Bayesian Statistical Inference in Ion-Channel Models with Exact Missed Event Correction

Dr M. Epstein

Dr B. Calderhead

Prof. M.A. Girolami

Prof. L.G. Sivilotti

May 31, 2016

Running Title: Ion-channel Inference with Missed Events
Keywords: [Markov chain Monte Carlo, Ion-channels, Markov processes, Model parameterisation]

Abstract

The stochastic behaviour of single ion channels is most often described as an aggregated continuous-time Markov process with discrete states. For ligand-gated channels each state can represent a different conformation of the channel protein or a different number of bound ligands. Single channel recordings show only whether the channel is open or shut: states of equal conductance are aggregated, so transitions between them have to be inferred indirectly. The requirement to filter noise from the raw signal further complicates the modelling process, as it limits the time resolution of the data. The consequence of the reduced bandwidth is that openings or shuttings that are shorter than the resolution cannot be observed; these are known as missed events. Postulated models fitted using filtered data must therefore explicitly account for missed events to avoid bias in the estimation of rate parameters and in order to assess parameter identifiability accurately.

In this paper we present the first Bayesian modelling of ion-channels with exact missed events correction. Bayesian analysis represents uncertain knowledge of the true value of model parameters by considering these parameters as random variables. This allows us to gain a full appreciation of parameter identifiability and uncertainty when estimating values for model parameters. However, Bayesian inference is particularly challenging in this context as the correction for missed events increases the computational complexity of the model likelihood. Nonetheless, we successfully implemented a two-step Markov chain Monte Carlo method which we called "BICME" which performs Bayesian inference in models of realistic complexity. The method is demonstrated on synthetic and real single channel data from muscle nicotinic acetylcholine channels. We show that parameter uncertainty can be more accurately characterised compared with Maximum Likelihood methods. Our code for performing inference in these ion channel models is publicly available.

Introduction

Ligand gated ion-channels are transmembrane proteins that enable fast cell to cell communication, which is crucial for the functioning of the nervous system and for the control of skeletal muscle. Conformational changes in the protein are induced by the binding of an agonist, such as a neurotransmitter, to the extracellular domain of the protein. These conformational changes lead to the opening of the gate of the channel pore within the protein, enabling the flow of ions down their electrochemical gradient. The resulting all-or-nothing current has an amplitude typically in the pA range at physiological membrane potential and ionic concentrations and can be recorded by patch clamp techniques [1].

The conformational changes that occur during the gating process cannot be observed directly and therefore must be inferred from such recordings. These single-channel data are used to fit and assess mechanistic Markov models that describe the binding events and conformational changes that occur during channel activation. Such mechanistic mathematical models consist of components that may be directly interpreted as part of the system being modelled, allowing investigation of the underlying physical process.

Ion channels are unique among proteins in allowing the prolonged recording of single molecule activity at high temporal resolution. This in principle allows the fitting of models that are unusually detailed and relatively close to the physical reality of activation. For instance, in a ligand gated ion channel, this reaction consists of the binding of several neurotransmitter molecules followed by conformational changes that eventually result in the channel opening. Fitting an appropriately parameterised model to estimate the rate constants for the different transitions allows us to establish the simplest model that adequately describes the observations and thus to probe the energy landscape of the channel. In turn this is useful to compare with structural information and with molecular dynamics simulations of protein dynamics.

Single channel records show only whether the channel is open or closed and realistic channel activation requires several closed and several open states. Ion-channel kinetics may be described mathematically using aggregated Markov models, since it is not possible to directly observe the conformational state of the protein, and their memoryless nature appears to match observed channel behaviour. There are obvious difficulties however. While temporal resolution can be very good for single-channel recordings, 10-30μs at best, it still is not infinite and this prevents the identification of shorter dwells of the channel in the open or closed states. These short dwells commonly occur in most ion-channels and, importantly, result in missing data, which we must also account for in our modelling approach.

Recordings are usually "idealised", that is, converted from digitised records to lists of intervals using various methods, such as time course fitting, threshold crossing and Hidden Markov Models to convert the recording to a putative sequence of open and closed intervals [2]. Models of increasing complexity are fitted to these idealised data until a reasonable description of the channel behaviour is obtained. Maximum Likelihood (ML) methods have been to date the most common inferential framework for estimating model parameters from single channel data. This approach has been found to be useful in examining the activation of channels in the nicotinic superfamily [3, 4, 5, 6, 7]. The main limitation of the ML approach however is that it is not straightforward to check the parameter non-identifiability and the impact of parameter uncertainty. Parameter non-identifiability can occur in two scenarios as outlined in [8]. First, there may exist a continuum of parameter values at the Maximum Likelihood point such that estimated values of the model parameters cannot be constrained within a finite range. Second, there may exist multiple well-constrained but discrete solutions that describe the observed data reasonably well. The first case, which we examine in this paper, can be a particular problem

for ion-channel models where the underlying structure of the physical process is not directly observable. The model may consequently be overparameterised and there may be great uncertainty in the parameter estimates. The ability to determine parameter identifiability and the impact of parameter uncertainty is vital for allowing physiologically meaningful conclusions to be drawn from a hypothesised ion-channel model, for example, when comparing the action of different channel agonists on a receptor, or when assessing the physical impact of mutations on channel behaviour. Theoretical work has explored the maximum number of parameters that can be fitted to single channel recordings [9] and investigated non-uniqueness of models [10] but these techniques are difficult to apply when fitting models to real experimental data. Assessing the impact of parameter uncertainty on model predictions also remains a challenging problem.

Our group used the approach pioneered by Colquhoun & Hawkes [11, 12] to examine parameter identifiability by assessing standard errors and correlations derived using the empirical covariance matrix calculated at the Maximum Likelihood estimate. Thus we verified the properties of the ML estimators for our main results on the nicotinic and glycine channel by extensive simulations [5, 12] showing our approach can reliably detect rate constants as fast as $130{,}000s^{-1}$ for glycine channels [5]. This technique has also robustly revealed intermediate conformational states in the nicotinic and glycine channels [7] and in the ELIC receptor in prokaryotes [13]. Despite these successes, there are limitations to the use of Maximum Likelihood inference. In particular, checking the validity of model parameter estimates by fitting simulations remains an *ad-hoc* laborious process.

There is increasing interest in Bayesian approaches in biophysics [14, 15] and in particular in single channel analysis e.g. [16, 17, 18]. In Bayesian methods, rate parameters are treated as random variables with a known prior probability distribution. This allows the assessment of parameter identifiability through the calculation of their posterior probability distributions. The uncertainty in these distributions may therefore be directly propagated through in order to examine uncertainty in the predictions of the model. Bayesian methods are more computationally expensive than ML estimation when the posterior probability distribution is not known analytically and so Markov chain Monte Carlo (MCMC) sampling schemes are required to facilitate Bayesian inference in such models. MCMC schemes are called "samplers" as they derive estimates of this density by defining probabilistic Markov processes that draw samples from such probability distributions.

However, to our knowledge, there is currently no Bayesian approach for ion-channel models that exactly corrects the model likelihood for the two important, inevitable technical constraints of the experimental data, namely the limited temporal resolution in the experimental record and the lack of knowledge about the number of channels in the experimental patch. This paper addresses this shortfall thus:

- We propose a practical approach specially tailored for performing highly efficient Bayesian inference in ion-channel models using multiplicative Metropolis-within-Gibbs (MWG) and Adaptive MCMC sampling in a package called "BICME" available at `https://github.com/miepstei/bicme`.

- We examine how well our approach assesses parameter identifiability and parameter uncertainty using the obtained posterior distribution. We examine how uncertainty in parameter values affects the uncertainty in model predictions and we apply our approach to both synthetic and real experimental data for muscle nicotinic acetylcholine receptors.

- We correct the model likelihood for missed events and compare the results of MCMC methods with those from existing ML and Bayesian methods [12, 19].

Materials and Methods

Ion-channel Stochastic Framework

The analysis aims to infer a continuous time Markov process from a discretely sampled signal which has been idealised by a time course fitting procedure [2]. Idealisation deconvolves the channel signal from the filtered output. Within a continuous time framework we seek an expression for the probability of observing an open or shut interval of length t in the experimental recording. To calculate the likelihood, we need the probability density of the length of time for which the channel appears to be open (though it may contain missed brief shuttings) or shut. Using the notation of Colquhoun and Hawkes [20] we outline the derivation for the required probability as follows. Consider a continuous time finite-state Markov process $S(t), t > 0$, such that $S(t) = i$ denotes that the process is in state i at time t. The state space, I, of this process represents the conformational states of the proposed mechanism. The possible transitions between states in this process are encoded and parameterised with a corresponding generator matrix $\mathbf{Q}$, which contain the rates of transition between the conformational states of the mechanism. Each state in I is either open (set A), or closed (set F).

We begin by assuming that there is perfect resolution in the record. From [20], the $\mathbf{Q}$ matrix is partitioned into conformations that produce one of two conductance levels, such that partition $\mathbf{Q}_{AA}$ represents the transition rates between states that that are open and $\mathbf{Q}_{AF}$ represents the rates of transitions from open to shut states. Partitions $\mathbf{Q}_{FF}$ and $\mathbf{Q}_{FA}$ are denoted similarly for transitions within shut states and from shut states to open states, respectively. The initial goal is to derive a probability for observing an open (or closed) interval of length t given that we cannot directly observe transitions within each conductance class A or F. We can subsequently derive a likelihood for the idealised recording from the probabilities of these individual sojourns [11]. Conditional on the process starting in an open state $i \in A$, the individual probabilities of the process remaining within the set of open states A for a sojourn t, and instantaneously transitioning to a shut state $j \in F$ are given by the elements of the matrix $\mathbf{G}_{AF}(t)$ in Equation 1 [20];

$$\mathbf{G}_{AF}(t) = \exp(\mathbf{Q}_{AA}t)\mathbf{Q}_{AF} \tag{1}$$

The overall probability of observing an opening of length t, where the process starts in any one of the open states and finishes in any one of the closed states, is given in Equation 2,

$$f_A(t) = \phi_{\mathbf{A}}\mathbf{G}_{AF}(t)\mathbf{u}_F \tag{2}$$

where the initial vector $\phi_{\mathbf{A}}$ denotes the probabilities of a sojourn starting in any one of the hidden open states of the process, and the final column vector $\mathbf{u}_F$ is a column of ones that sums up the probabilities of finishing in each shut state. The analogous probabilities for a closed sojourn are obtained by switching the partition labels A and F.

Accounting for Missed Events

Single channel recordings have finite time resolution, and channel events shorter than the resolution time are not observed in the idealised record. With the continuous time framework outlined above, these unobserved transitions need to be accounted for within the likelihood, and we used the exact missed events correction of [21, 22]. The probability of observing an open interval of length t, in the presence of missed events, can be calculated by noting that t can be broken into three parts given the recording resolution time τ. The first part is the open

interval of length $t-\tau$ which may contain zero, one or many shuttings of length $< \tau$ that are undetected. This first element, $\mathbf{R}_A(t-\tau)$, is known as the "survivor function" as it represents transitions from open states to closed states over a time $t-\tau$, during which shut intervals of duration $< \tau$ may occur but are not detected. The second component is the instantaneous transition from the open to the shut states, $\mathbf{Q}_{AF}$. The third component is a shut interval of length τ that must occur such that the open interval is brought to a close, which is simply $\exp(\mathbf{Q}_{FF}\tau)$. We denote these probabilities analogously to $\mathbf{G}_{AF}(t)$ as $\mathbf{G}^e_{AF}(t)$ [11];

$$\mathbf{G}^e_{AF}(t) = \mathbf{R}_A(t-\tau)\mathbf{Q}_{AF}\exp(\mathbf{Q}_{FF}\tau) \tag{3}$$

Calculating the correction for missed events (Equation 3) is much more involved than the ideal case (Equation 1) as it requires calculating the convolution of open intervals with an unknown number of short shut intervals, each less than τ in length, over the period $t-\tau$. The required convolution can be expressed as the following Laplace Transform [11];

$$\mathbf{R}^*_A(s) = [\mathbf{I} - \mathbf{G}^*_{AF}(s)\mathbf{S}^*_{FF}(s)\mathbf{G}^*_{FA}(s)]^{-1}(s\mathbf{I} - \mathbf{Q}_{AA})^{-1} \tag{4}$$

where $\mathbf{G}^*_{AF}(s)$ is the Laplace Transform of the $\mathbf{G}_{AF}(t)$ matrix defined in equation 1 and $\mathbf{S}^*_{FF}(s) = (\mathbf{I} - \exp(-(s\mathbf{I} - \mathbf{Q}_{FF})\tau))$ represents the Laplace Transform of a short sojourn in the shut states. A corresponding transform for observed shut intervals, $\mathbf{R}^*_F(s)$, is found again by reversing the partitioning labels.

