SUPPLEMENTARY MATERIALS
to
"A HIDDEN MARKOV MODEL APPROACH TO CHARACTERIZING THE PHOTO-SWITCHING BEHAVIOR OF FLUOROPHORES"

Lekha Patel[1], Nils Gustafsson[2], Yu Lin[3], Raimund Ober[4,5], Ricardo Henriques[2,6] and Edward Cohen[1*]

[1]Department of Mathematics, Imperial College London, UK.
[2]MRC Laboratory for Molecular Cell Biology, University College London, UK.
[3]Cell Biology and Biophysics, European Molecular Biology Laboratory, Heidelberg, Germany.
[4]Department of Biomedical Engineering, Texas A&M, USA.
[5]Centre for Cancer Immunology, Faculty of Medicine, University of Southampton, UK
[6]The Francis Crick Institute, London, UK.

We have provided an array of supplementary materials supporting this paper. A brief outline of those which follow are (in order):

S1 A proof that the observation process is non-Markovian.

S2 A full derivation of the transmission probabilities needed in the computation of the PSHMM log-likelihood function, suitable for any number of hidden dark states.

S3 An algorithm for computing the transmission matrices, suitable for any number of hidden dark states.

S4 A detailed discussion on the statistical properties of the PSHMM estimator.

S5 A detailed discussion on the exponential fitting estimator.

S6 A detailed discussion on implementing the PSHMM estimator.

*Corresponding author. Dr Edward Cohen, Department of Mathematics, Imperial College London, South Kensington Campus, London SW7 2AZ. Email: e.cohen@imperial.ac.uk.

S7 A detailed methods section on simulating and analyzing image data. This followed by an analysis of further simulations studies, in the case of a single dark state. Plots and comparisons with the exponential fitting method are provided.

S1 Non-Markovianity of the observation process.

In this section, we will prove that the observation process $\{Y_n\}$ as defined by equation (2) in the main text 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 $m \in \mathbb{Z}_{\geq 0}$ and paths to the absorption state 2, as is depicted in Figure 4 of the main text. This property provides the crucial basis of the PSHMM inference (see Section 3 of the main text) we have presented and conducted in this paper.

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

Then fixing $\Delta > 0$ and any $\delta \in [0, \Delta)$, the process $\{Y_n : n \in \mathbb{Z}_{\geq 0}\}$ generated by $\{X(t) : t \in \mathbb{R}_{\geq 0}\}$ as defined in (2) of the main text from all models M_A^m, is not a Markov Chain of any order.

Proof. For all $m \in \mathbb{Z}_{\geq 0}$, any $\boldsymbol{\lambda} = \begin{pmatrix} \lambda_{01} & \lambda_{00_1} & \lambda_{0_1 1} & \lambda_{0_1 0_2} & \ldots & \lambda_{0_m 1} & \lambda_{10} \end{pmatrix}^\top \in \mathbb{R}_{>0}^{2m+2}$ and any $\boldsymbol{\mu} = \begin{pmatrix} \mu_0 & \ldots & \mu_{0_m} & \mu_1 \end{pmatrix}^\top \in \mathbb{R}_{\geq 0}^{m+2}$ (as characterized by the model M_A^m), we define for $i, j \in \mathcal{S}_X$ and $n \in \mathbb{N}$

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

We consider E_n^l to be the event that $l \in \mathcal{S}_Y = \{0, 1\}$ is observed in the nth 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 $n \in \mathbb{N}$, we will show that the quantity $S(n)$, describing the probability of observing a 1 in the nth frame given observations of 1s in all previous $n-1$ frames is dependent on the full history of the process $\{Y_n\}$ from time $n = 0$. Using the notation defined above, we have that $S(n) = \mathbb{P}(E_n^1 | \cap_{i=0}^{n-1} E_i^1)$, with $S(0) = \mathbb{P}(E_0^1)$. 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 absorption state 2 would result in no observation

of the fluorophore. Using the Markov property of $\{X(t)\}$ and Bayes' theorem, we obtain that

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

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

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

By letting $\boldsymbol{\nu}_X$ be the initial probability mass function for $\{X(t)\}$ whereby $(\boldsymbol{\nu}_X)_i = \mathbb{P}(X(0)=i)$ $(i\in\mathcal{S}_X)$, and considering iterating the conditional probabilities in (2) backwards in time, we obtain the relationship

$$S(n) = \frac{\sum_j \bar{b}_{j,\Delta}^{(1)}\sum_{k_{n-1}} b_{k_{n-1}j,\Delta}^{(1)}\sum_{k_{n-2}}\cdots\sum_{k_1}\left(\prod_{i=2}^{n-1} b_{k_{n-i}k_{n-i+1},\Delta}^{(1)}\right)\sum_{k_0} b_{k_0k_1,\Delta}^{(1)}(\boldsymbol{\nu}_X)_{k_0}}{\sum_j \bar{b}_{j,\Delta}^{(1)}\sum_{k_{n-2}} b_{k_{n-2}j,\Delta}^{(1)}\sum_{k_{n-3}}\cdots\sum_{k_1}\left(\prod_{i=3}^{n-1} b_{k_{n-i}k_{n-i+1},\Delta}^{(1)}\right)\sum_{k_0} b_{k_0k_1,\Delta}^{(1)}(\boldsymbol{\nu}_X)_{k_0}},$$

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

S2 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 $\{B_\Delta^{(l)}\}_{l=0}^1$ needed in the computation of the PSHMM likelihood. Here, we note from the main text that for observations $l=0,1$

$$B_\Delta^{(l)} = \begin{pmatrix} b_{00,\Delta}^{(l)} & b_{00_1,\Delta}^{(l)} & \cdots & b_{00_m,\Delta}^{(l)} & b_{01,\Delta}^{(l)} & b_{02,\Delta}^{(l)} \\ b_{0_10,\Delta}^{(l)} & b_{0_10_1,\Delta}^{(l)} & \cdots & b_{0_10_m,\Delta}^{(l)} & b_{0_11,\Delta}^{(l)} & b_{0_12,\Delta}^{(l)} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ b_{0_m0,\Delta}^{(l)} & b_{0_m0_1,\Delta}^{(l)} & \cdots & b_{0_m0_m,\Delta}^{(l)} & b_{0_m1,\Delta}^{(l)} & b_{0_m2,\Delta}^{(l)} \\ b_{10,\Delta}^{(l)} & b_{10_1,\Delta}^{(l)} & \cdots & b_{10_m,\Delta}^{(l)} & b_{11,\Delta}^{(l)} & b_{12,\Delta}^{(l)} \\ 0 & 0 & 0 & 0 & \cdots & b_{22,\Delta}^{(l)} \end{pmatrix}, \tag{3}$$

where

$$
\begin{aligned}
b^{(l)}_{ij,\Delta} &:= \mathbb{P}(Y_n = l, X((n+1)\Delta) = j | X(n\Delta) = i) \\
&= \mathbb{P}(Y_0 = l, X(\Delta) = j | X(0) = i) \qquad i, j \in \mathcal{S}_X,\ l \in \mathcal{S}_Y, \\
b^{(l)}_{22,\Delta} &= \mathbb{1}_{\{0\}}(l).
\end{aligned} \tag{4}
$$

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 $m \in \mathbb{Z}_{\geq 0}$.

In particular, when $\{Y_n\}$ is recorded over exposure times Δ for each frame, we will consider deriving the following transmission probability function in (4), holding for any $i, j \in \mathcal{S}_X = \{0, 0_1, \ldots, 0_m, 1, 2\}$ 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 1 (provided in Section S3) that computes the elements of matrices in (3) will be provided. Finally, we will consider specifically deriving the form of transmission probabilities when $m = 0$, an experimentally useful and mathematically complete exercise.

On the notation we will use in this chapter, we will make extensive use of the following: Firstly, $\mathcal{S}_X = \{0, 0_1, \ldots, 0_m, 1, 2\}$ 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 absorption 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 $(M)_{(i_1:i_2),(j_1:j_2)}$ to be the matrix filled with rows i_1 to i_2 and columns j_1 to j_2 of any matrix M, and $(M)_{i_1,j_1}$ to be the (i_1, j_1)th entry of M.

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_0, unless stated otherwise.

S2.1 Overview.

In the following, we will describe key mathematical concepts, with particular reference to multivariate counting processes and the idea of *labeling sets* that are needed to compute (4).

In the most general setting when $m \in \mathbb{Z}_{\geq 0}$, we compute (4) by conditioning on the number of transitions made by $\{X(t)\}$ between states in a labeling set $\mathcal{R}$ before time Δ. Here, $\mathcal{R}$

considers transitions from state p to state q and denotes a set of *ordered* index pairs (p, q) with $p, q \in \mathcal{S}_X$, (see Minin and Suchard (2007)). For example, labeling all transitions from state 0 would imply that $\mathcal{R}$ is equal to the set $\{(0, 0_1), (0, 1), (0, 2)\}$.

We 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 labeling set $\mathcal{R}$ that have occurred by $\{X(t)\}$ before time t. This process has state space $\mathbb{Z}_{\geq 0}$.

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

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

$$q_{ij}(\mathbf{k}, \Delta) = \mathbb{P}(\mathbf{N}_{\boldsymbol{\mathcal{R}}_n^{ij}}(\Delta) = \mathbf{k}, X(\Delta) = j | X(0) = i)$$
$$\xi_{ij}(l, \mathbf{k}, \Delta) = \mathbb{P}(Y_0 = j | \mathbf{N}_{\boldsymbol{\mathcal{R}}_n^{ij}}(\Delta) = \mathbf{k}, X(0) = i, X(\Delta) = j),$$

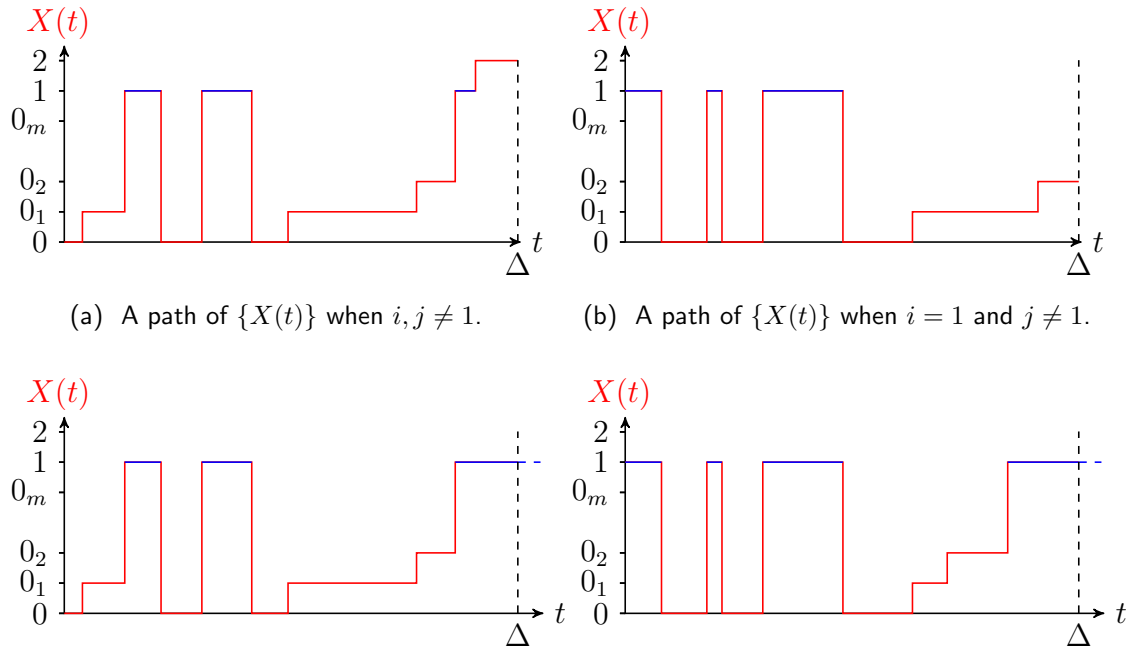
and write (by conditioning n times) the transmission probabilities in the form

$$b_{ij,\Delta}^{(l)} = \sum_{\mathbf{k} \in \mathbb{Z}^n_{\geq 0}} q_{ij}(\mathbf{k}, \Delta) \xi_{ij}(l, \mathbf{k}, \Delta). \tag{5}$$

Although we are free to choose any $\boldsymbol{\mathcal{R}}_n^{ij}$, poor choices may lead to intractability of the above probabilities. We will thus now describe a method as to how one can choose $\boldsymbol{\mathcal{R}}_n^{ij}$ for effective computation of (5).

The first term $q_{ij}(\mathbf{k}, \Delta)$ over any $\boldsymbol{\mathcal{R}}_n^{ij}$ can be computed using Laplace transforms. However, identifying which $\boldsymbol{\mathcal{R}}_n^{ij}$ is needed to compute (5) 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 $\boldsymbol{\mathcal{R}}_n^{ij}$. An observation of a fluorophore (as defined in (2) of the main text) is dependent on the total time spent in the On state 1 and as such we endeavor to characterize this time using the transitions that have occurred within the interval $[0, \Delta)$.

Figures S1a - S1d show four possible paths of $\{X(t)\}$ within this time interval. Firstly, Figures S1a and S1b highlight two paths when $X(\Delta) \neq 1$. In Figure S1a when $X(0) \neq 1$, each of the three time pieces in state 1 are exponentially independently and identically distributed



(a) A path of $\{X(t)\}$ when $i, j \neq 1$.

(b) A path of $\{X(t)\}$ when $i = 1$ and $j \neq 1$.

(c) A path of $\{X(t)\}$ when $i \neq 1$ and $l = 1$.

(d) A path of $\{X(t)\}$ when $i, l = 1$.

Figure S1: Figures S1a and S1b 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 S1c and S1d 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.

(iid) (with scale parameter σ_1), and thus the total time spent in this state is characterized by an Erlang$(3, \sigma_1)$ density (truncated over the interval $[0, \Delta)$). This is also true for the path shown in Figure S1b when $X(0) = 1$, since the memoryless 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 $b^{(l)}_{ij,\Delta}$ 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 $(n = 1)$ labeling set $\mathcal{R}^0 = \{(0, 1), (0_1, 1), \ldots, (0_m, 1)\}$ and setting $\boldsymbol{\mathcal{R}}^{ij}_1 = \mathcal{R}^0$, 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 (5) as

$$b^{(l)}_{ij,\Delta} = \sum_{k \in \mathbb{Z}_{\geq 0}} q_{ij}(k, \Delta)\xi_{ij}(l, k, \Delta). \tag{6}$$

Figures S1c and S1d highlight two paths when $X(\Delta) = 1$. In Figure S1c 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 S1d 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 $X(\Delta) = 1$ (absorption has almost surely not occurred), we can consider the holding times in each of the dark states $0_0, \ldots, 0_m$. Using similar arguments to before, the holding time in each dark state 0_p $(p = 0, \ldots, m)$ is exponentially distributed (with scale parameter σ_{0_p}), and the sum of all times spent in state 0_p is characterized by a truncated Erlang(k_p, σ_{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 $0_p \to 1$ and $0_p \to 0_{p+1}$ (for all $p \neq m$) transitions over the interval $[0, \Delta)$. Subtracting the sum of all times spent in any dark state (characterized by the truncated sum of $m + 1$ independent but non-identical Erlang(k_p, σ_{0_p}) densities) from Δ, recovers the total time spent in the On state, as depicted in Figure S2.

The utilization of $m + 1$ Erlang densities in this setting thus invokes conditioning on a multivariate $(n = m + 1)$ counting process $\boldsymbol{\mathcal{R}}^{i1}_{m+1}$. We can determine k_p by considering the labeling sets $\mathcal{R}^1_p = \{(0_p, 1), (0_p, 0_{p+1})\}$ for $p = 0, \ldots, m-1$ and $\mathcal{R}^1_m = \{(0_m, 0_{m+1})\}$. By defining the vector

$$\mathbf{k} = \begin{pmatrix} k_0 & k_1 & \ldots & k_{m-1} \end{pmatrix}^\top, \tag{7}$$

we thus consider conditioning on $\{\mathbf{N}_{\boldsymbol{\mathcal{R}}^{i1}_{m+1}}(t) = \mathbf{k}\}$, using $\boldsymbol{\mathcal{R}}^1_{m+1} := \{\mathcal{R}^1_0, \mathcal{R}^1_1, \ldots, \mathcal{R}^1_m\}$, where we set $\boldsymbol{\mathcal{R}}^{i1}_{m+1} = \boldsymbol{\mathcal{R}}^1_{m+1}$ for each $i \in \bar{\mathcal{S}}_X$.

This case, which we refer to as *Case* $j = 1$ allows (5) to be written as

$$b^{(l)}_{i1,\Delta} = \sum_{\mathbf{k} \in \mathbb{Z}^{m+1}_{\geq 0}} q_{i1}(\mathbf{k}, \Delta) \xi_{i1}(l, \mathbf{k}, \Delta). \tag{8}$$

We will now formally discuss the mathematical framework needed to compute $q_{ij}(\mathbf{k}, \Delta)$ and $\xi_{ij}(l, \mathbf{k}, \Delta)$ in both cases $j \neq 1$ and $j = 1$.

S2.2 Case $j \neq 1$.

For the $j \neq 1$ case, we will derive the forms of $q_{ij}(k, \Delta)$ using Laplace transforms and $\xi_{ij}(l, k, \Delta)$ using the distributions of state holding times as outlined in Section S2.1.

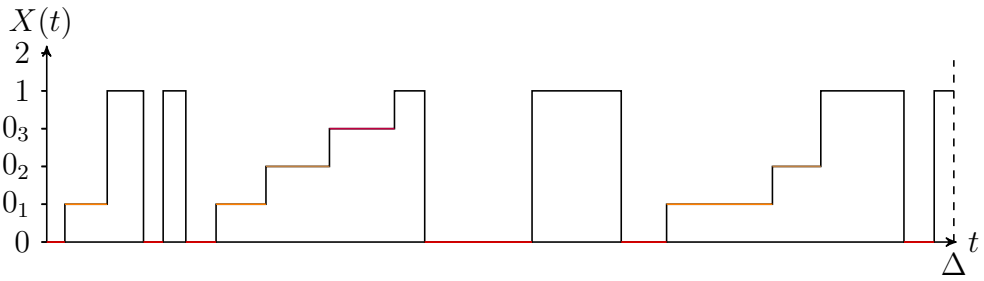


Figure S2: Highlighted holding times in each dark state: 0 (red), 0_1 (orange), 0_2 (brown) and 0_3 (purple) from a possible path when $X(\Delta) = 1$ and $m = 3$. All individual time pieces in each dark state are distinctly exponentially distributed. Subtracting the sum of all these (highlighted) time pieces from Δ gives the total time spent in state 1.

S2.2.1 Computation of $q_{ij}(k, \Delta)$.

