}(0, 1)$

**if**  $U < h_s$  **then**

$C[:, :, i] \leftarrow C'$   $\triangleright$  Accept shift

**else**

$C[:, :, i] \leftarrow C[:, :, i-1]$   $\triangleright$  Reject shift

**if**  $i \geq 100$  **then**  $\triangleright$  //Update  $\boldsymbol{\theta}'^{(d)}$

$\hat{\Sigma} \leftarrow \text{Cov}(\boldsymbol{\theta}'^{(d)}[:, 1 : i-1])$   $\triangleright$  Compute current covariance

$\tilde{\Sigma}_u \leftarrow \frac{(2.38)^2}{(3(d+2)+|A|)} \hat{\Sigma} + 10^{-3} \frac{(0.1)^2}{(3(d+2)+|A|)} I_{(3(d+2)+|A|)}$

Sample  $\boldsymbol{\theta}' \sim \mathcal{N}(\boldsymbol{\theta}'^{(d)}[:, i-1], \tilde{\Sigma}_u)$   $\triangleright$  Propose update

$h_u \leftarrow \frac{f_{C_{N_F} | Z^{(N_F)}, \Theta_\alpha^{(d)}}(C[:, :, i] | Z^{(N_F)}, g^{-1}(\boldsymbol{\theta}'), M) \pi_g(\boldsymbol{\theta}')}{f_{C_{N_F} | Z^{(N_F)}, \Theta_\alpha^{(d)}}(C[:, :, i] | Z^{(N_F)}, g^{-1}(\boldsymbol{\theta}'^{(d)}[:, i-1]), M) \pi_g(\boldsymbol{\theta}'^{(d)}[:, i-1])}$

Sample  $U \sim \text{Unif}(0, 1)$

**if**  $U < h_u$  **then**

$\boldsymbol{\theta}'^{(d)}[:, i] \leftarrow \boldsymbol{\theta}'$   $\triangleright$  Accept update

**else**

$\boldsymbol{\theta}'^{(d)}[:, i] \leftarrow \boldsymbol{\theta}'^{(d)}[:, i-1]$   $\triangleright$  Reject update

**return**  $C, \boldsymbol{\theta}'^{(d)}$

---

### 4.4.3 Parameter estimation

After running the MCMC algorithm, each chain will output  $N_{MC}$  samples of the parent RFS  $C_{N_F}$  and the transformed parameter vector  $\boldsymbol{\theta}'_{\alpha}(d)$ . Maximum a posteriori (MAP) values for the components of  $\hat{\boldsymbol{\theta}}'_{\alpha}(d)$  can thereby be determined via the modes of the corresponding posterior distributions. Moreover, a  $(1 - \alpha_C)\%$  credible interval for the  $i$ th component  $(\boldsymbol{\theta}'_{\alpha}(d))_i$  can be calculated as (Chen and Shao, 1998)

$$[(\boldsymbol{\theta}'_{\alpha}(d))_i^{(\alpha_C/2N_{MC})} \quad (\boldsymbol{\theta}'_{\alpha}(d))_i^{((1-\alpha_C/2)N_{MC})}],$$

where  $(\boldsymbol{\theta}'_{\alpha}(d))_i^{(\alpha_C/2N_{MC})}$  and  $(\boldsymbol{\theta}'_{\alpha}(d))_i^{((1-\alpha_C/2)N_{MC})}$  denote the  $(\alpha_C/2)N_{MC}$ th and  $(1 - \alpha_C/2)N_{MC}$ th ordered samples of  $(\boldsymbol{\theta}'_{\alpha}(d))_i$ , respectively. The MAP values and credible intervals for each component can then be mapped back to their untransformed versions through the function  $g^{-1}$ .

Modal estimates of parent locations  $\hat{C}_{N_F}$  can be obtained from finding the  $M$  peaks of a kernel smoothed density applied on the posterior samples of  $C_{N_F}$ . We note that credible intervals are not well defined here since  $C_{N_F}$  represents a random set, and therefore has no ordering of its points.

### 4.4.4 Model selection

After several MCMC runs have been conducted under different models parameterised by  $M$  and  $\mathcal{M}_A^d$ , using samples obtained from the respective posterior distributions, model selection can be conducted to select the most likely model.

There exist many methods in the Bayesian model selection literature (Newton and Raftery, 1994, Chib, 1995, Chib and Jeliazkov, 2001) that attempt to utilise the  $N_{MC}$  posterior MCMC samples  $(C_{N_F}, \boldsymbol{\theta}'_{\alpha}(d))^{(1)}, \dots, (C_{N_F}, \boldsymbol{\theta}'_{\alpha}(d))^{(N_{MC})}$

to approximate the *model evidence* or marginal likelihood

$$f(Z^{(N_F)}|M, \mathcal{M}_A^d) = \int_{\mathcal{C}^M \times \Theta_\alpha^{(d)}} f(Z^{(N_F)}|C_{N_F} = \{\mathbf{c}_1, \dots, \mathbf{c}_M\}, \boldsymbol{\theta}_\alpha^{(d)}, M, \mathcal{M}_A^d) \\ \times \pi(C_{N_F} = \{\mathbf{c}_1, \dots, \mathbf{c}_M\}, \boldsymbol{\theta}_\alpha^{(d)}) d\mathbf{c}_1 \dots d\mathbf{c}_M d\boldsymbol{\theta}_\alpha^{(d)},$$

required to compute the posterior model probabilities  $\mathbb{P}(M, \mathcal{M}_A^d|Z^{(N_F)}) \propto f(Z^{(N_F)}|M, \mathcal{M}_A^d)\mathbb{P}(M, \mathcal{M}_A^d)$ . However, the majority of these methods are difficult to apply in our setting. For example, the method of Chib (1995) is not applicable since it requires the MCMC samples to be generated from a Gibbs sampler, with prior and posterior distributions to be known in closed form. Nevertheless, while its extension, as derived in Chib and Jeliazkov (2001), to Metropolis-Hastings type samplers can be used here, computing the estimate is computational intensive and can exhibit high variance. On the other hand, while the method described in Newton and Raftery (1994) is easy to implement, the resulting estimator suffers from infinite variance and is therefore rarely used in practice.

For this problem, we therefore choose to use the Deviance Information Criterion (DIC) which is defined through the *deviance* function (Spiegelhalter et al., 2002)

$$D(\boldsymbol{\theta}_\alpha^{(d)}, C_{N_F}, M) = -2 \log f_{C_{N_F}, \Theta_\alpha^{(d)}|Z^{(N_F)}}(C_{N_F}, \boldsymbol{\theta}_\alpha^{(d)}|Z^{(N_F)}, M).$$

Here, we define the DIC as

$$DIC = \overline{2D(\boldsymbol{\theta}_\alpha^{(d)}, C_{N_F}, M)} - D(\hat{\boldsymbol{\theta}}_\alpha^{(d)}, \hat{C}_{N_F}, M),$$

where  $\hat{\boldsymbol{\theta}}_\alpha^{(d)}$  denotes the modal value of the parameter vector  $\boldsymbol{\theta}_\alpha^{(d)}$ ,  $\hat{C}_{N_F}$  denotes the modal parent locations from the RFS with  $M$  centres, and  $\overline{D(\cdot)}$  denotes the mean deviance. Similar to the model-robust version of the AIC, the DIC measures a model's goodness of fit and penalises over-fitting. Note that this criterion is particularly useful since the first expression can be easily calculated as the average deviance over the MCMC samples, and the second as the deviance evaluated at the MAP estimates of these samples. Similarly to the BIC, the model with the smallest DIC value is favoured.

## 4.5 Simulations

In this section, we present the results of three simulated dSTORM experiments that have been conducted to assess the performance of the MCMC algorithm described previously. For each dataset, several MCMC chains were run over different values of  $M$  and  $d$ , and the DIC applied to select the most likely model. For the selected model, the output of the simulations are presented through posterior densities, where 95% credible intervals and posterior modes are also noted.

In order to fully control and account for the random number of false positive observations in the model, the simulations in this section were conducted under a simplified procedure to that presented in Section 1.F of Appendix 1.7. This procedure is described in Section 4.B of Appendix 4, and requires simulation of the observed process  $\{Y_n\}$  for each of the  $M$  molecules imaged. It should be noted here that we held out the observation traces  $\mathcal{Y}$  for each simulated dataset, in order to compare the PSHMM rate estimation procedure described in Chapter 1 with that outputted by the MCMC algorithm developed in this chapter.

The first simulation was conducted under the ( $d = 0$ )  $\mathcal{M}_{\{1\}}^0$  model with  $\mathcal{C} = [0.1, 0.7] \times [0.25, 0.7]$ , the second under the ( $d = 1$ )  $\mathcal{M}_{\{1\}}^1$  model with  $\mathcal{C} = [0.3, 0.8] \times [0.3, 0.8]$  and the third under the ( $d = 2$ )  $\mathcal{M}_{\{1\}}^2$  model with  $\mathcal{C} = [0.2, 0.8] \times [0.3, 0.8]$ . The unit of measurement used was nanometres (nm)  $\times 10^4$ . The global parameter values used for the three simulations are presented in Table 4.1. In particular, to align for (d)STORM experimental conditions,  $\nu_1^* = 1$ , enabling the form of the more computationally efficient density in (4.13) of Remark 12 to be used in the MCMC algorithm.

In all three studies, MCMC chains were tested for values of  $M \in \{3, \dots, 10\}$  over three photo-switching models  $\mathcal{M}_{\{1\}}^0$ ,  $\mathcal{M}_{\{1\}}^1$  and  $\mathcal{M}_{\{1\}}^2$ . Each algorithm for was run for  $3 \times 10^5$  iterations, with a burn-in period of  $3 \times 10^4$  iterations. Moreover, in order to reduce the autocorrelation in the output, we sub-sampled each chain at every 20th iteration. These sub-samples form the basis of the (posterior) distributional summaries we will present for each

study.

Parameter	$d$	$\lambda_{001}$ ( $s^{-1}$ )	$\lambda_{01}$ ( $s^{-1}$ )	$\lambda_{010_2}$ ( $s^{-1}$ )	$\lambda_{011}$ ( $s^{-1}$ )	$\lambda_{021}$ ( $s^{-1}$ )	$\lambda_{10}$ ( $s^{-1}$ )	$\mu_1$ ( $s^{-1}$ )	$\Delta^{-1}$ ( $s^{-1}$ )	$\delta^{-1}$ ( $s^{-1}$ )	$\lambda_M$	$T^*$ (s)	$M$	$N_F$
Study														
1	0		1				3.162	0.001	30	300	10	$10\Delta$	7	$10^4$
2	1	0.35	1		0.3		2.3	0.001	30	300	10	$25\Delta$	8	$10^4$
3	2	0.35	1	0.1	0.3	0.1	3.162	0.001	30	300	10	$50\Delta$	6	$10^4$

**Table 4.1:** Global parameter values for the stimulation studies conducted in this section. All studies have been conducted with the model format of  $M_{\{1\}}^d$  for  $d = 0, 1, 2$ .

For each simulation study 1-3, the outputted DIC values are shown in Tables 4.6-4.8 in Section 4.C of Appendix 4, with the predicted model highlighted in red. Remarkably, it was able to identify the exact model under simulation for all three studies, providing us with primary evidence of the power this method has in estimating both the number of images molecules  $M$  and the number of multiple dark states  $d$  under observation.

The posterior intensity maps in Figures 4.5a, 4.6a and 4.7a show the distribution of samples gained from the respective posterior distributions for each study 1-3. The coordinate estimates (after applying kernel smoothing) are plotted against true positions and observation sets in Figures 4.5b, 4.6b and 4.7b. For most of the molecules, it is seen that the sampler is able to accurately estimate their exact positions. However, when molecules are very spatially close to one another, as is seen in studies 1 and 2, the posterior intensities have the tendency to overlap within the neighbouring region, causing the resulting spatial estimates to be slightly biased.

Under study 1, trace plots (without sub-sampling) for the transformed photo-switching parameters  $\log \lambda_{01}$ ,  $\log \lambda_{10}$ ,  $\log \mu_1$ ,  $-\log(\Delta/\delta - 1)$  and noise parameter  $\log \alpha$  is shown in Figures 4.8 and 4.9. The fast mixing about the posterior modes that can be observed in these plots is likely due to the inclusion of the adaptive step in the corresponding update moves of the algorithm. Furthermore, the posterior distributions of the transformed parameter  $\theta_\alpha^{(d)}$  are shown in Figures 4.10 and 4.11. Here, MAP values and 95% credible bounds are presented for each transformed parameter against the true simulated values. It is interesting to observe that posterior samples for all parameters but

$\mu_1$  centre close to the true parameter values and are well within the credible intervals. The resulting MAP estimate for  $\mu_1$  is slightly biased and its 95% credible interval does not contain the true value. The specific MAP values and 95% credible intervals of all untransformed parameters, together with rate estimates gained from the PSHMM are presented in Table 4.2, and generally show agreement between both estimators.

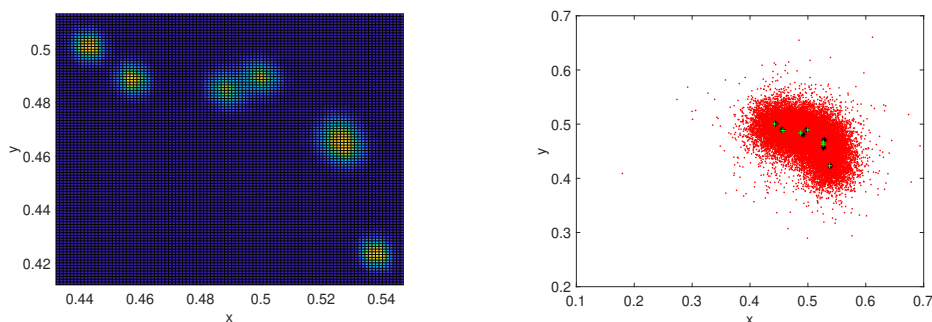
Under study 2, the posterior distributions of  $\theta'_\alpha^{(d)}$  are shown in Figures 4.12, 4.13 and 4.14. Here, it is again observed that the MAP estimates for all transformed photo-switching parameters  $\log \lambda_{00_1}$ ,  $\log \lambda_{0_1}$ ,  $\log \lambda_{0_1_1}$ ,  $\log \lambda_{10}$  are able to recover the underlying rates, with all parameters' credible intervals containing the true values (also shown in Table 4.3). In this study, however, the MAP estimates for  $\log \lambda_{00_1}$  and  $\log \lambda_{0_1_1}$  are seen to be more biased than those of  $\log \lambda_{0_1}$  and  $\log \lambda_{10}$ . Table 4.3 also shows that the PSHMM estimator is also more biased for these estimates; this is especially highlighted for the estimated value of  $\lambda_{00_1}$ , which is seen more poorly estimated by the PSHMM. Using the analysis presented in Chapter 2, the most likely reason for the bias exhibited by both methods is the small sample size (determined by  $M$ ) of the data.