Although the Laplace Transform in Equation 4 is intuitively simple to understand, it must be inverted to provide expressions for probability densities of open and closed sojourns as a function of time. Fortunately, an exact inversion of the transform in Equation 4 was found by [21] in the form of a piecewise solution in multiples of the resolution time, such that a different solution is obtained for $\tau < t \leq 2\tau$, $2\tau < t \leq 3\tau$, and so on. The solution relies on evaluating a matrix polynomial of increasing order based on multiples of the resolution time, τ. While this calculation becomes numerically unstable for large t,we can employ an asymptotic form of the solution (found by [22]), which is accurate for periods of $t > 3\tau$ [11]. Calculating the asymptotic solution in part relies on a numerical root finding procedure that adds to the computational burden of evaluating the corrected likelihood.

Accounting for the Number of Channels in the Patch

In general it is not known how many channels are present in the patch from which we are recording. Counting the number of channels simultaneously open provides only a minimum estimate. Our analysis therefore requires that records be broken up into stretches where it can be assumed almost certainly that the gating of a single channel molecule is being observed, because the gating of multiple channels would be detectable. These stretches of open and shut times will be referred to as groups. At low concentrations, channel openings occur in groups termed as "bursts", during which it is possible to assume that only a single channel is operating, since multiple channel openings during this interval would almost surely result in multiple conductance levels observable in the record. Bursts are separated by long shut times that are the expression of the time taken for the channel to rebind agonist and contain information on the binding steps in the mechanism. The record can be broken up into groups by choosing a time interval value, t_{crit}, on the basis of the dependence of the shut time distribution on the agonist concentration. Shut intervals longer than t_{crit} are deemed to separate the record into bursts. Effectively, the long shuttings can be only shortened by the presence of more than one channel in the patch. As a result, the analysis must take into account the fact that the

real shut sojourn before the first opening of the burst is equal to or longer than t_{crit}. This is done by employing corrected initial vectors, known as CHS vectors, [11] denoted ${\phi_{\mathbf{A}}}^{chs}$. Their use has been shown to increase the precision of estimates of rate constants from low concentration records [12]. In contrast, a feature of channel activity at high concentrations is the presence of channel desensitisation; long stretches of channel activity are separated by desensitised shuttings where there is no observed opening. Given that desensitised states are not typically incorporated in the model, these shuttings are excised from the record and are not used for fitting [12]. The analysis uses the groups (clusters) of openings separated by the desensitised gaps. The open probability in the clusters is high enough to be sure that only a single molecule is active. In this instance, equilibrium vectors denoted ${\phi_{\mathbf{A}}}^{eq}$ are used to provide the probability of starting in any open state at the start of the group.

Accounting for an unknown number of channels in the patch requires an alteration to the likelihood calculation. At a given concentration of agonist, $[x_i]$, time interval, t_{crit}, and resolution time, τ, the log-likelihood of the series of bursts or clusters may be defined as the sum of the log probabilities of N individual groups, each of varying length m_j, observed in the record. The initial vector, $\phi_{\mathbf{A}}$ is either the CHS vector ${\phi_{\mathbf{A}}}^{chs}$ or the equilibrium vector ${\phi_{\mathbf{A}}}^{eq}$ depending on the agonist concentration in the experiments, as outlined above.

$$\mathscr{L}(y \mid \mathbf{Q}, t_{crit,i}, \tau_i, [x_i]) = \sum_{j=1}^{N} \log[\phi_{\mathbf{A}} \mathbf{G}^e_{AF}(t_{1_j}) \mathbf{G}^e_{FA}(t_{2_j}) \ldots \mathbf{G}^e_{AF}(t_{m_j}) \mathbf{u}_F] \tag{5}$$

Equation 5 denotes the calculation of the corresponding log-likelihood over a series of recordings at different agonist concentrations which may similarly be defined as the sum of the log-likelihood at each concentration. The code for this likelihood calculation is publicly available at https://github.com/DCPROGS/HJCFIT.

An Introduction to Bayesian Inference and MCMC sampling

Unknown parameters in the model are specified as random variables in the Bayesian framework (for further discussion see Supplementary Materials S1 and S2). Bayes Theorem for continuous variables can be stated as per Equation 6 below:

$$p(\theta \mid y) = \frac{p(y \mid \theta) p(\theta)}{p(y)} = \frac{p(y \mid \theta) p(\theta)}{\int p(y \mid \theta) p(\theta) d\theta} \tag{6}$$

which requires the specification of a prior probability distribution, $p(\theta)$, referred to as the "prior", which captures what is known about the model rate constants before any data is observed. The combination of the model likelihood, $p(y \mid \theta)$ as calculated in Equation 5 , and the prior $p(\theta)$ results in a posterior probability distribution, $p(\theta \mid y)$, referred to as the "posterior", that may be calculated only pointwise up to a normalising constant. This means an analytical description of the posterior as a normalised probability distribution is not available, but the posterior distribution of the rate constants can be estimated using MCMC sampling.

We now provide a basic outline of MCMC sampling. We define a separate discrete-time Markov chain, noting here that this is a different mathematical object from the aggregated Markov process defined previously, which was used to define the likelihood function. The aim of this Markov chain is to sample values from the posterior distribution and has the desirable property that its stationary distribution is the posterior distribution under estimation. The Markov chain is initialised at an initial set of parameters, θ^c, and the posterior probability density at this point $p(\theta^c \mid y)$ is calculated. In each MCMC step, we define another probability

distribution $Q(. \mid \theta^c)$, known as the proposal distribution, to propose a new set of parameters, θ^*, given the current parameter values, for example a Gaussian distribution with mean equal to the current parameters. The posterior probability density of the proposed parameters is then calculated, $p(\theta^* \mid y)$. The move is accepted according to the Metropolis-Hastings acceptance ratio [23], which means accepting the move with a probability given in Equation 7:

$$min(1, \frac{p(\theta^* \mid y)q(\theta^c \mid \theta^*)}{p(\theta^c \mid y)q(\theta^* \mid \theta^c))} \tag{7}$$

When the sampler uses a symmetric proposal distribution, Equation 7 simplifies to min(1, $p(\theta^* \mid y)/p(\theta^c \mid y)$), the minimum of 1 and the ratio of the values of the posterior distribution at θ^* and θ^c. If the move is accepted, we set the current parameters equal to the proposed parameters, $\theta^c = \theta^*$, else we retain the current parameters. At each iteration we record the current parameters as another sample from the posterior distribution. The Metropolis-Hastings ratio ensures that the samples we obtain are in fact distributed according to the posterior distribution. We note that there is an initial period during which the chain is converging to the correct distribution; this is known as the "burn in" phase of the chain. These initial samples generated by our sampler are therefore typically discarded in subsequent analysis and an example of this convergence is shown in Figure 3a.

The main aims when choosing different MCMC samplers are 1) to maximise the speed of convergence of the chain and 2) to minimise the autocorrelations of the MCMC samples generated by the algorithm. The first can be assessed by plotting the parameters or the value of the log-posterior as the MCMC chain iterates and collects the parameter samples. The second can be assessed by calculating the autocorrelation lags of samples taken after the chain has converged. A commonly used metric of sampling efficiency that incorporates the autocorrelation of the samples is the Effective Sample Size (ESS), which gives an indication of the number of equivalent independent samples drawn per sampling iteration or per unit of computational time. Optimising an MCMC sampler often involves assessing different choices of proposal distribution $Q(. \mid \theta^c)$. For a brief introduction to Bayesian methods and MCMC sampling algorithms, please refer to the Supplementary Materials S1 and S2 and for a full primer in the context of biophysics see [15]. We now outline a strategy which demonstrates an efficient sampling approach for evaluating the ion-channel model posterior distribution.

MCMC Sampling of Ion-channel Models with Missed Events Correction

We assume with all our examples that we know little about the values of our rate constants, so that all rate constants have uniform priors. Opening and closing rates and dissociation rates have a prior, $U[0, 10^6]$, whereas association rates are limited by the theoretical rate of diffusion of the agonist, for which we use a prior of $U[0, 10^{10}]$. It should be noted that these are the same bounds that are in place during a typical Maximum Likelihood model fitting, but within the Bayesian framework they are specified as the prior probability distributions.

We propose a two-step MCMC sampling strategy which we present in our software package BICME. First, we employ a pilot MCMC based on a Metropolis-within-Gibbs (MWG) sampling scheme [24] in order to locate quickly the approximate mode of the posterior distribution [25]. We then switch to an Adaptive MCMC sampler [26], that learns a covariance matrix based on the empirical covariance between the model parameters at this mode. This relies on the assumption that the posterior distribution is unimodal. In fact analysis of single channel records at a single concentration can result in bi-modal distributions [12], although in practice the use of multiple concentrations for fitting often removes this second "false" mode [27]. Nonethe-

less, a different starting position for the pilot MCMC sampler can be found by sampling initial parameters from the prior distribution to assess convergence to the same mode. Our two step approach greatly speeds up the inference, since the second Adaptive sampler would converge much more slowly during the initial transient phase of the Markov chain's exploration of the parameter space.

The pilot MCMC chain (Algorithm S1) is based on a Multiplicative MWG algorithm [24], which updates each parameter individually in log space with proposal distribution $q^* = qe^y$, where $y \sim N(0, \Sigma)$ (Algorithm S1). The proposal distribution is scaled during the burn-in phase of the chain to account for varying parameter magnitudes. The benefit of this approach is that the multiplicative proposals in the original parameter space speed up convergence of the Markov chain compared to additive proposals. After convergence and location of the posterior mode, an Adaptive MCMC algorithm [26] is employed (Algorithm S2). The second algorithm learns an appropriate covariance structure from the sample history of the chain in such a way that it still converges to the correct stationary distribution, see [28] for full details.

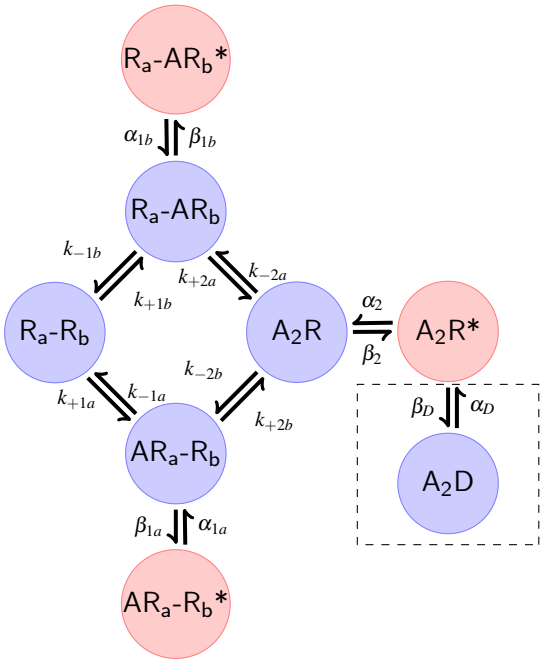
We consider the Effective Sample Size (ESS) (see Supplementary Materials S2) of each sampler as a metric for sampling efficiency. For each sampling algorithm we can assess both the ESS generated per iteration and the ESS per second of runtime, giving an assessment of overall computational efficiency. In practice, it is the second measurement that is often more important, since it gives a direct measure of computational efficiency across algorithms, although this may be highly dependent on specific implementations in code.