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

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

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

$$\begin{aligned}
q_{i0}(k, t+h) &= (1 - \sigma_0 h) q_{i0}(k, t) + \lambda_{10} q_{i1}(k, t) h + o(h) \\
q_{i0_p}(k, t+h) &= (1 - \sigma_{0_p} h) q_{i0_p}(k, t) + \lambda_{0_{p-1}0_p} q_{i0_{p-1}}(k, t) h + o(h) \quad p = 1 \ldots m \\
q_{i1}(k, t+h) &= (1 - \sigma_1 h) q_{i1}(k, t) + \sum_{p=0}^{m} \lambda_{0_p 1} q_{i0_p}(k-1, t) h + o(h) \\
q_{i2}(k, t+h) &= q_{i2}(k, t) + \sum_{p=0}^{m} \mu_{0_p} q_{i0_p}(k, t) h + \mu_1 q_{i1}(k, t) h + o(h).
\end{aligned} \tag{9}$$

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

$$Q_t(k) = \begin{pmatrix}
q_{00}(k, t) & q_{00_1}(k, t) & \ldots & q_{01}(k, t) \\
q_{0_1 0}(k, t) & q_{0_1 0_1}(k, t) & \ldots & q_{0_1 1}(k, t) \\
\ldots & \vdots & \ddots & \ldots \\
q_{10}(k, t) & q_{10_1}(k, t) & \ldots & q_{11}(k, t)
\end{pmatrix}. \tag{10}$$

Moreover, we define the *end-absorbed state* vector as $\bar{\mathbf{Q}}_t(k) = \begin{pmatrix} q_{02}(k,t) & \dots & q_{12}(k,t) \end{pmatrix}^\top$. It is clear from (9) that

$$\bar{\mathbf{Q}}_t(k) = \left(\int_0^t Q_s(k) \mathrm{d}s \right) \boldsymbol{\mu}, \tag{11}$$

with $\boldsymbol{\mu} = \begin{pmatrix} \mu_0 & \dots & \mu_{0_m} & \mu_1 \end{pmatrix}^\top \equiv (G)_{(1:m+2),m+3}$ as the $(m+2)$ vector of absorbing rates. Recovering $\bar{\mathbf{Q}}_t(k)$ will therefore require an expression for $Q_t(k)$.

Lemma 1. *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 λ_{ij}. Letting $Q_t(k)$ denote the matrix with (i,j)th entry $q_{ij}(k,t) = \mathbb{P}(N_{\mathcal{R}}(t) = k, X(t) = j | X(0) = i)$, which counts transitions in the some labeling set $\mathcal{R}$ in the time interval $[0,t)$, the Laplace transformed matrix (for all $k \in \mathbb{Z}_{\geq 0}$) $F_s(k) = \int_0^\infty \mathrm{e}^{-st} Q_t(k) \mathrm{d}t$ takes the form*

$$F_s(k) = (sI_n - G_{\bar{\mathcal{R}}})^{-1} \left(G_{\mathcal{R}} (sI_n - G_{\bar{\mathcal{R}}})^{-1} \right)^k. \tag{12}$$

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 $F_s(k) = \int_0^\infty \mathrm{e}^{-st} Q_t(k) \mathrm{d}t$ can be obtained by leveraging the result from Minin and Suchard (2007) given in Lemma 1. In particular, the result on $F_s(k)$ in (12) 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 (12), we can use the sub-Markovian generator of $\{X(t)\}$: $G_S \equiv (G)_{(1:m+2),(1:m+2)}$ which is an $(m+2) \times (m+2)$ matrix gained by deleting the $(m+3)$th row and column from the generator G in (1) from the main text [1]. We define $G_{S,\mathcal{R}^0}$ to be the $(m+2) \times (m+2)$ matrix filled with the transition rates *only* in the labeling set of interest $\mathcal{R}^0$. $G_{S,\mathcal{R}^0}$ thus has (i,j)th entry

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

[1] 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 σ_p with σ_q in the diagonal entries of G.

Letting $G_{S,\bar{\mathcal{R}}^0} = G_S - G_{S,\mathcal{R}^0}$, we use Lemma 1 to obtain that for any $k \in \mathbb{Z}_{\geq 0}$,

$$F_s(k) = (sI_{m+2} - G_{S,\bar{\mathcal{R}}^0})^{-1} \left(G_{S,\mathcal{R}^0}(sI_{m+2} - G_{S,\bar{\mathcal{R}}^0})^{-1}\right)^k. \tag{13}$$

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 $G_{S,\mathcal{R}^0}$ and G_S commute. Nonetheless, the non-commutative properties of $G_{S,\mathcal{R}^0}$ 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 the computational form $Q_\Delta(k)$.

Remark 1. *When* $k = 0$, $Q_\Delta(0) = \mathrm{e}^{G_{S,\bar{\mathcal{R}}^0}\Delta}$. *To compute,* $\bar{\mathbf{Q}}_\Delta(0)$, *we note from Van Loan (1978) that* $\bar{\mathbf{Q}}_\Delta(0) = \left(\int_0^\Delta \mathrm{e}^{G_{S,\bar{\mathcal{R}}^0}s}\mathrm{ds}\right)\boldsymbol{\mu} = (\mathrm{e}^{A\Delta})_{(i_1:i_2),(i_2+1,i_3)}\boldsymbol{\mu}$ *with* $i_1 = 2m+5, i_2 = 3(m+2), i_3 = 4(m+2)$; $A = \begin{bmatrix} A_1 & \mathbf{0}_{2(m+2)}\mathbf{0}_{2(m+2)}^\top \\ \mathbf{0}_{2(m+2)}\mathbf{0}_{2(m+2)}^\top & A_2 \end{bmatrix}$, $A_1 = \begin{bmatrix} -G_{S,\bar{\mathcal{R}}}^\top & I_{m+2} \\ \mathbf{0}_{m+2}\mathbf{0}_{m+2}^\top & -G_{S,\bar{\mathcal{R}}}^\top \end{bmatrix}$ *and* $A_2 = \begin{bmatrix} G_{S,\bar{\mathcal{R}}} & I_{m+2} \\ \mathbf{0}_{m+2}\mathbf{0}_{m+2}^\top & \mathbf{0}_{m+2}\mathbf{0}_{m+2}^\top \end{bmatrix}$.

S2.2.2 Computation of $\xi_{ij}(l, k, \Delta)$.

In this section, we will relate $\{X(t)\}$ to a renewal sequence. This allows us to characterize 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 1.

Definition 1. *Let* $\mathbf{U} = (U_1, U_2, \ldots)$ *denote the successive lengths of time* $\{X(t)\}$ *is in the On state* 1 *and let* $\mathbf{D} = (D_1, D_2, \ldots)$ *denote the successive lengths of time that the process is not in state* 1 *before absorption. We define the* **photo-switching alternating renewal process (PSARP)** *as being characterized by the sequence of iid random vectors* Γ, *where*

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

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* $R_0, R_1, \ldots$.

For full characterization, we need to consider the distributions of $\{U_n\}$ and $\{D_n\}$. Clearly each $U_n \overset{iid}{\sim} \exp(\sigma_1)$. 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 $\mathcal{S}_J = \{0, 1, \ldots, m\}$, which counts the number of jumps between dark states $(0_p \to 0_{p+1} \;\; p = 0, \ldots m-1)$ during the nth renewal. We have that

$$D_0|J_0 \overset{d}{=} \begin{cases} \sum_{s=0}^{J_0} D_0^s & \text{if } X(0) = 1 \\ \sum_{s=p}^{J_0} D_0^s & \text{if } X(0) = 0_p \;\; p = 0, \ldots, m, \end{cases}$$

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

Once Y_0 has been recorded, $\{X(t)\}$ continues after time Δ without observation. Section S2.1 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 1) for some $p = 0, \ldots, m$, 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 $(U_0, U_1, \ldots U_{k-1})$. By construction, these k pieces are iid exponentially distributed and their *sum* has an Erlang distribution with the shape and rate parametrization $\Upsilon(k) = \sum_{i=0}^{k-1} U_i \sim \text{Erlang}(k, \sigma_1)$. 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 $(U_0, U_1, \ldots, U_{k-1}, U_k)$ and the total time spent in state 1 is governed by $\Upsilon(k+1)$. Since $Y_0 = 0$ if and only if the total time spent in the On state within the interval $[0, \Delta)$ is less than or equal to δ, for $k \in \mathbb{N}$, $i \in \bar{\mathcal{S}}_X$ and $j \neq 1$ we have

$$\begin{aligned} \xi_{ij}(0, k, \Delta) &= \mathbb{P}(\Upsilon(k + \mathbb{1}_{\{1\}}(i)) \leq \delta | \Upsilon(k + \mathbb{1}_{\{1\}}(i)) \leq \Delta) \\ &= \frac{1 - \sum_{m=0}^{k+\mathbb{1}_{\{1\}}(i)-1} \frac{(\sigma_1 \delta)^m}{m!} \mathrm{e}^{-\sigma_1 \delta}}{1 - \sum_{m=0}^{k+\mathbb{1}_{\{1\}}(i)-1} \frac{(\sigma_1 \Delta)^m}{m!} \mathrm{e}^{-\sigma_1 \Delta}} \qquad (14) \\ \xi_{1j}(0, 0, \Delta) &= \frac{1 - \mathrm{e}^{-\sigma_1 \delta}}{1 - \mathrm{e}^{-\sigma_1 \Delta}} \qquad (15) \\ \xi_{ij}(1, k, \Delta) &= 1 - \xi_{ij}(0, k, \Delta). \qquad (16) \end{aligned}$$

For computational purposes [2], we form (analogously to $Q_\Delta(k)$) the $(m+2) \times (m+1)$ matrix

[2] The matrix representations $Q_\Delta(k)$ and $\Xi_\Delta^l(k)$ are used in Algorithm 1 in Section S3 for computing transmission matrices in the form (3).

$\Xi^l_\Delta(k)$ with elements gained from (14)-(16). For $i, j \in \bar{\mathcal{S}}_X, j \neq 1$ and $l \in \mathcal{S}_Y$ we define

$$\Xi^l_\Delta(k) = \begin{pmatrix} \xi_{00}(l,k,\Delta) & \xi_{00_1}(l,k,\Delta) & \dots & \xi_{00_m}(l,k,\Delta) \\ \xi_{0_10}(l,k,\Delta) & \xi_{0_10_1}(l,k,\Delta) & \dots & \xi_{0_10_m}(l,k,\Delta) \\ \dots & \vdots & \ddots & \dots \\ \xi_{10}(l,k,\Delta) & \xi_{10_1}(l,k,\Delta) & \dots & \xi_{10_m}(l,k,\Delta) \end{pmatrix}, \tag{17}$$

and $\bar{\mathbf{\Xi}}^l_\Delta(k)$ to be the $(m+2)$ end state vector

$$\bar{\mathbf{\Xi}}^l_\Delta(k) = \begin{pmatrix} \xi_{02}(l,k,\Delta) & \dots & \xi_{12}(l,k,\Delta) \end{pmatrix}^\top. \tag{18}$$

S2.3 Case $j = 1$.

Using the same mathematical tools as in the case $j \neq 1$ (Section S2.2), 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 $b^{(l)}_{i1,\Delta}$.

S2.3.1 Equivalent definition.

For computational feasibility, we will in this section seek an alternative method for computing the sum in (8).

Using the form of $\mathbf{k}$ in (7), we let $k \in \mathbb{Z}_{>0}$ be equal to $\mathbf{k}^\top \mathbf{1}_{m+1}$: 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 $1 \to 0 \to 0_1 \to \cdots \to 0_p$ for some $p = 0, \dots, m$ 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 $k_p \geq k_{p+1} \geq \cdots \geq k_m$ 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 $k_0 \geq k_1 \geq \cdots \geq k_{p-1} \geq k_p - 1$. When $i = 0_p$, the set of feasible transitions which we denote as $\mathcal{C}^i_k$, is of the form

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

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

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

probabilities

$$b^{(l)}_{i1,\Delta} = \sum_{k=0}^{\infty} \sum_{\mathbf{k}\in\mathcal{C}^i_k} q_{i1}(\mathbf{k},\Delta)\xi_{i1}(l,\mathbf{k},\Delta). \tag{19}$$

S2.3.2 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 $\mathbf{k}\in\mathbb{Z}^{m+1}_{\geq 0}$

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

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

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

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

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

$$(s+\sigma_0)f_{i0}(\mathbf{k},s) = \lambda_{10}f_{i1}(\mathbf{k},s) \tag{20}$$

$$(s+\sigma_{0_p})f_{i0_p}(\mathbf{k},s) = \lambda_{0_{p-1}0_p}f_{i0_{p-1}}(\mathbf{k}-\mathbf{e}^p_{m+1},s) \quad p=1\ldots m, \tag{21}$$

$$(s+\sigma_1)f_{i1}(\mathbf{k},s) = \sum_{p=0}^{m}\lambda_{0_p1}f_{i0_p}(\mathbf{k}-\mathbf{e}^{p+1}_{m+1},s), \tag{22}$$

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

For $\mathbf{k}\in\mathbb{Z}^{m+1}_{\geq 0}$, this yields the recursion [3]

$$f_{i1}(\mathbf{k},s) = \frac{\lambda_{10}}{s+\sigma_1}\sum_{p=0}^{m}\frac{\lambda_{0_p1}\prod_{q=0}^{p-1}\lambda_{0_q0_{q+1}}}{\prod_{q=0}^{p}(s+\sigma_{0_q})}f_{i1}\left(\mathbf{k}-\sum_{r=0}^{p}\mathbf{e}^{r+1}_{m+1},s\right), \tag{23}$$

[3] To avoid division by zero, when $\sigma_p = \sigma_q$ for any $p\neq q$, σ_p must be replaced with σ_q in the Laplace transforms.

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

Remark 2. *Using equations (20) - (22), we can obtain the initializations:* $f_{i1}(\mathbf{0}_{m+1}, s) = \frac{\mathbb{1}_{\{1\}}(i)}{s+\sigma_1}$, $f_{0_p 1}(\mathbf{e}_{m+1}^{p+1}, s) = \frac{\lambda_{0_p 1}}{(s+\sigma_{0_p})(s+\sigma_1)}$ *for* $p = 1, \ldots, m$ *and* $f_{11}(\mathbf{e}_{m+1}^{1}, s) = \frac{\lambda_{10}\lambda_{01}}{(s+\sigma_0)(s+\sigma_1)^2}$.

The inverse Laplace transform of (23), and its initializations to recover $q_{i1}(\mathbf{k}, t)$ and thus $q_{i1}(\mathbf{k}, \Delta)$ are left as a computational exercise.

Remark 3. *When* $k = 0$, $\mathbf{k}$ *can only be the vector* $\mathbf{0}_{m+1}$ *and hence* $q_{i1}(\mathbf{0}_{m+1}, \Delta) = 0$ *for all* $i \neq 1$*; it is however easily seen that* $q_{11}(\mathbf{0}_{m+1}, \Delta) = \mathrm{e}^{-\sigma_1 \Delta}$.

S2.3.3 Computation of $\xi_{i1}(l, \mathbf{k}, \Delta)$.

Following on from the analysis presented in Sections S2.1 and S2.2.2, we will in this section endeavor to compute $\xi_{i1}(l, \mathbf{k}, \Delta)$. In particular, by using the exponential time pieces in the dark states $0 \ldots, 0_m$, we can recover the total time spent in state 1 by subtracting the total time spent in the dark states from Δ.

If $X(0) = i$ for some $i \in \bar{\mathcal{S}}_X$, $X(\Delta) = 1$ and $\mathbf{N}_{\boldsymbol{\mathcal{R}}_{m+1}^{i1}}(\Delta) = \mathbf{k}$ as defined in (7), we study the number of transitions $k_0, \ldots, k_m$ from each dark state $0_0, \ldots, 0_m$. There are exactly k_p exponential time pieces in each dark state 0_p for $p = 0, \ldots, m$ with a total time $\upsilon_p^i(k_p)$ characterized by an Erlang distribution. For any p, we therefore have (using the PSARP construction in Definition 1) that $\upsilon_p^i(k_p) = \sum_{s=0}^{k_p} D_s^p \sim \text{Erlang}(k_p, \sigma_{0_p})$.

By defining the event $A_i := \{Y_0 = 0 | \mathbf{N}_{\boldsymbol{\mathcal{R}}_{m+1}^{i1}}(\Delta) = \mathbf{k}, X(0) = i, X(\Delta) = 1\}$, we equivalently have

$$A_i = \left\{ \sum_{p=0}^{m} \upsilon_p^i(k_p) \geq \Delta - \delta \middle| \sum_{p=0}^{m} \upsilon_p^i(k_p) \leq \Delta \right\},$$

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

By further defining $\boldsymbol{\sigma} = \begin{pmatrix} \sigma_0 & \sigma_{0_1} & \ldots & \sigma_{0_m} \end{pmatrix}^\top$, we let $\Phi(\mathbf{k}, \boldsymbol{\sigma})$ denote the sum of $m+1$ independent Erlang distributions, each with shape parameter k_p and rate parameter σ_{0_p} for $p = 0, \ldots m$. For each i, we have $\sum_{p=0}^{m} \upsilon_p^i(k_p) \sim \Phi(\mathbf{k}, \boldsymbol{\sigma})$. If $F_\Phi(\phi | \mathbf{k}, \boldsymbol{\sigma})$ denote its cumulative

distribution function [4], we have for $k \in \mathbb{Z}_{>0}, i \in \bar{\mathcal{S}}_X$ and $\mathbf{k} \in \mathcal{C}_k^i$

$$\xi_{i1}(1, \mathbf{k}, \Delta) = \frac{F_\Phi(\Delta - \delta | \mathbf{k}, \boldsymbol{\sigma})}{F_\Phi(\Delta | \mathbf{k}, \boldsymbol{\sigma})} \tag{24}$$

$$\xi_{i1}(0, \mathbf{k}, \Delta) = 1 - \xi_{i1}(1, \mathbf{k}, \Delta) \tag{25}$$

$$\xi_{11}(1, \mathbf{0}, \Delta) = 1. \tag{26}$$

S2.4 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 (3).

When $i \in \bar{\mathcal{S}}_X, j \neq 1$ and $l \in \mathcal{S}_Y$ we can compute the inverse Laplace transform of (13) to obtain $Q_\Delta^0(k)$, and therefore $\bar{\mathbf{Q}}_\Delta^0(k)$ from (11). Furthermore using (14)-(16) as entries in matrices $\Xi_\Delta^l(k)$ and $\bar{\boldsymbol{\Xi}}_\Delta^l(k)$ (for $l \in \mathcal{S}_Y$) as defined in (17) and (18) enables the computation of transmission probabilities (6). Moreover, we use (24)-(26) with inverse Laplace transforms $q_{i1}(\mathbf{k}, \Delta)$ from (23) in the computation of (8).

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

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

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

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

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

Then considering the transmission matrix representation in (3), we have that each

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

In practice, as is described in Algorithm 1 of Section S3, the above sum, for small rates and Δ,

[4] $F_\Phi(\cdot|\cdot,\cdot)$ can be computed in a recursive fashion using the algorithm presented in Moschopoulos (1985).

will converge extremely quickly. At least in this practical application, computations for $k \geq 3$ are unlikely to be needed.

S2.5 Study: Exact solution 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 $m = 0$.

With no multiple dark states, the counting process which we must consider conditioning on uses the sole labeling set $\{(0,1)\}$, and is therefore univariate. Using the same notation as that from Section S2.1, this labeling set is $\mathcal{R} := \boldsymbol{\mathcal{R}}_1^1 = \mathcal{R}^0 = \{(0,1)\}$. We therefore consider $\{N(t) : t \in \mathbb{R}_{\geq 0}\} := \{N_{\boldsymbol{\mathcal{R}}_1^1}(t)\} = \{N_{\mathcal{R}^0}(t)\}$ 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

$$b_{ij,\Delta}^{(l)} = \sum_{k=0}^{\infty} \xi_{ij}(l,k,\Delta) q_{ij}(l,\Delta), \tag{27}$$

with $q_{ij}(k,\Delta) = \mathbb{P}(N(\Delta) = k, X(\Delta) = j | X(0) = i)$ and $\xi_{ij}(l,k,\Delta) = \mathbb{P}(Y_0 = l | N(\Delta) = k, X(0) = i, X(\Delta) = j)$.