Under study 3, the posterior distributions of  $\theta'_\alpha^{(d)}$  are shown in Figures 4.15-4.18, and (untransformed) parameter estimates in Table 4.4. While the algorithm is seen to perform well in estimating the parameters  $\lambda_{0_1}$ ,  $\lambda_{0_1_1}$ ,  $\lambda_{10}$ , MAP estimates for  $\log \lambda_{00_1}$ ,  $\log \lambda_{0_1_0_2}$ ,  $\log \lambda_{0_2_1}$  are poorer and are not contained in the respective credible intervals. Furthermore, the posterior variance under these parameters is higher. This is highlighted most along the directions of  $\lambda_{00_1}$  and  $\lambda_{0_1_1}$ . Although invoking proper priors enables posterior modes to be identified, these observations indicate that the resulting MAP values are not wholly indicative of the true parameter values. While most rate estimates (especially for  $\lambda_{0_1_0_2}$ ) under the PSHMM seem less biased, this method still struggles in properly identifying  $\lambda_{00_1}$ , which aligns with the results presented from study 2. We may conclude that a dataset of this size ( $M = 6$ ) is not sufficiently large to produce more meaningful estimates. Insufficiently large data was also shown as the reason for the identification issues of  $\lambda_{0_1_0_2}$  in Chapter 2. We would therefore expect to see (under these parameter val-



ues) sharper peaked and lower variance posteriors given datasets consisting of more photo-switching fluorophores.

Under all studies, with the exception of  $\mu_1$  under study 1 as previously noted, the posterior distributions and credible intervals of the parameters  $\log \mu$  and  $\log \alpha$  indicate accurate estimation of the true values. Furthermore,  $-\log(\Delta/\delta - 1)$  is also seen to be well estimated, albeit producing a higher variance posterior. Interestingly, the sampler being at higher values of  $\delta$  may also justify the larger posterior variance of the photo-switching parameters in studies 2 and 3, in that these parameters could be forced into regions of the parameter space which may otherwise not be visited, thereby corroborating the correlation and identifiability analysis presented in Chapter 2. This may, for example, be owed to the choice of Uniform prior invoked on  $\delta$ , of which a change in prior would lead to a different posterior being observed.



(a) Posterior intensity map.

(b) Estimates against true positions.

**Figure 4.5:** Posterior spatial analysis for study 1. 4.5a: Posterior intensity map of MCMC samples. 4.5b: Coordinate estimates from the Bayes filter (red crosses) plotted with the true positions (black stars) and the superimposition of offspring observation sets and false positive observations (red dots). Unit of measurement is  $\text{nm} \times 10^4$ .

Parameter	True	PSHMM MLE	MAP	95% Credible Interval
$\lambda_{01}$	1	1.0234	1.0226	(0.7496, 1.5338)
$\lambda_{10}$	3.162	3.1710	2.4509	(1.691, 3.8339)
$\mu_1$	0.001	0.0034	0.0006	(0.0003, 0.0009)
$\delta_s$	0.0033	0.0026	0.0094	(0.0003, 0.0305)
$\alpha$	0.001	Not estimated	$8.74 \times 10^{-4}$	(0.0004, 0.0016)
$\nu_0$	Unknown	0.1264	0.0280	(0.0062, 0.1413)
$\nu_1$	Unknown	0.8736	0.8925	(0.7303, 0.9604)
$\nu_2$	Unknown	0	0.0795	(0.0245, 0.1784)

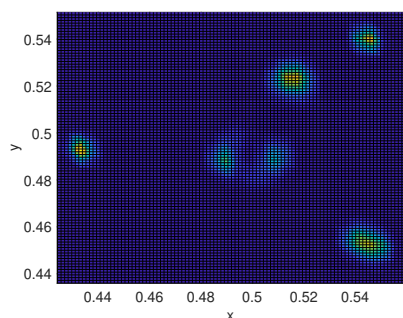
**Table 4.2:** Parameter estimates and 95% credible intervals under study 1.

Parameter	True	PSHMM MLE	MAP	95% Credible Interval
$\lambda_{001}$	0.35	4.5493	0.8970	(0.1308, 6.2179)
$\lambda_{01}$	1	1.4480	0.9359	(0.1646, 4.4325)
$\lambda_{0_11}$	0.3	0.5312	0.9675	(0.0490, 4.3652)
$\lambda_{10}$	2.3	2.2103	1.8110	(0.2862, 7.4025)
$\mu_1$	0.001	0.0021	0.0008	(0.0005, 0.0012)
$\delta$	0.0033	0.0003	0.0110	(0.0005, 0.0317)
$\alpha$	0.001	Not estimated	0.0010	(0.0005, 0.0018)
$\nu_0$	Unknown	0.1587	0.0808	(0.0112, 0.2538)
$\nu_{0_1}$	Unknown	0.2636	0.1182	(0.0272, 0.2931)
$\nu_1$	Unknown	0.4937	0.5659	(0.3925, 0.7500)
$\nu_2$	Unknown	0	0.0839	(0.0155, 0.1130)

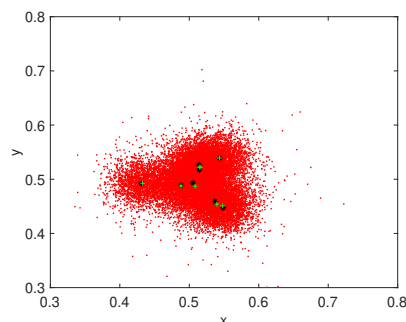
**Table 4.3:** Parameter estimates and 95% credible intervals under study 2.

Parameter	True	PSHMM MLE	MAP	95% Credible Interval
$\lambda_{001}$	0.35	2.6383	1.2883	(0.4852, 9.3418)
$\lambda_{01}$	1	1.4907	1.6139	(0.2178, 5.4257)
$\lambda_{0_10_2}$	0.1	0.2557	1.3724	(0.2356, 6.2320)
$\lambda_{0_11}$	0.3	0.8468	1.0663	(0.1094, 3.8521)
$\lambda_{0_21}$	0.1	0.1401	0.1477	(0.1091, 0.2722)
$\lambda_{10}$	3.162	2.9065	1.4605	(0.7364, 5.3808)
$\mu_1$	0.001	0.0019	0.0010	(0.0005, 0.0014)
$\delta$	0.0033	0.0199	0.0114	(0.0011, 0.0326)
$\alpha$	0.001	Not estimated	$7.66 \times 10^{-4}$	(0.0003, 0.0014)
$\nu_0$	Unknown	0.1561	0.0778	(0.0212, 0.3755)
$\nu_{0_1}$	Unknown	0.1819	0.1802	(0.0126, 0.3107)
$\nu_{0_2}$	Unknown	0.2076	0.2328	(0.0180, 0.2915)
$\nu_1$	Unknown	0.4544	0.4427	(0.3230, 0.7025)
$\nu_2$	Unknown	0	0.0665	(0.0517, 0.2167)

**Table 4.4:** Parameter estimates and 95% credible intervals under study 3.

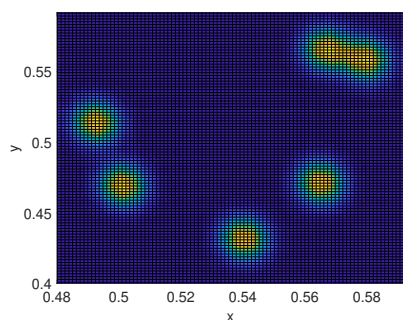


(a) Posterior intensity map.

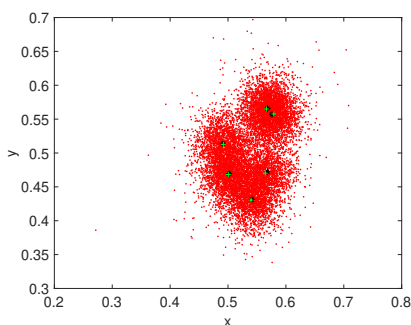


(b) Estimates against true positions.

**Figure 4.6:** Posterior spatial analysis for study 2. 4.6a: Posterior intensity map of MCMC samples. 4.6b: Coordinate estimates from the Bayes filter (red crosses) plotted with the true positions (black stars) and the superimposition of offspring observation sets and false positive observations (red dots). Unit of measurement is  $\text{nm} \times 10^4$ .



(a) Posterior intensity map.

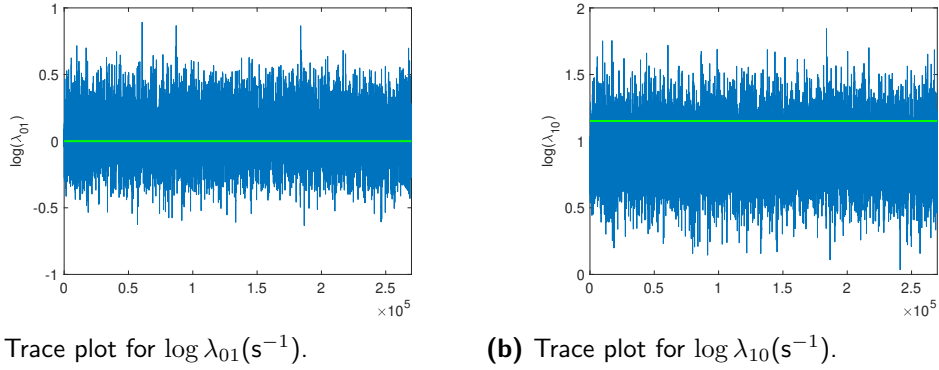


(b) Estimates against true positions.

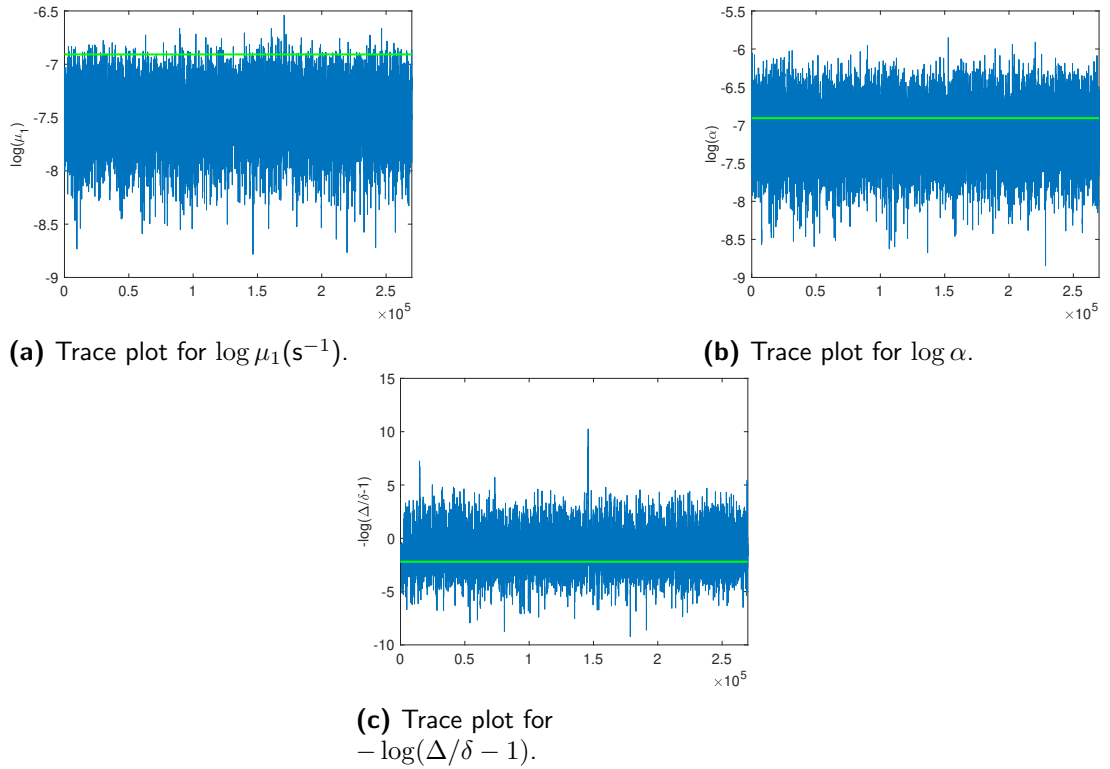
**Figure 4.7:** Posterior spatial analysis for study 3. 4.7a: Posterior intensity map of MCMC samples. 4.7b: Coordinate estimates from the Bayes filter (red crosses) plotted with the true positions (black stars) and the superimposition of offspring observation sets and false positive observations (red dots). Unit of measurement is  $\text{nm} \times 10^4$ .

## 4.6 Application to Alexa Fluor 488 data

In this section, we apply the MCMC method presented in this chapter to a dSTORM experiment imaging Alexa Fluor 488 fluorophores. Here, we select two small subsets of the overall (superimposed) image and apply the MCMC algorithm described by testing different models which vary the cardinality of molecules  $M$  and the number of multiple dark states  $d$  of the underlying photo-switching process. We present the spatial posterior locations of the estimated positions of fluorophores, and MAP values (with 95% credible



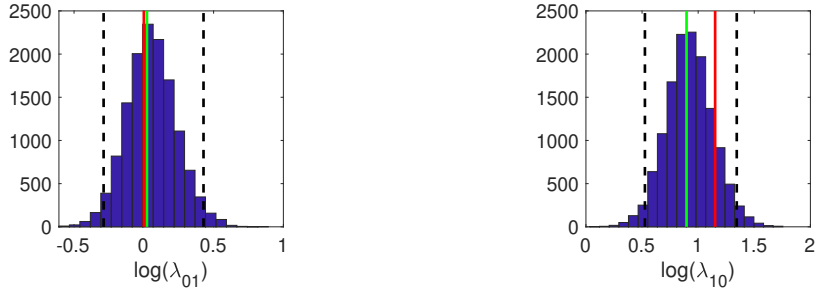
**Figure 4.8:** Trace plots of the transformed photo-switching rates  $\log \lambda_{01}$ ,  $\log \lambda_{10}$  under study 1 with true values in green.