Results

Bayesian inference with synthetic data

We initially evaluate our method by examining whether the dual MCMC sampling strategy we propose works on simulated data produced with a well characterised ion-channel model. Simulated data at three different agonist concentrations were fitted simultaneously. The model we chose (shown in Figure 1a) was first proposed by Colquhoun and Sakmann in 1985 [29] to describe single channel activity of the muscle nicotinic receptor and validated by other labs [30]. Hatton et al. [3] investigated parameter identifiability when this model was used to fit muscle nicotinic data using Maximum Likelihood inference [12]. This makes it ideal to test whether Bayesian inference can also successfully identify the model parameters. In contrast to Maximum Likelihood inference, we provide estimates of parameter uncertainty derived from the posterior distributions. We display how the uncertainty over the estimated parameters affects the predictions of the observable features of the data by drawing sets of parameters from the posterior distribution and assessing the corresponding variability of the predictions.

The muscle nicotinic ion-channel model used for fitting (Figure 1a) has three open states, four closed states and two binding sites, A and B. The dashed box contains a "desensitised" state used only to simulate high concentration data [12]. This state is not fitted. We assume that the binding sites are independent for this model, i.e. the presence of agonist bound to one site (e.g. site A) does not impact the rates of binding or unbinding of agonist at the other site (site B) and vice versa. As the channel is at steady-state, we can assume that the principle of microscopic reversibility holds in the mechanism. This means that, for the mechanism with a cyclic component in Figure 1a, the product of the rate constants going clockwise around the cycle is equal to the product of the rate constants in the anti-clockwise direction [20]. This allows an additional rate constant to be constrained [31]. The assumption of independent binding sites, together with the assumption of microscopic reversibility in the cycle of the



(a) Candidate model

Rate	Units	Value
α_2	s^{-1}	2000
β_2	s^{-1}	52000
α_{1a}	s^{-1}	6000
β_{1a}	s^{-1}	50
α_{1b}	s^{-1}	50000
β_{1b}	s^{-1}	150
β_D	s^{-1}	5
α_D	s^{-1}	1.4
k_{-2a}	s^{-1}	1500
k_{+2a}	$M^{-1}s^{-1}$	2.0×10^8
k_{-2b}	s^{-1}	10000
k_{+2b}	$M^{-1}s^{-1}$	4.0×10^8
k_{-1a}	s^{-1}	1500
k_{+1a}	$M^{-1}s^{-1}$	2.0×10^8
k_{-1b}	s^{-1}	10000
k_{+1b}	$M^{-1}s^{-1}$	4.0×10^8

(b) Rate constants used to simulate data for the mechanism in Figure 1a

Figure 1: **1a**: Example model used for fitting. This seven state model has previously been used for the analysis of parameter identifiability in acetylcholine receptor gating [3] with limited time resolution using maximum likelihood estimation [12]. This model has three open states (in red) and four closed states (in blue) and two agonist binding sites A and B. The data are simulated assuming independent binding sites for the agonist. An additional eighth state inside the dashed box is used to simulate data at high concentrations (10μM acetylcholine) but is not used for fitting. **1b**: Parameter values used to simulate data from the model [12].

mechanism, reduces the number of rate constants to be estimated from 14 to 10, since, as in [12], $k_{+1a} = k_{+2a}$, $k_{+1b} = k_{+2b}$, $k_{-1a} = k_{-2a}$ and $k_{-1b} = k_{-2b}$.

Our design for the experiments reproduce that in the previous ML identifiability study with this model [12], with respect to the number of events, concentration values, the t_{crit} times chosen to divide the records into groups and the resolution imposed onto the raw records (25μs). The same rate parameters were used to generate raw records (Table 1b). Figure 2 shows the workflow for the experiments. Two sets of intervals using low agonist concentrations of 30nM and 100nM were simulated from the basic seven state model in Figure 1a. In addition, a set of intervals at a high agonist concentration (10 μM) was simulated from the same model that includes the additional desensitised state A_2D. The additional state is attached to the doubly liganded opening A_2R^* with transition rate constants $\alpha_D = 1.4$, $\beta_D = 5$ and is represented by the dashed box in Figure 1a. It introduces into the simulated data long shut times that mimic the long silent periods that appear in real experiments at high acetylcholine concentrations. For each of the three agonist concentrations, 20,000 intervals were generated reflecting the number of events that can be gathered in a typical single channel experiment with the muscle nicotinic channel. The t_{crit} time was set at 3.5ms for the low concentration recordings and 5 ms for the high concentration recording. This t_{crit} value is used to break up the resolved record into groups of openings that almost certainly all originate from the same individual channel. For the low concentrations, CHS vectors were used in the calculation of the likelihood (see Methods). We

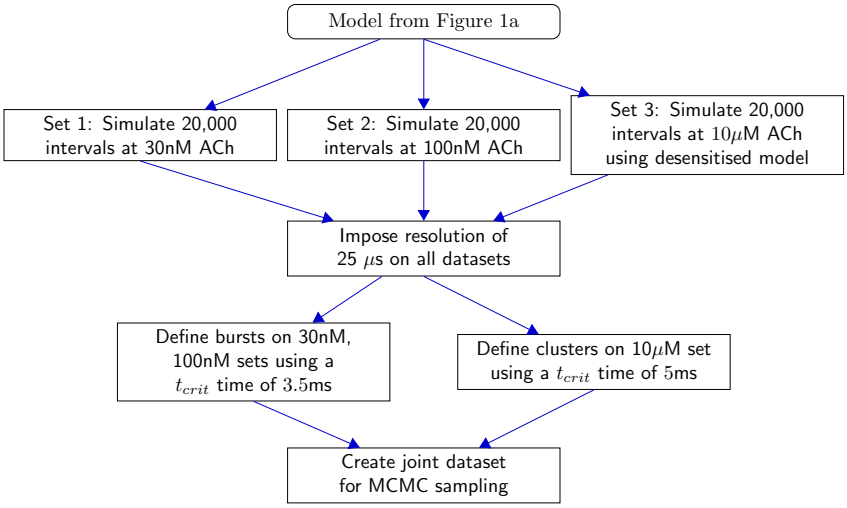


Figure 2: An experimental workflow to perform MCMC sampling using the muscle nicotinic receptor in Figure 1a. Two low concentration datasets (30nM and 100nM) are simulated and these records are separated into groups (bursts), to account for the lack of knowledge of the number of channels in the patch. A high concentration recording is also separated into groups discarding the intervals where the channel is desensitised. See Methods for the definition of groups. The joint dataset comprising the groups for all three sets is then used for MCMC sampling. The first sampler (MWG) is used to locate the posterior mode. The Adaptive sampler, started at the mode, learns the covariances of the posterior distribution and then draws samples for analysis.

apply the two-step MCMC sampling approach, described in the Methods, to this synthetic data. For the initial pilot MWG chain, 10,000 samples were drawn from the posterior distribution, of which 5,000 were discarded as burn-in samples. Subsequently, the Adaptive MCMC was started at the posterior mode evaluated from the pilot chain output. 100,000 samples were then drawn from the posterior distribution using this second chain and again half of these samples were discarded as burn-in, since the chain is initially learning the correlation structure of the posterior distribution and takes time to converge to the distribution of interest.

Whenever we employ MCMC to draw samples from a distribution, it is important to examine diagnostics of the Markov chain to ensure its convergence. We initially consider some diagnostic plots of the pilot chain, the Multiplicative MWG sampler (Algorithm S1), used to locate the posterior mode. One such diagnostic is to examine visually how quickly the chain converges to the posterior distribution given its starting position. This can be done by observing how quickly the value of the log-posterior of the model approaches a stable, stationary time-series using a trace plot of how the log posterior value changes, with each set of sampled parameters, as the chain iterates. In Figure 3a, we observe the sampler converging quickly to the posterior mode after fewer than 500 iterations. A second possible metric involves calculating and plotting the autocorrelations of the samples drawn by the MCMC chain after it has converged. Although samples from the MCMC chain are invariably autocorrelated, we aim to deploy an MCMC chain that exhibits as little autocorrelation as possible. The autocorrelations of each series of parameter samples are considered a measure of the sampling efficiency of the chain. This is distinct from any correlations which may be observable between parameters in the posterior distribution. Higher levels of autocorrelation mean that the sampler takes a

larger number of iterations to produce one independent sample from the posterior distribution. In such cases we say that our Markov chain mixes poorly. This is seen to be the case in our example with the MWG sampler, which displays very inefficient mixing after convergence. Although it was observed in Figure 3a that the MCMC chain converged to the posterior mode in around 500 iterations, however the proposal distribution was allowed to adjust according to the acceptance rate until 5,000 samples had been drawn. We therefore still discard the first 5000 of the drawn samples under the assumption that these samples are drawn when the MCMC chain is still converging. We considered the next 5,000 of samples to be drawn from the posterior distribution, and found that these 5,000 samples exhibit high levels of autocorrelation (shown in Figure 3b for parameter α_2). This is shown by the persistence of the sample autocorrelation, even after the 100th lag. The reason behind this can be examined by illustrating the distribution of proposed points relative to those points that are accepted as samples. Figure 3c plots the 5,000 posterior samples of α_2 and β_2. This shows that many proposals (red points) are made in areas of lower posterior probability density, away from the ridge of high posterior density. The posterior density is formed by the accepted samples (in blue). The proposals clearly have an uncorrelated Gaussian shape, quite different from the correlated shape of the accepted parameter values. We can conclude from this graph that this algorithm learns the required proposal scales per parameter through adjusting the proposal variance, but cannot take into account the correlations between parameters in the posterior distribution. This results in inefficient sampling after convergence.

The mode of the posterior distribution represents the point of highest probability density in the posterior distribution and has been located using the MWG algorithm. However, we have shown that this sampler draws samples inefficiently once it has converged to this mode. This is why we use the second sampler, the Adaptive MCMC algorithm (Algorithm S2), which is started at the posterior mode obtained by the first sampler. The Adaptive sampler learns the parameter correlation from the history of samples drawn by the chain and therefore, after the burn-in period, makes correlated proposals that are more likely to be accepted as new parameter samples in the Metropolis-Hastings acceptance step. This leads to a chain that exhibits a lower level of autocorrelation and thus needs to be run for a shorter number of iterations to produce the equivalent number of independent samples as the initial pilot MWG algorithm.

The bottom panel of Figure 3 demonstrates the benefit of using this Adaptive sampler initiated at the posterior mode (Figure 3d). This sampler learns the posterior correlations, and this reduces the autocorrelation of the resulting samples (shown again by plotting parameter α_2 in Figure 3e). Figure 3f demonstrates the improved efficiency of sampling using an Adaptive Markov chain. This is shown by the tight overlap of the proposed parameter values (in red) with the samples that are actually accepted (in blue).

In Table 1 we compare the efficiency of the two samplers based on the Expected Sample Size (ESS). This provides a measure of sampling efficiency per iteration and per unit of computational time, as defined in the Methods. The Adaptive sampler is superior to the pilot MWG sampler both in terms of equivalent independent samples per iteration and in terms of equivalent independent samples generated per minute of computational time. This further validates the approach of using a two-step sampling technique - firstly, the MWG sampler to quickly locate the posterior mode and secondly, the Adaptive sampler to sample more efficiently from the posterior distribution once this mode has been found. Although the Adaptive sampler can be used in isolation, it takes longer to converge to the mode of the posterior. The run time (in minutes) for the two algorithms is also shown and each sampler takes approximately an hour to complete on an Intel 2.5 GHz core i7 Macbook Pro with 16GB of RAM.