The matrix representation of the transmission matrices $B_\Delta^{(0)}$ and $B_\Delta^{(1)}$ as in (3) 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 3×3 matrices $Q_\Delta(k)$ and $\Xi_\Delta^l(k)$ as

$$Q_\Delta(k) = \begin{pmatrix} q_{00}(k,\Delta) & q_{01}(k,\Delta) & q_{02}(k,\Delta) \\ q_{10}(k,\Delta) & q_{11}(k,\Delta) & q_{12}(k,\Delta) \\ 0 & 0 & \mathbb{1}_{\{0\}}(k) \end{pmatrix} \tag{28}$$

$$\Xi_\Delta^l(k) = \begin{pmatrix} \xi_{00}(l,k,\Delta) & \xi_{01}(l,k,\Delta) & \xi_{02}(l,k,\Delta) \\ \xi_{10}(l,k,\Delta) & \xi_{11}(l,k,\Delta) & \xi_{12}(l,k,\Delta) \\ 0 & 0 & \mathbb{1}_{\{0\}}(k+j) \end{pmatrix}. \tag{29}$$

S2.5.1 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 S2.2.1 and S2.3.2, 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 $a \in \mathbb{R}$, $\mathcal{L}_s^{-1}\left[\frac{1}{(s+a)^{n+1}}\right](t) = \frac{t^n \mathrm{e}^{-at}}{n!}$, we obtain that

$$
q_{ij}(0,t) = \frac{(1-j)\lambda_{10}^{i-j}}{(\sigma_1-\sigma_0)^{i-j}} \mathrm{e}^{-\sigma_0 t} + \frac{i\lambda_{10}^{i-j}}{(\sigma_0-\sigma_1)^{i-j}} \mathrm{e}^{-\sigma_1 t} \quad \sigma_0 \neq \sigma_1 \tag{30}
$$

$$
q_{ij}(0,t) = \frac{\lambda_{10}^{i-j} t^{i-j} \mathrm{e}^{-\sigma_1 t}}{(i-j)!} \mathbb{1}_{\geq 0}(i-j) \quad \sigma_0 = \sigma_1. \tag{31}
$$

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

$$
q_{ij}(k,t) = \frac{\lambda_{01}^k \lambda_{10}^{k+i-j} t^{2k+i-j} \mathrm{e}^{-\sigma_1 t}}{(2k+i-j)!}. \tag{32}
$$

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

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

where

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

We are thus able to obtain the following closed form expressions

$$
q_{ij}(k,t) = \sum_{p=1}^{k+1-j} \frac{{}_{(ij)}A^{0}_{k:(k-p+1-j)}}{(p-1)!} t^{p-1}\mathrm{e}^{-\sigma_0 t} + \sum_{p=1}^{k+i} \frac{{}_{(ij)}A^{1}_{k:(k-p+i)}}{(p-1)!} t^{p-1}\mathrm{e}^{-\sigma_1 t}. \tag{33}
$$

We now consider two cases to compute $q_{i2}(k,t) = \sum_{p=0}^{1}\mu_p \int_0^t q_{ip}(k,s)\mathrm{ds}$. Firstly, when $\sigma_0 = \sigma_1$, we have that

$$
\begin{aligned}
q_{i2}(0,t) &= \sum_{p_1=0}^{1} \frac{\mu_{p_1}\lambda_{10}^{i-p_1}\mathbb{1}_{\geq 0}(i-p_1)}{\sigma_1^{i-p_1+1}} \sum_{p_2=i-p_1+1}^{\infty} \frac{(\sigma_1 t)^{p_2}\mathrm{e}^{-\sigma_1 t}}{p_2!} \\
q_{i2}(k,t) &= \sum_{p_1=0}^{1} \frac{\mu_{p_1}\lambda_{01}^{k}\lambda_{10}^{k+i-p_1}}{\sigma_1^{2k+i-p_1+1}} \sum_{p_2=2k+i-p_1+1}^{\infty} \frac{(\sigma_1 t)^{p_2}\mathrm{e}^{-\sigma_1 t}}{p_2!}.
\end{aligned} \tag{34}
$$

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

$$
\begin{aligned}
q_{i2}(0,t) &= \frac{\mu_0\lambda_{10}^{i}}{(\mu_0+\lambda_{01})(\sigma_1-\sigma_0)^{i}}(1-\mathrm{e}^{-\sigma_0 t}) + \frac{i}{\sigma_1}\left[\frac{\mu_0\lambda_{10}}{(\sigma_0-\sigma_1)}+\mu_1\right](1-\mathrm{e}^{-\sigma_1 t}) \\
q_{i2}(k,t) &= \sum_{p_1=0}^{1}\mu_{p_1}\sum_{p_2=0}^{1}\sum_{p_3=1}^{k+1-p_2(1-p_1)} \frac{{}_{(ip_2)}A^{p_1}_{k:(k-p_3+ip_1+(1-p_2)(1-p_1))}}{\sigma_{p_1}^{p_3}} \sum_{m=p_3}^{\infty}\frac{(\sigma_{p_1}t)^m}{m!}\mathrm{e}^{-\sigma_{p_1}t}.
\end{aligned} \tag{35}
$$

Setting $t=\Delta$ in all equations (30)-(35) yields all entries of $Q_\Delta(k)$ as in (28), for $k \in \mathbb{Z}_{\geq 0}$.

S2.5.2 Computation of $\Xi^l_\Delta(k)$.

Using the same methodology as presented in Sections S2.2.2 and S2.3.3, we will demonstrate computation of $\Xi^l_\Delta(k)$. In particular, since $m=0$ we have for $s \in \mathbb{N}$ that $R_s = U_s + D_s$ is a

PSARP (see Definition 1) with inter-arrival times $\mathbf{R} = (R_1, R_2, \ldots)$. Here, $U_s \overset{iid}{\sim} \exp(\sigma_1)$ and $D_s \overset{iid}{\sim} \exp(\sigma_0)$.

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 $j \neq 1$

$$\xi_{ij}(0, k, \Delta) = \frac{1 - \sum_{m=0}^{k+i-1} \frac{(\sigma_1 \delta)^m}{m!} \mathrm{e}^{-\sigma_1 \delta}}{1 - \sum_{m=0}^{k+i-1} \frac{(\sigma_1 \Delta)^m}{m!} \mathrm{e}^{-\sigma_1 \Delta}} \tag{36}$$

$$\xi_{00}(0, 0, \Delta) = 1 \tag{37}$$

$$\xi_{10}(0, 0, \Delta) = \frac{1 - \mathrm{e}^{-\sigma_1 \delta}}{1 - \mathrm{e}^{-\sigma_1 \Delta}}. \tag{38}$$

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 $(D_1, D_2, \ldots, D_k)$. These k pieces form iid exponential random variables and their sum $\upsilon(k)$ also has an Erlang distribution $\upsilon(k) = \sum_{i=1}^{k} D_i \sim \text{Erlang}(k, \sigma_0)$. Now since the random event

$$\{Y_0 = 0 | N(\Delta) = k, X(0) \in \{0, 1\}, X(\Delta) = 1\} = \{\upsilon(k) \geq \Delta - \delta | \upsilon(k) \leq \Delta\},$$

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

$$\xi_{i1}(1, k, \Delta) = \frac{1 - \sum_{m=0}^{k-1} \frac{(\sigma_0(\Delta - \delta))^m}{m!} \mathrm{e}^{-\sigma_0(\Delta - \delta)}}{1 - \sum_{m=0}^{k-1} \frac{(\sigma_0 \Delta)^m}{m!} \mathrm{e}^{-\sigma_0 \Delta}} \tag{39}$$

$$\xi_{11}(1, 0, \Delta) = 1. \tag{40}$$

We may fill the matrices $\Xi_\Delta^0(k)$ and $\Xi_\Delta^1(k)$ in (29) by using (36)-(40) and the fact that for $i, j \in \bar{\mathcal{S}}_X$, $\xi_{ij}(1, k, \Delta) = 1 - \xi_{ij}(0, k, \Delta)$.

S3 Algorithm to compute transmission matrices.

Algorithm 1 presents the method for computing transmission matrices $B_\Delta^{(0)}$ and $B_\Delta^{(1)}$ as detailed in Section S2, suitable for any $m \geq 0$. Here, we denote $\mathbf{0}_n$ and $\mathbf{1}_n$ to be the $n \times 1$ vectors of zeros and ones respectively and I_n to be the $n \times n$ identity matrix. We denote $(M)_{(i_1:i_2),(j_1:j_2)}$ to be the matrix filled with rows i_1 to i_2 and columns j_1 to j_2 of any matrix M, and $(M)_{i_1,j_1}$ to be the (i_1, j_1)th entry of M. We use the $\odot$ notation to denote the Hadamard (element wise) product between two matrices.

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

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

$G_S = (G)_{(1:m+2),(1:m+2)}$ [a]

$\boldsymbol{\mu} = (G)_{(1:m+2),m+3}$

$\sigma_1 = -(G)_{m+2,m+2}$

$\boldsymbol{\sigma} = -\mathsf{diag}((G)_{(1:m+1),(1:m+1)})$

for $i = 1 : m+1$ **do**

$(G_{S,\mathcal{R}^0})_{i,m+2} = (G_S)_{i,m+2}$

end for

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

$A = \begin{bmatrix} A_1 & \mathbf{0}_{2(m+2)}\mathbf{0}_{2(m+2)}^\top \\ \mathbf{0}_{2(m+2)}\mathbf{0}_{2(m+2)}^\top & A_2 \end{bmatrix}$, where $A_1 = \begin{bmatrix} -G_{S,\bar{\mathcal{R}}^0}^\top & I_{m+2} \\ \mathbf{0}_{m+2}\mathbf{0}_{m+2}^\top & -G_{S,\bar{\mathcal{R}}^0}^\top \end{bmatrix}$ and $A_2 = \begin{bmatrix} G_{S,\bar{\mathcal{R}}^0} & I_{m+2} \\ \mathbf{0}_{m+2}\mathbf{0}_{m+2}^\top & \mathbf{0}_{m+2}\mathbf{0}_{m+2}^\top \end{bmatrix}$

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

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

$\Xi_\Delta^0(0) = \begin{bmatrix}\mathbf{1}_{m+1}\mathbf{1}_{m+1}^\top & c\mathbf{1}_{m+1}\end{bmatrix}^\top$, $\Xi_\Delta^1(0) = \mathbf{1}_{m+2}\mathbf{1}_{m+1}^\top - \Xi_\Delta^0(0)$ where $c = \frac{1-e^{-\sigma_1\delta}}{1-e^{-\sigma_1\Delta}}$

$\bar{\boldsymbol{\Xi}}_\Delta^0(0) = \begin{bmatrix}\mathbf{1}_{m+1}^\top & c\end{bmatrix}^\top$, $\bar{\boldsymbol{\Xi}}_\Delta^1(0) = \mathbf{1}_{m+2} - \bar{\boldsymbol{\Xi}}_\Delta^0(0)$

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

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

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

$k = 1$
while $B_\Delta^{(0)}$ and $B_\Delta^{(1)}$ have not converged **do**
Compute inverse Laplace transform matrices

$Q_\Delta^0(k) = \mathcal{L}_s^{-1}[(sI_{m+2} - G_{S,\bar{\mathcal{R}}^0})^{-1}\left(G_{S,\mathcal{R}^0}(sI_{m+2} - G_{S,\bar{\mathcal{R}}^0})^{-1}\right)^k](\Delta)$
$\bar{\mathbf{Q}}_\Delta^0(k) = \left(\int_0^\Delta Q_s^0(k)\mathrm{ds}\right)\boldsymbol{\mu}$
for $i = 1 : m+1, j = 1 : m+1$ **do**
$(\Xi_\Delta^0(k))_{i,j}, (\bar{\boldsymbol{\Xi}}_\Delta^0(k))_i = \frac{F_\Upsilon(\delta,k,\sigma_1)}{F_\Upsilon(\Delta,k,\sigma_1)}$
where $F_\Upsilon(u,k,\sigma_1) = \mathbb{P}(\Upsilon \leq u)$ and $\Upsilon \sim$ Erlang(k,σ_1)

$(\Xi_\Delta^1(k))_{i,j} = 1 - (\Xi_\Delta^0(k))_{i,j}$
$(\bar{\boldsymbol{\Xi}}_\Delta^1(k))_i = 1 - (\bar{\boldsymbol{\Xi}}_\Delta^0(k))_i$
$(\Xi_\Delta^0(k))_{m+2,j}, (\bar{\boldsymbol{\Xi}}_\Delta^0(k))_{m+2} = \frac{F_\Upsilon(\delta,k+1,\sigma_1)}{F_\Upsilon(\Delta,k+1,\sigma_1)}$
$(\Xi_\Delta^1(k))_{m+2,j} = 1 - (\Xi_\Delta^0(k))_{m+2,j}$
$(\bar{\boldsymbol{\Xi}}_\Delta^1(k))_{m+2} = 1 - (\bar{\boldsymbol{\Xi}}_\Delta^0(k))_{m+2}$
end for
$B_\Delta^{(0)} = B_\Delta^{(0)} + \begin{bmatrix} (Q_\Delta^0(k))_{(1:m+2),(1:m+1)} \odot \Xi_\Delta^0(k) & \mathbf{0}_{m+2} & \bar{\mathbf{Q}}_\Delta^0(k) \odot \bar{\boldsymbol{\Xi}}_\Delta^0(k) \\ \mathbf{0}_{m+1}^\top & 0 & 0 \end{bmatrix}$
$B_\Delta^{(1)} = B_\Delta^{(1)} + \begin{bmatrix} (Q_\Delta^0(k))_{(1:m+2),(1:m+1)} \odot \Xi_\Delta^1(k) & \mathbf{0}_{m+1} & \bar{\mathbf{Q}}_\Delta^0(k) \odot \bar{\boldsymbol{\Xi}}_\Delta^1(k) \\ \mathbf{0}_{m+1}^\top & 0 & 0 \end{bmatrix}$
for $i = 1 : m+2$ **do**
Find all $\mathbf{k} \in \mathcal{C}_k^{0_{i-1}}$ where $\mathbf{k} = \begin{pmatrix} k_0 & k_1 & \dots & k_m \end{pmatrix}^\top$ where
$\mathcal{C}_k^{0_{i-1}} := \left\{\mathbf{k} : \mathbf{k}^\top \mathbf{1}_{m+1} = k, k_{i-1} > 0, k_0 \geq \dots \geq k_{i-1} - 1 \geq \dots \geq k_m - 1\right\}$ and
$\mathcal{C}_k^{0_{m+1}} = \mathcal{C}_k^0$
For all $\mathbf{k}$ compute inverse Laplace transforms $q_{0_{i-1}1}^1(\mathbf{k},\Delta)$; $q_{0_{m+1}1}^1(\mathbf{k},\Delta) \equiv q_{11}^1(\mathbf{k},\Delta)$
$\xi_{0_{i-1}1}^1(0,\mathbf{k},\Delta) = \frac{F_\Phi(\Delta|\mathbf{k},\boldsymbol{\sigma}) - F_\Phi(\Delta-\delta|\mathbf{k},\boldsymbol{\sigma})}{F_\Phi(\Delta|\mathbf{k},\boldsymbol{\sigma})}$ where $F_\Phi(\phi|\mathbf{k},\boldsymbol{\sigma}) = \mathbb{P}(\Phi \leq \phi)$
and $\Phi = \sum_{p=0}^m W_p,\ W_p \overset{indep}{\sim}$ Erlang(k_p, σ_{0_p})

$\xi_{0_{m+1}1}^1(0,\mathbf{k},\Delta) \equiv \xi_{01}^1(0,\mathbf{k},\Delta)$

$(B_\Delta^{(0)})_{i,m+2} = (B_\Delta^{(0)})_{i,m+2} + \sum_{\mathbf{k}\in\mathcal{C}_k^{0_{i-1}}} q_{0_{i-1}1}^1(\mathbf{k},\Delta)\xi_{0_{i-1}1}^1(0,\mathbf{k},\Delta)$

$(B_\Delta^{(1)})_{i,m+2} = (B_\Delta^{(1)})_{i,m+2} + \sum_{\mathbf{k}\in\mathcal{C}_k^{0_{i-1}}} q_{0_{i-1}1}^1(\mathbf{k},\Delta)(1 - \xi_{0_{i-1}1}^1(0,\mathbf{k},\Delta))$
end for
$k = k + 1$
end while

S4 Discussion on the quality of the PSHMM rate estimates.

In this section, we discuss the quality of the PSHMM rate estimates. Identifiability of model parameters will be thoroughly investigated via analyses of the observed Fisher information matrices, correlation structures and contour plots of the log-likelihood surface as the noise parameter δ increases from 0 to Δ. Furthermore, fundamental statistical properties of the PSHMM-ML estimator such as asymptotic consistency will be shown through the convergence of mean-square errors in further simulation studies. A discussion of these properties for when m increases, will also be provided.

The results presented have been collected from nine further simulation studies that were conducted exclusively for this section. In particular, we investigate the effects of slow, medium and fast switching parameters on the three models $M^0_{\{1\}}, M^1_{\{1\}}$ and $M^2_{\{1\}}$. The nine studies, which will be referenced to throughout this section, were driven by their parameter set-ups as is presented in Table S1. All studies were executed with ω (false positive rate) equal to zero. Unless stated otherwise, the values shown of m, $\boldsymbol{\theta}, \Delta^{-1}, N_F$ and N_E 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.

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

Table S1: Global parameter values for the stimulation studies conducted in this section. All studies have been conducted with the model format of $M^m_{\{1\}}$ for $m = 0, 1, 2$.

S4.1 Identification of model parameters.

In order to analyze the properties of the PSHMM model and its advancement in characterizing fluorophore photo-switching behavior, it is necessary to understand the reliability of the parameter estimates. 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. Furthermore, while a set of parameters may be strictly identifiable, correlations between their estimates may be such as to make numerical estimation troublesome (Jacquez and Greif, 1985).

It is well understood that highly noisy data can make parameter estimation and identification more difficult. In the PSHMM, the noise comes in the form of $\delta \in [0, \Delta)$, the time that a fluorophore needs to be in the On state for it to be detected. This primarily results from a noise floor that the fluorophore signal needs to penetrate. As $\delta \to \Delta$, 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 $\{X(t)\}$. Furthermore, the accretion of data does not necessarily result in better estimation of parameters. Therefore, to understand the effects of parameter δ in estimation, we varied the ratio δ/Δ for all nine simulation parameters shown in Table S1.

S4.1.1 Local identifiability.

As stated in the main text, identifiability analysis is typically carried out by analyzing the Fisher Information matrix (e.g. Colquhoun et al., 2003). Here, a non-singular Fisher information matrix (all eigenvalues are strictly positive) implies local identifiability, whereas a singular Fisher information matrix (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). Since the true Fisher Information matrix $\mathcal{I}(\boldsymbol{\theta})$ cannot be computed for our problem, we initially test singularity of the observed Fisher information or Hessian matrix $\mathcal{J}(\hat{\boldsymbol{\theta}}) = -\nabla\nabla^{\top}\ell(\boldsymbol{\theta})|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$ evaluated at the maximum likelihood estimate $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ (Colquhoun et al., 2003), which is computed using the method of finite differences. In particular, singular observed Fisher Information matrices would indicate unidentifiable parameters within the model. We adopt the threshold of Viallefont et al. (1998) which states that if the smallest eigenvalue $\xi_{\min}$ of $\mathcal{J}(\hat{\boldsymbol{\theta}})$ is larger than $p \cdot \xi_{\max} \cdot 1 \times 10^{-9}$, where p is the dimension of $\boldsymbol{\theta}$ and $\xi_{\max}$ is the largest eigenvalue, then it can be considered to be non-zero and hence $\mathcal{J}(\hat{\boldsymbol{\theta}})$ non-singular.