**Figure 4.9:** Trace plots of the transformed parameters  $\log \mu_1$ ,  $\log \alpha$ ,  $-\log(\Delta/\delta - 1)$  under study 1 with true values in green.

intervals) of  $\theta_\alpha^{(d)}$  under the photo-kinetic model estimated for each dataset.

In this experiment, chromosomes labelled with Alexa Fluor 488 fluorophores embedded in Polyvinyl Alcohol (PVA) resin were imaged over  $10^4$  frames, at



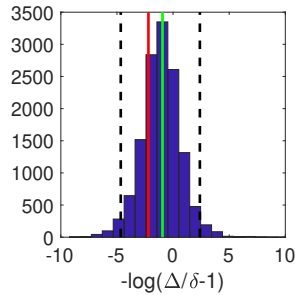
(a) Posterior histogram for  $\log \lambda_{01}(\text{s}^{-1})$ . (b) Posterior histogram for  $\log \lambda_{10}(\text{s}^{-1})$ .

**Figure 4.10:** Posterior histograms of transformed photo-switching rates  $\log \lambda_{01}$ ,  $\log \lambda_{10}$  under study 1, with MAP values (green), true values (red) and 95% credible intervals (black, dotted).



(a) Posterior histogram for  $\log \mu_1(\text{s}^{-1})$ .

(b) Posterior histogram for  $\log \alpha$ .

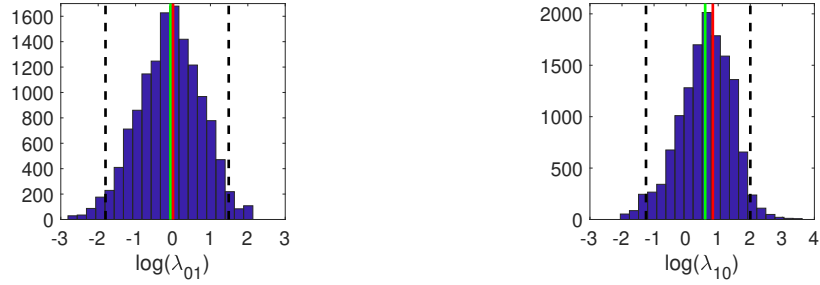


(c) Posterior histogram for  $-\log(\Delta/\delta - 1)$ .

**Figure 4.11:** Posterior histograms of the transformed parameters  $\log \mu_1$ ,  $\log \alpha$ ,  $-\log(\Delta/\delta - 1)$  under study 1, with MAP values (green), true values (red) and 95% credible intervals (black, dotted).

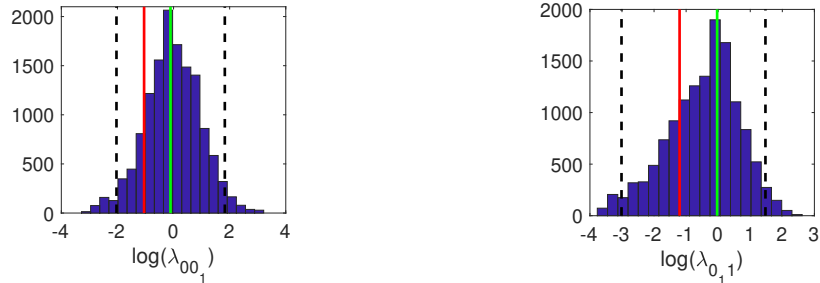
a frame acquisition time of  $\Delta = 0.03$  seconds, with dSTORM<sup>§</sup>. In each raw image obtained, the algorithm in Ovesný et al. (2014) was applied to find

<sup>§</sup>This dataset was made available for our use by Professor Paul French, Imperial College London.



(a) Posterior histogram for  $\log \lambda_{01}(\text{s}^{-1})$ . (b) Posterior histogram for  $\log \lambda_{10}(\text{s}^{-1})$ .

**Figure 4.12:** Posterior histograms of the transformed photo-switching rates  $\log \lambda_{01}$ ,  $\log \lambda_{10}$  under study 2, with MAP values (green), true values (red) and 95% credible intervals (black, dotted).

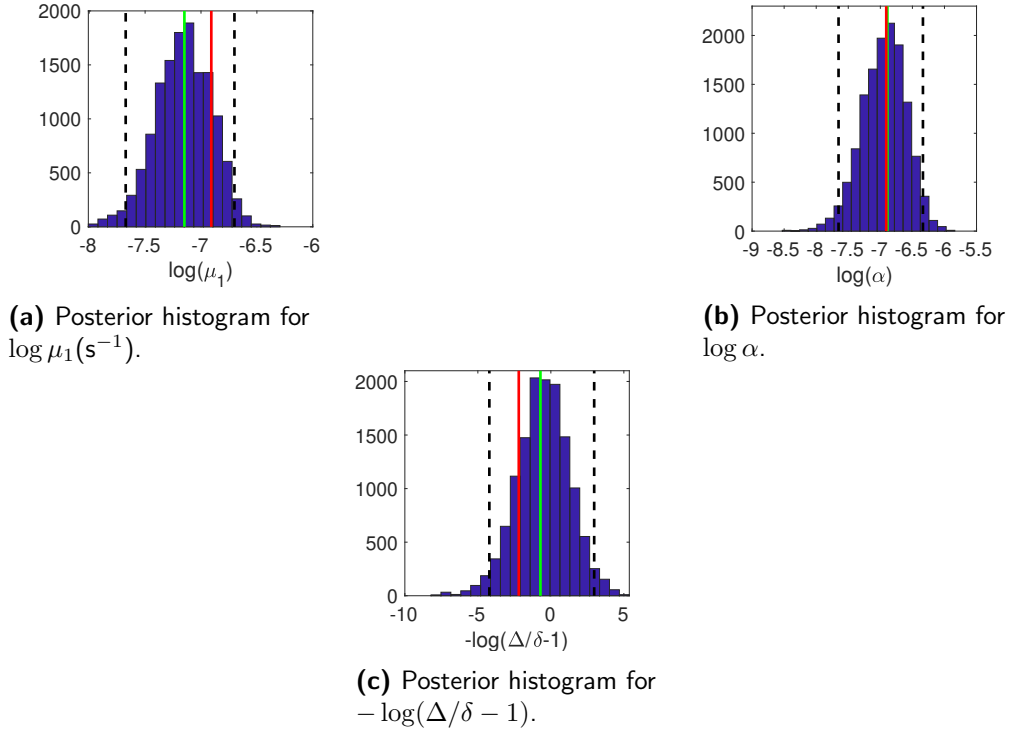


(a) Posterior histogram for  $\log \lambda_{001}(\text{s}^{-1})$ . (b) Posterior histogram for  $\log \lambda_{011}(\text{s}^{-1})$ .

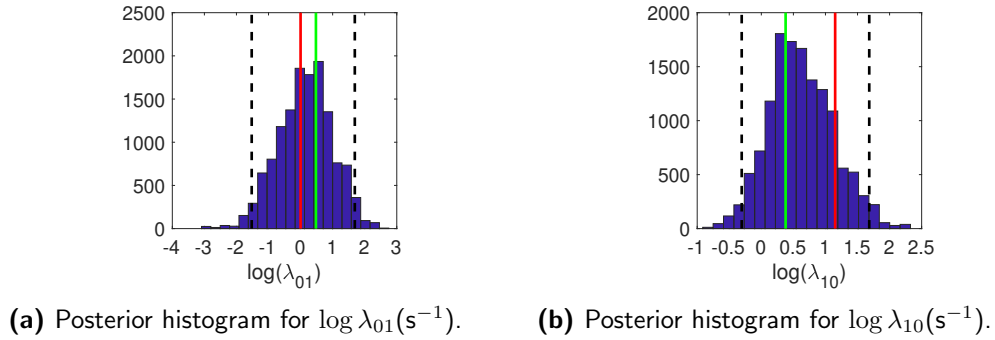
**Figure 4.13:** Posterior histograms of the transformed photo-switching rates  $\log \lambda_{001}$ ,  $\log \lambda_{011}$  under study 2, with MAP values (green), true values (red) and 95% credible intervals (black, dotted).

high photon intensity spots, and localisations determined through a Gaussian fitting algorithm. More information about the experimental set-up for this study can be found in Section 4.D.1 of Appendix 4. In this experiment, the number of molecules present, their true spatial positions and the underlying photo-kinetic model were unknown.

In order to test the MCMC algorithm constructed in this chapter, we located two subsets of the superimposed imaged (with no edge effects). These areas within the superimposed image of localisations derived from this experiment are presented in Figure 4.19. For each of the two datasets, we ran several MCMC algorithms (in parallel) over values of  $M \in \{5, \dots, 25\}$  and models  $\mathcal{M}_{\{1\}}^d$  with  $d = 0, 1, 2$ , and subsequently calculated the DIC for each model. The prior distributions for all (transformed) parameters were the same as

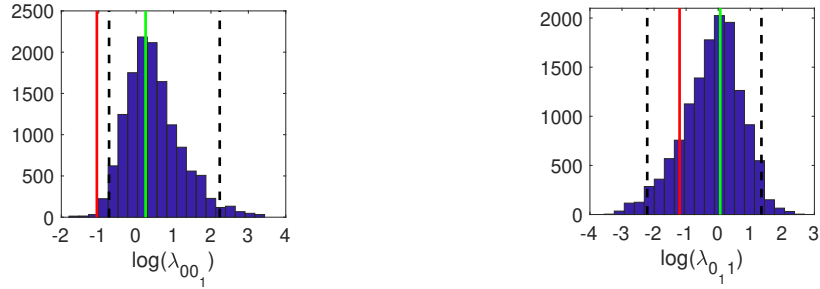


**Figure 4.14:** Posterior histograms of the transformed parameters  $\log \mu_1$ ,  $\log \alpha$ ,  $-\log(\Delta/\delta - 1)$  under study 2, with MAP values (green), true values (red) and 95% credible intervals (black, dotted).



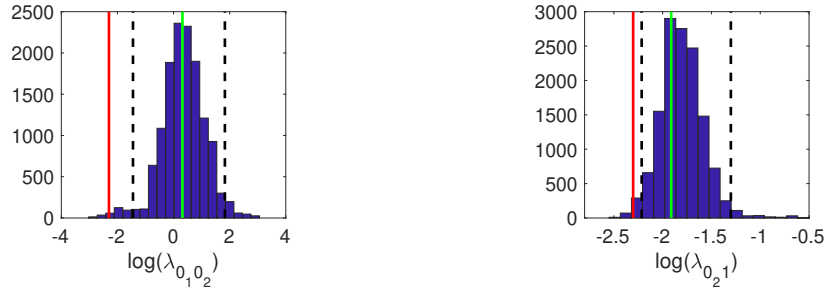
**Figure 4.15:** Posterior histograms of the transformed photo-switching rates  $\log \lambda_{01}$ ,  $\log \lambda_{10}$  under study 3, with MAP values (green), true values (red) and 95% credible intervals (black, dotted).

those described in Section 4.4, with  $\lambda_M = 10$ . Each algorithm for was run for  $3 \times 10^5$  iterations, with a burn-in period of  $3 \times 10^4$  iterations. Furthermore, we sub-sampled each chain at every 20th iteration.



(a) Posterior histogram for  $\log \lambda_{00_1} (\text{s}^{-1})$ . (b) Posterior histogram for  $\log \lambda_{01_1} (\text{s}^{-1})$ .

**Figure 4.16:** Posterior histograms of the transformed photo-switching rates  $\log \lambda_{00_1}$ ,  $\log \lambda_{01_1}$  under study 3, with MAP values (green), true values (red) and 95% credible intervals (black, dotted).



(a) Posterior histogram for  $\log \lambda_{01_0_2} (\text{s}^{-1})$ . (b) Posterior histogram for  $\log \lambda_{02_1} (\text{s}^{-1})$ .

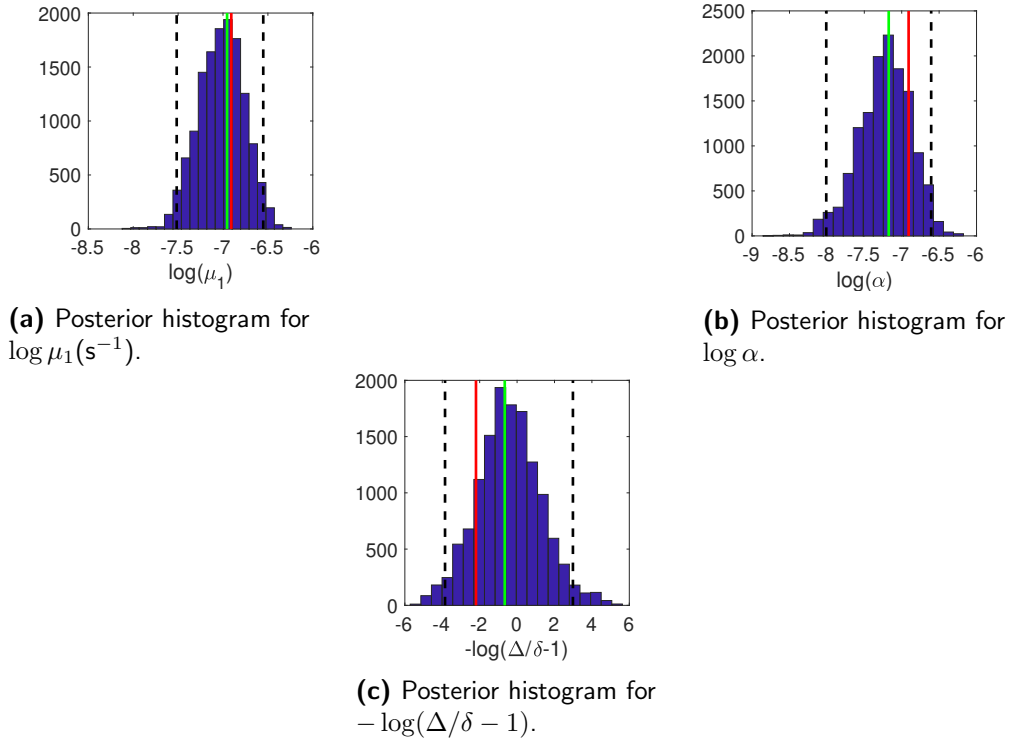
**Figure 4.17:** Posterior histograms of the transformed photo-switching rates  $\log \lambda_{01_0_2}$ ,  $\log \lambda_{02_1}$  under study 3, with MAP values (green), true values (red) and 95% credible intervals (black, dotted).