We now examine the posterior samples generated by the Adaptive sampler. The individual

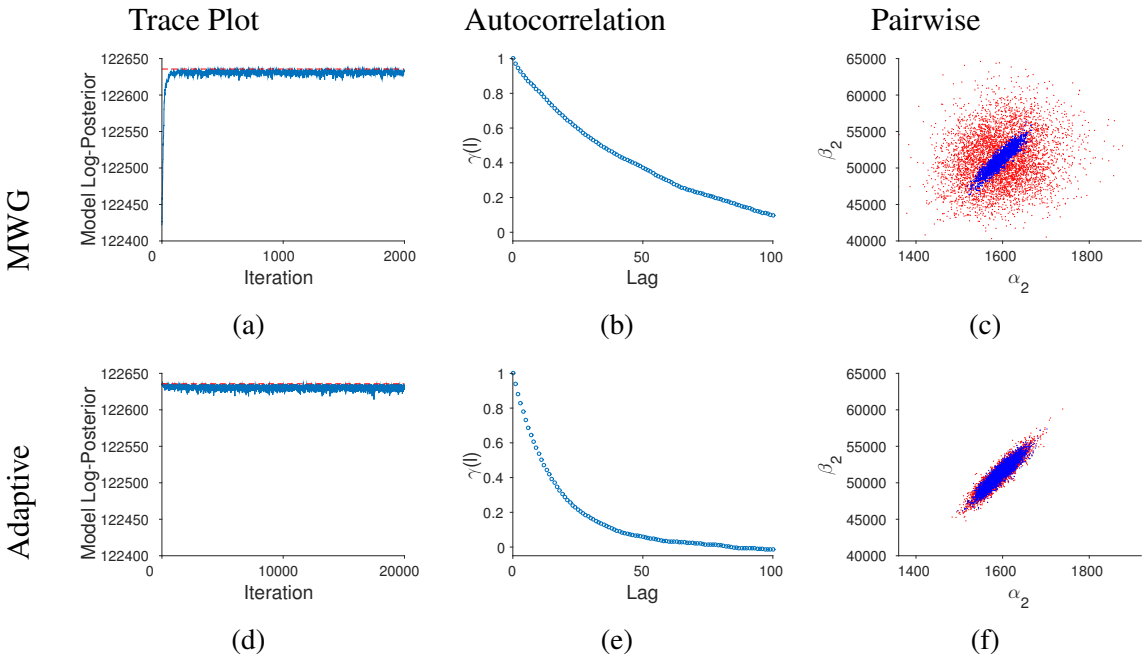


Figure 3: Chain diagnostics for MWG and Adaptive algorithms. **Top Panel**: MWG Pilot sampler. **3a**: The sampler converges quickly to the posterior mode. **3b**: The pilot MCMC chain exhibits high levels of autocorrelation in the posterior samples of parameter α_2. Autocorrelation within the α_2 sample is defined as correlation between values of α_2 at the nth iteration and the value at the $n-l$th iteration where l is the autocorrelation lag. **3c**: 5,000 samples drawn from the posterior distribution show the strong pairwise correlation between the rate constants β_2 and α_2. A similar correlation is observed using experimental repetition in [12]. The random-walk nature of the MWG sampler results in proposals (red points) away from the ridge of high posterior density, and this increases the number of rejected samples. The accepted samples are shown in blue. This is corrected by the Adaptive MCMC sampler (**Bottom Panel**). **3d**: The chain is started at the posterior mode and continues to sample from it. **3e**: The level of autocorrelation of α_2 has been reduced as the sampler has learned the posterior correlation between β_2 and α_2. **3f**: The proposals become much more efficient. This is seen in the substantial overlap between the 5,000 accepted samples (blue) and the 5,000 proposed samples (red).

	MWG	Adaptive
Number of significant lags	137	81
ESS/sample	0.01	0.03
ESS/minute	0.99	23.65
Runtime, mins	57	64

Table 1: Effective sample sizes and run times for the converged MWG pilot and Adaptive samplers for distribution of the α_2 parameter. Adaptive sampling is three times as efficient as the MWG sampler, in terms of samples per iteration (see second line). Furthermore, it is much more computationally efficient in terms of samples per minute. The number of significant lags is the number of lags for which the autocorrelation coefficient is significantly greater than zero and so these lags are used in the calculation of the ESS estimates.

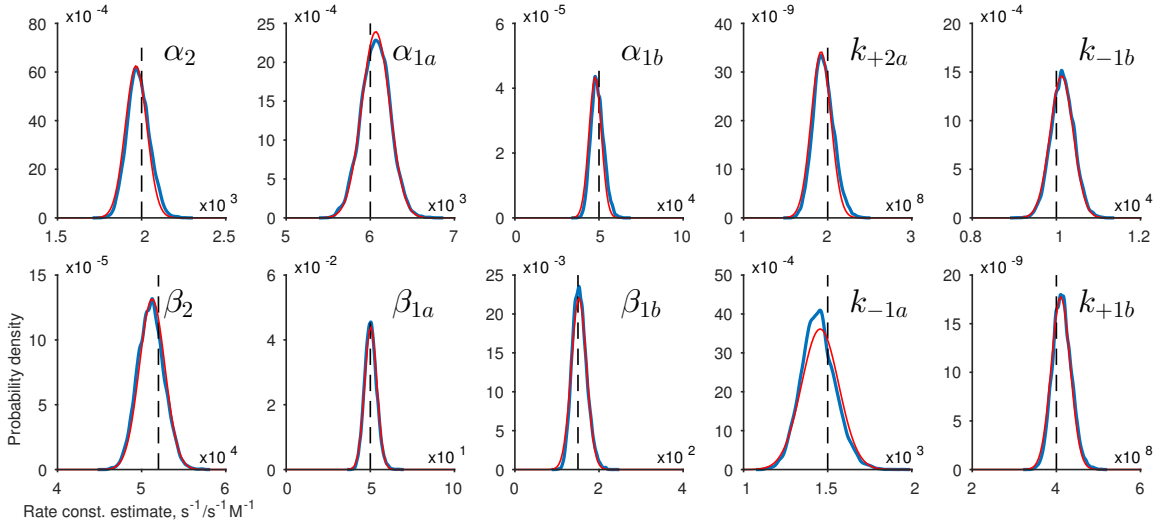


Figure 4: Marginal posterior distributions for model parameters obtained from synthetic data are shown in blue. The black vertical lines indicate the parameter values used to generate the data. The association rate constants k_{+2a} and k_{+1b} are in $s^{-1}\mathrm{M}^{-1}$ units, otherwise the units are s^{-1}. The posterior parameter distributions enclose the rate constants and show that the Bayesian approach can recover the rate parameters used to generate the data. Plotted in red are Gaussian distributions obtained from the Hessian at the mode. This shows that the posterior distributions in this example are approximated well by Gaussian distributions.

posterior distributions for the ten rate parameters, known as marginal distributions, are shown in Figure 4 (in blue). The rate parameters used to generate the synthetic data are shown by the dashed black lines. The fact that each distribution is constrained and encloses the true rate constant demonstrates that the correct rate constants originally used to generate this data can be identified and recovered as in the original ML study [12]. Next, we observe that the shape of the marginal distributions appears Gaussian. Given our choice of uniform priors (see Methods), we note that the location of the posterior mode has to be the same as the location of the ML estimates and therefore we can directly compare each marginal distribution to the Gaussian distribution that would be used to estimate asymptotic standard errors for parameter estimates in a ML estimation. We obtain estimates of the standard errors by inverting an estimate of the Hessian at the posterior mode. Comparing the shape of the error distributions (in red, Figure 4) with the obtained marginal distributions for the parameters reveals that the shape of the posterior distributions are indeed approximately Gaussian. Hence in this case, similar conclusions would be drawn about parameter identifiability and uncertainty using Bayesian inference and ML.

We can now illustrate the impact of the uncertainty in the estimated parameter posterior distribution on the uncertainty of model predictions, by taking samples from the posterior distribution and using them to simulate predictions of open and shut interval distributions and the correlation between the durations of adjacent open and shut intervals from the model. All predictions are corrected for missed events so that they can be compared directly to the observed data. 100 parameter samples were taken at random from the posterior distributions of the Adaptive sampler (shown in Figure 4). Predicted distributions were then calculated using these samples in order to assess how well they reproduce the observed data and how variable the model predictions are. These predictions are shown in Figure 5.

The variability in the predictive open time distributions, at low concentration (30 nM) and

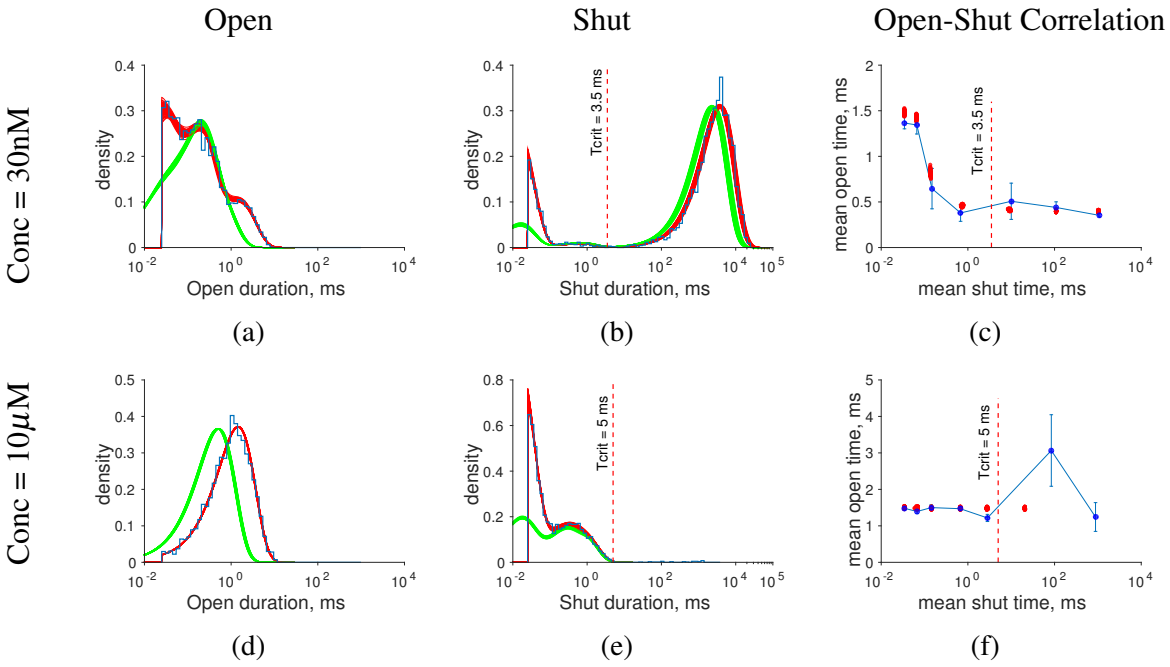


Figure 5: Predicted distributions of model behaviour at low (30nM, **Top Panel**) and high acetylcholine concentrations (10μM, **Bottom Panel**). 100 samples taken at random from the posterior distributions of Model 1a, fitted to synthetic data, are used to calculate predictive open time distributions (**5a, 5d**), shut time distributions (**5b, 5e**) and correlations between adjacent closed-to-open sojourns (**5c, 5f**). The open and shut time distributions show the predicted durations of intervals with the imposed resolution of 25μs (**red**), and with perfect resolution (**green**) and are overlaid by the observed durations summarised as a histogram (**blue**). Open-shut correlations are examined by calculating the mean of the succeeding open interval against the mean of the preceding shut interval, conditional on range of preceding shut interval durations [11]. Mean durations predicted by the model are denoted by red points. These are consistent with the conditional mean calculated from the observed experimental data (blue points with standard deviation error bars, connected with a blue line). The shut range intervals used to calculate the conditional means are [0.025-0.05, 0.05-0.1, 0.1-0.2, 0.2-2, 2-20, 20-200, 200-2000] ms. The plots in Figures **5a, 5d**, **5b and 5e** show that there is excellent agreement of the predictions from the fit with the time resolved open and shut distributions and that it becomes more precise as the agonist concentration increases. In practice, the shut distributions and correlations can be interpreted only up to the value of the t_{crit} interval (marked on the graph) as it is not known how many channels are in the patch. The model can also recover the correlations observed in the adjacent shut to open intervals up to the t_{crit} value.

high concentration (10 μM), are shown in Figures 5a, 5d. As expected from previous identifiability analysis of the channel [12], the model fit is very good across all concentrations. This Figure demonstrates that uncertainty in the model predictions decreases as the agonist concentration increases. This is apparent as the range of the curves evaluated from the parameter samples becomes narrower as the thickness of the superimposed predicted curves (red) is reduced. Note the higher variability in the prediction of short open times at low concentration. This is to be expected as low concentration records contain sparse information compared to high concentration records, since they have fewer usable shut times because of the short t_{crit}

required to separate the record into individual channel activations. They are nonetheless the predominant source of information on single-liganded openings in the model. At high concentrations, the record is rich with predominantly diliganded openings and so this aspect of the channel is very accurately predicted. Shut time distributions (Figures 5b, 5e) are predicted well across both concentrations, since although we have treated the recordings so that there may be multiple channels in the patch, in reality, we have simulated the data with only one channel. In practice, we would not be able to predict accurately the shut time distribution beyond t_{crit}. In all open and shut distributions, the impact of not taking missed events into account is illustrated by the difference between the green probability densities (which are the open or closed probability densities assuming perfect resolution calculated from Equation 2) and the red densities (which correctly assume the finite recording resolution using Equation 3). This illustrates the importance of applying the missed events correction in the open and closed probability density functions in order to compare the model predictions to what is actually observed.