500 independent datasets were generated for various values of δ for each study $1-9$ as described in Table S1. In all cases the Hessian matrices were determined to be non-singular, and hence the model formally 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 formally locally identifiable for all parameter values studied.

S4.1.2 Correlations.

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

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

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

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 (plus some other mild regulatory conditions) these two matrices should approximately align. Therefore, when the two mismatch this is evidence that the numerical optimization procedure we deploy is struggling to properly identify the parameters (Colquhoun et al., 2003).

Synopsis. We will show that the correlation between the parameters λ_{01} and δ increases as $\delta/\Delta \to 1$, and there 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. When δ/Δ is low, i.e. the signal to noise ratio is high, the correlation between λ_{01} and δ 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 practice identifiable. Contour plots will be shown to support this statement.

Secondly, we also explore the effects of changing the sampling time Δ. As $\Delta \to 0$, 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 Δ decreases. We first show that the previously highlighted correlation structures start to diminish and secondly, that estimates from selected simulation studies (especially those with faster switching parameters) become unbiased as $\Delta \to 0$, for all values of δ/Δ.

Details. In the tables that follow: for each study, 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 are averaged over these 500 datasets.

Tables S2, S3 and S4 show the correlation estimates for studies 1,2 and 3 (parameters provided in Table S1) as δ/Δ increases to 1. The correlations obtained from the Hessians and those obtained empirically agree quite well even when δ/Δ is large. The parameters, however which are most affected are λ_{01} and δ, whose correlation increases as δ/Δ increases. While this pattern is seen for all three switching scenarios, the correlation coefficients $\rho(\lambda_{01}, \delta)$ become 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 λ_{10} and δ, and μ_1 and δ, whose correlations becomes increasingly negative as δ/Δ increases. While these correlations do not become close to -1, the correlation pattern observed with all photo-switching parameters and δ suggest issues with model identification with large values of δ/Δ.

Tables S5-S7 show the correlation estimates for studies $4, 5$ and 6, as δ/Δ 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 λ_{01} and δ, 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 δ/Δ, 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. These correlations inherent to the PSHMM are not affected by the noise parameter δ and do not appear to pose issues to identifiability.

Tables S8-S13 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 λ_{01} and δ, 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 $m = 1$ case, the increased correlations between photo-switching rates from the multiple dark states $(\lambda_{00_1}, \lambda_{01}, \lambda_{0_1 0_2}, \lambda_{0_1 1}, \lambda_{0_2 1})$, is owed to the increased number of zeros the datasets.

To look more closely at how the change in δ affects the likelihood surface between δ and its highest correlated parameters, Figures S4 and S5 show the likelihood contour surfaces for 3 different values of δ, $\delta/\Delta = 0.1, 0.4, 0.7$, 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 δ/Δ. These contours show the change between λ_{01} and δ about their maximum likelihood estimates with all other parameters of the model fixed at their maximum likelihood estimates. Figure S4 highlights the smaller correlation between λ_{01} and δ with the contours appearing to extend vertically. Moreover, the MLE for λ_{01} appears to be computed around the 0.3 (true value) mark for all ranges of δ/Δ.

This is contrasted to the faster switching scenario in Figure S5. The greater correlation and dependence between parameters can be observed through the contours lying in diagonal direction which becomes more extreme as δ/Δ increases. While these closed contours show definitive maxima exist on the surfaces (it is structural unidentifiable), the confidence area as depicted by the elliptical contours around the maximum likelihood estimates (plotted in asterisks) is seen to increase as δ/Δ increases. The larger area and flatness highlight the difficultly in correctly estimating/identifying these parameters. As previously stated, in the case of slow switching (Figure S4), the estimate of λ_{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 S5), as is depicted by a greater bias in the estimate.

S4.1.3 Global identifiability.

To get a handle on global identifiability we assess whether the likelihood surface has a unique global maximum or if there are further modes the optimization method is determining to be the maximum. For the 3 state case ($m = 0$), an approximation scheme is used to find a suitable starting point for the Nelder-Mead simplex (see Section S6) and we ensure we locate the correct mode. In the 4 and 5 state cases ($m = 1$ and $m = 2$, 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 have found a unimodal distribution for the estimates of all parameters under all simulation studies (figures omitted) with one exception. Figure S3 shows histograms of estimates of $\lambda_{0_1 0_2}$ in Study 7 (slow switching) the 5 state ($m = 2$) model. When the number of frames $N_F = 10,000$, it is clear that the optimization procedure is finding two dominant modes. However, this global identifiability issue diminishes as N_F increases.

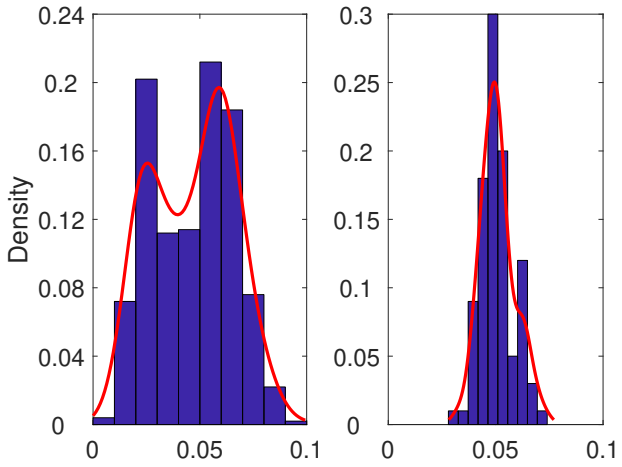


Figure S3: Rate estimates for $\lambda_{0_1 0_2} = 0.05$ under the slow switching scenario when $m = 2$, estimates show multi-modality when $N_F = 10^4$ (left) and uni-modality when $N_F = 5 \times 10^4$.

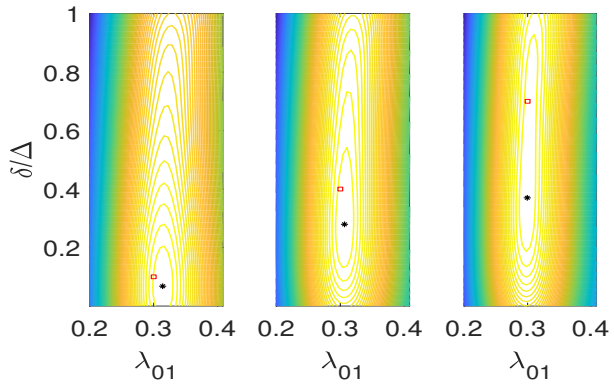


Figure S4: Contour plots showing the log-likelihood surface of λ_{01} against δ when $m = 1$ under study 4 (of Table S1) with other rate parameters $\lambda_{00_1}, \lambda_{10}$ and μ_1 fixed at their maximum likelihood values. The left most plot shows the surface when $\frac{\delta}{\Delta} = 0.1$, 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.

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

Table S2: Simulation results from $m = 0$ slow switching (study 1 of Table S1): Mean approximate correlations obtained from the Hessian matrices compared with the values calculated directly from 500 fits.

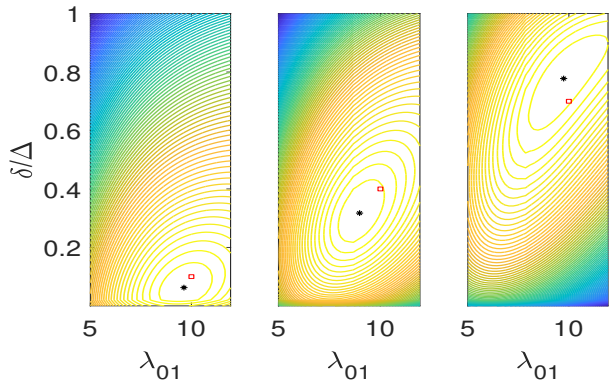


Figure S5: Contour plots showing the log-likelihood surface of λ_{01} against δ when $m = 1$ under study 6 (of Table S1) with other rate parameters $\lambda_{00_1}, \lambda_{10}$ and μ_1 fixed at their maximum likelihood values. The left most plot shows the surface when $\frac{\delta}{\Delta} = 0.1$, 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.

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

Table S3: Simulation results from $m = 0$ medium switching (study 2 of Table S1): 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
ρ (Correlation) $\downarrow$											
$\rho(\lambda_{01}, \lambda_{10})$	0.1114	0.0740	0.0213	−0.0728	−0.1774	−0.2962	−0.4209	−0.5151	−0.2016	0.1049	0.2962
	0.1120	0.0446	0.0515	−0.0529	−0.3221	−0.3005	−0.4783	−0.7272	−0.5883	0.0973	0.2722
$\rho(\lambda_{01}, \mu_1)$	0.0004	−0.0108	−0.0239	−0.0536	−0.0962	−0.1585	−0.2457	−0.4131	−0.4409	−0.3805	−0.3079
	0.0041	−0.0285	−0.0050	−0.1103	−0.1199	−0.1078	−0.2672	−0.6345	−0.6690	−0.3813	−0.2918
$\rho(\lambda_{10}, \mu_1)$	0.0374	0.0714	0.0909	0.1180	0.1478	0.1866	0.2413	0.3232	0.1791	0.0268	−0.0521
	−0.0232	0.0787	0.0687	0.1526	0.1492	0.2669	0.2172	0.4944	0.5065	0.0598	−0.0299
$\rho(\lambda_{01}, \delta)$	0.0549	0.1372	0.1997	0.3169	0.4464	0.5870	0.7207	0.8599	0.9173	0.9086	0.8664
	0.0606	0.1019	0.1897	0.2986	0.5680	0.6007	0.7456	0.9531	0.9603	0.9186	0.8544
$\rho(\lambda_{10}, \delta)$	−0.2615	−0.4980	−0.5683	−0.6140	−0.6454	−0.6736	−0.7026	−0.6837	−0.3339	−0.0428	0.1172
	−0.3324	−0.5800	−0.5959	−0.5643	−0.6925	−0.6900	−0.6598	−0.8279	−0.7023	−0.0835	0.0202
$\rho(\mu_1, \delta)$	−0.0543	−0.1133	−0.1423	−0.1782	−0.2169	−0.2668	−0.3344	−0.4679	−0.4719	−0.4061	−0.3354
	−0.0502	−0.1062	−0.1356	−0.2344	−0.2217	−0.2767	−0.2647	−0.6435	−0.7002	−0.4043	−0.3414

Table S4: Simulation results from $m = 0$ fast switching (study 3 of Table S1): 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
ρ (Correlation) $\downarrow$											
$\rho(\lambda_{00_1}, \lambda_{01})$	0.5926	0.5915	0.5916	0.5896	0.5901	0.5898	0.5857	0.5832	0.5823	0.5828	0.5855
	0.5884	0.5734	0.5523	0.5645	0.5193	0.5791	0.5347	0.5674	0.5770	0.5978	0.5820
$\rho(\lambda_{00_1}, \lambda_{0_11})$	0.6754	0.6761	0.6752	0.6764	0.6769	0.6798	0.6807	0.6807	0.6834	0.6828	0.6844
	0.6668	0.6706	0.7146	0.6833	0.6407	0.6643	0.5911	0.6740	0.6488	0.6971	0.6636
$\rho(\lambda_{01}, \lambda_{0_11})$	0.2854	0.2856	0.2845	0.2841	0.2868	0.2910	0.2928	0.2953	0.3004	0.2973	0.2958
	0.2414	0.2930	0.2821	0.2798	0.2420	0.3352	0.2373	0.3489	0.2777	0.3682	0.3157
$\rho(\lambda_{00_1}, \lambda_{10})$	0.0158	0.0152	0.0145	0.0118	0.0078	0.0033	−0.0013	−0.0041	−0.0057	−0.0044	−0.0024
	0.0084	−0.0103	−0.0214	0.0478	0.0254	−0.0038	0.0382	−0.0681	−0.1221	−0.0332	0.0114
$\rho(\lambda_{01}, \lambda_{10})$	0.0275	0.0259	0.0238	0.0166	0.0046	−0.0083	−0.0193	−0.0163	−0.0098	0.0028	0.0164
	−0.0013	0.0515	−0.0358	0.0359	−0.0607	−0.0905	−0.0734	−0.1606	−0.1049	0.0456	−0.0530
$\rho(\lambda_{0_11}, \lambda_{10})$	0.0038	0.0032	0.0025	0.0001	−0.0037	−0.0079	−0.0111	−0.0102	−0.0077	−0.0037	0.0012
	−0.0295	−0.0067	−0.0331	0.0511	0.0147	−0.0397	0.0191	−0.0667	−0.0750	−0.0068	−0.0099
$\rho(\lambda_{00_1}, \mu_1)$	0.0108	0.0112	0.0108	0.0113	0.0105	0.0098	0.0090	0.0082	0.0072	0.0077	0.0079
	0.0321	0.0828	0.0191	0.0008	−0.0010	0.0935	0.0375	−0.0906	0.0402	0.0407	0.0007
$\rho(\lambda_{01}, \mu_1)$	0.0097	0.0099	0.0096	0.0093	0.0070	0.0039	−0.0000	−0.0032	−0.0064	−0.0036	−0.0011
	−0.0011	0.0755	0.0649	−0.0289	0.0220	0.1154	−0.0762	−0.0721	−0.0061	0.0963	−0.0413
$\rho(\lambda_{0_11}, \mu_1)$	0.0326	0.0336	0.0328	0.0338	0.0325	0.0313	0.0301	0.0297	0.0283	0.0292	0.0293
	0.0539	0.1016	0.0092	−0.0256	−0.0366	0.0745	0.0557	0.0025	0.0154	0.0880	0.0240
$\rho(\lambda_{10}, \mu_1)$	−0.0069	−0.0061	−0.0052	−0.0037	−0.0014	0.0007	0.0016	−0.0005	−0.0035	−0.0060	−0.0086
	−0.0321	−0.0395	0.0123	−0.0677	−0.0500	−0.0191	−0.0013	−0.0007	0.0720	−0.0001	−0.0310
$\rho(\lambda_{00_1}, \delta)$	−0.0089	−0.0092	−0.0084	−0.0045	0.0049	0.0177	0.0365	0.0568	0.0773	0.0748	0.0707
	−0.0515	0.0053	0.0697	0.0345	−0.0290	0.0115	0.0218	0.1857	0.0605	0.1193	0.0816
$\rho(\lambda_{01}, \delta)$	−0.0115	−0.0076	−0.0011	0.0196	0.0589	0.1086	0.1792	0.2436	0.3084	0.2958	0.2733
	0.0189	0.0246	0.0555	0.0912	0.1491	0.2846	0.3455	0.5210	0.3632	0.4294	0.2540
$\rho(\lambda_{0_11}, \delta)$	−0.0013	0.0005	0.0031	0.0097	0.0225	0.0389	0.0609	0.0811	0.1017	0.0952	0.0874
	−0.0417	0.0392	0.0722	−0.0074	0.0139	0.0545	0.1099	0.1687	0.0769	0.0984	0.0659
$\rho(\lambda_{10}, \delta)$	−0.1354	−0.1663	−0.1876	−0.2165	−0.2343	−0.2425	−0.2265	−0.1626	−0.0920	−0.0420	0.0033
	0.0256	−0.1309	−0.1557	−0.1972	−0.2832	−0.3085	−0.2948	−0.3686	−0.3085	−0.1145	−0.0364
$\rho(\mu_1, \delta)$	−0.0124	−0.0159	−0.0190	−0.0241	−0.0307	−0.0376	−0.0439	−0.0456	−0.0462	−0.0393	−0.0322
	0.0139	−0.0036	0.0458	−0.0462	0.0043	0.0336	−0.1000	−0.0441	−0.1469	0.0150	−0.0453

Table S5: Simulation results from $m = 1$ slow switching (study 4 of Table S1): 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
ρ (Correlation) $\downarrow$											
$\rho(\lambda_{00_1}, \lambda_{01})$	0.5729	0.5758	0.5748	0.5717	0.5663	0.5603	0.5570	0.5502	0.5493	0.5529	0.5555
	0.5214	0.5593	0.5484	0.5422	0.5174	0.5094	0.4324	0.4277	0.5939	0.5298	0.4957
$\rho(\lambda_{00_1}, \lambda_{0_11})$	0.6802	0.6824	0.6823	0.6850	0.6839	0.6856	0.6899	0.6918	0.7003	0.7023	0.7008
	0.7081	0.6599	0.6531	0.6503	0.7029	0.7018	0.6401	0.6296	0.6645	0.7089	0.6922
$\rho(\lambda_{01}, \lambda_{0_11})$	0.2924	0.2947	0.2943	0.2937	0.2903	0.2917	0.3038	0.3105	0.3204	0.3199	0.3094
	0.2581	0.2379	0.2772	0.2552	0.2548	0.2888	0.2839	0.2906	0.3816	0.2970	0.2636
$\rho(\lambda_{00_1}, \lambda_{10})$	0.0489	0.0475	0.0439	0.0360	0.0253	0.0100	−0.0072	−0.0157	−0.0152	−0.0069	0.0015
	0.0073	0.0871	0.0330	0.0024	−0.0025	0.0689	−0.0672	−0.1155	−0.1029	−0.0185	0.0112
$\rho(\lambda_{01}, \lambda_{10})$	0.0874	0.0823	0.0735	0.0518	0.0195	−0.0318	−0.0778	−0.0651	−0.0341	0.0167	0.0638
	0.0537	0.1334	0.1303	0.0098	0.0352	−0.0818	−0.3192	−0.3792	−0.2396	−0.0334	0.0212
$\rho(\lambda_{0_11}, \lambda_{10})$	0.0240	0.0229	0.0205	0.0151	0.0065	−0.0071	−0.0216	−0.0205	−0.0119	0.0019	0.0150
	−0.0204	0.0454	−0.0349	0.0126	−0.0303	0.0054	−0.1004	−0.1949	−0.0769	−0.0142	0.0592
$\rho(\lambda_{00_1}, \mu_1)$	0.0010	0.0011	0.0011	0.0006	−0.0010	−0.0052	−0.0140	−0.0208	−0.0238	−0.0214	−0.0158
	−0.0044	−0.0558	−0.0203	0.0269	−0.0651	0.0270	−0.0735	−0.0301	−0.1018	−0.0145	0.0048
$\rho(\lambda_{01}, \mu_1)$	0.0015	0.0013	0.0004	−0.0028	−0.0109	−0.0300	−0.0602	−0.0828	−0.0918	−0.0807	−0.0581
	0.0206	−0.0976	−0.0460	0.0222	−0.0486	−0.0313	−0.1469	−0.1258	−0.1143	−0.1881	−0.0141
$\rho(\lambda_{0_11}, \mu_1)$	0.0004	0.0004	0.0002	−0.0006	−0.0026	−0.0076	−0.0173	−0.0242	−0.0263	−0.0235	−0.0168
	−0.0289	−0.0396	0.0082	0.0505	−0.0552	0.0147	−0.0762	−0.0785	−0.1116	−0.0469	0.0222
$\rho(\lambda_{10}, \mu_1)$	0.0110	0.0167	0.0211	0.0297	0.0402	0.0572	0.0710	0.0526	0.0345	0.0128	−0.0032
	0.0153	−0.0353	0.0207	−0.0192	0.0741	−0.0107	0.0750	0.1038	0.0899	0.0152	0.1020
$\rho(\lambda_{00_1}, \delta)$	−0.0137	−0.0149	−0.0129	−0.0060	0.0069	0.0292	0.0633	0.1106	0.1342	0.1364	0.1191
	0.0136	−0.0547	−0.0208	−0.0296	0.0109	−0.0215	0.0945	0.1694	0.2768	0.0965	0.0551
$\rho(\lambda_{01}, \delta)$	−0.0185	−0.0117	0.0013	0.0371	0.0956	0.1903	0.3050	0.4493	0.5195	0.5192	0.4580
	0.0276	−0.0156	−0.0346	0.0512	0.1403	0.3237	0.6120	0.6889	0.6827	0.5565	0.4915
$\rho(\lambda_{0_11}, \delta)$	−0.0058	−0.0043	−0.0009	0.0078	0.0226	0.0474	0.0828	0.1281	0.1464	0.1470	0.1254
	−0.0104	−0.0162	−0.0217	−0.0425	−0.0006	0.0478	0.1873	0.3027	0.2810	0.1391	0.0776
$\rho(\lambda_{10}, \delta)$	−0.2139	−0.2795	−0.3157	−0.3601	−0.3925	−0.4246	−0.4181	−0.2805	−0.1610	−0.0570	0.0222
	−0.0514	−0.2808	−0.2899	−0.3803	−0.3834	−0.5160	−0.5869	−0.5994	−0.4065	−0.1904	−0.0170
$\rho(\mu_1, \delta)$	−0.0371	−0.0518	−0.0618	−0.0793	−0.0991	−0.1296	−0.1616	−0.1713	−0.1673	−0.1464	−0.1137
	0.0300	−0.0303	−0.0464	−0.0858	−0.1312	−0.1113	−0.2092	−0.2037	−0.1956	−0.2347	−0.1104