For both datasets, the algorithm predicted the  $\mathcal{M}_{\{1\}}^2$  model; this is the photo-kinetic model with three dark states and is such that the photo-bleached state is only accessible from the photo-emitting On state 1. Although the fluorophore used here is different than that considered in Chapters 1 and 3, this observation supports the analysis presented in these chapters, in that the fluorophores used in (d)STORM experiments are likely to give rise to a photo-kinetic model with three dark states.

For datasets 1 and 2, the DIC predicted  $M = 18$  and  $M = 16$  fluorophores in the respective clusters. The posterior intensity maps against estimated locations of fluorophores are depicted in Figures 4.20 and 4.21, and show clear structures of the chromosomes under observation.

The MAP values of the photo-switching rates, with their respective 95%

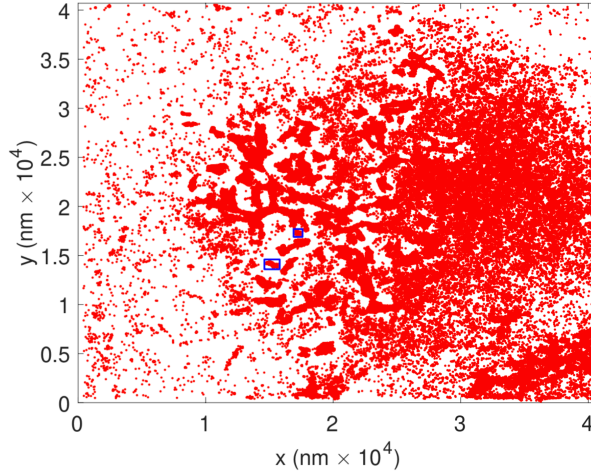




**Figure 4.18:** Posterior histograms of the transformed parameters  $\log \mu_1$ ,  $\log \alpha$ ,  $-\log(\Delta/\delta - 1)$  under study 3, with MAP values (green), true values (red) and 95% credible intervals (black, dotted).

credible intervals is shown in Table 4.5, for both datasets. Furthermore, the distributions of the transformed parameter vector  $\boldsymbol{\theta}'_{\alpha}{}^{(d)}$  are shown in Figures 4.22 - 4.30 of Section 4.D.1 in Appendix 4. It is generally seen that the posterior distributions under most parameters are similar between the two datasets. The parameters which are observed to be most different are  $\lambda_{01}$  and  $\lambda_{10}$ , with dataset 1 indicating a much higher rate for  $\lambda_{01}$ , and a lower rate for  $\lambda_{10}$ . Nevertheless, all rates are shown to be of the same order of magnitude between the two studies. In particular, the photo-switching rates determined by the MAP values indicate that  $\lambda_{10}$  is higher than  $\lambda_{01}$ , and the furthest dark state  $O_2$  is characterised by the longest lifetime. This observation aligns with the kinetic behaviour that fluorophores should undergo in STORM applications, as is dictated by the photo-physics.

In addition, the posterior variances are seen to be lower than from the same parameters studied in study 3 of Section 3.4, of which fewer molecules were



**Figure 4.19:** Superimposed localisations obtained from a dSTORM experiment observing chromosomes labelled with Alexa Fluor 488 fluorophores. Blue squares correspond to datasets 1 and 2 analysed.

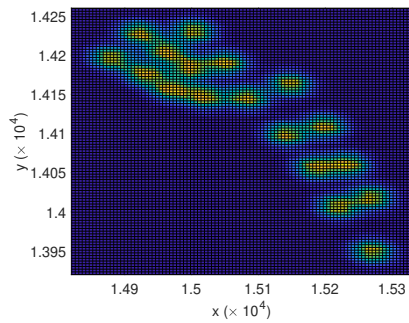
observed. This corroborates our intuition that analysing datasets containing more molecules are likely to yield lower variance parameter estimates.

Parameter	MAP dataset 1	95% Credible Interval dataset 1	MAP dataset 2	95% Credible Interval dataset 2
$\lambda_{001}$	0.0159	(0.0058, 0.0443)	0.0651	(0.0165, 0.1765)
$\lambda_{01}$	0.0796	(0.0665, 0.1564)	0.1395	(0.0447, 0.3534)
$\lambda_{0_1 0_2}$	0.2492	(0.0298, 1.0120)	0.1294	(0.0291, 0.7184)
$\lambda_{0_1 1}$	4.3836	(0.9010, 17.7112)	1.4429	(0.7667, 11.0597)
$\lambda_{0_2 1}$	0.0060	(0.0048, 0.0088)	0.0089	(0.0036, 0.0153)
$\lambda_{10}$	0.6003	(0.5798, 0.8803)	1.6030	(0.9402, 4.5359)
$\mu_1$	0.0258	(0.0197, 0.0277)	0.0144	(0.0067, 0.0250)
$\delta$	0.0155	(0.0006, 0.0292)	0.0151	(0.0010, 0.0286)
$\alpha$	0.1947	(0.1851, 0.2042)	0.1636	(0.1562, 0.1720)
$\nu_0$	0.0291	(0.0048, 0.1014)	0.1307	(0.0112, 0.2538)
$\nu_{0_1}$	0.0440	(0.0171, 0.1167)	0.1462	(0.0272, 0.2931)
$\nu_{0_2}$	0.7694	(0.5365, 0.8441)	0.5843	(0.3925, 0.7500)
$\nu_1$	0.0586	(0.0138, 0.0955)	0.0389	(0.0155, 0.1130)
$\nu_2$	0.0989	(0.0304, 0.2593)	0.0998	(0.0517, 0.2924)

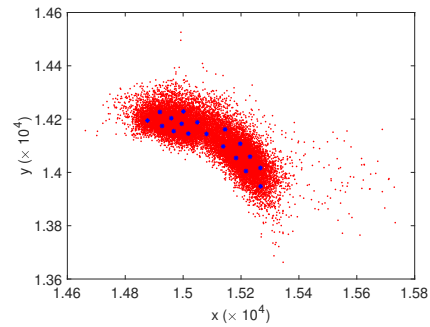
**Table 4.5:** Parameter estimates and 95% credible intervals for the Alexa Fluor 488 datasets 1 and 2.

## 4.7 Conclusions

In this chapter, we have proposed a novel spatio-temporal model describing the evolution of photo-switching fluorophores imaged in fluorescence microscopy. By using a pure birth process to model when fluorophores first reach their photo-emission states, we have formulated a state-space model

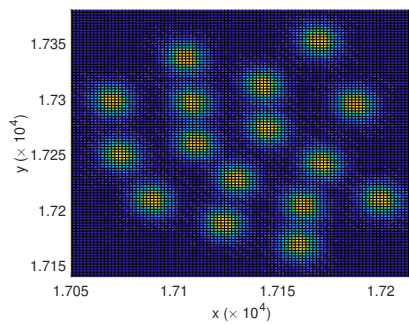


(a) Posterior intensity map.

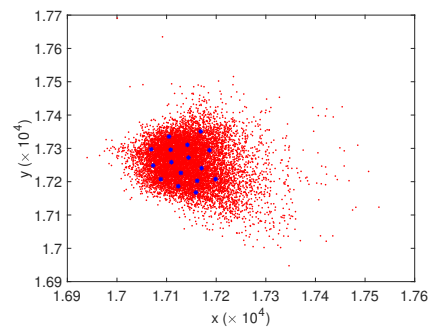


(b) Estimates against true positions.

**Figure 4.20:** Posterior spatial analysis for dataset 1 (Alexa Fluor 488). 4.20a: Posterior intensity map of MCMC samples. 4.20b: Coordinate estimates from the Bayes filter (blue crosses) plotted with the superimposed dataset. Measurement unit is nanometres.



(a) Posterior intensity map.



(b) Estimates against true positions.

**Figure 4.21:** Posterior spatial analysis for dataset 2 (Alexa Fluor 488). 4.21a: Posterior intensity map of MCMC samples. 4.21b: Coordinate estimates from the Bayes filter (blue crosses) plotted with the superimposed dataset. Measurement unit is nanometres.

which links the hidden point process of true molecular positions,  $C_n$  with observation sets collected at each time frame  $n$ . Under this model, we have derived a molecule’s birth and detection probabilities, and showed that these are functions of the unknown photo-switching parameters  $\theta^{(d)}$ . This model, furthermore, is able to account for the common scenario in which observation sets are corrupted by false positive observations. Using this, we have carefully derived the Bayes filter which is used to update the true point pattern configuration of molecules across time. Inference of the resulting posterior distribution has been made possible through the application of a Metropolis-Hastings type algorithm and the computation of the Deviance Information Criterion, which was shown to be effective in selecting the most likely cardinality of parent points  $M$  and number of dark states  $d + 1$  in a given photo-switching model. We have further shown, through simulations, that the true locations of molecules and most parameters of the underlying photo-switching parameter can, in general, be recovered well. Additionally, we have discussed, through a comparison with the PSHMM estimator, that the most likely reason for the poor estimation of other parameters (specifically those of  $\lambda_{0_11}, \lambda_{0_10_2}, \lambda_{0_11}$  in the  $d = 1, 2$  scenarios) is due to the limited number of molecules studied. An application to two subsets of a real dSTORM image dataset, consisting of more molecules than those studied in simulations, has also supported this. In both of these datasets, the algorithm determined a three dark state model of the fluorophore and the resulting posterior distributions of the photo-switching parameters approximately aligned, thereby verifying the suitability of this method to real data.

While our method cannot easily be extended to account for a greater number of fluorophores, the sequential Monte Carlo approximations discussed in the introduction of this chapter, provide a solid basis for the extension of this work to larger datasets.

## Appendix 4

### 4.A Proof of Proposition 9

In this section, we detail the proof of Proposition 9 described in Section 4.3.2.

*Proof.* Fixing  $d \geq 0$ , when the photo-bleaching state 2 is accessible from at least one of the other  $d + 2$  states of  $\{X(t)\}$ , i.e. under some model  $\mathcal{M}_A^d$  with  $|A| > 0$ , we can write  $B_{\Delta, \delta=0}^{(0)} = \begin{bmatrix} A_0 & \mathbf{d}_0 \\ \mathbf{0}_{d+2}^\top & 1 \end{bmatrix}$ , and  $B_{\Delta, \delta=0}^{(1)} = \begin{bmatrix} A_1 & \mathbf{d}_1 \\ \mathbf{0}_{d+2}^\top & 0 \end{bmatrix}$ , where for each  $j = 0, 1$ ,  $A_j = (B_{\Delta, \delta=0}^{(j)})_{(1:d+2, 1:d+2)}$  denotes the  $(d+2) \times (d+2)$  sub-matrix gained by deleting the last row and column of  $B_{\Delta, \delta=0}^{(j)}$ ,  $\mathbf{d}_j = (B_{\Delta, \delta=0}^{(j)})_{(1:d+2, d+3)}$  and  $\mathbf{0}_{d+2}$  denotes the  $(d+2) \times 1$  vector of zeros.

Since  $B_{\Delta, \delta=0}^{(0)} + B_{\Delta, \delta=0}^{(1)} = e^{G\Delta}$  is a stochastic matrix, it is easy to see that the row sums of both  $A_0$  and  $A_1$  are less than one when  $\Delta > 0^\natural$ . In particular if  $(\phi, \mathbf{v})$  denotes an eigenpair of  $A_0$ , let  $v_k$  be the entry of  $\mathbf{v}$  with maximum absolute value and  $a_{ij}$  the  $i, j$ th entry of  $A_0$ . Then  $\phi v_k = \sum_j a_{kj} v_j$  and  $|\phi v_k| \leq \sum_j |a_{kj}| |v_j| < |v_k|$  since the row sums of  $A_0$  are strictly less than one. This implies that every eigenvalue  $\phi$  of  $A_0$  satisfies  $|\phi| < 1$  and thus that  $\lim_{n \rightarrow \infty} \phi^n = 0$ .

Now let  $J$  be the Jordan-canonical form for matrix  $A_0$  composed of its  $p$  (distinct eigenvalues) Jordan blocks, so that  $A_0 = QJQ^{-1}$  for some invertible matrix  $Q$ . Each Jordan block  $J_i(\phi)$  corresponding to eigenvalue  $\phi$  can be written in the form  $J_i(\phi) = \phi I_s + N$ , where  $s$  is the algebraic multiplicity of  $\phi$  and  $N$  is the nilpotent matrix whose entries are ones on its super-diagonal and zero otherwise. By the binomial theorem  $J_i(\phi)^n = (\phi I_s + N)^n = \sum_{k=0}^n \binom{n}{k} \phi^{n-k} N^k$ , from which it is easy to see that each Jordan block will go to the zero matrix of its dimension as  $n \rightarrow \infty$ . Note also that if  $A_0$  is diagonalisable, then  $J = D$  (its diagonal matrix) and  $D^n \rightarrow \mathbf{0}_{d+2} \mathbf{0}_{d+2}^\top$  mirrors.

---

<sup>‡</sup>This result actually holds for any set of transmission matrices computed when  $\delta \in [0, \Delta)$ , so  $\delta = 0$  is non-restrictive in this proof.

We have  $A_0^n = QJ^nQ^{-1} \rightarrow \mathbf{0}_{d+2}\mathbf{0}_{d+2}^\top$  and hence that  $(B_{\Delta,\delta=0}^{(0)})^n \rightarrow \begin{bmatrix} \mathbf{0}_{d+2}\mathbf{0}_{d+2}^\top & \mathbf{f} \\ \mathbf{0}_{d+2}^\top & 1 \end{bmatrix}$ ,  
for some column vector  $\mathbf{f}$ . Now since the last row of  $B_{\Delta,\delta=0}^{(1)}$  consists of just zeros, we have  $\lim_{n \rightarrow \infty} (B_{\Delta,\delta=0}^{(0)})^n B_{\Delta,\delta=0}^{(1)} = \mathbf{0}_{d+3}\mathbf{0}_{d+3}^\top$  and hence that  $p_{B,n}(\boldsymbol{\theta}^{(d)}) \rightarrow 0$ .