One of the key considerations of any proposed ion-channel model is the ability of the postulated model to reproduce the degree of correlations between adjacent open and shut intervals observed in experimental data. A feature of nicotinic muscle channel behaviour is that at low concentrations there is a negative correlation between the length of the open interval and the length of the preceding shut interval. Predictions of such model correlations are shown alongside correlations observed in the empirical data in Figure 5c for low concentration and Figure 5f for high concentration. The model correctly recovers, with high accuracy, the negative correlation between the duration of shut intervals and the duration of the following open intervals observed at low concentrations, as does the original ML fit [12]. At high concentrations, no correlation is predicted or observed. This is because diliganded openings are predominantly observed at this concentration. This represents only one kind of opening in the channel and so correlations are absent [32].

Sampling with experimental data results in a non-Gaussian posterior distribution

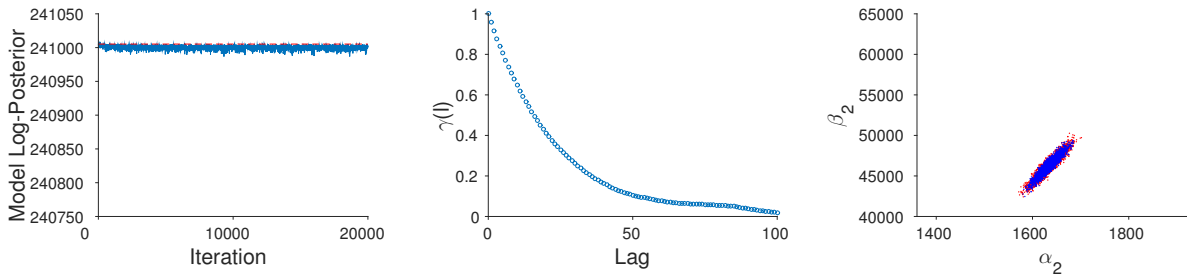
Real data from a nicotinic acetylcholine receptor [3] were used in order to test our MCMC approach. As with the simulated data, the experimental recordings were taken at three concentrations. A summary of these three datasets and their experimental conditions are described in Table 2.

Set	ACh Conc (μM)	# Resolved intervals	t_{crit} (ms)	# groups	use CHS vectors
1	0.05	14,056	2	4134	yes
2	0.1	24,230	3.5	8471	yes
3	10	13,822	35	134	no

Table 2: Experimental conditions and summary data for real acetylcholine receptor recordings from [3]. The resolution for all recordings was 25μs.The table summarises the number of resolved intervals in the experimental patch, the t_{crit} interval used to separate the resolved intervals into groups, the number of resulting individual groups (see Methods) and whether CHS vectors are required for initial openings in each group.

Similarly to the synthetic example, we begin our analysis by considering the diagnostic output from our MCMC sampler. The pilot MWG sampler successfully located the posterior mode and at this stage we consider only the output of the Adaptive sampler. Figure 6a shows

how the Adaptive chain samples starting from the posterior mode, Figure 6b shows how the Adaptive sampler draws samples efficiently from the posterior distribution of α_2 given the correlation between rate parameters α_2 and β_2 that was observed also in the synthetic example (Figure 6c). This demonstrates that our MCMC approach successfully samples from both real and synthetic data.



(a) (b) (c)

Figure 6: Chain diagnostics for the Adaptive algorithm with experimental data (cf. Figures 3d-3f). **6a**: The Adaptive chain is started at the posterior mode, as shown by the red horizontal dotted line. **6b**: The Adaptive sampler produces efficient samples from α_2 parameter as it has learned the pairwise posterior correlation between α_2 and β_2 shown in **6c**. This is shown by the tight overlap between the proposed samples (red) and the accepted samples (blue) between the α_2 and β_2 parameters.

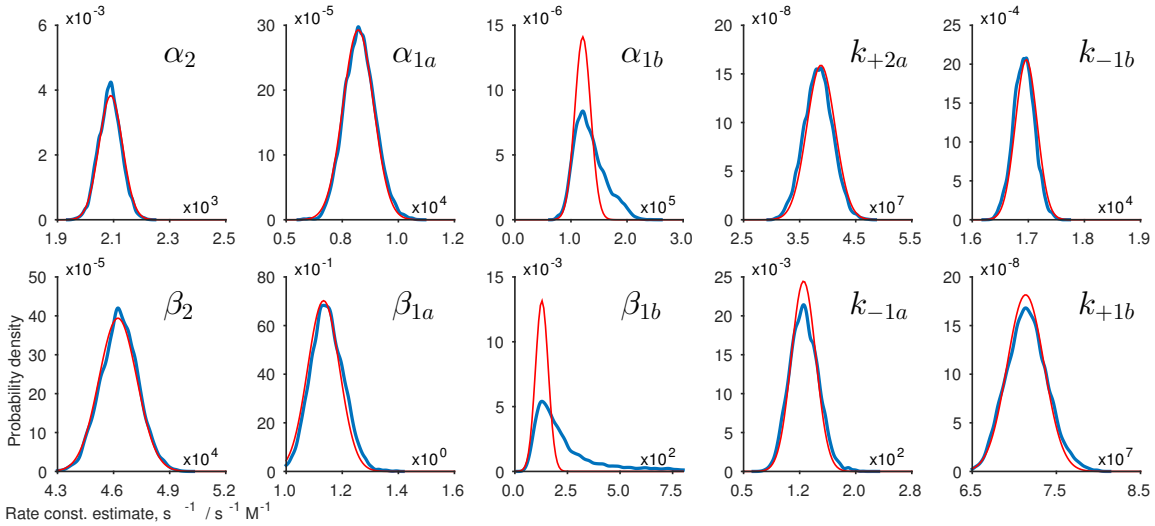


Figure 7: Marginal Posterior parameter distributions calculated using experimental data using Adaptive MCMC are shown in blue. The Gaussian approximation obtained from the Hessian at the mode are shown in red. It is clearly seen that the posterior of parameters α_{1b} and β_{1b} exhibit non-Gaussian distributions. This uncertainty is more accurately captured by the Bayesian approach. The association rate constants k_{+2a} and k_{+1b} again are in $s^{-1}\mathrm{M}^{-1}$ units, otherwise the units are s^{-1}.

Figure 7 shows the marginal posterior distribution of the rate constants. Most rate constants are close in shape to a Gaussian distribution, and hence approximated well by the Hessian derived at the mode. However, that is not so for the faster monoliganded opening and shutting

rates, those for the B binding site, β_{1b} and α_{1b}. For these rates the posterior distributions are non-symmetrical and consequently their uncertainty is not directly captured well by the Gaussian error approximation used in a typical ML approach.

We can again assess the impact of parameter uncertainty on the predicted output of the model (Figure 8). 100 samples were again taken at random from the posterior distribution in Figure 7 and used to calculate summary distributions of open, closed and conditional mean open times. Predictions of open time distributions again become more certain with increasing concentration, with the short openings at low concentration being most variably predicted (Figures 8a and 8d). Shut interval distributions at low concentration (Figure 8b) illustrate the ambiguity posed by not knowing how many channels are in the experimental patch. Although the short shut interval components are well predicted, longer shut interval durations are not, since the presence of many channels in the low concentration experimental patches makes the observed lifetime in the long shut state appear to be shorter than it actually is. In practice, it is accepted that due to this limitation, only short shut events up to t_{crit} can be inferred reliably from the experimental data.

In determining the model prediction of the mean open time conditional on the preceding mean shut time, model predictions are fairly accurate across agonist concentrations (Figures 8c and 8f). The negative correlation with the preceding shut interval duration is reasonably accounted for by the model at low concentrations, although predictions of mean intervals after t_{crit}, indicated on each chart, for the experimental patch must be taken with caution given the unknown number of channels in the patch [3].

Inferential approaches should account accurately for the limitations of raw recordings

We have shown, using our synthetic example, that a model likelihood based on a continuous-time Markov model can recover rate constants used to generate datasets from a physiologically realistic model of a muscle nicotinic receptor. The key to the success of this approach is an accurate idealisation of the noisy filtered record and the correction for missed events which arise from the filtering of the raw recordings, as described in the Methods. In practice, the incomplete detection of opening and shut intervals is a very real problem for accurately inferring model parameters from single-channel recordings. Indeed, it was estimated in a real channel record that even with good resolution as many as 88% of short shuttings may be missed [5].

An alternative practical Bayesian approach to single channel analysis has recently been described [18, 19]. It has been claimed that this method can detect model overparameterisation more robustly than a Maximum Likelihood inferential method that provides an approximate correction for missed events [33]. Broadly, the method of [19] relies on estimating a model likelihood from sampled points in the single-channel record. The model likelihood makes no probabilistic statement about the states of the channel between sampling points and it is claimed therefore that this method does not need to correct for missed events. It is therefore a natural step to compare our Bayesian method with missed events correction with the Bayesian approach of [19] using datasets which have missed events incorporated into the channel record to mimic the effect of filtering the raw trace.

The method of [19] can briefly be defined as follows. Raw sampling points are taken at time intervals δt from an experimental trace. These points are classed into open or closed observations in the trace by half thresholding each sampling point. These are the data on which inference of the model rate constants is performed. Note that experimental noise and missed events may cause the classification of the open and closed points to be wrong.