Table S6: Simulation results from $m = 1$ medium switching (study 5 of Table S1): 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
ρ (Correlation) $\downarrow$											
$\rho(\lambda_{00_1}, \lambda_{01})$	0.3405	0.3349	0.3459	0.3774	0.4155	0.4888	0.6258	0.6706	0.6555	0.6224	0.5765
	0.3628	0.3507	0.4301	0.3108	0.3824	0.5229	0.8100	0.8409	0.6629	0.5735	0.6091
$\rho(\lambda_{00_1}, \lambda_{0_11})$	0.3386	0.3368	0.3371	0.3358	0.3261	0.3366	0.3980	0.4339	0.4531	0.4611	0.4640
	0.3038	0.3429	0.3030	0.2940	0.3251	0.3622	0.4578	0.5063	0.4640	0.4623	0.4496
$\rho(\lambda_{01}, \lambda_{0_11})$	0.1793	0.1758	0.1734	0.1687	0.1635	0.1793	0.2466	0.2773	0.2723	0.2553	0.2375
	0.2272	0.2038	0.2124	0.0852	0.1339	0.1729	0.3222	0.3821	0.2721	0.2359	0.2230
$\rho(\lambda_{00_1}, \lambda_{10})$	0.0566	−0.0493	−0.1066	−0.1763	−0.2428	−0.3332	−0.4572	−0.4535	−0.3202	−0.1847	−0.0797
	0.0602	−0.0620	−0.1243	−0.2787	−0.2494	−0.3709	−0.6973	−0.6761	−0.3627	−0.0915	−0.1371
$\rho(\lambda_{01}, \lambda_{10})$	0.5471	0.4469	0.3567	0.1720	−0.0583	−0.2968	−0.5318	−0.5674	−0.3987	−0.1932	0.0010
	0.4775	0.3501	0.2732	0.1298	−0.1502	−0.3976	−0.7826	−0.7792	−0.4798	−0.1459	−0.0967
$\rho(\lambda_{0_11}, \lambda_{10})$	0.0783	0.0583	0.0426	0.0158	−0.0177	−0.0650	−0.1467	−0.1678	−0.1171	−0.0579	−0.0119
	0.0802	0.0715	0.0561	−0.0467	0.0820	−0.0290	−0.2728	−0.2393	−0.1500	−0.0298	−0.0373
$\rho(\lambda_{00_1}, \mu_1)$	−0.0112	−0.0310	−0.0462	−0.0707	−0.1018	−0.1655	−0.2615	−0.2681	−0.2104	−0.1575	−0.1178
	0.0160	−0.0869	0.0047	−0.1182	−0.0918	−0.1394	−0.4245	−0.4593	−0.2448	−0.0935	−0.1078
$\rho(\lambda_{01}, \mu_1)$	0.0406	0.0408	0.0282	−0.0127	−0.0801	−0.1876	−0.3309	−0.3646	−0.3174	−0.2661	−0.2224
	0.0659	−0.0173	0.0020	−0.0232	−0.0253	−0.2111	−0.5190	−0.5624	−0.4151	−0.2480	−0.2524
$\rho(\lambda_{0_11}, \mu_1)$	0.0046	0.0036	0.0011	−0.0049	−0.0152	−0.0413	−0.0927	−0.1072	−0.0872	−0.0658	−0.0515
	0.0045	0.0090	0.0858	−0.0460	0.0048	−0.1148	−0.1870	−0.1522	−0.1387	−0.0166	0.0270
$\rho(\lambda_{10}, \mu_1)$	0.0922	0.1494	0.1779	0.2144	0.2438	0.3051	0.3727	0.3394	0.2195	0.1180	0.0422
	0.0813	0.0900	0.1089	0.2483	0.2469	0.2569	0.5041	0.5874	0.2500	0.1503	0.0054
$\rho(\lambda_{00_1}, \delta)$	0.1052	0.1923	0.2366	0.2874	0.3344	0.4121	0.5496	0.5971	0.5611	0.5027	0.4321
	0.0726	0.2080	0.2657	0.3300	0.3623	0.4415	0.7632	0.7753	0.5765	0.4441	0.4436
$\rho(\lambda_{01}, \delta)$	−0.0966	−0.0979	−0.0471	0.0956	0.2808	0.4847	0.7125	0.8244	0.8510	0.8399	0.7972
	−0.0648	−0.0544	0.0780	0.1247	0.3901	0.5974	0.9035	0.9103	0.8719	0.8292	0.8121
$\rho(\lambda_{0_11}, \delta)$	−0.0082	−0.0042	0.0050	0.0237	0.0498	0.0939	0.1867	0.2345	0.2294	0.2082	0.1870
	−0.0179	−0.0416	−0.0140	0.0748	−0.0006	0.0760	0.2925	0.3215	0.2336	0.1871	0.1832
$\rho(\lambda_{10}, \delta)$	−0.4631	−0.7184	−0.7860	−0.8158	−0.7972	−0.8155	−0.8300	−0.7621	−0.5873	−0.3941	−0.1998
	−0.5135	−0.7308	−0.7868	−0.8493	−0.8411	−0.8743	−0.9301	−0.9184	−0.6647	−0.3895	−0.2742
$\rho(\mu_1, \delta)$	−0.0964	−0.1701	−0.2031	−0.2426	−0.2726	−0.3368	−0.4184	−0.4186	−0.3587	−0.3059	−0.2603
	−0.0937	−0.1357	−0.1218	−0.2575	−0.2991	−0.3191	−0.5532	−0.6135	−0.4383	−0.2904	−0.2848

Table S7: Simulation results from $m = 1$ fast switching (study 6 of Table S1): 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
ρ (Correlation) $\downarrow$											
$\rho(\lambda_{00_1}, \lambda_{01})$	−0.1610	−0.1158	−0.1025	−0.1637	−0.0808	−0.1255	−0.0980	−0.0860	−0.1826	−0.1768	−0.0720
	0.4636	0.3054	0.3245	0.4173	0.5135	0.3869	0.4449	0.4316	0.3913	0.4898	0.3780
$\rho(\lambda_{00_1}, \lambda_{0_10_2})$	0.3396	0.3015	0.2835	0.3388	0.2714	0.3261	0.2891	0.2844	0.3623	0.3685	0.2800
	−0.1316	−0.0657	−0.0279	−0.1000	−0.0148	−0.0698	−0.0362	−0.1160	−0.1319	−0.0965	−0.0631
$\rho(\lambda_{01}, \lambda_{0_10_2})$	−0.7763	−0.7605	−0.7673	−0.7897	−0.7430	−0.7548	−0.7661	−0.7743	−0.8030	−0.8119	−0.7640
	−0.6749	−0.5007	−0.6502	−0.6557	−0.5843	−0.6500	−0.6122	−0.6609	−0.6682	−0.6017	−0.6394
$\rho(\lambda_{00_1}, \lambda_{0_11})$	0.0027	0.0574	0.0735	0.0146	0.0939	0.0663	0.0995	0.1137	0.0163	0.0198	0.1254
	0.5237	0.3805	0.5875	0.5831	0.2380	0.5986	0.5905	0.6323	0.5082	0.6147	0.6844
$\rho(\lambda_{01}, \lambda_{0_11})$	0.7683	0.7494	0.7518	0.7716	0.7377	0.7506	0.7473	0.7553	0.7828	0.7884	0.7483
	0.7166	0.8141	0.6065	0.6667	0.6780	0.6503	0.6476	0.6628	0.6814	0.6770	0.6096
$\rho(\lambda_{0_10_2}, \lambda_{0_11})$	−0.7324	−0.7136	−0.7292	−0.7438	−0.6997	−0.6879	−0.7144	−0.7195	−0.7415	−0.7529	−0.7105
	−0.5998	−0.4808	−0.5624	−0.5811	−0.6077	−0.5032	−0.5283	−0.5261	−0.5840	−0.5094	−0.5109
$\rho(\lambda_{00_1}, \lambda_{0_21})$	−0.4348	−0.3985	−0.3920	−0.4337	−0.3760	−0.4084	−0.3922	−0.3824	−0.4514	−0.4481	−0.3744
	0.0908	−0.0216	−0.0453	−0.0207	−0.0235	−0.0908	−0.0088	0.0799	0.0372	0.0485	−0.0510
$\rho(\lambda_{01}, \lambda_{0_21})$	0.7245	0.7024	0.7040	0.7385	0.6843	0.7144	0.7134	0.7226	0.7601	0.7738	0.7155
	0.5706	0.5451	0.4805	0.4705	0.4554	0.4899	0.5016	0.5154	0.5428	0.4880	0.4971
$\rho(\lambda_{0_10_2}, \lambda_{0_21})$	−0.8287	−0.8049	−0.8244	−0.8402	−0.7972	−0.8054	−0.8246	−0.8330	−0.8537	−0.8639	−0.8268
	−0.7142	−0.7098	−0.7551	−0.7281	−0.7510	−0.6973	−0.7406	−0.7352	−0.7373	−0.6680	−0.7415
$\rho(\lambda_{0_11}, \lambda_{0_21})$	0.7449	0.7287	0.7287	0.7503	0.7140	0.7299	0.7273	0.7299	0.7556	0.7669	0.7302
	0.5949	0.5811	0.5247	0.5173	0.5580	0.4431	0.5270	0.5238	0.5570	0.5020	0.4588
$\rho(\lambda_{00_1}, \lambda_{10})$	0.4647	0.4306	0.4233	0.4575	0.4072	0.4354	0.4158	0.4101	0.4715	0.4679	0.4006
	−0.0491	0.0298	0.0261	−0.0113	0.0706	0.0172	0.0422	0.0175	−0.0017	−0.0025	0.1008
$\rho(\lambda_{01}, \lambda_{10})$	−0.6628	−0.6388	−0.6383	−0.6866	−0.6215	−0.6596	−0.6584	−0.6667	−0.7177	−0.7327	−0.6606
	−0.5564	−0.2211	−0.5275	−0.5618	−0.4047	−0.5259	−0.5020	−0.4994	−0.5451	−0.4701	−0.4685
$\rho(\lambda_{0_10_2}, \lambda_{10})$	0.7532	0.7308	0.7380	0.7723	0.7111	0.7427	0.7500	0.7631	0.8001	0.8221	0.7652
	0.7268	0.7349	0.7866	0.7284	0.7298	0.6084	0.7387	0.7241	0.6935	0.7056	0.7207
$\rho(\lambda_{0_11}, \lambda_{10})$	−0.6547	−0.6326	−0.6310	−0.6698	−0.6124	−0.6387	−0.6380	−0.6408	−0.6875	−0.7033	−0.6450
	−0.4979	−0.2404	−0.5101	−0.5017	−0.4507	−0.4887	−0.4657	−0.3815	−0.4739	−0.4260	−0.3933
$\rho(\lambda_{0_21}, \lambda_{10})$	−0.7887	−0.7671	−0.7734	−0.8041	−0.7503	−0.7830	−0.7865	−0.7956	−0.8318	−0.8475	−0.8000
	−0.6504	−0.5952	−0.6845	−0.6518	−0.6771	−0.6168	−0.6430	−0.6223	−0.6434	−0.5618	−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
	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	0.6893	0.6901	0.7297	0.6654	0.7005	0.6989	0.7075	0.7507	0.7644	0.7011
	0.6626	0.3925	0.6148	0.6245	0.5208	0.6370	0.5776	0.6132	0.6428	0.5733	0.6022
$\rho(\lambda_{0_10_2}, \mu_1)$	−0.8275	−0.8067	−0.8203	−0.8412	−0.7915	−0.8145	−0.8246	−0.8362	−0.8593	−0.8756	−0.8328
	−0.8870	−0.8837	−0.8959	−0.8767	−0.8818	−0.8643	−0.8927	−0.8835	−0.8770	−0.8559	−0.8819
$\rho(\lambda_{0_11}, \mu_1)$	0.7067	0.6868	0.6878	0.7176	0.6662	0.6896	0.6897	0.6919	0.7279	0.7405	0.6931
	0.6151	0.4014	0.5766	0.5828	0.5771	0.5499	0.5319	0.5133	0.5949	0.5246	0.5236

Table S8: Simulation results from $m = 2$ slow switching (study 7 of Table S1): 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
ρ (Correlation) $\downarrow$											
$\rho(\lambda_{0_2 1}, \mu_1)$	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
$\rho(\lambda_{00_1}, \delta)$	−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
$\rho(\lambda_{0_1 0_2}, \delta)$	−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
$\rho(\lambda_{0_1 1}, \delta)$	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
$\rho(\lambda_{0_2 1}, \delta)$	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 S9: Simulation results from $m = 2$ slow switching (study 7 of Table S1) 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
ρ (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
$\rho(\lambda_{00_1}, \lambda_{0_10_2})$	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
$\rho(\lambda_{01}, \lambda_{0_10_2})$	−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
$\rho(\lambda_{00_1}, \lambda_{0_11})$	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
$\rho(\lambda_{01}, \lambda_{0_11})$	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
$\rho(\lambda_{0_10_2}, \lambda_{0_11})$	−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
$\rho(\lambda_{00_1}, \lambda_{0_21})$	−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
$\rho(\lambda_{01}, \lambda_{0_21})$	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
$\rho(\lambda_{0_10_2}, \lambda_{0_21})$	−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
$\rho(\lambda_{0_11}, \lambda_{0_21})$	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
$\rho(\lambda_{0_10_2}, \lambda_{10})$	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
$\rho(\lambda_{0_11}, \lambda_{10})$	−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
$\rho(\lambda_{0_21}, \lambda_{10})$	−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 −0.1973	−0.2455 −0.1578	−0.2151 −0.1436	−0.1915 −0.0563	−0.1745 −0.1368	−0.1658 −0.1149	−0.1752 −0.0934	−0.1719 −0.2910	−0.1454 −0.0182	−0.1232 −0.1163	−0.1102 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
$\rho(\lambda_{0_10_2}, \mu_1)$	−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
$\rho(\lambda_{0_11}, \mu_1)$	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

Table S10: Simulation results from $m = 2$ medium switching (study 8 of Table S1): 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
ρ (Correlation) $\downarrow$											
$\rho(\lambda_{0_2 1}, \mu_1)$	0.7892	0.7931	0.7682	0.7430	0.7238	0.7036	0.6845	0.6461	0.6193	0.6136	0.6031
	0.7308	0.7359	0.6911	0.6600	0.5947	0.5151	0.4457	0.2904	0.2393	0.2314	0.2140
$\rho(\lambda_{10}, \mu_1)$	−0.1696	−0.1516	−0.1213	−0.0932	−0.0673	−0.0352	0.0128	0.0072	−0.0516	−0.0883	−0.0981
	−0.1419	−0.1207	−0.1143	0.0096	−0.0065	−0.0526	0.1267	0.2870	0.0659	0.0484	0.0283
$\rho(\lambda_{00_1}, \delta)$	0.0097	0.0251	0.0392	0.0637	0.0929	0.1395	0.2312	0.3502	0.3348	0.2749	0.2391
	0.0344	0.0906	0.1098	0.1030	0.1259	0.1979	0.3136	0.6346	0.3826	0.2479	0.1893
$\rho(\lambda_{01}, \delta)$	−0.0247	0.0104	0.0460	0.1360	0.2580	0.4269	0.6343	0.8224	0.8432	0.7971	0.7568
	0.0371	0.0456	0.0871	0.1160	0.3167	0.4716	0.7803	0.9485	0.8742	0.7884	0.7136
$\rho(\lambda_{0_1 0_2}, \delta)$	−0.0009	−0.0027	0.0045	0.0112	0.0186	0.0274	0.0407	0.0517	0.0462	0.0366	0.0336
	−0.0917	−0.0066	−0.0172	0.0377	0.0262	0.0610	0.0370	0.1253	0.1161	0.0567	−0.0179
$\rho(\lambda_{0_1 1}, \delta)$	−0.0095	−0.0073	−0.0073	0.0006	0.0130	0.0362	0.0876	0.1690	0.1751	0.1494	0.1338
	0.0148	0.0379	0.0230	0.0587	0.0403	0.0344	0.1317	0.4143	0.2533	0.1249	0.1302
$\rho(\lambda_{0_2 1}, \delta)$	0.0071	0.0150	0.0095	0.0072	0.0046	0.0043	0.0111	0.0204	0.0195	0.0168	0.0161
	0.0087	0.0584	−0.0125	0.0114	0.0028	0.0583	0.0296	0.0512	0.1283	−0.0333	−0.0154
$\rho(\lambda_{10}, \delta)$	−0.3872	−0.5744	−0.6153	−0.6602	−0.6947	−0.7337	−0.7813	−0.5962	−0.3004	−0.0994	0.0205
	−0.2960	−0.6377	−0.6105	−0.6919	−0.7027	−0.7630	−0.8379	−0.8906	−0.4383	−0.0889	−0.0137
$\rho(\mu_1, \delta)$	−0.0088	−0.0105	−0.0234	−0.0378	−0.0562	−0.0808	−0.1196	−0.1437	−0.1233	−0.0956	−0.0854
	0.0353	0.0577	0.0100	−0.0561	−0.0918	−0.0357	−0.1748	−0.3398	−0.1752	−0.2138	−0.1197

Table S11: Simulation results from $m = 2$ medium switching (study 8 of Table S1) 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
ρ (Correlation) $\downarrow$											
$\rho(\lambda_{00_1}, \lambda_{01})$	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 −0.1227	−0.1340 −0.1989	−0.1544 −0.1324	−0.2243 −0.1101	−0.3015 −0.4881	−0.2857 −0.3325	−0.2128 0.0572	−0.1656 0.0196	−0.1327 0.1341	−0.1052 0.2721	−0.0906 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