Under model  $\mathcal{M}_\emptyset^d$ , i.e. when the photo-bleaching state is not included in the model,  $p_{B,n}(\boldsymbol{\theta}^{(d)}) = \boldsymbol{\nu}_X^\top A_0^n A_1 \mathbf{1}_{d+2} \rightarrow 0$ .  $\square$

## 4.B Imaging simulation

In this section we provide the simulation methods for the simulation studies presented in Section 4.5.

Each study (1-3), was simulated using the parameter values listed in Table 4.1 for each corresponding study. Firstly, a random number of  $M \sim \text{Poi}(\lambda_M)$  molecules was generated and then uniformly distributed over  $\mathcal{C}$  to obtain the true point pattern  $\{\mathbf{c}_1, \dots, \mathbf{c}_M\}$ . For each point  $\mathbf{c}_i$ , the hidden photo-switching process  $\{X(t)\}$  was simulated (with  $\nu_1^* = 1$ ) up to time  $N_F\Delta + T^*$ . From this, both the time series  $\{Y_n\}$  and the total times spent in the On state  $T_n$  in each frame  $[(n-1)\Delta, n\Delta], n \in \mathbb{Z}_{>0}$  were extracted. When the molecule was detected, i.e. when  $Y_n = 1$ , its localisation  $\mathbf{z}_n$  was then generated via

$$\mathbf{z}_n \sim \mathcal{N}\left(\mathbf{c}_i, \frac{\Delta}{N_p T_n} I_2\right),$$

with  $\frac{N_p T_n}{\Delta}$  denoting the number of photons emitted by the molecule in that frame. Here,  $N_p$  represents the expected number of photons a molecule emits if it occupies the On state 1 for an entire frame of length  $\Delta$ , and is therefore user defined. Under the three simulations studied in this chapter,  $N_p = 2000$ . Note also that this covariance aligns with standard experimental conditions, as is discussed in Section 1.F.1, and is such that each localisation's standard deviance is inversely proportional to the time the molecule spends in the frame.

Secondly, in each frame  $n$ , a random number of  $N_{A_n} \sim \text{Poi}(\alpha)$  false positive observations were generated and then uniformly distributed over  $\mathcal{C}$ . The observed covariance matrix of each false positive was then calculated as  $\frac{\Delta}{N_p T_n} I_2$ , with  $T_n \sim \text{Unif}(\delta, \Delta)$ .

The observation set  $Z_n$  for frame  $n$  was collected as the superposition of all vectors collected from the detected molecules and false positives. The localisation standard deviations were also stored for the calculations of the posterior density.

Lastly, to align with STORM experiments, the first  $\lceil T^*/\Delta \rceil$  frames were

removed from the video, allowing for  $Z^{(N_F)} = \{Z_1, \dots, Z_{N_F}\}$  to be used.



## 4.C DIC outputs

In this section, we present the DIC values outputted from the three MCMC simulations conducted in Section 3.4.

Model → M ↓	$\mathcal{M}_{\{1\}}^0$	$\mathcal{M}_{\{1\}}^1$	$\mathcal{M}_{\{1\}}^2$
3	-1.2209	-1.2204	-1.2190
4	-1.3082	-1.3081	-1.3079
5	-1.3428	-1.3426	-1.3424
6	-1.3534	-1.3530	-1.3528
7	<b>-1.3550</b>	-1.3548	-1.3546
8	-1.3549	-1.3548	-1.3546
9	-1.3544	-1.3542	-1.3539
10	-1.3538	-1.3534	-1.3530

**Table 4.6:** Table showing the computed DIC values ( $\times 10^5$ ) under simulation study 1, from three candidates  $\mathcal{M}_{\{1\}}^0$ ,  $\mathcal{M}_{\{1\}}^1$  and  $\mathcal{M}_{\{1\}}^2$  with varying values of  $M$ . The predicted model is indicated in red.

Model → M ↓	$\mathcal{M}_{\{1\}}^0$	$\mathcal{M}_{\{1\}}^1$	$\mathcal{M}_{\{1\}}^2$
3	-1.09555	-1.09493	-1.09461
4	-1.17717	-1.17664	-1.17610
5	-1.21563	-1.21526	-1.21521
6	-1.22936	-1.22911	-1.22878
7	-1.23382	-1.23347	-1.23325
8	-1.23385	<b>-1.23391</b>	-1.23372
9	-1.23384	-1.23345	-1.23333
10	-1.23196	-1.23245	-1.23216

**Table 4.7:** Table showing the computed DIC values ( $\times 10^5$ ) under simulation study 2, from three candidates  $\mathcal{M}_{\{1\}}^0$ ,  $\mathcal{M}_{\{1\}}^1$  and  $\mathcal{M}_{\{1\}}^2$  with varying values of  $M$ . The predicted model is indicated in red.

Model → M ↓	$\mathcal{M}_{\{1\}}^0$	$\mathcal{M}_{\{1\}}^1$	$\mathcal{M}_{\{1\}}^2$
3	-5.2202	-5.2183	-5.2143
4	-5.4526	-5.4536	-5.4523
5	-5.6610	-5.6687	-5.6749
6	-5.6974	-5.7872	<b>-5.7874</b>
7	-5.7645	-5.7475	-5.7473
8	-5.7734	-5.7751	-5.7656
9	-5.7735	-5.7741	-5.7767
10	-5.7614	-5.7566	-5.7617

**Table 4.8:** Table showing the computed DIC values ( $\times 10^5$ ) under simulation study 3, from three candidates  $\mathcal{M}_{\{1\}}^0$ ,  $\mathcal{M}_{\{1\}}^1$  and  $\mathcal{M}_{\{1\}}^2$  with varying values of  $M$ . The predicted model is indicated in red.

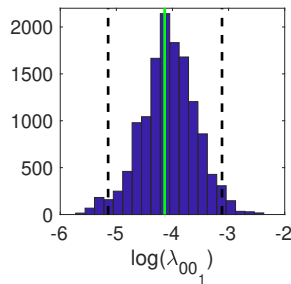
## 4.D Alexa Fluor 488 results

In this section, we first describe the experimental set-up used to obtain the Alexa Fluor 488 data used in Section 4.6 and then provide the posterior histograms of the parameter vector  $\theta_\alpha^{(2)}$  under both datasets.

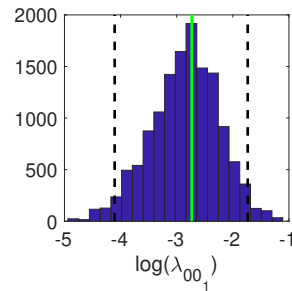
### 4.D.1 Experimental set-up

In this experiment, the EMCCD camera lens used to obtain the Alexa Fluor 488 data had a Numerical Aperture (NA) of 1.49. The electron count per pixel was calculated after applying an EMCCD gain of 100 counts and adding it to a base offset of 170 digital camera counts and a readout noise of  $1.8e^-$  per pixel. This was then divided by an analogue to digital conversion sensitivity to give the digital camera count. This produced raw images (of length 127 micrometres) with physical pixel sizes of 106 nanometres. These raw images were then post-processed using ThunderSTORM (Ovesný et al., 2014). Using this software, localisation measurements were recorded by using a Gaussian fitting algorithm. In particular, a Gaussian PSF was applied to high photon intensity spots binned to grids of size 11 by 11 pixels (squares of length 1166 nanometres), and run over each image of size 1200 by 1200 pixels. When fitting the point spread function (PSF), the photon positions were distributed according to a 2D Gaussian distribution, with parameters determined by the NA and the observed emission wavelengths.

### 4.D.2 Posterior histograms

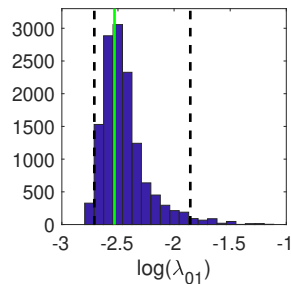


(a) Posterior histogram of  $\log \lambda_{00_1} (\text{s}^{-1})$  under dataset 1.

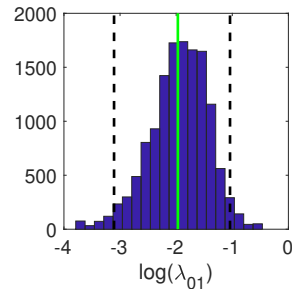


(b) Posterior histogram of  $\log \lambda_{00_1} (\text{s}^{-1})$  under dataset 2.

**Figure 4.22:** Posterior histograms of transformed photo-switching rates  $\log \lambda_{00_1}$  for the Alexa Fluor 488 datasets 1 (Figure 4.22a) and 2 (Figure 4.22b), plotted with MAP values (green) and 95% credible intervals (black, dotted).

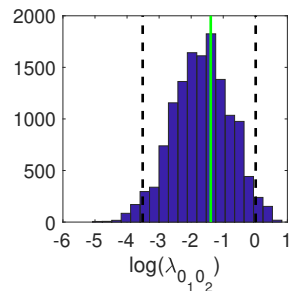


(a) Posterior histogram of  $\log \lambda_{01} (\text{s}^{-1})$  under dataset 1.

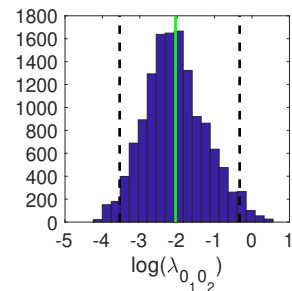


(b) Posterior histogram of  $\log \lambda_{01} (\text{s}^{-1})$  under dataset 2.

**Figure 4.23:** Posterior histograms of transformed photo-switching rates  $\log \lambda_{01}$  for the Alexa Fluor 488 datasets 1 (Figure 4.23a) and 2 (Figure 4.23b), plotted with MAP values (green) and 95% credible intervals (black, dotted).

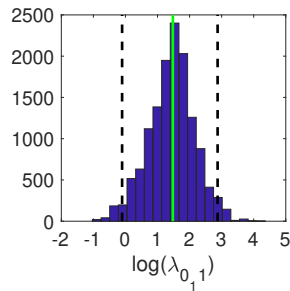


(a) Posterior histogram of  $\log \lambda_{01_0_2} (\text{s}^{-1})$  under dataset 1.

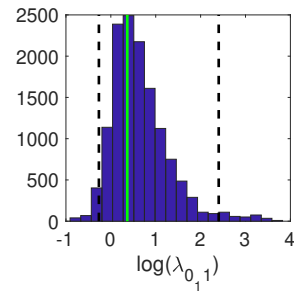


(b) Posterior histogram of  $\log \lambda_{01_0_2} (\text{s}^{-1})$  under dataset 2.

**Figure 4.24:** Posterior histograms of transformed photo-switching rates  $\log \lambda_{01_0_2}$  for the Alexa Fluor 488 datasets 1 (Figure 4.24a) and 2 (Figure 4.24b), plotted with MAP values (green) and 95% credible intervals (black, dotted).

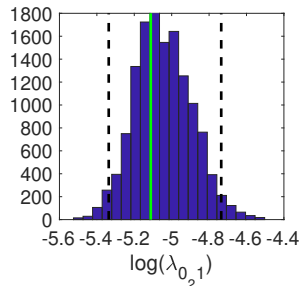


(a) Posterior histogram of  $\log \lambda_{0,1}(\text{s}^{-1})$  under dataset 1.

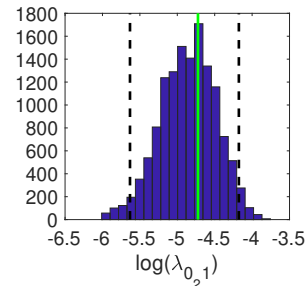


(b) Posterior histogram of  $\log \lambda_{0,1}(\text{s}^{-1})$  under dataset 2.

**Figure 4.25:** Posterior histograms of transformed photo-switching rates  $\log \lambda_{0,1}$  for the Alexa Fluor 488 datasets 1 (Figure 4.25a) and 2 (Figure 4.25b), plotted with MAP values (green) and 95% credible intervals (black, dotted).

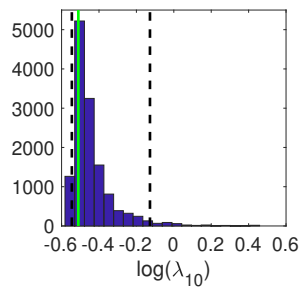


(a) Posterior histogram of  $\log \lambda_{0,2}(\text{s}^{-1})$  under dataset 1.

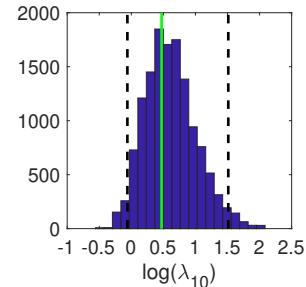


(b) Posterior histogram of  $\log \lambda_{0,2}(\text{s}^{-1})$  under dataset 2.

**Figure 4.26:** Posterior histograms of transformed photo-switching rates  $\log \lambda_{0,2}$  for the Alexa Fluor 488 datasets 1 (Figure 4.26a) and 2 (Figure 4.26b), plotted with MAP values (green) and 95% credible intervals (black, dotted).

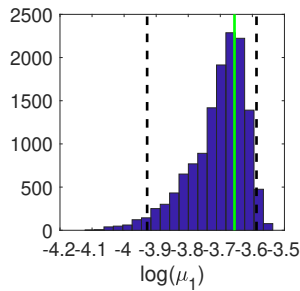


(a) Posterior histogram of  $\log \lambda_{10}(\text{s}^{-1})$  under dataset 1.

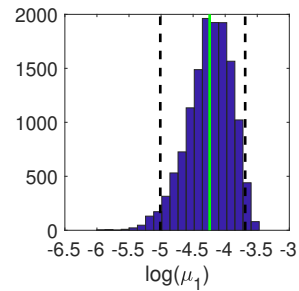


(b) Posterior histogram of  $\log \lambda_{10}(\text{s}^{-1})$  under dataset 2.