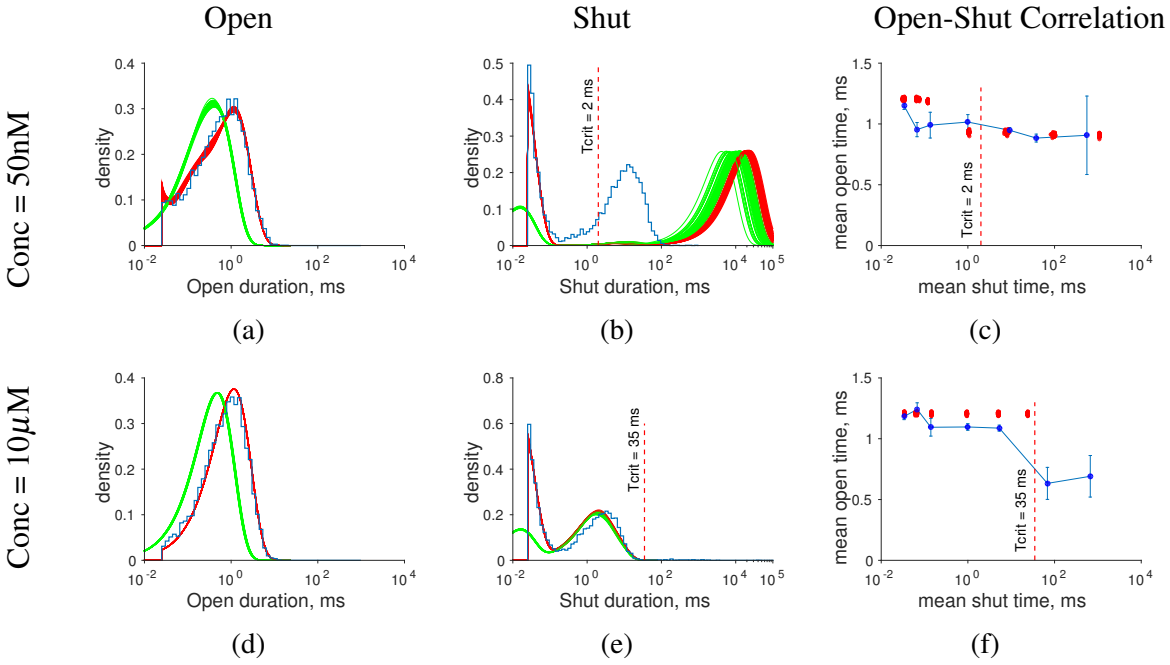


Figure 8: Predicted distributions of model behaviour at low (50nM, top row) and high Acetylcholine concentrations (10μM, bottom row). 100 samples from the posterior distributions of the model (Figure 1a), fitted with real data, are used to calculate predicted distributions, analogous to those generated in Figure 5. The predicted durations of intervals with the imposed resolution of 25μs are again shown in **red**, and with perfect resolution in **green**. The predicted distributions of the model are less precise with this real dataset than for the synthetic dataset. The open time distributions are still generally well predicted, but show variability near the resolution time. The long component of the low concentration shut intervals (Figure 8b) is poorly predicted as expected as no information is available about the numbers of channels in the patch. This results in accurate inferences about the shut time distribution being restricted to the dwell times up to the t_{crit} interval.

The rate constants in the $\mathbf{Q}$ matrix are then used to calculate a Markov transition matrix $\mathbf{T} = \exp(\mathbf{Q}\delta t)$ that denotes the individual probabilities of moving from one state at the start of the sampling interval to another at the end of the sampling interval. This uses the whole $\mathbf{Q}$ matrix (as for macroscopic currents), and considers only the state of the system at the sampling points [20]. There may be any number of transitions between the sampling points. A discrete likelihood for the entire record is then calculated using this transition matrix and projections for each sampling point that restrict the entry and exit states of the process as defined by the thresholding. The likelihood is calculated using a forward algorithm commonly used to estimate likelihoods in discrete HMM models. This likelihood, in combination with a weakly informative prior distribution for the rate constants forms the posterior distribution which is sampled using an MCMC algorithm. [19].

We reproduced the synthetic experiment from [19] in conditions of perfect resolution and after imposing progressively worse resolution in order to emulate the process by which filtering raw experimental traces results in missed events in the outputted signal. We did this to test empirically whether the analysis method from [19] and from the present paper are equivalent in retrieving the rate constants of the model.

We have used the 4-state ion-channel model and rates described in [19] (Figure 9), to sim-

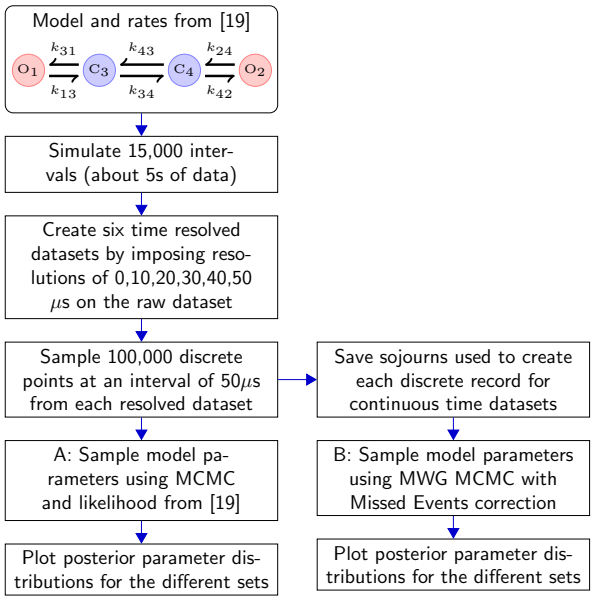


Figure 9: Experimental workflow in order to examine the requirement for missed events correction. Raw data was initially simulated from the 4-state ion-channel model and rates from [19]. $k_{13} = 3500$, $k_{31} = 7000$, $k_{34} = 400$, $k_{43} = 500$, $k_{42} = 100$, $k_{24} = 50$. Rates in s^{-1}.

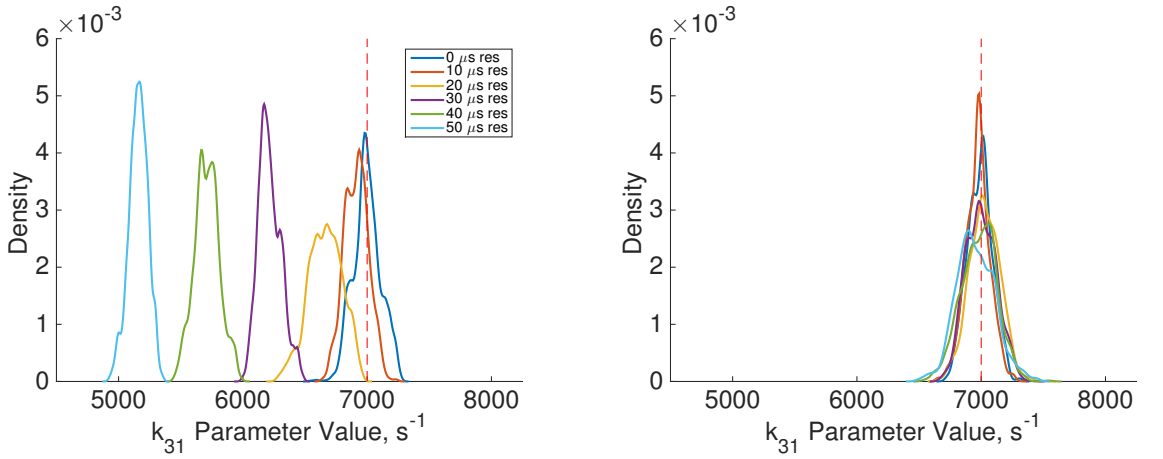
ulate 15,000 sojourns assuming unrealistically that there is only one channel in the patch. This perfect resolution dataset, R, was sampled at 50μs to produce 100,000 points, as in [19]. The original sojourn intervals were also stored for comparison with the likelihood and sampler described in the Methods.

This ideal, perfect resolution record R, was then subjected to increasing coarsening of time resolution. Separate datasets of continuous records were generated from R with time resolutions of 10 μs, 20 μs, 30 μs, 40 μs and 50 μs. This concatenates shorter intervals than the resolution with the adjacent resolved intervals. This distorts the channel record to an increasing degree as the resolution time worsens and mimics the act of filtering in real experimental traces. Next, for each dataset, the data was again sampled at 50μs to produce 100,000 discrete data points. Posterior parameter samples were generated using these datasets with the discrete likelihood MCMC algorithm and codebase (`https://github.com/merlinthemagician/ahmm/`) of [19]. The continuous time sojourns for each dataset were then analysed separately using the initial MWG pilot MCMC algorithm described in the Methods. Thus the performance of the two methods can be compared over with the same amount of information. The prior for the rate constants in both sets of experiments was the same as [19]. Posterior distributions were estimated after both MCMC samplers had converged.

The posterior distributions from the Siekmann likelihood [19] and that from our method, which uses the likelihood of [11], were examined at each resolution in order to establish whether they could recover the rate constant values that generated the initial dataset, R. The comparison for the faster rate k_{31} is shown in Figure 10a. It is clear that estimates obtained with the method of [19] become increasingly biased as the resolution worsens. Even with an optimistic resolution of 20μs , the posterior distribution of rate constant k_{31} is biased away from the correct value (dashed red line). In contrast, the correct rate constant can still be recovered using our analysis (Figure 10b) even at the worst resolution (50 μs). It is interesting to note that even in these limited MCMC runs the inference over the rate constant k_{31} in Figure 10b becomes less precise. This is seen by a broadening of the posterior and would be expected

despite the missed events correction, as more information is removed from the record.

It is worth noting also that a rate constant value of 7000 s^{-1} is relatively slow for an ion channel. At a resolution of 50μs, the fraction of events that can be expected to be missed is this example is only 29% of openings and 16% of shuttings. It is to be anticipated that this problem will worsen with faster channels with their quicker rate constants.



(a) Posterior distributions of k_{31} obtained with worsening time resolution using the likelihood of [19]

(b) Posterior distributions of k_{31} are identified using the likelihood of [11] and the pilot MCMC sampler.

Figure 10: A simulation based on an example from [19] highlights the requirement for missed events correction. **10a**: Posterior distributions of rate constant k_{31} using the likelihood of [19] sampled from data with worsening resolution times imposed (0μs - 50μs). These distributions become more biased as the resolution worsens. **10b**: The same analysis with our likelihood [11] and Pilot MWG MCMC sampler. This true parameter value is shown as the red dashed line on both charts.

Discussion

In the context of modern single molecule biophysics, the modelling of single ion-channels using Markov processes has a comparatively long history. Inferential techniques for rate parameters have progressed from the fit of exponential components to dwell-time distributions [29] to maximum likelihood approaches that use the sequence of the observed signal [11, 33]. Current "full likelihood" approaches broadly consist of two alternative methods. The first involves fitting a continuous time Markov process to a signal idealised from the raw, filtered data typically by the use of threshold crossing or by time course fitting idealisation [2]. Filtering of the original signal is needed to enable event detection, but events of short duration are absent in the filtered record and subsequent idealisation. This requires an explicit correction for missed events, as the model likelihood makes probabilistic statements regarding the state of the process through time (Equation 1). The most widely used modelling approaches employ maximum likelihood estimation with a continuous time model with either an approximate [33] or an exact correction [11, 21, 22] in the likelihood for missed events.

A second approach using discrete Hidden Markov Models has been applied to extract information directly from the raw data, without the idealisation step [34]. HMMs often require the estimation of the distribution over hidden states at each sampling point during the likelihood

calculation. Calculating the likelihood directly from raw recordings in practice can require implementing higher order Markov models, to account for correlated experimental noise [35, 36], and this increases the computational complexity of calculating the model likelihood. More recent methods avoid this issue by classifying points as either open or closed [19]. Maximum likelihood inference of rate constants in HMM models is often achieved by implementing bespoke Forward-Backward algorithms combined with Baum's re-estimation [37] or direct optimisation [38]. We note that likelihoods defined in this manner make no probabilistic statement about channel activity between sampling points and it has been claimed that no missed events correction is necessary [18]. However, we have demonstrated that even for a simple ion channel model with four hidden states, HMM approaches that do not account for the unavoidable consequences of filtering and experimental noise, for example by fixed rather than probabilistic classifications of sampling points into conductance classes, result in severely biased inferences when using realistically sampled data.