Table S12: Simulation results from $m = 2$ fast switching (study 9 of Table S1): 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
ρ (Correlation) $\downarrow$											
$\rho(\lambda_{0_2 1}, \mu_1)$	0.3356	0.3797	0.3948	0.4513	0.4571	0.4090	0.3590	0.3293	0.3053	0.2799	0.2662
	0.3900	0.3971	0.3593	0.3572	0.4253	0.1470	−0.1956	−0.2335	−0.3394	−0.4054	−0.3861
$\rho(\lambda_{10}, \mu_1)$	0.1818	0.1730	0.1932	0.2468	0.3011	0.2370	0.1326	0.0710	0.0281	−0.0017	−0.0196
	0.2083	0.2031	0.1800	0.2156	0.5718	0.5468	0.2826	0.3285	0.2487	0.2693	0.3157
$\rho(\lambda_{00_1}, \delta)$	0.2862	0.3072	0.3486	0.4606	0.6247	0.6987	0.6506	0.5851	0.5300	0.4784	0.4323
	0.3317	0.3202	0.4446	0.5140	0.8560	0.8311	0.6868	0.6312	0.5043	0.5148	0.3463
$\rho(\lambda_{01}, \delta)$	−0.1171	−0.0230	0.0611	0.3069	0.6540	0.8472	0.8578	0.8361	0.8164	0.7897	0.7638
	0.0380	0.0149	0.0767	0.3770	0.9100	0.9211	0.8556	0.8544	0.8037	0.7933	0.6842
$\rho(\lambda_{0_1 0_2}, \delta)$	0.1187	0.1268	0.1365	0.1683	0.2037	0.1889	0.1459	0.1197	0.1041	0.0933	0.0893
	0.1516	0.2215	0.2373	0.2975	0.4224	0.4253	0.2382	0.2510	0.2036	0.1671	0.1251
$\rho(\lambda_{0_1 1}, \delta)$	−0.0705	−0.0611	−0.0536	−0.0292	0.0721	0.1701	0.1878	0.1755	0.1664	0.1568	0.1447
	−0.0607	−0.0930	−0.0387	−0.1213	0.1667	0.2773	0.1776	0.1937	0.1175	0.1607	0.0881
$\rho(\lambda_{0_2 1}, \delta)$	−0.1494	−0.1557	−0.1653	−0.1846	−0.1743	−0.1195	−0.0639	−0.0414	−0.0227	−0.0045	0.0117
	−0.1951	−0.2923	−0.3103	−0.2549	−0.4483	−0.3200	−0.2484	−0.2362	−0.1840	−0.1652	−0.0590
$\rho(\lambda_{10}, \delta)$	−0.7786	−0.7828	−0.8115	−0.8588	−0.8723	−0.7580	−0.6020	−0.4491	−0.3053	−0.1681	−0.0732
	−0.7288	−0.7674	−0.8005	−0.8653	−0.9448	−0.9320	−0.7030	−0.4571	−0.2589	−0.1178	0.0043
$\rho(\mu_1, \delta)$	−0.2283	−0.2331	−0.2570	−0.3149	−0.3713	−0.3353	−0.2602	−0.2192	−0.1924	−0.1693	−0.1622
	−0.2079	−0.2474	−0.2130	−0.2348	−0.5936	−0.5430	−0.1652	−0.1521	0.0032	0.0379	−0.0300

Table S13: Simulation results from $m = 2$ fast switching (study 9 of Table S1) continued: Mean approximate correlations obtained from the Hessian matrices compared with the values calculated directly from 500 fits.

S4.1.4 The effect of frame length.

In order to test the effect of decreasing the frame length Δ (or increasing the frame rate) on estimates, we simulated 500 datasets for the fast switching scenarios (studies $3, 6$ and 9 in Table S1) under the models $m = 0, 1, 2$. The fast switching parameters were used in this analysis due to the fact that the correlation structures between the noise parameter δ and photo-switching parameters was highest, as is explained in Section S4.1.2. In particular, small values of Δ should yield a 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 $N_E = 100$ molecules over a period of 3×10^5 seconds. Figures S6 - S10 show box-plots of rate results for all simulations. Under all models, it is evident that rate estimates become unbiased as $\Delta \to 0$, under most values of the noise fraction δ/Δ. 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 Δ, this bias is significantly seen to decrease. This intuitively reflects the fact that $\Delta \to 0$ implies the observed process $\{Y_n\}$ provides greater information of the process $\{X(t)\}$, with the $\Delta, \delta \to 0$ case trivially becoming completely informative of the hidden process. When $\Delta^{-1} = 30\text{s}^{-1}$ for the $m = 2$ 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 Δ^{-1} is not large enough to capture the transitions between On and dark states.

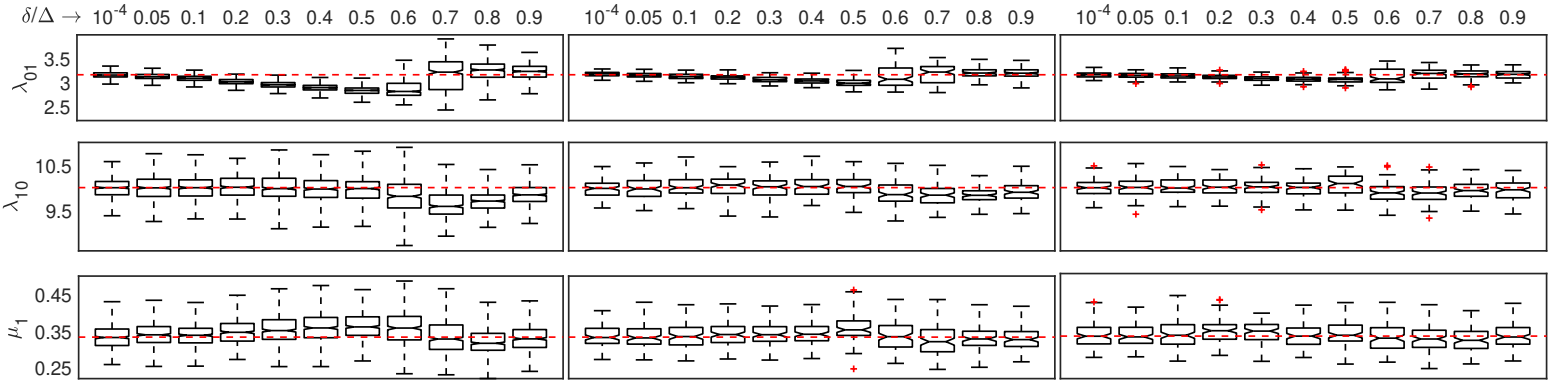


Figure S6: Box-plots showing rate estimates when $m = 0$ under study 3 of Table S1 when $\Delta^{-1} = 30$ (left), $\Delta^{-1} = 65$ (middle) and $\Delta^{-1} = 100$ (right). True rates indicated by the red dashed line.

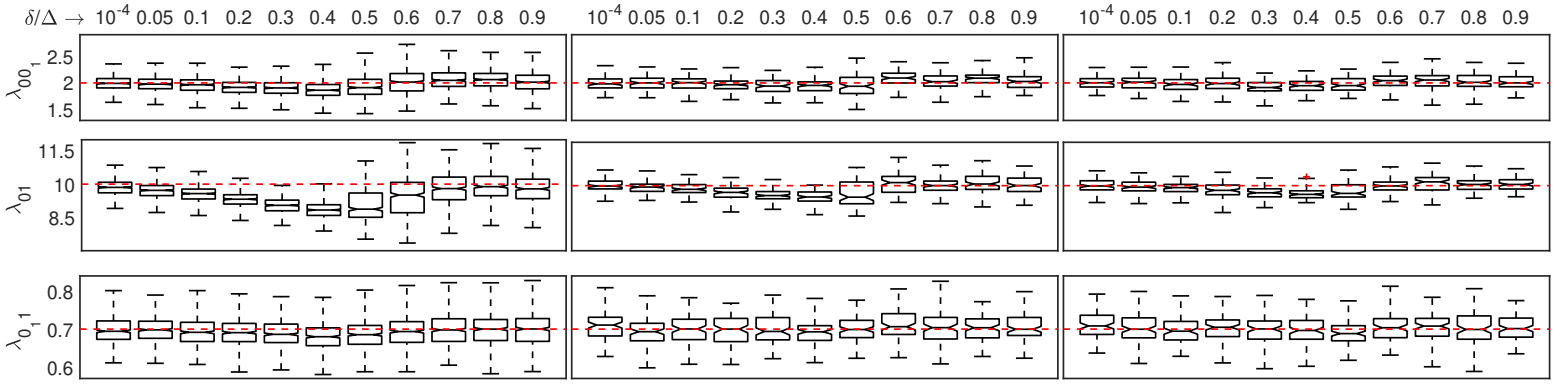


Figure S7: Box-plots showing rate estimates when $m = 1$ under study 4 of Table S1 when $\Delta^{-1} = 30$ (left), $\Delta^{-1} = 65$ (middle) and $\Delta^{-1} = 100$ (right). True rates indicated by the red dashed line.

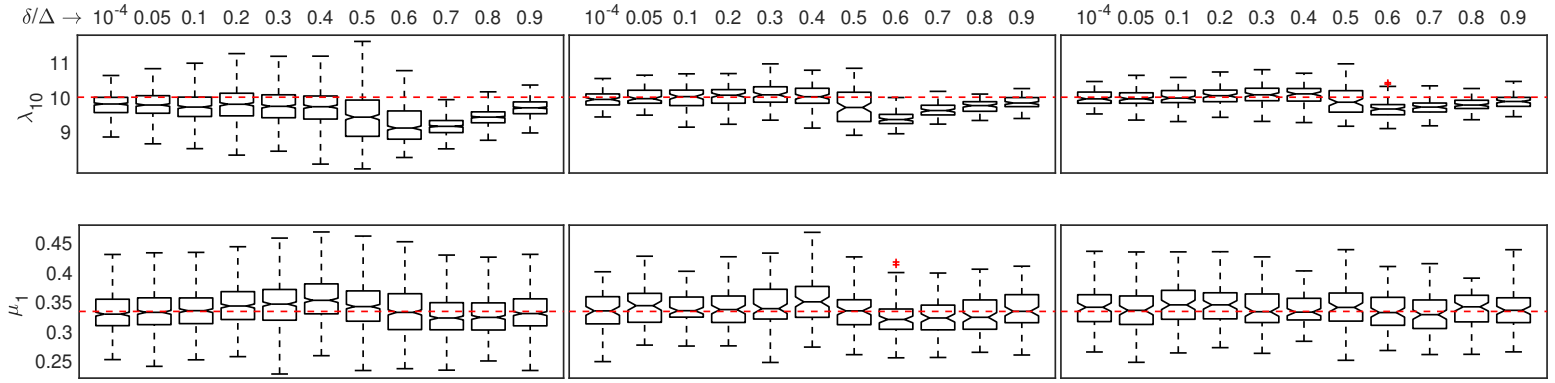


Figure S8: Box-plots showing rate estimates when $m = 1$ under study 4 of Table S1 when $\Delta^{-1} = 30$ (left), $\Delta^{-1} = 65$ (middle) and $\Delta^{-1} = 100$ (right). True rates indicated by the red dashed line.

To look more closely at how the change in Δ affects the likelihood surface between δ and its highest correlated parameter λ_{01}, Figure S11 shows the likelihood contour surfaces for 3 different values of Δ, $\Delta^{-1} = 30, 65, 100\text{s}^{-1}$, 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 δ/Δ. These contours show the change between λ_{01} and δ 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 λ_{01} and δ reduces as Δ decreases.

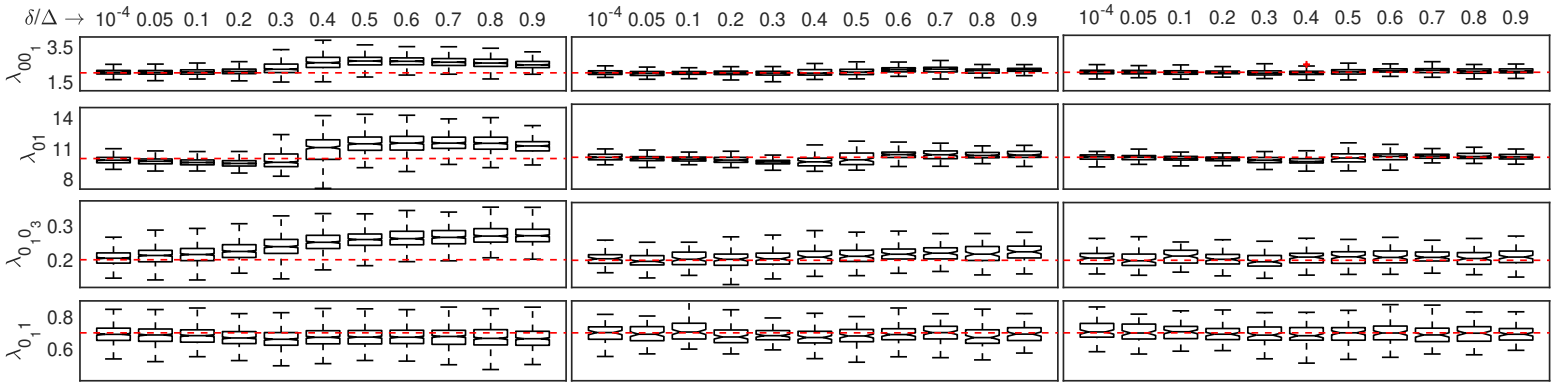


Figure S9: Box-plots showing rate estimates when $m = 2$ under study 9 of Table S1 when $\Delta^{-1} = 30$ (left), $\Delta^{-1} = 65$ (middle) and $\Delta^{-1} = 100$ (right). True rates indicated by the red dashed line.

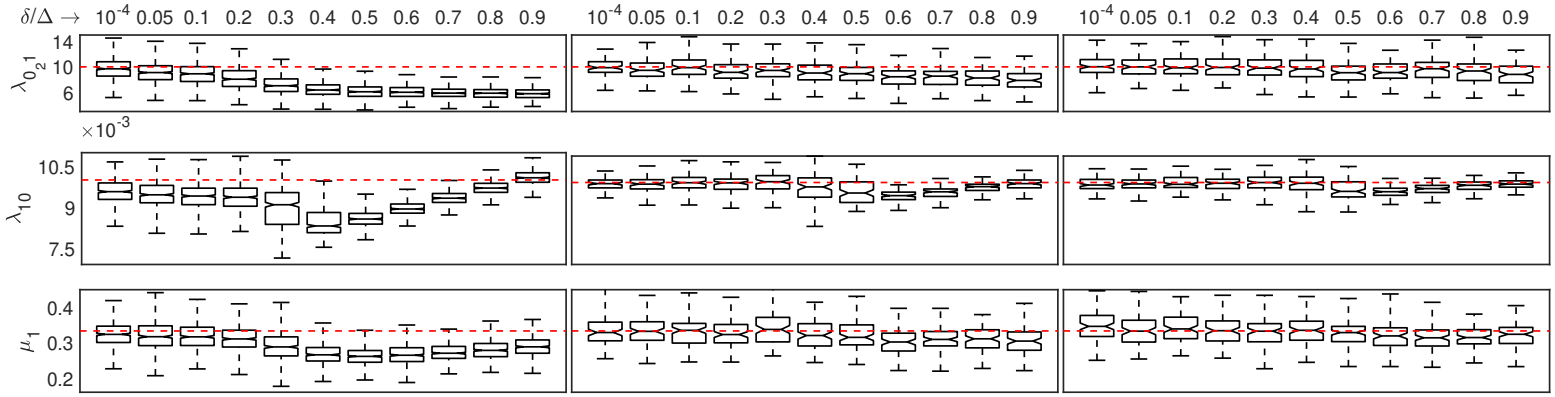


Figure S10: Box-plots showing rate estimates when $m = 2$ under study 9 of Table S1 when $\Delta^{-1} = 30$ (left), $\Delta^{-1} = 65$ (middle) and $\Delta^{-1} = 100$ (right). True rates indicated by the red dashed line. True rates indicated by the red dashed line.

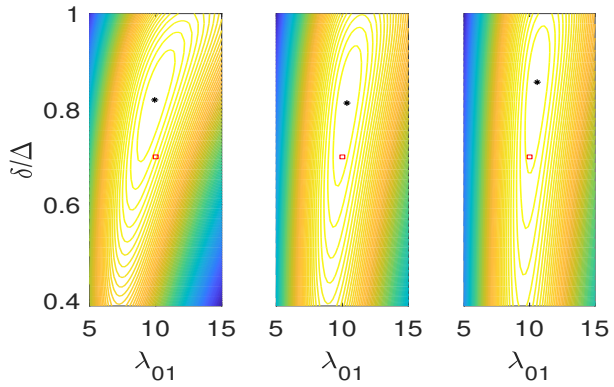


Figure S11: Contour plots showing the log-likelihood surface of λ_{01} against δ/Δ when $m = 1$ under study 6 (of Table S1) with other rate parameters $\lambda_{00_1}, \lambda_{10}$ and μ_1 fixed at their maximum likelihood values. The left most plot shows the surface when $\Delta^{-1} = 30\text{s}^{-1}$, 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 δ/Δ 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.

S4.2 Consistency.

In this section, we will use the identifiability analysis from the previous section to discuss when the PSHMM-ML estimator is consistent. Rate estimates will be computed as N_E is changed. The variance of the PSHMM estimator is shown to decrease as N_E increases, whilst consistency is shown through the convergence of mean-square errors only when δ/Δ is small and unknown transition rates are identifiable given the sampling time Δ.

Specifically, the PSHMM estimator $\hat{\boldsymbol{\theta}}_{N_F,N_E} := \arg\max_{\theta\in\Theta}\ \ell\left(\mathcal{Y};\boldsymbol{\theta}\right)$ of the true parameter vector $\boldsymbol{\theta}^*$ of the model is consistent if it converges in probability to $\boldsymbol{\theta}^*$, i.e. that $\operatorname{plim}_{N_F,N_E\to\infty} \hat{\boldsymbol{\theta}}_{N_F,N_E} = \boldsymbol{\theta}^*$. From equation (3.3) of the main text, $\mathcal{Y} = \begin{pmatrix}\mathbf{y}^1 & \mathbf{y}^2 & \dots & \mathbf{y}^{N_E}\end{pmatrix}$ is the data matrix with each column being an $N_F \times 1$ data vector from each independent imaging experiment.

Consistency is difficult to show theoretically for the PSHMM estimator given the form of the log-likelihood function in (3.3) of the main text. However, since convergence in mean square implies convergence in probability, we can loosely show that convergence in probability is attained by analyzing if the mean-squared errors of parameter estimates tends to zero as N_E 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 absorption state. In this section, we have therefore solely considered the effects of changing N_E.

Using the simulation studies $1-9$ from the previous section (considered in Table S1), we were also able to compute the means and mean squared errors of the estimates, the results of which are shown for different δ/Δ in Tables S14 - S22. To compare the mean squared errors as the sample size N_E grows, we additionally show these errors under the same simulation studies executed with $N_E = 1000$. The mean squared errors, on the whole, are seen to decrease as the sample size N_E increases, for all switching parameters and values of δ/Δ. This reduction is highlighted more when δ/Δ is smaller, especially for faster switching rates. For example, in Table S16, while the mean squared error for parameter λ_{10} reduces for larger δ/Δ, 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 δ/Δ is between 0.4 and 0.8. This pattern is also observed from other parameters, especially λ_{01} and λ_{00_1} in the faster switching scenarios. The fast reduction in MSE for parameters under slowing switching scenarios and with lower noise δ/Δ is reflected by better model identification

(as shown in Section S4.1) and is therefore a good indication of consistency.