**Figure 4.27:** Posterior histograms of transformed photo-switching rates  $\log \lambda_{10}$  for the Alexa Fluor 488 datasets 1 (Figure 4.27a) and 2 (Figure 4.27b), plotted with MAP values (green) and 95% credible intervals (black, dotted).

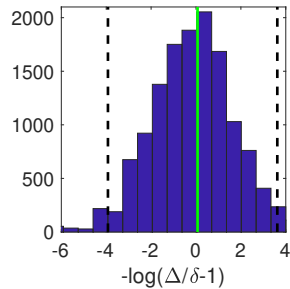


(a) Posterior histogram of  $\log \mu_1 (\text{s}^{-1})$  under dataset 1.

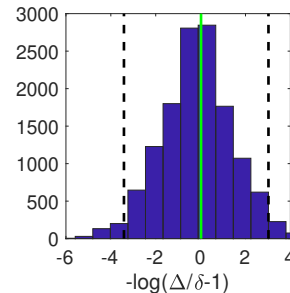


(b) Posterior histogram of  $\log \mu_1 (\text{s}^{-1})$  under dataset 2.

**Figure 4.28:** Posterior histograms of transformed photo-switching rates  $\log \mu_1$  for the Alexa Fluor 488 datasets 1 (Figure 4.28a) and 2 (Figure 4.28b), plotted with MAP values (green) and 95% credible intervals (black, dotted).

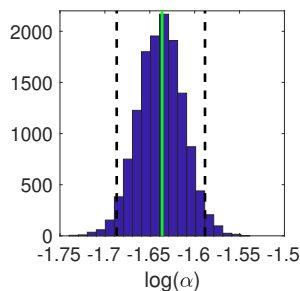


(a) Posterior histogram of  $-\log(\Delta/\delta - 1)$  under dataset 1.

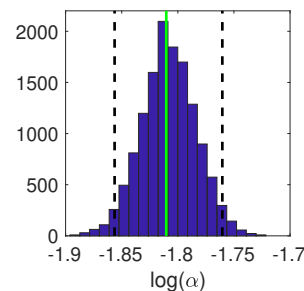


(b) Posterior histogram of  $-\log(\Delta/\delta - 1)$  under dataset 2.

**Figure 4.29:** Posterior histograms of transformed photo-switching rates  $-\log(\Delta/\delta - 1)$  for the Alexa Fluor 488 datasets 1 (Figure 4.29a) and 2 (Figure 4.29b), plotted with MAP values (green) and 95% credible intervals (black, dotted).



(a) Posterior histogram of  $\log \alpha$  under dataset 1.



(b) Posterior histogram of  $\log \alpha$  under dataset 2.

**Figure 4.30:** Posterior histograms of transformed photo-switching rates  $\log \alpha$  for the Alexa Fluor 488 datasets 1 (Figure 4.30a) and 2 (Figure 4.30b), plotted with MAP values (green) and 95% credible intervals (black, dotted).

# 5

## CONCLUSIONS AND FUTURE WORK

Accurate measurement of fluorophore photo-switching rates has the potential to enable tailored design of single molecule localisation microscopy (SMLM) experiments to specific requirements. For instance, one may wish to select a fluorophore and photo-switching environment to achieve the rapid photo-switching at low laser intensities required for live-cell samples. Alternatively, one may wish to promote long off times required for densely packed samples. Furthermore, precise estimates of photo-switching rates has both the potential to advance data processing methods used in SMLM imaging, enabling more accurate image reconstruction, and can also aid proper quantitative analysis of molecular stoichiometries. For this purpose, the main idea that we have presented in this thesis is a method for fully characterising the photo-switching kinetics of fluorophores imaged in SMLM.

In Chapter 1, we carefully defined a continuous time Markov process  $\{X(t)\}$  suitable for modelling fluorophores that stochastically photo-switch between a photon emission state,  $d + 1$  dark states and a photo-bleached state. Motivated by the imperfections of most imaging systems, we subsequently derived the observation process  $\{Y_n\}$  indicating when a molecule is detected in a given frame  $n$  of the video. From this, we formulated the photo-switching

hidden Markov model (PSHMM) and utilised a novel adaptation to the forward-backward algorithm to derive its likelihood function. Maximum likelihood parameter estimates were then obtained via numerical optimisation and shown to be accurate in estimating the unknown parameters  $\theta^{(d)}$  and underlying photo-kinetic model (characterised by  $d$ ), frequently outperforming estimates obtained from an existing method. While this model was constructed specifically for (d)STORM applications, we also noted (see Remark 3) that the form of  $\{X(t)\}$  is also well suited to modelling the photo-switching behaviour of molecules imaged in PALM.

In Chapter 2, we investigated the properties of the PSHMM maximum likelihood estimator under different experimental conditions. Although we found that certain photo-switching parameters are subject to inherent correlation structures dictated by the model, correlations with the noise parameter  $\delta$  were also observed. This had the ability to effectuate parameter identification issues via decreased curvature of the PSHMM log-likelihood surface in two situations: as  $\delta$  becomes large relative to the frame acquisition rate  $\Delta^{-1}$ , and under faster photo-switching scenarios. Furthermore, this was also seen to be the reason for poorer convergence of mean squared errors in the resulting parameter estimates. Although we concluded that this situation does not frequently arise in experimental set-ups, and hence the PSHMM generally well suited for parameter estimation in this application, the two problematic situations derived from this analysis provide a useful basis for improved experimental design.

Nevertheless, while it may not always be possible to change the conditions of imaging, especially to those concerning the background noise, the analysis conducted here may additionally provide a scope for improvements in the PSHMM likelihood estimation procedure. For poorly conditioned data, alternative approaches such as pre-conditioning (Penrose, 1955) or regularisation/penalisation strategies (Golub et al., 1999) are available and have seen application within many fields, including neuroscience (Monti et al., 2015, 2017). These methods can be implemented to improve the invertibility of the Hessian matrix, and therefore the curvature of the (log-)likelihood surface at the maximum likelihood estimate, thereby having the ability to reduce



potential identification issues arising from collinearity between parameters. Future work to improve the PSHMM could therefore study potential avenues for combining such strategies with our current methodology. Such extensions may also be combined with a novel method to estimate both first and second derivatives of the PSHMM (log-)likelihood function from the aforementioned forward-backward algorithm developed for this problem for improved parameter estimation in the PSHMM via Newton’s algorithm.

In Chapter 3, we used the PSHMM to derive the exact (posterior) distribution of the unknown number of  $M$  fluorescing molecules given the total number of localisations obtained in an imaging experiment. Using the resulting mode as an estimate for  $M$ , we demonstrated the utility this method has in addressing large scale molecular counting problems arising in SMLM. While our method requires a plug-in estimate for  $\boldsymbol{\theta}_\omega^{(d)}$ , the resulting (modular) inference technique allows for thousands of molecules to be counted at once and at computational ease.

Constructing a method suitable for (d)STORM experiments which is able to dually estimate the number of molecules  $M$  and their photo-switching rates  $\boldsymbol{\theta}_\omega^{(d)}$  using count localisation data, is difficult and non-trivial. One such attempt could be via the inference of these parameters from the *count time series*  $\{\tilde{Y}_n : n \in \mathbb{Z}_{>0}\}$  defined by

$$\tilde{Y}_n = \sum_{m=1}^M Y_{n,m}.$$

Here, a reversible-jump Metropolis-within-Gibbs sampler could be implemented to sample from the posterior distribution

$$p(\boldsymbol{\theta}_\omega^{(d)}, M, \mathbf{x}_0, \dots, \mathbf{x}_{N_F} | \tilde{y}_1, \dots, \tilde{y}_{N_F}),$$

with  $\mathbf{x}_i \in \mathbb{R}^M$  denoting the  $M$  molecules’ hidden states at time instance  $i\Delta$ . In this setting, these hidden states could be (block) updated through the (aforementioned) forward-backward algorithm, the parameter  $\boldsymbol{\theta}_\omega^{(d)}$  updated through a standard Metropolis-Hastings kernel and  $M$  updated via a reversible-jump move. However, such a method is unlikely to guarantee

feasible scalability with respect to  $M$ . For example, in the PALM setting, [Rollins et al. \(2014\)](#) attempts to count and do rate estimation simultaneously. While having a single procedure avoids the problem of a separate training experiment, the authors recognise that the computational burden of such a procedure is extreme and drastically limits the numbers of molecules that can be counted at any one time. Furthermore, it requires careful extraction of the time traces from crowded environments, which is in itself problematic and challenging. Nevertheless, recent advances in machine learning algorithms applied to SMLM data ([Nehme et al., 2018](#), [Davis et al., 2019](#)), would indicate opportunities to do so. For example, a recurrent neural network (RNN), which can learn patterns in temporal information, could be trained to identify all parameters of interest. However, the success of these algorithms lies in the ability to perfectly train a network and may therefore require a huge amount of training data to reflect the numerous types of experiments conducted in SMLM.

In Chapter 4, we formulated a novel spatio-temporal point process that links the true unobserved point process of molecular positions with spatial localisation sets obtained when an unknown number of  $M$  fluorescing molecules are detected during an experiment. Using the PSHMM, we derived the full Bayes filter appropriate for inference and applied a constructed Metropolis-within-Gibbs MCMC algorithm that was shown in simulations to recover the true point process and underlying photo-kinetic model  $\theta_{\alpha}^{(d)}$  of the imaged molecules.

Although we demonstrated that the MCMC algorithm suitable for this problem can be used to infer the parameters of interest in the model, inference through this Bayes filter is only computationally feasible for a small number of imaged molecules. When dealing with larger datasets, which is common in fluorescence experiments, approximations of this filter will be required. The sequential Monte Carlo (SMC) approximations ([Mahler, 2007a](#), [Vo et al., 2003, 2005](#), [Whiteley et al., 2010](#)) to Mahler’s full Bayes filter, which instead of inferring the full point process, propagate its first order intensity, therefore provide us with a useful insight that the Bayes filter developed here can be approximated to deal with larger datasets. As a scope for fu-

ture research, for example, the number of molecules  $M$ , their true locations  $C = \{\mathbf{c}_1, \dots, \mathbf{c}_M\}$  and the static parameters of the model  $\boldsymbol{\theta}_\alpha^{(d)}$  could be inferred in a particle MCMC framework. Such an approach could utilise an (unbiased) approximation to the likelihood of (the first order intensity of)  $C$  and  $\boldsymbol{\theta}_\alpha^{(d)}$  via an SMC algorithm, with parameter updates conducted through a Metropolis-Hastings kernel. Furthermore, inference on  $d$  could be subsequently be performed using a particle estimate of the marginal likelihood and then applying the method described in [Chib and Jeliazkov \(2001\)](#) as a model selection tool.

While the work presented here focuses on SMLM, the type of kinetic models we have utilised are unlikely to be unique to photo-switching fluorophores and super-resolution applications. Certainly, stochastic processes in which the observed signal depends on both the current and past states of a hidden process are likely to be a general feature of digital, discretised measurements of stochastic signals. This is particularly true in image processing, where images are inevitably formed by exposing the camera’s sensor over a non-zero length time window. The relationship we observed between the emission and transition probabilities of the PSHMM is a direct consequence of this exposure time, and therefore it is likely that the presented methodology will find use in imaging applications that have not been considered in this thesis.



## **A.1 Copyright statements**

This section first highlights (see Figure A.1) the copyright transfer agreement for the paper entitled “A hidden Markov model approach to characterizing the photo-switching behavior of fluorophores” (Patel et al., 2019), which is presented in Chapters 1 and 2. The relevant section for this thesis is highlighted in yellow.

Secondly, we provide the the copyright agreement (see Figure A.2) for the paper entitled “Bayesian filtering for spatial estimation of photo-switching fluorophores imaged in Super-resolution fluorescence microscopy” (Patel and Cohen, 2018), which the work presented in Chapter 4 is based upon.

## COPYRIGHT TRANSFER AGREEMENT

INSTITUTE OF MATHEMATICAL STATISTICS  
BRAZILIAN STATISTICAL ASSOCIATION

ASSOCIATION PUBLICATIONS DE L'INSTITUT HENRI POINCARÉ  
BERNOULLI SOCIETY FOR MATHEMATICAL STATISTICS & PROBABILITY

*This signed Agreement must be received before the manuscript can be scheduled for publication. Please fill in and sign as indicated. Fax, email or mail to: Mattson Publishing Services, 2002 Holly Neck Road, Baltimore, MD 21221 USA Email: [assistant@mattsonpublishing.com](mailto:assistant@mattsonpublishing.com), Fax: 410-391-2542*

Transmittal #: AOAS1240

Author(s): Lekha Patel, Nils Gustafsson, Yu Lin, Raimund Ober, Ricardo Henriques, and Edward Cohen

Title: A hidden Markov model approach to characterizing the photo-switching behavior of fluorophores

Copyright of the above work is hereby transferred to (please check the box next to the journal in which your article will appear).

**Institute of Mathematical Statistics**

- Annals of Applied Probability
- Annals of Probability
- Annals of Applied Statistics
- Annals of Statistics
- IMS Collections
- Statistical Science

**Association Publications de l'Institut Henri Poincaré**

- Annales de l'Institut Henri Poincaré (B) Probabilités et Statistiques

**Bernoulli Society for Mathematical Statistics & Probability**

- Bernoulli

**Brazilian Statistical Association**


- Brazilian Journal of Probability and Statistics

Effective if and when the work is accepted for publication. This copyright applies to the extent transferable for U.S. Government employees, see item B, below.

However, the author(s) reserve the following:


1. All proprietary rights other than copyright, such as patent rights.
2. The rights to use all or part of this work (with appropriate reference to the journal and article) in future works of their own, such as lectures, press releases, reviews, textbooks, or reprint books.
3. The right to place the final version of this article (exactly as published in the journal) on their own homepage or in a public digital repository, provided there is a link to the official journal site.
4. The right to provide a copy of the final peer-reviewed manuscript to the NIH upon acceptance for publication, for public archiving in PubMed Central as soon as possible but no later than 12 months after publication.