We must seek to evaluate the quality of our estimates whenever we infer parameter values from data. This should involve establishing as realistically as possible that they are unbiased and measuring the extent of our uncertainty in their values. For single channel modelling, this has been achieved in examples where ML point estimates were used by using simulation [5, 12] although this simulation can be a laborious approach. The use of Bayesian inference, where by definition we obtain the posterior distribution rather than point estimates of the parameters, is beneficial in making these checks efficient and systematic. Bayesian approaches to date have either used MCMC sampling in combination with either the continuous time [39, 40] or discrete likelihoods [16, 18, 19] in order to compute posterior distributions. We have demonstrated in this paper that the correction for missed events is still vitally important in order to recover the correct rate constants in even small model examples. We therefore investigated MCMC sampling approaches that use a model likelihood that both corrects exactly for missed events and takes into account the unknown numbers of channels in each experimental patch [11].

We have proposed a two-stage sampling approach for performing Bayesian inference in these computationally expensive models, and made the package "BICME" available to download as MATLAB code. The first step relies on a MWG pilot algorithm to efficiently locate the posterior mode. This is followed by an Adaptive MCMC sampler, which learns the covariance structure of the joint posterior distribution in order to sample the posterior distributions more efficiently. With these tools we performed our Bayesian analysis on a physiologically realistic ion-channel model for a muscle nicotinic acetylcholine receptor. This model's identifiability has previously been evaluated using Maximum Likelihood estimation [12]. The sampling approach was initially evaluated using synthetic data and demonstrates how parameter identifiability can be established and how uncertainty in model parameterisation can help examine directly the uncertainty in model predictions. A subsequent application using experimental data confirms the proposed approach can implement Bayesian inference in these models by efficiently sampling from the resulting posterior distribution. We show that, with real data, some posterior distributions are non-Gaussian in shape.

As in a typical Bayesian analysis, we report the full posterior distributions of the rate parameters rather than point estimates. We do note, however, that for our choice of prior distribution, the point estimates of rate constants as defined by the posterior mode would be the same as those that would be found by ML estimation. In the example with synthetic data, the posterior distributions are close to Gaussian and their shape well approximated by a covariance matrix calculated at the ML estimate. This is not the case with experimental data. In this instance the Bayesian approach has the advantage in cases where the posterior distributions are non-symmetrical, in which case the posterior distribution gives a more realistic estimate of the

uncertainty in the parameter estimates than error estimates derived from ML estimation.

Our approach was specifically designed to sample effectively for higher dimensional ion-channel models and will form the basis for future work examining candidate models with larger numbers of rate parameters. In such models it has been shown that robust model parameterisation and comparison remains statistically challenging [7]. In the latter case, this sampling method will help measure uncertainty in the competing mechanistic models through the estimation of marginal likelihoods, which will form the basis of future research.

Author contributions

ME performed the research and contributed analytic tools. ME, BC, LGS and MAG designed the research. ME, BC and LGS wrote the manuscript.

Acknowledgments

The authors would like to thank Dr Remigijus Lape for extremely helpful advice and extensive discussions and providing experimental data for the muscle nicotinic acetylcholine receptor from the original study of [3]. They would also like to thank Prof. David Colquhoun for helpful comments on the draft.

ME was funded by the CoMPLEx Doctoral Training Programme (2010-2014) at University College London and was funded by an EPSRC Doctoral Prize Fellowship (2014-2015) at Imperial College.

References

[1] Hamill, O., A. Marty, E. Neher, B. Sakmann and F. Sigworth. Improved patch-clamp techniques for high-resolution current recording from cells and cell-free membrane patches. *Pflügers Archiv*, 391(2):85–100, 1981.

[2] Colquhoun, D. and F. Sigworth. Fitting and statistical analysis of single-channel records. In *Single-channel recording*, pages 483–587. Springer, 1995.

[3] Hatton, C., C. Shelley, M. Brydson, D. Beeson and D. Colquhoun. Properties of the human muscle nicotinic receptor, and of the slow-channel myasthenic syndrome mutant εL221F, inferred from maximum likelihood fits. *J. Physiol.*, 547(3):729–760, 2003.

[4] Mukhtasimova, N., W. Lee, H. Wang and S. Sine. Detection and trapping of intermediate states priming nicotinic receptor channel opening. *Nature*, 459(7245):451–454, 2009.

[5] Burzomato, V., M. Beato, P. Groot Kormelink, D. Colquhoun and L. Sivilotti. Single-channel behavior of heteromeric $\alpha 1\beta$ glycine receptors: An attempt to detect a conformational change before the channel opens. *J. Neurosci.*, 24(48):10924–10940, 2004.

[6] Lape, R., D. Colquhoun and L. Sivilotti. On the nature of partial agonism in the nicotinic receptor superfamily. *Nature*, 454(7205):722–727, 2008.

[7] Lape, R., A. Plested, M. Moroni, D. Colquhoun and L. Sivilotti. The α1K276E startle disease mutation reveals multiple intermediate states in the gating of glycine receptors. *J. Neurosci.*, 32(4):1336–1352, 2012.

[8] Milescu, L., G. Akk and F. Sachs. Maximum likelihood estimation of ion channel kinetics from macroscopic currents. *Biophys. J.*, 88(4):2494–2515, 2005.

[9] Fredkin, D., M. Montal and J. Rice. Identification of aggregated Markovian models: application to the nicotinic acetylcholine receptor. In *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer*, volume 1, pages 269–289. 1985.

[10] Bruno, W., J. Yang and J. Pearson. Using independent open-to-closed transitions to simplify aggregated Markov models of ion channel gating kinetics. *Proc. Natl. Acad. Sci. U. S. A.*, 102(18):6326–6331, 2005.

[11] Colquhoun, D., A. Hawkes and K. Srodzinski. Joint distributions of apparent open and shut times of single-ion channels and maximum likelihood fitting of mechanisms. *Phil, Trans. R. Soc, Lond. A*, 354(1718):2555–2590, 1996.

[12] Colquhoun, D., C. Hatton and A. Hawkes. The quality of maximum likelihood estimates of ion channel rate constants. *J. Physiol.*, 547(3):699–728, 2003.

[13] Marabelli, A., R. Lape and L. Sivilotti. Mechanism of activation of the prokaryotic channel ELIC by propylamine: A single-channel study. *J. Gen. Physiol.*, 145(1):23–45, 2015.

[14] Hines, K., T. Middendorf and R. Aldrich. Determination of parameter identifiability in nonlinear biophysical models: A Bayesian approach. *J. Gen. Physiol.*, 143(3):401–416, 2014.

[15] Hines, K. A Primer on Bayesian inference for biophysical systems. *Biophys. J.*, 108(9):2103–2113, 2015.

[16] Rosales, R. MCMC for hidden Markov models incorporating aggregation of states and filtering. *Bull. Math. Biol*, 66(5):1173–1199, 2004.

[17] Calderhead, B., M. Epstein, L. Sivilotti and M. Girolami. Bayesian approaches for mechanistic ion channel modeling. In *In Silico Systems Biology*, pages 247–272. Springer, 2013.

[18] Siekmann, I., L. Wagner, D. Yule, C. Fox, D. Bryant, E. Crampin and J. Sneyd. MCMC estimation of Markov models for ion channels. *Biophys. J.*, 100(8):1919–1929, 2011.

[19] Siekmann, I., J. Sneyd and E. Crampin. MCMC can detect nonidentifiable models. *Biophys. J.*, 103(11):2275–2286, 2012.

[20] Colquhoun, D. and A. Hawkes. On the stochastic properties of bursts of single ion channel openings and of clusters of bursts. *Phil, Trans. R. Soc, Lond. B*, 300(1098):1–59, 1982.

[21] Hawkes, A., A. Jalali and D. Colquhoun. The distributions of the apparent open times and shut times in a single channel record when brief events cannot be detected. *Phil, Trans. R. Soc, Lond. A*, 332(1627):511–538, 1990.

[22] Hawkes, A., A. Jalali and D. Colquhoun. Asymptotic distributions of apparent open times and shut times in a single channel record allowing for the omission of brief events. *Phil, Trans. R. Soc, Lond. B*, 337(1282):383–404, 1992.

[23] Metropolis, N., A. Rosenbluth, M. Rosenbluth, A. Teller and E. Teller. Equation of state calculations by fast computing machines. *J. Chem. Phys.*, 21(6):1087–1092, 1953.

[24] Sherlock, C., P. Fearnhead and G. Roberts. The random walk metropolis: Linking theory and practice through a case study. *Statist. Sci.*, pages 172–190, 2010.

[25] Geyer, C. Practical Markov chain Monte Carlo. *Statist. Sci.*, pages 473–483, 1992.

[26] Roberts, G. and J. Rosenthal. Examples of adaptive MCMC. *J. Comp. Graph. Stat.*, 18(2):349–367, 2009.

[27] Ball, F., S. Davies and M. Sansom. Single-channel data and missed events: Analysis of a two-state markov model. *Proc. R. Soc. B*, 242(1303):61–67, 1990.

[28] Haario, H., E. Saksman and J. Tamminen. An adaptive Metropolis algorithm. *Bernoulli*, pages 223–242, 2001.

[29] Colquhoun, D. and B. Sakmann. Fast events in single-channel currents activated by acetylcholine and its analogues at the frog muscle end-plate. *J. Physiol.*, 369:501, 1985.

[30] Milone, M., H. Wang, K. Ohno, T. Fukudome, J. Pruitt, N. Bren, S. Sine and A. Engel. Slow-channel Myasthenic Syndrome caused by enhanced activation, desensitization, and agonist binding affinity attributable to mutation in the M2 domain of the acetylcholine receptor α subunit. *J. Neurosci.*, 17(15):5651–5665, 1997.

[31] Colquhoun, D., K. Dowsland, M. Beato and A. Plested. How to impose microscopic reversibility in complex reaction mechanisms. *Biophys. J.*, 86(6):3510–3518, 2004.

[32] Colquhoun, D. and A. Hawkes. A note on correlations in single ion channel records. *Proc. R. Soc. B*, 230(1258):15–52, 1987.

[33] Qin, F., A. Auerbach and F. Sachs. Maximum likelihood estimation of aggregated Markov processes. *Proc. R. Soc. B*, 264(1380):375–383, 1997.

[34] Chung, S., J. Moore, L. Xia, L. Premkumar and P. Gage. Characterization of single channel currents using digital signal processing techniques based on hidden markov models. *Phil, Trans. R. Soc, Lond. B*, 329(1254):265–285, 1990.

[35] Venkataramanan, L., J. Walsh, R. Kuc and F. Sigworth. Identification of hidden Markov models for ion channel currents. I. colored background noise. *IEEE Trans. Signal Process*, 46(7):1901–1915, 1998.

[36] Qin, F., A. Auerbach and F. Sachs. Hidden Markov modeling for single channel kinetics with filtering and correlated noise. *Biophys. J.*, 79(4):1928–1944, 2000.

[37] Venkataramanan, L. and F. Sigworth. Applying hidden Markov models to the analysis of single ion channel activity. *Biophys. J.*, 82(4):1930–1942, 2002.

[38] Qin, F., A. Auerbach and F. Sachs. A direct optimization approach to hidden Markov modeling for single channel kinetics. *Biophys. J.*, 79(4):1915–1927, 2000.

[39] Ball, F., Y. Cai, J. Kadane and A. O'Hagan. Bayesian inference for ion-channel gating mechanisms directly from single-channel recordings, using Markov chain Monte Carlo. In *Proc. R. Soc. A*, volume 455, pages 2879–2932. The Royal Society, 1999.

[40] Gin, E., M. Falcke, L. E. Wagner, D. I. Yule and J. Sneyd. Markov chain Monte Carlo fitting of single-channel data from inositol trisphosphate receptors. *J. Theor. Biol.*, 257(3):460–474, 2009.

Supplementary Materials

S1 Bayesian Overview

Bayesian approaches in Systems Biology are increasingly popular for characterising the inherent uncertainty observed in biological systems. Such a formalism provides a mathematically consistent framework for reasoning with and propagating uncertainty, from model parameterisation through to model prediction.