To more clearly visualize the effects of an increase in N_E on maximum likelihood estimation, we show rate estimates as δ/Δ increases when $N_E = 100, 1000, 5000$ for the fast switching scenario (studies $3, 6$ and 9). N_F is fixed at $10,000$ under all simulations. Figures S12 - S16 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 δ/Δ. Furthermore, strong correlations between parameter estimates can be seen as δ/Δ increases, leading to inconsistency. For example, the estimates for λ_{01} under all $m = 0, 1, 2$ models remain biased as N_E increase, albeit at a lower variance. This is also true for λ_{10} under studies 6 and 9.

From Figures S12 - S16, the pattern in bias that is captured by the majority of estimates under $N_E = 100$ as δ/Δ increases persists as N_E increases. This is replicated by the mismatch in correlations that indicates identifiability issues. However, for δ/Δ, 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 S16, whereby estimates for λ_{10}, although obtaining a smaller variance, remain biased for even the smallest values of δ/Δ. This is related to the identifiability analysis previously presented, as the sampling interval Δ is too large to informatively detect and estimate faster switching rates. As $\Delta \to 0$, however, using the analysis from Section S4.1.4, these faster rates should become identifiable and consistency holds.

$\frac{\delta}{\Delta} \to$	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 S14: $m = 0$ slow switching (study 1 of Table S1): Means and Mean Squared Errors (MSE) from datasets with $N_E = 100$ (100) and $N_E = 1000$ (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 S15: $m = 0$ medium switching (study 2 of Table S1): Means and Mean Squared Errors (MSE) from datasets with $N_E = 100$ (100) and $N_E = 1000$ (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 S16: $m = 0$ fast switching (study 3 of Table S1): Means and Mean Squared Errors (MSE) from datasets with $N_E = 100$ (100) and $N_E = 1000$ (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	0.1522	0.1499	0.1521	0.1491	0.1501	0.1500	0.1493	0.1505	0.1517	0.1508	0.1515
$\hat{\lambda}_{00_1}$ 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
$\hat{\lambda}_{00_1}$ 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
$\hat{\lambda}_{0_11}$ Mean	0.1002	0.0995	0.1000	0.0994	0.0994	0.0996	0.0997	0.0999	0.1004	0.0999	0.1001
$\hat{\lambda}_{0_11}$ 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
$\hat{\lambda}_{0_11}$ 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 S17: $m = 1$ slow switching (study 4 of Table S1): Means and Mean Squared Errors (MSE) from datasets with $N_E = 100$ (100) and $N_E = 1000$ (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 S18: $m = 1$ medium switching (study 5 of Table S1): Means and Mean Squared Errors (MSE) from datasets with $N_E = 100$ (100) and $N_E = 1000$ (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 S19: $m = 1$ fast switching (study 6 of Table S1): Means and Mean Squared Errors (MSE) from datasets with $N_E = 100$ (100) and $N_E = 1000$ (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
$\hat{\lambda}_{01}$ Mean	0.2935	0.2979	0.2974	0.2927	0.2952	0.2915	0.2952	0.2933	0.2905	0.2920	0.2938
$\hat{\lambda}_{01}$ 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
$\hat{\lambda}_{01}$ 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
$\hat{\lambda}_{0_10_2}$ Mean	0.0541	0.0527	0.0523	0.0568	0.0525	0.0547	0.0540	0.0549	0.0577	0.0564	0.0537
$\hat{\lambda}_{0_10_2}$ 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
$\hat{\lambda}_{0_10_2}$ 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 S20: $m = 2$ slow switching (study 7 of Table S1): Means and Mean Squared Errors (MSE) from datasets with $N_E = 100$ (100) and $N_E = 1000$ (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
$\hat{\lambda}_{01}$ Mean	3.9984	3.9547	3.9089	3.8364	3.7578	3.6906	3.6473	3.9421	4.1152	4.1213	4.0996
$\hat{\lambda}_{01}$ 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
$\hat{\lambda}_{01}$ 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 S21: $m = 2$ medium switching (study 8 of Table S1): Means and Mean Squared Errors (MSE) from datasets with $N_E = 100$ (100) and $N_E = 1000$ (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
$\hat{\lambda}_{01}$ Mean	9.8536	9.7290	9.6305	9.5301	9.9097	10.9416	11.4587	11.5278	11.5300	11.5069	11.1842
$\hat{\lambda}_{01}$ 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
$\hat{\lambda}_{01}$ 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 S22: $m = 2$ fast switching (study 9 of Table S1) of Table S1): Means and Mean Squared Errors (MSE) from datasets with $N_E = 100$ (100) and $N_E = 1000$ (1000) are shown for all parameter estimates.

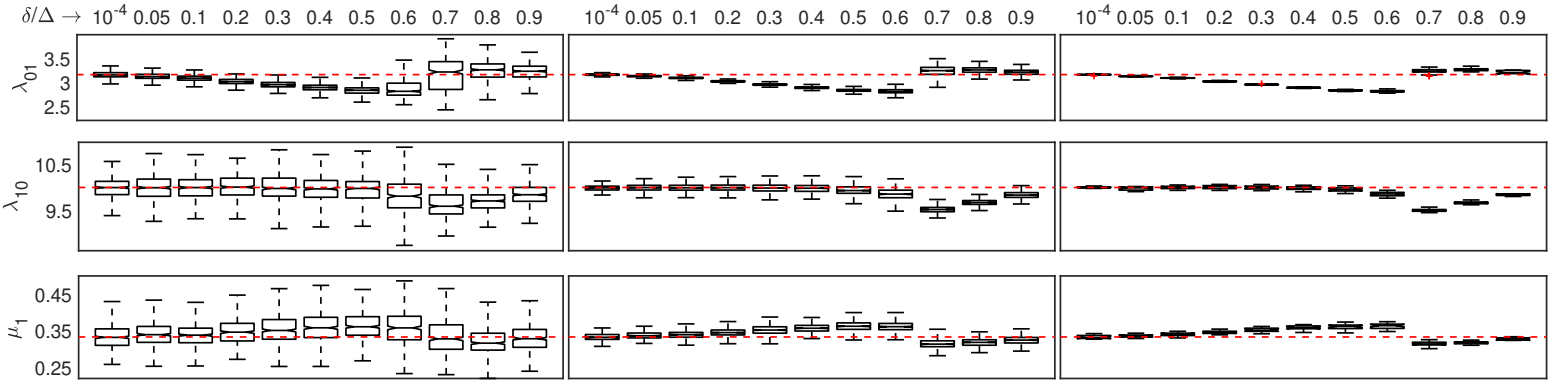


Figure S12: Box-plots showing rate estimates when $m = 0$ under study 3 of Table S1 when $N_E = 100$ (left), $N_E = 1000$ (middle) and $N_E = 5000$ (right). True rates indicated by the red dashed line. True rates indicated by the red dashed line.

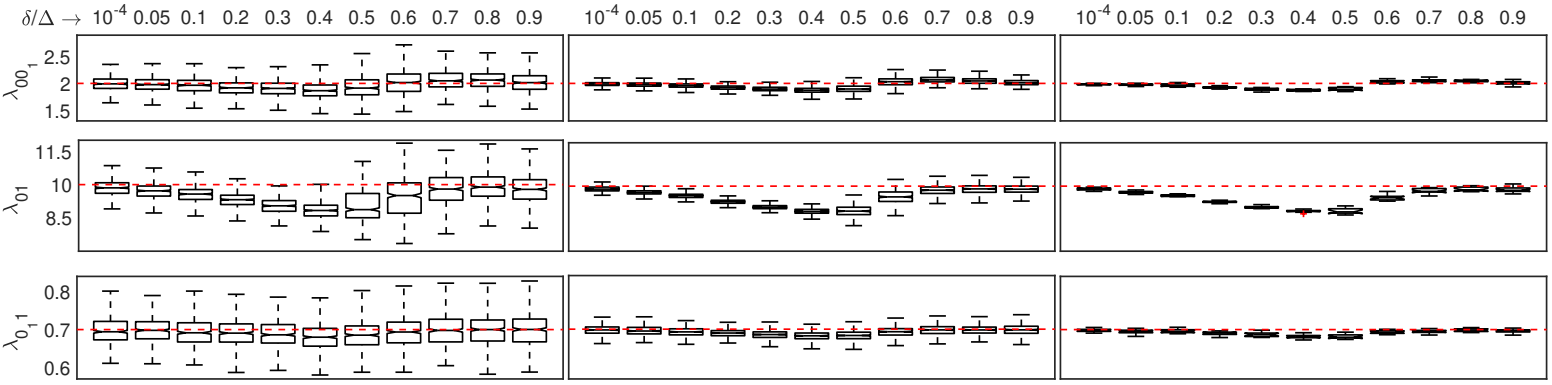


Figure S13: Box-plots showing rate estimates when $m = 1$ under study 6 of Table S1 when $N_E = 100$, $N_E = 1000$ and $N_E = 5000$ (right). True rates indicated by the red dashed line. True rates indicated by the red dashed line.

S5 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 the paper. We begin with an outline of the original method, demonstrating how maximum likelihood rate estimates of the switching rates under models of type $M_{\emptyset}^{m}$ $(m \in \mathbb{Z}_{\geq 0})$, are obtained from the photo-switching data. We then demonstrate how it can be extended to models of type $M_{\{1\}}^{m}$ $(m \in \mathbb{Z}_{\geq 0})$ that include an absorbing state 2.

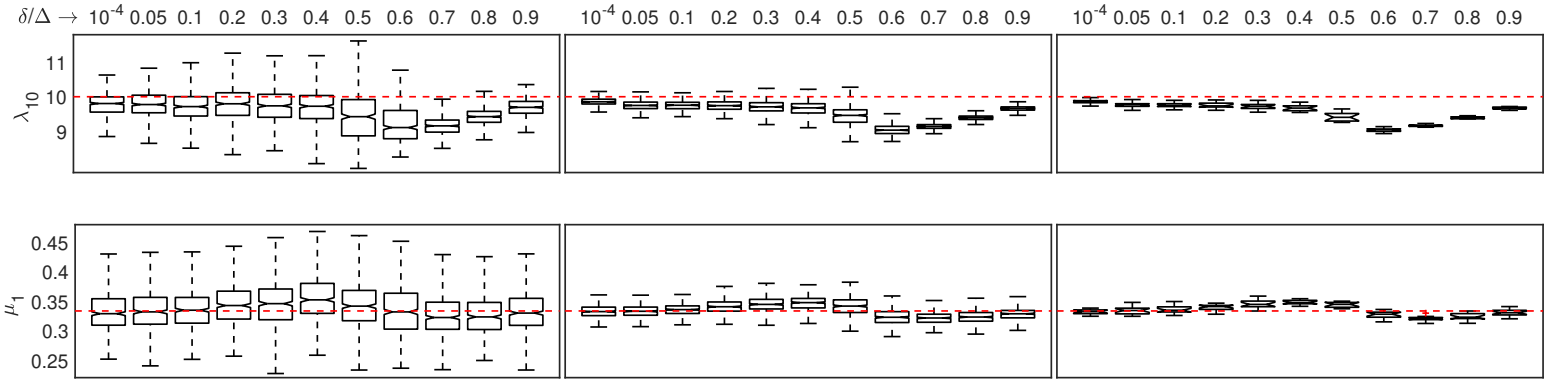


Figure S14: Box-plots showing rate estimates when $m = 1$ under study 6 of Table S1 when $N_E = 100$, $N_E = 1000$ and $N_E = 5000$ (right). True rates indicated by the red dashed line. True rates indicated by the red dashed line.

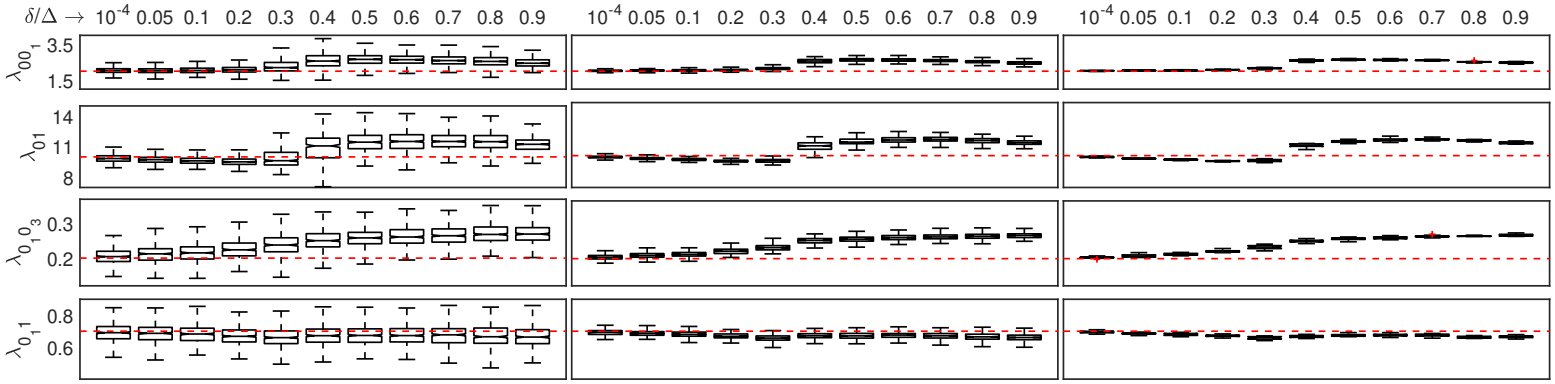


Figure S15: Box-plots showing rate estimates when $m = 2$ under study 9 of Table S1 when $N_E = 100$, $N_E = 1000$ and $N_E = 5000$ (right). True rates indicated by the red dashed line. True rates indicated by the red dashed line.

S5.1 Original method.

Consider the irreducible Markov process $\{X(t) : t \in [0, \infty)\}$ on the state space $\mathcal{S}_X = \{0, 0_1, \ldots, 0_m, 1\}$ equipped with the generator G and initial probability mass $\boldsymbol{\nu}_X$ as shown in equation (2.1) of the main text. We note here that the inclusion of the absorption 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 $m + 2$ 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 (following the same notation as the main text) $\sigma_{0_m} = \lambda_{0_m 1}, \sigma_1 = \lambda_{10}$ and when $m > 0, \sigma_{0_i} = \lambda_{0_i 0_{i+1}} + \lambda_{0_i 1}$, for $i = 0, \ldots, m + 1$. Maximum likelihood estimation of each σ_s is subsequently straightforward. Specifically, if N realizations

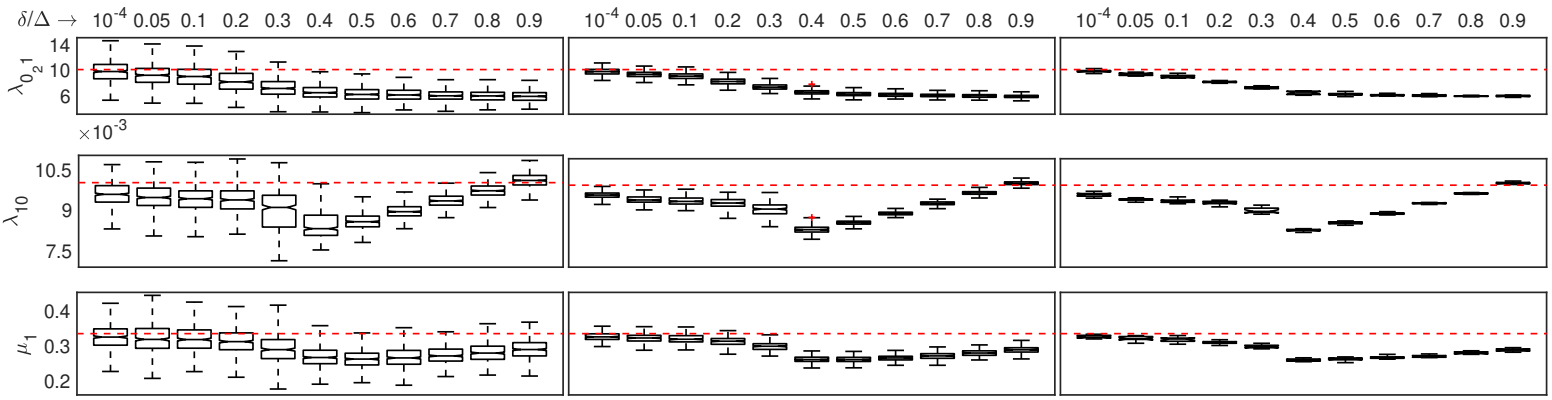


Figure S16: Box-plots showing rate estimates when $m = 2$ under study 9 of Table S1 when $N_E = 100$, $N_E = 1000$ and $N_E = 5000$ (right). True rates indicated by the red dashed line. True rates indicated by the red dashed line.

$t_s^1, t_s^2, \ldots, t_s^N$ from T_s are obtained, then the maximum likelihood estimator $\hat{\sigma}_s$ of σ_s is given by

$$\hat{\sigma}_s = \frac{N}{\sum_{i=1}^{N} t_s^i}. \tag{41}$$

Suppose now that $\{X(t)\}$ is not directly observable but instead the times spent in the Dark-On cycle $\phi_0 = 0 \to 0_1 \to \ldots 0_m \to 1$ and the On-Dark cycle $\phi_1 = 1 \to 0$ are observed. Firstly, we note that T_1 is the dwell time in an On-Dark cycle ϕ_1. Now let random variable $\bar{T}_0$ be a dwell time in the Dark-On cycle ϕ_0; that is, $\bar{T}_0$ is the time taken to reach state 1 from state 0 along the path ϕ_0. One can express $\bar{T}_0 = \sum_{i=0}^{m} T_{0_i}$, 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 $m = 0$.

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

$$f_{\bar{T}_0}(t) = \sum_{i=0}^{m} k_i \sigma_{0_i} \mathrm{e}^{-\sigma_{0_i} t} \qquad t > 0. \tag{42}$$

When $m = 1$, it can be shown $k_0 = 1 + \frac{\lambda_{00_1}}{\sigma_{0_1} - \sigma_0}$ and $k_1 = 1 - k_0$. When $m = 2$, 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.

Recognizing that the each row of G must sum to zero, the density in (42) enables the unknown photo-switching parameters

$$
\boldsymbol{\theta} = \begin{pmatrix} \lambda_{00_1} & \dots & \lambda_{0_{m-1}0_m} & \lambda_{01} & \dots & \lambda_{0_m 1} \end{pmatrix}^\top
$$

to be estimated via maximization of the log-likelihood function

$$
\ell(\bar{t}_0^1, \bar{t}_0^2, \dots, \bar{t}_0^N | \sigma_{0_0}, ..., \sigma_{0_m}) = \sum_{i=0}^{N} \log\left(\sum_{i=0}^{m} k_i \sigma_{0_i} \mathrm{e}^{-\sigma_{0_i}\bar{t}_0^i} \right), \tag{43}
$$

where $\bar{t}_0^1, \bar{t}_0^2, \dots, \bar{t}_0^N$ are N realizations from $\bar{T}_0$. Numerical optimization of (43), together with maximum likelihood estimator $\hat{\sigma}_1 = N/(\sum_{i=1}^{N} t_1^i)$ from equation (41) can be used to obtain the maximium likelihood estimate $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$.