**Figure A.1:** Copyright agreement for [Patel et al. \(2019\)](#).



RightsLink®

[Home](#)
[?](#)
[Email Support](#)
[Sign in](#)
[Create Account](#)



Requesting permission to reuse content from an IEEE publication

Bayesian filtering for spatial estimation of photo-switching fluorophores imaged in Super-resolution fluorescence microscopy

Conference Proceedings: 2018 52nd Asilomar Conference on Signals, Systems, and Computers

Author: Lekha Patel

Publisher: IEEE

Date: Oct. 2018

Copyright © 2018, IEEE

Thesis / Dissertation Reuse

The IEEE does not require individuals working on a thesis to obtain a formal reuse license, however, you may print out this statement to be used as a permission grant:

*Requirements to be followed when using any portion (e.g., figure, graph, table, or textual material) of an IEEE copyrighted paper in a thesis:*

- 1) In the case of textual material (e.g., using short quotes or referring to the work within these papers) users must give full credit to the original source (author, paper, publication) followed by the IEEE copyright line © 2011 IEEE.
- 2) In the case of illustrations or tabular material, we require that the copyright line © [Year of original publication] IEEE appear prominently with each reprinted figure and/or table.
- 3) If a substantial portion of the original paper is to be used, and if you are not the senior author, also obtain the senior author's approval.

*Requirements to be followed when using an entire IEEE copyrighted paper in a thesis:*

- 1) The following IEEE copyright/ credit notice should be placed prominently in the references: © [year of original publication] IEEE. Reprinted, with permission, from [author names, paper title, IEEE publication title, and month/year of publication]
- 2) Only the accepted version of an IEEE copyrighted paper can be used when posting the paper or your thesis on-line.
- 3) In placing the thesis on the author's university website, please display the following message in a prominent place on the website: In reference to IEEE copyrighted material which is used with permission in this thesis, the IEEE does not endorse any of [university/educational entity's name goes here]'s products or services. Internal or personal use of this material is permitted. If interested in reprinting/republishing IEEE copyrighted material for advertising or promotional purposes or for creating new collective works for resale or redistribution, please go to [http://www.ieee.org/publications\\_standards/publications/rights/rights\\_link.html](http://www.ieee.org/publications_standards/publications/rights/rights_link.html) to learn how to obtain a License from RightsLink.

If applicable, University Microfilms and/or ProQuest Library, or the Archives of Canada may supply single copies of the dissertation.

BACK
CLOSE

**Figure A.2:** Copyright agreement for Patel and Cohen (2018).

## REFERENCES

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 19(6), 716–723.
- Atchinson, J. and S. Shen (1980). Logistic-Normal distributions: Some properties and uses. *Biometrika* 67(2), 261–272.
- Bar-Shalom, Y., T. Kirubarajan, and X. R. Li (2001). *Estimation with Applications to Tracking and Navigation: Theory Algorithms and Software*. Wiley.
- Baum, L. and J. Eagon (1967). An inequality with applications to statistical estimation for probabilistic functions of Markov processes and to a model for ecology. *Bulletin of the American Mathematical Society* 73(3), 360–363.
- Baum, L. and T. Petrie (1966). Statistical Inference for Probabilistic Functions of Finite State Markov Chains. *The Annals of Mathematical Statistics* 37(6), 1554–1563.
- Baum, L. and G. Sell (1968). Growth transformations for functions on manifolds. *Pacific Journal of Mathematics* 27(2), 211–227.
- Baum, L., G. S. T. Petrie, and N. Weiss (1970). A Maximization Technique Occurring in the Statistical Analysis of Probabilistic Functions of Markov Chains. *The Annals of Mathematical Statistics* 41(1), 164–171.
- Betzig, E., G. H. Patterson, R. Sougrat, O. W. Lindwasser, S. Olenych, J. S. Bonifacino, M. W. Davidson, J. Lippincott-Schwartz, and H. F. Hess (2006). Imaging Intracellular Fluorescent Proteins at Nanometer Resolution. *Science* 313(5793), 1642–1645.

- Boyd, N., E. Jonas, and B. R. H. Babcock (2018). DeepLoco: Fast 3D Localization Microscopy Using Neural Networks. *bioRxiv* <https://doi.org/10.1101/267096>.
- Buchholz, P., J. Kriege, and I. Felko (2014). *Input Modeling with Phase-Type Distributions and Markov Models*. Springer eBooks.
- Chen, M.-H. and Q.-M. Shao (1998). Monte Carlo Estimation of Bayesian Credible and HPD Intervals. *Journal of Computational and Graphical Statistics* 8(1), 69–92.
- Chib, S. (1995). Marginal Likelihood from the Gibbs Output. *Journal of the American Statistical Association* 90(432), 1313–1321.
- Chib, S. and I. Jeliazkov (2001). Marginal Likelihood From the Metropolis - Hastings Output. *Journal of the American Statistical Association* 96(453), 270–281.
- Ching, W., E. Fung, and M. Ng (2003). *Higher-Order Hidden Markov Models with Applications to DNA Sequences*. Springer Berlin Heidelberg.
- Cohen, E., A. Abraham, S. Ramakrishnan, and R. Ober (2019). Resolution limit of image analysis algorithms. *Nature Communications* 10, 793.
- Colquhoun, D., C. J. Hatton, and A. G. Hawkes (2003). The quality of maximum likelihood estimates of ion channel rate constants. *The Journal of Physiology* 547(3), 699–728.
- Colquhoun, D. and A. G. Hawkes (1981). On the stochastic properties of single ion channels. *Proceedings of the Royal Society of London. Series B, Biological sciences* 211, 205–235.
- Colquhoun, D., A. G. Hawkes, and K. Srodzinski (1996). Joint Distributions of Apparent Open and Shut Times of Single-Ion Channels and maximum likelihood Fitting of Mechanisms. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 354, 2555–2590.



- Cox, S., E. Rosten, J. Monypenny, T. Jovanovic-Taliman, D. T. Burnette, J. Lippincott-Schwartz, G. E. Jones, and R. Heintzmann (2011). Bayesian localization microscopy reveals nanoscale podosome dynamics. *Nature Methods* 9(2), 195–200.
- Daley, D. J. and D. Vere-Jones (1998). *An Introduction to the Theory of Point Processes*. Springer Series in Statistics. Springer-Verlag New York.
- Davis, S. P. X., S. Kumar, Y. Alexandrov, A. Bhargava, G. da Silva Xavier, G. A. Rutter, P. Frankel, S. Erik, S. Flaxman, P. French, and J. McGinty (2019). Convolutional neural networks for reconstruction of undersampled optical projection tomography data applied to in vivo imaging of zebrafish. *Journal of Biophotonics* 0, e201900128.
- Dempsey, G. T., J. C. Vaughan, K. H. Chen, M. Bates, and X. Zhuang (2011). Evaluation of fluorophores for optimal performance in localization-based super-resolution imaging. *Nature Methods* 8(12), 1027–1036.
- Du Preez, J. (1998). Efficient training of high-order hidden Markov models using first-order representations. *Computer Speech & Language* 12(1), 23–39.
- Efron, B. and R. Tibshirani (1993). *An introduction to the Bootstrap*. Chapman & Hall.
- Epstein, M., B. Calderhead, M. A. Girolami, and L. Sivilotti (2016). Bayesian Statistical Inference in Ion-Channel Models with Exact Missed Event Correction. *Biophysical Journal* 111(2), 333–348.
- Fricke, F., J. Beaudouin, R. Eils, and M. Heilemann (2015). One, two or three? Probing the stoichiometry of membrane proteins by single-molecule localization microscopy. *Scientific Reports* 5, 14072.
- Geyer, C. J. and J. Møller (1994). Simulation Procedures and Likelihood Inference for Spatial Point Processes. *Scandinavian Journal of Statistics* 21(4), 359–373.

- Godfrey, K. and J. DiStefano (1985). Identifiability of Model Parameter. *IFAC Proceedings Volumes 18(5)*, 89–114.
- Golub, G. H., P. C. Hansen, and D. P. O’Leary (1999). Tikhonov Regularization and Total Least Squares. *SIAM Journal on Matrix Analysis and Applications 21(1)*, 185–194.
- Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika 82(4)*, 711–732.
- Greenfeld, M., D. S. Pavlichin, H. Mabuchi, and D. Herschlag (2015). Single Molecule Analysis Research Tool (SMART): An Integrated Approach for Analysing Single Molecule Data. *PLoS ONE 7(2)*, e30024.
- Ha, T. and P. Tinnefeld (2012). Photophysics of Fluorescent Probes for Single-Molecule Biophysics and Super-Resolution Imaging. *Annual Review of Physical Chemistry 63(1)*, 595–617.
- Haario, H., E. Saksman, and T. J (2001). An adaptive Metropolis algorithm. *Bernoulli 7(2)*, 223–242.
- Hawkes, A. G., A. Jalali, and D. Colquhoun (1990). The Distributions of the Apparent Open Times and Shut Times in a Single Channel Record when Brief Events Cannot Be Detected. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 332*, 511–538.
- Hawkes, A. G., A. Jalali, and D. Colquhoun (1992). Asymptotic Distributions of Apparent Open Times and Shut Times in a Single Channel Record Allowing for the Omission of Brief Events. *Philosophical Transactions of the Royal Society B: Biological Sciences 337*, 383–404.
- Heilemann, M., S. Van de Linde, M. Schüttelpeiz, R. Kasper, B. Seefeldt, A. Mukherjee, P. Tinnefeld, and M. Sauer (2008). Subdiffraction - Resolution Fluorescence Imaging with Conventional Fluorescent Probes. *Angewandte Chemie International Edition 47(33)*, 6172–6176.
- Henriques, R., M. Lelek, E. Fornasiero, F. Valtorta, C. Zimmer, and M. Mhlanga (2010). QuickPALM: 3D real-time photoactivation nanoscopy image processing in ImageJ. *Nature Methods 7(5)*, 339–340.

- Hess, S. T., T. P. K. Girirajan, and M. D. Mason (2006). Ultra-high resolution imaging by fluorescence photoactivation localization microscopy. *Biophysical Journal* 91(11), 4258–4272.
- Hirsch, M., R. Wareham, M. Martin-Fernandez, M. Hobson, and D. Rolfe (2013). A stochastic model for electron multiplication charge-coupled devices-from theory to practice. *PLoS ONE* 8(1), 1–13.
- Hsu, C. and T. Baumgart (2011). Spatial Association of Signaling Proteins and F-Actin Effects on Cluster Assembly Analyzed via Photoactivation Localization Microscopy in T Cells. *PLoS ONE* 6(8), e23586.
- Huang, B., M. Bates, and X. Zhuang (2009). Super-Resolution Fluorescence Microscopy. *Annual Review of Biochemistry* 78(1), 993–1016.
- Hyndman, R. J. (1996). Computing and Graphing Highest Density Regions. *The American Statistician* 50(2), 120–126.
- Jacob, P. E., L. M. Murray, C. C. Holmes, and C. P. Robert (2017). Better together? Statistical learning in models made of modules. *arXiv* <https://arxiv.org/abs/1708.08719>.
- Jacquez, J. A. and P. Greif (1985). Numerical Parameter Identifiability and Estimability: Integrating Identifiability, Estimability, and Optimal Sampling Design. *Mathematical Biosciences* 77, 201–227.
- Jiang, L., S. S. Singh, and S. Yildirim (2015). Bayesian Tracking and Parameter Learning for Non-Linear Multiple Target Tracking Models. *IEEE Transactions on Signal Processing* 63(21), 5733–5745.
- Jungmann, R., C. Steinhauer, M. Scheible, A. Kuzyk, P. Tinnefeld, and F. C. Simmel (2010). Single-molecule kinetics and super-resolution microscopy by fluorescence imaging of transient binding on DNA origami. *Nano Letters* 10(11), 4756–4761.
- Lee, L.-M. and J.-C. Lee (2006). *A Study on High-Order Hidden Markov Models and Applications to Speech Recognition*. Springer Berlin Heidelberg.

- Lee, S. H., J. Y. Shin, A. Lee, and C. Bustamante (2012). 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.
- Lehmann, M., G. Lichtner, H. Klenz, and J. Schmoranzer (2016). Novel organic dyes for multicolor localization-based super-resolution microscopy. *Journal of Biophotonics* 9(1-2), 161–170.
- Levinson, S., L. Rabiner, and M. Sondhi (1983). An Introduction to the Application of the Theory of Probabilistic Functions of a Markov Process to Automatic Speech Recognition. *Bell Labs Technical Journal* 62(4), 1035–1074.
- Lin, Y., 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 (2015). Quantifying and Optimizing Single-Molecule Switching Nanoscopy at High Speeds. *PLoS ONE* 10(5), e0128135.
- Little, M. P., W. F. Heidenreich, and G. Li (2010). Parameter Identifiability and Redundancy: Theoretical Considerations. *PLoS ONE* 5(1), 1–6.
- Liu, Y.-Y., S. Li, F. Li, L. Song, and J. Rehg (2015). Efficient Learning of Continuous-Time Hidden Markov Models for Disease Progression. *Advances in Neural Information Processing Systems* 28(1), 3599–3607.
- MacDonald, I. L. and W. Zucchini (1997). *Hidden Markov and other models for discrete-valued time series*. Monographs on Statistics and Applied Probability. Chapman & Hall.
- Maggio, E., M. Taj, and A. Cavallaro (2008). Efficient multi-target visual tracking using random finite sets. *IEEE Transactions on Circuits and Systems for Video Technology* 18(8), 1016–1027.
- Mahler, R. (2003). Multitarget Bayes filtering via first-order multitarget moments. *IEEE Transactions on Aerospace and Electronic Systems* 39(4), 1152–1178.