Bayesian reasoning may be implemented through the specification of a probabilistic model assumed to be the generative process of the data and the specification of prior probability distributions that incorporate existing knowledge regarding the model parameters which are to be estimated are unknown. In the Bayesian framework we model the parameters and the uncertainty regarding their true, fixed values as random variables. The aim is to update our knowledge of the model parameters by gathering additional experimental data and applying Bayes' theorem, which combines the prior and the model likelihood to produce a posterior distribution over the parameter values. These distributions express our updated knowledge of the model's parameters after taking into account the experimental data. Our use of Uniform prior distributions results in point estimates being derived from estimating the mode of the posterior being the same values as the point estimates derived from an ML analysis.

Mathematically, Bayes' rule is simply the application of the rules of joint and conditional probability, which may be stated as,

$$p(\theta \mid y) = \frac{p(y \mid \theta)p(\theta)}{p(y)} = \frac{p(y \mid \theta)p(\theta)}{\int p(y \mid \theta)p(\theta)d\theta} \tag{1}$$

where θ is the vector of unknown parameters, and y is the observed data. The main benefit of the Bayesian approach is that uncertainty about parameter estimates is directly encoded within the well defined posterior probability distribution obtained after examining the experimental data, rather than estimating a single "true" parameter value with accompanying assumptions about estimation errors. Posterior distributions can then be sampled to propagate parameter uncertainty through to the predictions that the model makes about observable quantities. This can be achieved by sampling parameter values from the posterior distribution and calculating model predictions from these samples to see how they vary. A specified model can be judged not only through its predictive power of the "best" set of model parameters but also through the certainty of its predictions given the uncertainty of the parameter values. The posterior distribution represents the uncertainty in our

knowledge regarding the true values of the rate constants.

S2 Sampling from the Posterior Distribution of Ion-channel models

Computing posterior probability distributions, as opposed to maximum likelihood estimation, often faces significant statistical challenges as the posterior distributions of interest are often known pointwise up to a normalising constant. Many challenges in Bayesian statistics, for example the estimation of the normalising constant for model evidence, arise from the need to calculate expectations with respect to probability distributions whose analytic form may not be known.

Consider an arbitrary probability distribution $g(\theta)$ from which we would want to estimate an expectation of the form $\mathbb{E}_{g(\theta)}(f(\theta)) = \int f(x)g(x)dx$. If we do not know the analytical distribution of $g(\theta)$ but are able to draw realisations from $g(\theta)$, we can calculate the function $f(\theta)$ and hence the integral can be numerically estimated. In such instances we can use the weak law of large numbers, that is, for any ϵ,

$$P(\mid \bar{f}_n - \mu_f \mid < \epsilon) \to 1 \quad as \quad n \to \infty \tag{2}$$

where μ_f is the true expectation of $f(\theta)$ and $\bar{f}_n$ is the estimator for the expectation after the nth sample. An ideal estimator for $\bar{f}_n$ is the Monte Carlo integrator $\bar{f}_n = \frac{1}{n}\sum_{i=1}^{n} f(\theta_i)$. This requires the ability to draw independent and identically distributed samples from $g(\theta)$. In practice, it is very difficult to draw such samples but the Monte Carlo estimator remains the ideal estimator as the rate of convergence of the estimator scales favourably with the dimensionality of the sampled probability distribution.

Fortunately, Markov chain Monte Carlo methods are an invaluable tool that can be used to perform sampling in this context. These methods employ a probabilistic process whose stationary distribution is our target probability distribution and hence allows us to draw samples from it. The most widely known algorithm for such a process is the Metropolis-Hastings algorithm [1, 2].

MCMC methods typically require significantly more computational power than optimisation methods for Maximum Likelihood inference. This means that the efficient design of these sampling algorithms is of crucial importance particularly as the dimension of models increase. Although the converged Markov chain defined by the Metropolis-Hastings algorithm will draw samples from the exact target distribution, they will not be uncorrelated as the next proposed point depends on the current position for the chain. The pres-

ence of autocorrelation within the samples drawn increases the variance of the Monte Carlo estimator. We can assess the impact of autocorrelation on our sampling by considering the Effective Sample Size (ESS) of our samples. This is the equivalent number of independent samples given the autocorrelation present in the samples. The ESS for a given parameter θ is given by:

$$ESS(\theta) = \frac{N}{(1 + 2\sum_{l=1}^{L} \gamma(l))} \tag{3}$$

where $\sum_{l=1}^{L} \gamma(l)$ is the sum of the L significant autocorrelation lag coefficients of the converged chain and N is the number of samples drawn. A lag is determined to be significant if its coefficient is different from zero. The total number of significant lags can be established by visually examining the autocorrelation plot or with by calculating a confidence interval for the autocorrelation sequence of a white noise process with which to compare with the empirical coefficients.

The auxiliary proposal distribution $Q(. \mid x)$ is often manipulated to speed up the convergence of the chain and reduce the level of autocorrelation in collected samples. Such schemes often use information that is available at the current point in the parameter space. These schemes often integrate gradient information about the log-target density e.g. HMC[3] or MALA [4] or utilise Riemannian geometry [5] in order to make more informed proposals. These schemes work well when analytic derivatives of the log-likelihood are available to use, or at least can be approximated in a computationally timely fashion. Unfortunately the missed events correction in Equation 3 (main text) does not have analytical expressions for the derivatives and the computational expense of evaluating derivative information numerically outweighs the gain in sampling efficiency for these models.

S3 Algorithms

Algorithm S1 Multiplicative Metropolis-within-Gibbs algorithm [6] with scaling during burn-in

$i = 0$
$burnin = N/2$ ▷ Set number of burn-in samples
$AC(1:K) = 0$ ▷ Number of acceptances over each adjust period
$adjust(1:K) = X$ ▷ Consider adjusting SF every X samples during burn-in
$SF(1:K) = 1$ ▷ Initialise scale factor for proposal covariance
$\theta_i = \theta$
$\Sigma = \mathbf{I}$
while $i < N$ **do**
  for each k in K **do**
    $q = \mathcal{N}(\theta(k)_i, SF(k) * \Sigma(k,k))$
    $Y \sim q(\theta \mid \theta(k)_i)$
    $\theta(k)^* = \theta(k)e^Y$
    $\alpha = min\left(1, \frac{p(\theta(k)^*)q(\theta(k)_i|\theta(k)^*)[\prod_{j\neq k}^{K}\theta(j)]\theta(k)^*}{p(\theta(k))q(\theta(k)^*|\theta(k)_i)\prod_{j}^{K}(\theta(j)_i)}\right)$
    if $u \sim U[0,1] < \alpha$ **then**
      $\theta(k)_{i+1} = \theta(k)^*$
      $AC = AC + 1$
    else
      $\theta(k)_{i+1} = \theta(k)_i$
    end if
    if $mod(i, adjust(k)) == 0$ and $i <= burnin$ **then**
      if $AC/adjust(k) < 0.1$ **then** ▷ reduce the step size
        $SF(k) = SF(k) * 0.9$
      else if $AC/adjust(k) > 0.5$ **then** ▷ increase the step size
        $SF(k) = SF(k) * 1.1$
      end if
      $AC(k) = 0$
    end if
  end for
  $i = i + 1$
end while

In Algorithm S1, an iteration of the MCMC sampler proceeds as follows. At the ith step for parameter k, the proposal distribution q is constructed as a univariate Gaussian distribution with mean $\theta(k)$, the current parameter

value, and standard deviation $\Sigma(k,k)$ where $\Sigma(k,k)$ is the diagonal element of the covariance matrix Σ scaled by the individual scaling factor for parameter k.

A parameter value for $\theta(k)^*$ is then proposed by exponentiating a sample drawn from the proposal distribution and multiplying it by the current parameter value $\theta(k)_i$. This proposal is accepted with probability α according to the Metropolis-Hastings acceptance ratio which has been adjusted to account for the fact that proposals are being made in log-space. If the move is accepted, the value of the $i+1$th value of parameter k is set to the proposed parameter $\theta(k)*$ and the acceptance counter for parameter k, $AC(k)$ is iterated. If the move is not accepted, the value of the kth parameter set to the current value. At this point, if we are still in the burnin phase and the current iteration is a multiple of the adjustment period (denotes by the modulus of i and $adjust(k)$), then a decision is made as to whether to adjust the scale factor for the proposal distribution for parameter k. Briefly, if the proportion of acceptances in the last X samples is less than 0.1 then the scale factor for parameter k is reduced and if it is greater than 0.5 the scale factor is lengthened.

Regardless of whether an adjustment is made, the acceptance counter is reset to 0 for the next set of X samples. If we are not in the burn-in phase or the current iteration is not a multiple of the adjustment period then the algorithm updates to the $i+1$th iteration and the next sample is drawn. The process continues until N iterations have been performed.

Algorithm S2 Adaptive algorithm of [7]

$i = 0$
$\theta = \theta_i$
while $i < N$ **do**
  if $i <= 2K$ **then**
    $q = N(\theta, \frac{(0.1^2)}{K}\mathbf{I})$ ▷ where K is number of params
  else
    $q = (1-\beta)N(\theta, \frac{(2.38^2)}{K}\hat{\mathbf{\Sigma}} + \beta N(\theta, \frac{(0.1^2)}{K}\mathbf{I})$ ▷ where β is the mixture parameter
  end if
  $\theta^* \sim q(\theta \mid \theta_i)$
  $\alpha = min\left(1, \frac{p(\theta^*)q(\theta_i|\theta^*)}{p(\theta)q(\theta^*|\theta_i)}\right)$
  if $u \sim U[0,1] < \alpha$ **then**
    $\theta_{i+1} = \theta^*$
  else
    $\theta_{i+1} = \theta_i$
  end if
  $\hat{\mathbf{\Sigma}} = cov(\theta(.))$ ▷ Update the covariance matrix with all the samples
  $i = i + 1$
end while

In Algorithm S2, an iteration of the MCMC algorithm proceeds as follows. If the current iteration i is less that $2K$ where K is the number of parameters, then the proposal distribution q is set to a multivariate Gaussian distribution with mean vector of θ, the current set of parameter values, and a scaled identity matrix for the covariance matrix. This is to ensure that the process proposes moves with a small step size at the start, in order to build up the initial estimation of the sample covariance. If the current iteration is greater than $2K$ then the proposal distribution is made up of a multivariate Gaussian mixture distribution composed of an estimate of the current covariance distribution $\hat{\mathbf{\Sigma}}$ and an uncorrelated Gaussian. The fraction of each is decided by the β parameter which is fixed at $\beta = 0.05$ as per [7]. A vector of parameters θ is then proposed from this distribution, and the vector is accepted according to the probability derived from the Metropolis-Hastings ratio. The final step of the iteration is to update the sample covariance $\hat{\mathbf{\Sigma}}$ with all of the samples derived so far. The process terminates after N samples have been drawn. Similarly to Algorithm S1, a burn-in phase was also incorporated where a global scaling factor is used to increase or decrease the proposal step size according to the global parameter acceptance rate. This has been omitted from Algorithm S2 for clarity.

References

[1] Metropolis, N., A. Rosenbluth, M. Rosenbluth, A. Teller and E. Teller. Equation of state calculations by fast computing machines. *J. Chem. Phys.*, 21(6):1087–1092, 1953.

[2] Hastings, W. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97–109, 1970.

[3] Neal, R. MCMC using Hamiltonian dynamics. In *Handbook of Markov Chain Monte Carlo*, pages 113–162. CRC Press, 2011.

[4] Roberts, G. and R. Tweedie. Exponential convergence of Langevin distributions and their discrete approximations. *Bernoulli*, pages 341–363, 1996.

[5] Girolami, M. and B. Calderhead. Riemann manifold Langevin and Hamiltonian Monte Carlo methods. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 73(2):123–214, 2011.

[6] Sherlock, C., P. Fearnhead and G. Roberts. The random walk metropolis: Linking theory and practice through a case study. *Statist. Sci.*, pages 172–190, 2010.

[7] Roberts, G. and J. Rosenthal. Examples of adaptive MCMC. *J. Comp. Graph. Stat.*, 18(2):349–367, 2009.