Uncertainty characterisation in action potential modelling for cardiac drug safety

Ross H. Johnstone

Hertford College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Hilary 2018

Acknowledgements

Personal

First of all, I must (consciously, not contractually) thank my supervisors Gary, Dave, Rémi, Mark, and Liudmila for all of their support and help throughout my DPhil. In particular, thanks to Gary for the daily encouragement and support, for fixing Chaste when I broke it, and for documenting the transformation of a Northerner into a pseudo-Southerner. Thanks to Dave for the limitless optimism and enthusiasm, and for the 2k advice. Thanks to Rémi for knowing seemingly all of statistics and teaching the locally-relevant parts to me. Thanks to Liudmila and Mark for the Roche-side support, and helping me to visit Basel to perform experiments.

I would also like to thank Teun de Boer and Birgit Goversen at UMC Utrecht for teaching me how to perform patch-clamp experiments. Thanks also to Fabian Häusermann at Roche for teaching me about the Roche lab setup, and for helping me to improve (or perfect?) my patch-clamp technique. I should, and therefore will, also thank Jonathan Cooper for fixing Chaste when I broke it such that even Gary couldn't fix it, and James Grogan for helping me with PyChaste.

Thanks to my parents for continued support and encouragement, and other family members too, I suppose. A necessary but not sufficient list of other thankworthy people: Adam (CRB), Ali, Anna, Anna, Eleanor, Eloi, Joe (which one?), Катерина, Kevin, Kylie, Mustak, Robert, not Neesh.

Institutional

Thanks to the support provided by SABS-IDC, and the funding provided by the EPSRC with contributions from Roche.

Abstract

Drugs can interact with cardiac cells to produce dangerous effects on the heart's natural rhythm. Pharmaceutical companies and regulators are interested in predicting these unintended side effects during the drug development and safety testing process. One clinical trial is the *Thorough QT* (TQT) study, where a drug is administered to healthy patients and particular changes in their cardiac electrical behaviour are monitored. Clinical trials such as the TQT study are expensive and time-consuming. Following a workshop in July, 2013 at the US Food and Drug Administration (FDA), the *Comprehensive in-vitro Proarrhythmia Assay* (CiPA) initiative was proposed to improve safety testing, led by a number of drug safety and regulatory bodies around the world. The CiPA initiative aims to replace the TQT study with a combination of *in-vitro* ion channel screening and mathematical mechanistic models of single-cell cardiac electrophysiology to predict a drug's proarrhythmic risk. Typically, mathematical models of ion channel screening and of single-cell electrophysiology are published with single best-fit parameter values. These point estimates do not account for any uncertainty in the models, which can come from many different sources. Therefore, predictions from these models will provide single point estimates. While such a prediction may suggest that the 'most likely' outcome given the available data is safe, it will not describe the other likely outcomes which may be dangerous. We adopt a Bayesian statistical framework to characterise some of the uncertainty in both ion channel screening data and in action potential model parameters. We then propagate this uncertainty into predictions of how a drug will affect a cardiac cell's electrical behaviour. We obtain a probability distribution describing possible effects of a particular concentration of a drug, instead of a point estimate. Such probability distributions contain more information than single point estimates: for example, the probability of crossing some dangerous threshold, which may be omitted from the point estimate if the most likely output is evaluated as safe. This process of uncertainty characterisation and propagation allows for more-informed decision-making when assessing a pharmaceutical compound's potential proarrhythmic risk.

Contents

x

List of Abbreviations

AP . . . . . action potential.

APD . . . . . action potential duration.

BIC . . . . . Bayes information criterion.

BF . . . . . Bayes factor.

CDF . . . . . cumulative distribution function.

CDI . . . . . Cellular Dynamics International.

CHO . . . . . Chinese hamster ovary (cell).

CiPA . . . . . Comprehensive *in-vitro* Proarrhythmia Assay.