- Mahler, R. (2007a). PHD filters of higher order in target number. *IEEE Transactions on Aerospace and Electronic Systems* 43(4), 1523–1543.
- Mahler, R. (2007b). *Statistical Multisource-Multitarget information fusion*. Arctech house.
- Marsh, R., K. Pfisterer, P. Bennett, L. Hirvonen, M. Gautel, G. Jones, and S. Cox (2018). Artifact-free high-density localization microscopy analysis. *Nature Methods* 15, 689–692.
- Minin, V. N. and M. A. Suchard (2007). Counting labeled transitions in continuous-time Markov models of evolution. *Journal of Mathematical Biology* 56(3), 3600–3608.
- Møller, J. and R. Waagepetersen (2003). *Statistical Inference and Simulation for Spatial Point Processes*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability.
- Monti, R. P., C. Anagnostopoulos, and G. Montana (2017). Learning population and subject-specific brain connectivity networks via mixed neighborhood selection. *Annals of Applied Statistics* 11(4), 2142–2164.
- Monti, R. P., R. Lorenz, P. Hellyer, C. A. R. Leech, and G. Montana (2015). Graph Embeddings of Dynamic Functional Connectivity Reveal Discriminative Patterns of Task Engagement in HCP Data. In *2015 International Workshop on Pattern Recognition in NeuroImaging*.
- Moschopoulos, P. G. (1985). The distribution of the sum of independent gamma random variables. *Annals of the Institute of Statistical Mathematics* 37(3), 541–544.
- Mrkvička, T., M. Muška, and J. Kubečka (2014). Two step estimation for Neyman-Scott point process with inhomogeneous cluster centers. *Statistics and Computing* 24(7), 91–100.
- Mukamel, E., H. Babcock, and X. Zhuang (2012). Statistical deconvolution for superresolution fluorescence microscopy. *Biophysical Journal* 102(10), 2391–2400.

- Nehme, E., L. E. Weiss, T. Michaeli, and Y. Shechtman (2018). Deep-STORM: super-resolution single-molecule microscopy by deep learning. *Optica* 5(4), 458–464.
- Newton, M. A. and A. E. Raftery (1994). Approximate Bayesian Inference with the Weighted Likelihood Bootstrap. *Journal of the Royal Statistical Society: Series B (Methodological)* 56(1), 3–26.
- Nieuwenhuizen, R. P. J., M. Bates, A. Szymborska, K. A. Lidke, B. Rieger, and S. Stallinga (2015). Quantitative Localization Microscopy: Effects of Photophysics and Labeling Stoichiometry. *PLoS ONE* 10(5), e0127989.
- Nino, D., N. Rafiei, Y. Wang, A. Zilman, and J. N. Milstein (2017). Molecular Counting with Localization Microscopy: A Bayesian Estimate Based on Fluorophore Statistics. *Biophysical Journal* 112(9), 1777–1785.
- Ober, R., A. Tahmasbi, S. Ram, Z. Lin, and E. Ward (2015). Quantitative Aspects of Single-Molecule Microscopy: Information-theoretic analysis of single-molecule data. *IEEE Signal Processing Magazine* 32(1), 58–69.
- Ober, R. J., S. Ram, and E. S. Ward (2004). Localization accuracy in single-molecule microscopy. *Biophysical Journal* 87(2), 1185–1200.
- Olivo-Marin, J.-C. (2002). Extraction of spots in biological images using multiscale products. *Pattern Recognition* 35(9), 1989–1996.
- Ovesný, M., J. B. P. Křížek, Z. Švindrych, and G. Hagen (2014). Thunder-STORM: a comprehensive ImageJ plug-in for PALM and STORM data analysis and super-resolution imaging. *Bioinformatics* 30(16), 2389–2390.
- Owen, D., C. Rentero, J. Rossy, A. Magenau, D. Williamson, M. Rodriguez, and K. Gaus (2010). PALM imaging and cluster analysis of protein heterogeneity at the cell surface. *Journal of Biophotonics* 3(7), 446–454.
- Patel, L. and E. Cohen (2018). Bayesian filtering for spatial estimation of photo-switching fluorophores imaged in Super-resolution fluorescence microscopy. In *2018 52nd Asilomar Conference on Signals, Systems, and Computers*.

- Patel, L., N. Gustafsson, Y. Lin, R. Ober, R. Henriques, and E. Cohen (2019). A hidden Markov model approach to characterizing the photo-switching behavior of fluorophores. *Annals of Applied Statistics* 13(3), 1397–1429.
- Patel, L., D. M. Owen, and E. Cohen (2019). Blinking Statistics and Molecular Counting in direct Stochastic Reconstruction Microscopy (dSTORM). *bioRxiv* <https://doi.org/10.1101/834572>.
- Penrose, R. (1955). A generalized inverse for matrices. *Mathematical Proceedings of the Cambridge Philosophical Society* 51(3), 406–413.
- Qin, F., A. Auerbach, and F. Sachs (1996). Estimating single-channel kinetic parameters from idealized patch-clamp data containing missed events. *Biophysical Journal* 70(1), 264–280.
- Qin, F. and L. Li (2004). Model-Based Fitting of Single-Channel Dwell-Time Distributions. *Biophysical Journal* 87(3), 1657–1671.
- Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE* 77(2), 257–286.
- Rajarshi, M. (2013). Markov chains and their extensions. In *Statistical Inference for Discrete Time Stochastic Processes*. Springer India.
- Ram, S., E. S. Ward, and R. J. Ober (2012). A stochastic analysis of distance estimation approaches in single molecule microscopy - quantifying the resolution limits of photon-limited imaging systems. *Multidimensional Systems and Signal Processing* 24(3), 503–542.
- Rief, M., R. S. Rock, A. D. Mehta, M. S. Mooseker, R. E. Cheney, and J. A. Spudich (2000). Myosin-V stepping kinetics: a molecular model for processivity. *Proceedings of the National Academy of Sciences USA* 97(17), 9482–9486.
- Rieger, B. and S. Stallinga (2014). The lateral and axial localization uncertainty in super-resolution light microscopy. *ChemPhysChem* 15(4), 664–670.

- Rollins, G. C., J. Y. Shin, C. Bustamante, and S. Pressé (2014). 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.
- Rothenberg, T. J. (1971). Identification in parametric models. *Econometrica* 39(3), 577–591.
- Rust, M. J., M. Bates, and X. Zhuang (2006). Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nature Methods* 3(10), 793–795.
- Sage, D., H. Kirshner, T. Pengo, N. Stuurman, J. Min, S. Manley, and M. Usher (2015). Quantitative evaluation of software packages for single-molecule localization microscopy. *Nature Methods* 12(8), 717–724.
- Schwarz, G. (1978). Estimating the Dimension of a Model. *Annals of Statistics* 6(2), 461–464.
- Sharonov, A. and R. M. Hochstrasser (2006). Wide-field subdiffraction imaging by accumulated binding of diffusing probes. *Proceedings of the National Academy of Sciences of the United States of America* 103(50), 18911–18916.
- Spiegelhalter, D., J. David, N. Best, B. Carlin, and A. V. D. Linde (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 64(4), 583–639.
- Stoyan, D., W. Kendall, and J. Mecke (1987). *Stochastic geometry and its applications*. Wiley series in probability and mathematical statistics: Applied probability and statistics. Wiley.
- Thompson, R. E., D. R. Larson, and W. W. Webb (2002). Precise nanometer localization analysis for individual fluorescent probes. *Biophysical Journal* 82(5), 2775–2783.
- Van de Linde, S. and M. Sauer (2014). How to switch a fluorophore: from undesired blinking to controlled photoswitching. *Chemical Society reviews* 43(4), 1076–1087.



- Van de Linde, S., S. Wolter, M. Heilemann, and M. Sauer (2010). The effect of photoswitching kinetics and labeling densities on Super-Resolution fluorescence imaging. *Journal of Biotechnology* 149(4), 260–266.
- Van Loan, C. (1978). Computing integrals involving the matrix exponential. *IEEE Trans Automatic Control* 23(3), 395–404.
- Viallefont, A., J. Lebreton, G. A.M. Reboulet, and Gory (1998). Parameter Identifiability and Model Selection in Capture-Recapture Models: A Numerical Approach. *Biometrical Journal* 40, 313–325.
- Vo, B. and W. Ma (2005). A closed-form solution for the probability hypothesis density filter. In *2005 7th International Conference on Information Fusion*.
- Vo, B., S. Singh, and A. Doucet (2003). Sequential Monte Carlo Implementation of the PHD Filter for Multi-target Tracking. In *Proceedings of the Sixth International Conference on Information Fusion*.
- Vo, B., B. Vo, and A. Cantoni (2006). The Cardinalized Probability Hypothesis Density Filter for Linear Gaussian Multi-Target Models. In *2006 40th Annual Conference on Information Sciences and Systems*.
- Vo, B.-N., S. Singh, and A. Doucet (2005). Sequential Monte Carlo Methods for Multi-target Filtering with Random Finite Sets. *IEEE Transactions on Aerospace and Electronic Systems* 41(4), 1224–1245.
- Vogelsang, J., C. Steinhauer, C. Forthmann, I. H. Stein, B. Person-Skegro, T. Cordes, and P. Tinnefeld (2010). Make them Blink: Probes for Super-Resolution Microscopy. *ChemPhysChem* 11(12), 2475–2490.
- Whiteley, N., S. J. Godsill, and S. Singh (2009). An approximate likelihood method for estimating static parameters in multi-target tracking models. In *Inference and Learning in Dynamic Models*. Cambridge University Press.
- Whiteley, N., S. Singh, and S. Godsill (2010). Auxiliary Particle Implementation of Probability Hypothesis Density Filter. *IEEE Transactions on Aerospace and Electronic Systems* 46(3), 1437–1454.

- Wiemhöfer, M., C. Thiel, J. Klingauf, and R. Chow (2012). Probing Protein-Protein Interactions on the NM Scale using TIRF-PALM. *Biophysical Journal* 102(3), 182.
- Yıldırım, S., L. Jiang, S. Singh, and T. A. Dean (2015). Calibrating the Gaussian multi-target tracking model. *Statistics and Computing* 25(3), 595–608.

# LIST OF ABBREVIATIONS & SYMBOLS

SMLM	Single Molecule Localisation Microscopy
HMM	Hidden Markov Model
PSHMM	Photo-Switching Hidden Markov Model
MCMC	Markov Chain Monte Carlo
PALM	Photo-activated Localisation Microscopy
(d)STORM	(Direct) Stochastic Optical Reconstruction Microscopy
PSF	Point Spread Function
FFT	Fast Fourier Transform
EM	Expectation Maximisation
PSARP	Photo-Switching Alternating Renewal Process
BIC	Bayesian Information Criterion
(R)MSE	(Root) Mean Squared Error
PMF	Probability Mass Function
PGF	Probability Generating Function
DFT	Discrete Fourier Transform
MAP	Maximum a Posteriori
HDR	Highest Density Region
RFS	Random Finite Set
SMC	Sequential Monte Carlo
PDF	Probability Density Function
DIC	Deviance Information Criterion
$X(t)$	Continuous time photo-switching process
$Y_n$	Discrete time observation process
$\mathbb{R}_{\geq 0}$	Non-negative real numbers
$\mathbb{R}_{> 0}$	Positive real numbers
$\mathbb{Z}_{\geq 0}$	Non-negative integers
$\mathbb{Z}_{> 0}$	Positive integers
$-T^*$	Time at the beginning of an experiment
$d$	Number of multiple dark states
$\mathcal{S}_X$	State space of $X(t)$
$\bar{\mathcal{S}}_X$	State space of $X(t)$ without the photo-bleached state
$G$	Generator matrix of $X(t)$
$e^{Gt}$	Transition matrix of $X(t)$ over time $t$
$\lambda_G$	Vector of photo-switching rates in $G$
$\nu_X$	Probability mass of $X(t)$ at time 0
$\nu_X^*$	Probability mass of $X(t)$ at time $-T^*$
$\emptyset$	Empty set
$\mathcal{M}_S^d$	PSHMM model with $d + 1$ dark states restricted on the set $S \subseteq \mathcal{S}_X$ from where photo-bleaching is accessible

$\mathbb{1}(\cdot)$	Indicator function
$\Delta$	Frame acquisition time
$\delta$	Time threshold for a detection of a molecule in a frame
$\omega$	Probability a molecule is falsely observed in a frame
$\mathbb{S}^d$	$d$ -dimensional simplex of probability vectors
$\boldsymbol{\theta}^{(d)}$	Parameter vector containing $\lambda_G, \nu_X, \delta$ when $\omega = 0$
$\Theta^{(d)}$	Parameter space for $\boldsymbol{\theta}^{(d)}$
$\boldsymbol{\theta}_\omega^{(d)}$	Parameter vector when $\omega > 0$
$\Theta_\omega^{(d)}$	Parameter space for $\boldsymbol{\theta}_\omega^{(d)}$
$B_{\Delta}^{(0)}, B_{\Delta}^{(1)}$	Transmission matrices over a frame when $\omega = 0$
$B_{\Delta}^{*(0)}, B_{\Delta}^{*(1)}$	Transmission matrices over a frame when $\omega > 0$
$\mathbf{e}_n^p$	$p$ th canonical (standard) basis vector of $\mathbb{R}^n$
$\odot$	Hadamard (element-wise) product between two matrices
$N_F$	Number of imaged frames
$M$	Number of imaged molecules
$\mathcal{N}(\boldsymbol{\mu}, \Sigma)$	Multivariate Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance $\Sigma$
$N_l$	Cumulative number of observed localisations
$S_n$	Discrete time cumulative localisation process
$\mathcal{S}_{S_n}$	State space of $S_n$
$\gamma_{S_{N_F}}(\cdot)$	Characteristic function of $S_{N_F}$
$\hat{\boldsymbol{\theta}}_\omega^{(d)}$	Maximum likelihood estimate of $\boldsymbol{\theta}_\omega^{(d)}$
$\mathcal{F}(\mathcal{C})$	Space of all finite subsets of $\mathcal{C}$
$\bar{\mu}(\cdot)$	Reference measure for random finite sets
$v(\cdot)$	Intensity measure
$C_n$	Parent point process
$Z_n$	Offspring/observed point process
$Z^{(n)}$	All offspring/observation sets collected up to and including time $n$
$B_n$	Birth point process
$A_n$	False positive point process
$\lambda_M$	(Apriori) expected number of imaged molecules
$\alpha$	(Apriori) expected number of false positive localisations in a frame
$C^*$	Implicit (limiting) parent point process
$\boldsymbol{\theta}_\alpha^{(d)}$	Parameter vector when $\alpha > 0$
$\Theta_\alpha^{(d)}$	Parameter space for $\boldsymbol{\theta}_\alpha^{(d)}$
$\boldsymbol{\theta}'_\alpha^{(d)}$	Transformed parameter vector of $\boldsymbol{\theta}_\alpha^{(d)}$