In the context of this paper and that of Lin et al. (2015), the data 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 $\boldsymbol{y}$ (from a single emitter), can be written in block vector form

$$
\boldsymbol{y} = \begin{bmatrix} \mathbf{0}_{n_0^1}^\top & \mathbf{1}_{n_1^1}^\top & \mathbf{0}_{n_0^2}^\top & \mathbf{1}_{n_1^2}^\top & \dots & \mathbf{0}_{n_0^N}^\top & \mathbf{1}_{n_1^N}^\top \end{bmatrix}^\top, \tag{44}
$$

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 (44) 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 (41) and $\{\tilde{t}_0^i\}_{i=1}^N$ is used to obtain $\boldsymbol{\theta}$ by maximizing the log-likelihood in equation (43). The likelihood function for dwell times $\mathcal{T} = \{t_{0.k}^i, i = 1, ..., N_k, k = 1, ..., N_E\}$ recorded from $N_E \geq 1$ independent emitters becomes

$$\ell(\mathcal{T}|\sigma_{0_0}, ..., \sigma_{0_m}) = \sum_{k=1}^{N_E} \sum_{i=0}^{N_k} \log \left(\sum_{i=0}^{m} k_i \sigma_{0_i} \mathrm{e}^{-\sigma_{0_i} \bar{t}_{0,k}^i} \right), \tag{45}$$

which again can be numerically optimized, in our case using the Nelder-Mead simplex.

The key problems associated with this method are: 1. the observed dwell times can only take a discrete set of values, however, the true dwell times are continuous random variables. 2. observed dwell times for the On-Dark cycle ϕ_1 may be an over estimate of the true dwell times as short transitions to an Off state may not be detected. 3. observed dwell times for Dark-On cycle ϕ_0 may be an overestimate as short transitions to the On state may not be detected. 4. the assumed distribution of the dwell time for the Dark-On cycle may be incorrect if not all dark states on the path are reached.

S5.2 Extension to handling the absorbing state.

To gain estimates of these switching rates in the presence of the absorbing state 2, we consider *each* observation sequence before death: $\tilde{\mathbf{y}}^j = \{y_i^j\}_{i=0}^{o^j-1}$; where $o^j \in \mathbb{Z}_{\geq 0}$ denotes the last frame a fluorophore is seen in observation sequence $j = 1, \ldots, N_E$. Photo-switching estimates are gained by fitting to the truncated dataset $\tilde{\mathcal{Y}} = \begin{pmatrix} \tilde{\mathbf{y}}^1 & \tilde{\mathbf{y}}^2 & \ldots & \tilde{\mathbf{y}}^{N_E} \end{pmatrix}$.

We are able to infer upon the absorption parameter μ_1 (when $\mu_j = 0$ for all $j \neq 1$) by using the fact that the random variable N_{10} denoting the number of $1 \to 0$ 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 $\bar{n}_{10} = \frac{1}{N_E} \sum_{j=1}^{N_E} n_{10}^j$ and n_{10}^j denotes the number of observed $1 \to 0$ transitions in sequence j. Since the exponentially fitting method, by considering dwell time sequences from the On state, always yield a maximum likelihood estimate for $\hat{\sigma}_1 = \hat{\lambda}_{10} + \hat{\mu}_1$, we obtain that $\hat{\lambda}_{10} = \frac{n\bar{1}_0\hat{\sigma}_1}{1+\bar{n}_{10}}$ and $\hat{\mu}_1 = \frac{\hat{\sigma}_1}{1+\bar{n}_{10}}$.

S6 Discussion on implementing the PSHMM algorithm.

In this section, we discuss the implementation of PSHMM algorithm. Key mathematical ideas to suggest appropriate initial parameters for these optimization techniques, including a discussion

on the multimodality of the log-likelihood surfaces will be comprehensively reviewed.

S6.1 Convergence of transmission matrices.

In most practical aspects, $\Delta = \mathcal{O}(10^{-2})$ and $\sigma_i \Delta < 1$ for all $i \in \bar{\mathcal{S}}_X$. With these parameter values, convergence of the transmission matrices from Algorithm 1 (from Section S3) usually transpires up to and including $k = 2$, with $k \geq 3$ probabilities seldom needed in practice. For the simulated parameters in this paper, our Matlab implementation of this Algorithm thus approximates the matrices $B_\Delta^{(0)}$ and $B_\Delta^{(1)}$ using only $k = 0, 1$ and 2.

S6.2 Likelihood optimization.

The log-likelihood function in equation (3.3) of the main text is optimized with respect to $\boldsymbol{\theta}$ via the Nelder-Mead simplex to obtain maximum likelihood estimates. When $m = 0$, we are able to gain starting parameters for this search using the method described below. However when $m > 0$, we use deploy a stochastic search with many starting values to mitigate for multiple modes and find the global maximum.

S6.2.1 Estimating initial rate parameters.

To numerically maximize the log-likelihood function in (5) of the main text, a starting value $\hat{\boldsymbol{\theta}}^*$ needs to be determined primarily due to the potential multimodality of the likelihood function and further to reduce the computational time given the size of the parameter space Θ^*.

To do so, we consider a crude approximate estimate to start a maximum likelihood search. In the case $m = 0$, we have found that optimizing the likelihood function of a first order Markov chain $\{Z_n : n \in \mathbb{Z}^*\}$ (used as a partial likelihood function for $\{Y_n\}$) has yielded appropriate initial estimates for λ_{01} and λ_{10} using the dataset $\tilde{\mathcal{Y}}$. The chain $\{Z_n\}$ is constructed by using one-step transition probabilities from $\{Y_n\}$ through the setting $\mu_0 = \mu_1 = \delta = 0$ and $\boldsymbol{\nu}_X = \left(\frac{\lambda_{10}}{(\lambda_{01}+\lambda_{10})} \quad \frac{\lambda_{01}}{(\lambda_{01}+\lambda_{10})}\right)^\top$, the stationary distribution for $\{X(t)\}$.

Remark 4. *We can calculate that the transition probability matrix* $\mathbf{P}^Z$ *of* $\{Z_n\}$*, whereby* $(\mathbf{P}^Z)_{ij} = \mathbb{P}(Y_n = j | Y_{n-1} = i)$ *takes the form*

$$\mathbf{P}^Z = \begin{pmatrix} e^{-\lambda_{01}\Delta} & 1 - e^{-\lambda_{01}\Delta} \\ \frac{\pi_X^0 \mathrm{e}^{-\lambda_{01}\Delta}(1-\mathrm{e}^{-\lambda_{01}\Delta})}{1-\pi_X^0 \mathrm{e}^{-\lambda_{01}\Delta}} & \frac{\pi_X^1+\pi_X^0(1-\mathrm{e}^{-\lambda_{01}\Delta})^2}{1-\pi_X^0 \mathrm{e}^{-\lambda_{01}\Delta}} \end{pmatrix}, \tag{46}$$

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 λ_{01} and λ_{10} can be obtained by optimizing the restricted likelihood function for $\{Z_n\}$ using the truncated dataset $\tilde{\mathcal{Y}}$; this takes the form $L_Z(\tilde{\mathcal{Y}}; \lambda_{01}, \lambda_{10}) = \prod_{j=1}^{N_E} \prod_{i=0}^{o^j-1} p^Z_{y_i^j y_{i+1}^j}$. If $n^j_{s_1 s_2}$ 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 $j = 1, \ldots, N_E$, then it is shown in Rajarshi (2013) that the maximum likelihood estimators of the transition probabilities in $\mathbf{P}^Z$ are given by $\hat{p}^Z_{00} = \frac{\sum_{j=1}^{N_E} n^j_{00}}{\sum_{j=1}^{N_E} n^j_{00}+n^j_{01}}$ and $\hat{p}^Z_{10} = \frac{\sum_{j=1}^{N_E} n^j_{10}}{\sum_{j=1}^{N_E} n^j_{10}+n^j_{11}}$. By rearranging the expressions for λ_{01} and λ_{10} given in (46), we obtain the crude estimates

$$\hat{\lambda}_{01} = -\frac{\log(\hat{p}^Z_{00})}{\Delta} \qquad \hat{\lambda}_{10} = \frac{\hat{p}^Z_{10}\hat{\lambda}_{01}}{(1-\mathrm{e}^{\hat{\lambda}_{01}\Delta})(\mathrm{e}^{-\hat{\lambda}_{01}\Delta} - \hat{p}^Z_{10})}. \tag{47}$$

Moreover, gaining initial estimates of the death rates $\hat{\mu_0}$ and $\hat{\mu_1}$ (when at least one rate is non-zero) can be done by considering approximate bleached times: $t^j = o^j\Delta$ for $j = 1 \ldots N_E$. These times can be used to fit the absorption time distribution (see Buchholz et al. (2014)) $f_\tau(\tau|\lambda_{01}, \lambda_{10}, \mu_0, \mu_1) = \begin{pmatrix} 1-(\boldsymbol{\nu}_X)_1 & (\boldsymbol{\nu}_X)_1 \end{pmatrix} e^{\mathbf{T}\tau}\mathbf{t}$, with $(\boldsymbol{\nu}_X)_i = \mathbb{P}(X(0) = i)$ when $i \in \mathcal{S}_X$ and

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

Estimates are gained from maximizing the log-likelihood

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

where $\hat{\lambda}_{01}$ and $\hat{\lambda}_{10}$ are the crude estimates gained from (47). In the case that $m > 0$, we may choose to set $\hat{\mu}_{0_i} = \hat{\mu}_0$ for all $i = 1, \ldots, m$ where the model permits absorption from other such dark states.

The noise parameters δ and α 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 initialize with $(\boldsymbol{\nu}_X)_1 = \frac{1}{N_E}\sum_{j=1}^{N_E} {y_0}^j$ and $(\boldsymbol{\nu}_X)_0 = 1 - (\boldsymbol{\nu}_X)_1$. When $m > 0$, 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 $m \geq 0$. 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 optimization.

S7 Simulation methods and further results.

In this section we provide the methods for the simulation studies presented in Section 4 of the main paper, specifically Figure 7 and Table 2 (of Appendix A). We then present further simulation studies that explore the effects of varying the frame length Δ, threshold (proportional to δ), number of frames N_F and absorption parameter μ_1.

S7.1 Imaging simulation.

Simulated images of a fluorophore are produced from a sequence of continuous time state transitions realized from model $M^0_{\{1\}}$, discretized to give a sequence of fractional On times $q_0, ..., q_{N_F-1}$ for each frame. All other state information (dark-state / bleached-state identity) is discarded. Let λ_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 centered at the stationary position of the fluorophore $\boldsymbol{s} \in \mathbb{R}^2$. The photon positions are binned into a grid of N_{pix} 100×100nm pixels $\{C_k \subset \mathbb{R}^2; k = 1, ..., N_{pix}\}$ representing the EMCCD camera. The expected number of photons from the fluorophore in pixel k for frame i is

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

where $\mathcal{N}(\cdot; \boldsymbol{\mu}, \Sigma)$ denotes the probability density function of the Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance matrix Σ. 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 $\mu^e_{i,k}$ by multiplying it by the EMCCD quantum efficiency, 0.9, and adding a spurious electron dark current of $0.005e^-$ s^{-1}, i.e. $\mu^e_{i,k} = 0.9(\mu_{i,k} + 5) + 0.005$. The electron count $e_{i,k}$ for pixel k in frame i is sampled from a Poisson$(\mu^e_{i,k})$ distribution. The electron count after EMCCD gain $\epsilon_{i,k}$, is then sampled from a Gamma$(e_{i,k}, \beta)$ distribution, with β 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 digitization and the final count is discretized and truncated to the range $[0, 65535]$.

S7.2 Image analysis and trace idealization.

To generate idealized 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 position of molecules used in simulated data were known. The position of each molecule was used to extract 5×5 pixel regions from the raw data centered (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 5×5 pixel Gaussian kernel. A hard threshold was applied to the photon number per frame trace at multiples of the standard deviation of the background, σ_{BG}.

S7.3 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
Δ	1/30 s
λ_{01}	0.3162 s^{-1}
λ_{10}	3.1623 s^{-1}
μ_{10}	0.1 s^{-1}
N_F	9872 frames

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

S7.4 Further results.

We analyze simulated image traces based on a range of different scenarios. All datasets have $N_E = 100$ with known initial probability $\boldsymbol{\nu}_X = \begin{pmatrix} 0 & 1 & 0 \end{pmatrix}^\top$ and unknown false positive rate $\alpha > 0$. In Section 4 of the main paper, we explored the effect of different photo-switching rates $(\lambda_{01}, \lambda_{10})$. Here, we consider varying the frame length Δ (Figure S17), threshold (proportional to δ) (Figure S18), number of frames N_F (Figure S19) and absorption parameter μ_1 (Figure S20).

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 N_E (data not shown). Figure S20 finally highlights a reduction in bias with μ_1, due to a greater number of transitions between hidden states.

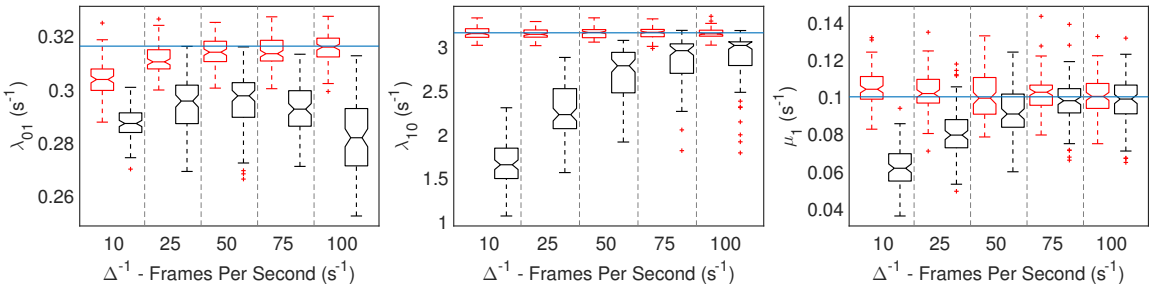


Figure S17: Boxplots showing quantiles from estimates of λ_{01}, λ_{10} and μ_1 from both exponential fitting (black) and PSHMM fitting (red) are plotted against $\log(\Delta)$. True rates given by the blue line.

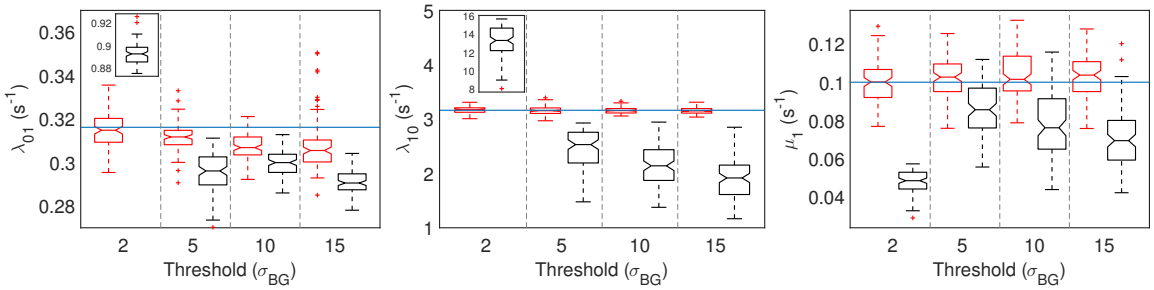


Figure S18: Boxplots showing quantiles from estimates of λ_{01}, λ_{10} and μ_1 from both exponential fitting (black) and PSHMM fitting (red) are plotted against $\log(\text{threshold})$. $N_F = 9872$ for all simulations. True rates given by the blue line.

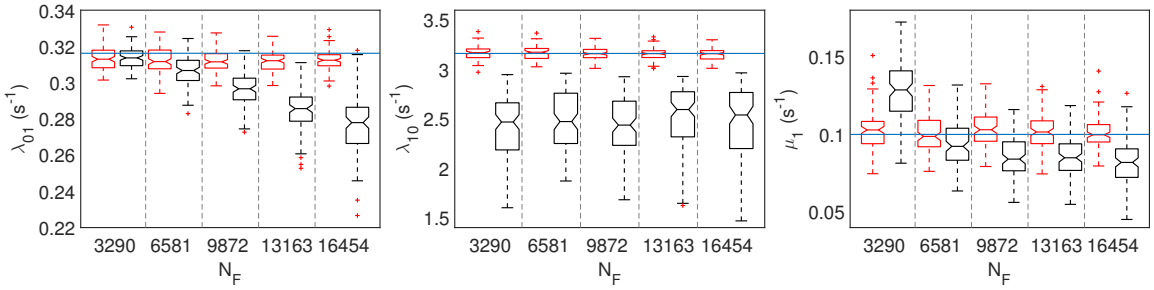


Figure S19: Boxplots showing quantiles from estimates of λ_{01}, λ_{10} and μ_1 from both exponential fitting (black) and PSHMM fitting (red) are plotted against N_F. True rates given by the blue line.

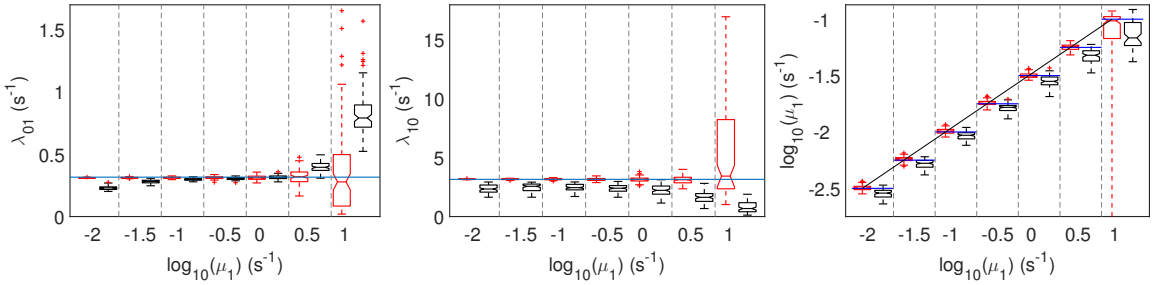


Figure S20: Boxplots showing quantiles from estimates of λ_{01}, λ_{10} and varying μ_1 from both exponential fitting (black) and PSHMM fitting (red) are plotted against $\log_{10}(\mu_1)$. True rates given by the blue lines.

REFERENCES

Buchholz, P., J. Kriege, and I. Felko (2014). Input modeling with phase-type distributions and Markov models. *Springer eBooks*.

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.

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)*, e53671.

Jacquez, J. A. and P. Greif (1985). Numerical parameter identifiability and estimability: Integrating identifiability, estimability, and optimal sampling design. *Mathematical Biosciences 77*, 201–227.

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). Quantifying and optimizing single-molecule switching nanoscopy at high speeds. *Plos One 5(1)*, e8915.

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.

Moschopoulos, P. G. (1985). The distribution of the sum of independent gamma random variables. *Annals of the Institute of Statistical Mathematics 37*, 541–544.

Rajarshi, M. (2013). Markov chains and their extensions. In *Statistical Inference for Discrete Time Stochastic Processes*, pp. 19–38. Oxford: Springer India.

Van Loan, C. (1978). Computing integrals involving the matrix exponential. *IEEE Trans Automatic Control 23*, 395–404.

Viallefont, A., J.-D. Lebreton, A.-M. Reboulet, and G. Gory (1998). Parameter identifiability and model selection in capture-recapture models: A numerical approach. *Biometrical Journal 40*, 313–325.