CMA-ES . . . Covariance Matrix Adaptation Evolution Strategy.

ECG . . . . . electrocardiogram.

FDA . . . . . (US) Food and Drug Administration.

GHK . . . . . Goldman-Hodgkin-Katz (equation).

hERG . . . . . human-Ether-à-go-go Related Gene.

IC50 . . . . . half maximal inhibitory concentration.

iPS . . . . . induced pluripotent stem (cell).

I-V . . . . . current-voltage.

MAPE . . . . mean absolute percentage error.

MCMC . . . . Markov chain Monte Carlo.

NLME . . . . . non-linear mixed effects.

ODE . . . . . ordinary differential equation.

PDF . . . . . probability density function.

SC-CM . . . . stem cell-derived cardiomyocyte.

TQT . . . . . thorough QT.

UQ . . . . . uncertainty quantification.

VVUQ . . . . verification, validation, and uncertainty quantification.

1 Introduction

The heart is a muscle that pumps blood around the body. Heart contractions are coordinated by electrical waves. These waves propagate through the heart by individual cells passing on electrical signals to their neighbours. Electrical charge is carried across the cell membranes by ions, and heart cell membranes have different channels that allow the passage of specific ions. Pharmaceutical compounds can interfere with the heart's electrical behaviour by binding to these membrane ion channels and stopping them from conducting ions as normal. In some cases, *e.g.* anti-arrhythmic drugs, this is the intended effect. However, this interference is often an unwanted side effect that is potentially fatal.

One particular type of potassium ion channel, known as hERG, is prone to being blocked by pharmaceutical compounds. hERG can have dangerous effects when blocked, leading to arrhythmias (Sanguinetti & Tristani-Firouzi, 2006). As such, using hERG block as a surrogate for proarrhythmic risk has been the predominant approach to cardiac safety for new drugs since the early 2000s when the link between risk and hERG block was established. However, drugs can interact with other ion channels which "may mitigate or exacerbate effects of hERG blockade" (Hancox *et al.*, 2008), and so, more recently, block of channels other than hERG is being considered.

In the 1940s and 1950s, Hodgkin and Huxley pioneered the development of mathematical models describing excitable cells. They modelled a squid giant axon membrane as being a capacitor with different ion currents flowing across the membrane. These ion currents had their own model formulations, which were based on experimental observations. As time went by, mathematical models of excitable heart cells followed the same formulation as the Hodgkin and Huxley squid giant axon model. Experimental methods improved so that more ion currents could be identified and modelled in a similar fashion, and known ion currents could be mapped onto specific ion channels. More recently, such models have been incorporated into the cardiac drug safety process. By measuring a drug's effect on several different ion channels *in vitro*, and using the mathematical model to predict changes to whole-cell electrical behaviour, we can therefore provide a risk prediction.

These mathematical models of heart cells are very complex. They are typically systems of ordinary differential equations (ODEs) with many state variables and even more parameters that need to be calibrated to produce realistic behaviour. This complexity leads to questions of the applicability of these models when the origin of particular parameter values is not clear, or when the models have been fitted to datasets obtained from different cell types, *etc.* When simulating the same experimental conditions, two different mathematical models can give very different predictions in whole-cell electrical behaviour, despite effectively modelling the same system. This variation in predictions can be a problem when trying to predict how a drug will affect a heart cell, since one model might provide a "safe" prediction while another might provide "dangerous".

Typically, when a new heart cell model is published, it includes single parameter values which provided a best fit to particular sets of experimental data that the modellers were considering. Any model with single best-fit parameter values will ultimately provide a single output when using the model to make predictions of future behaviour, *e.g.* after application of a drug. However, given the variable nature of biological systems, we want to predict a range of possible outputs. That is, we want to identify possible sources of uncertainty when constructing a model

and using the model to make predictions, and then propagate this uncertainty into model outputs. This process is illustrated in Figure 1.1.

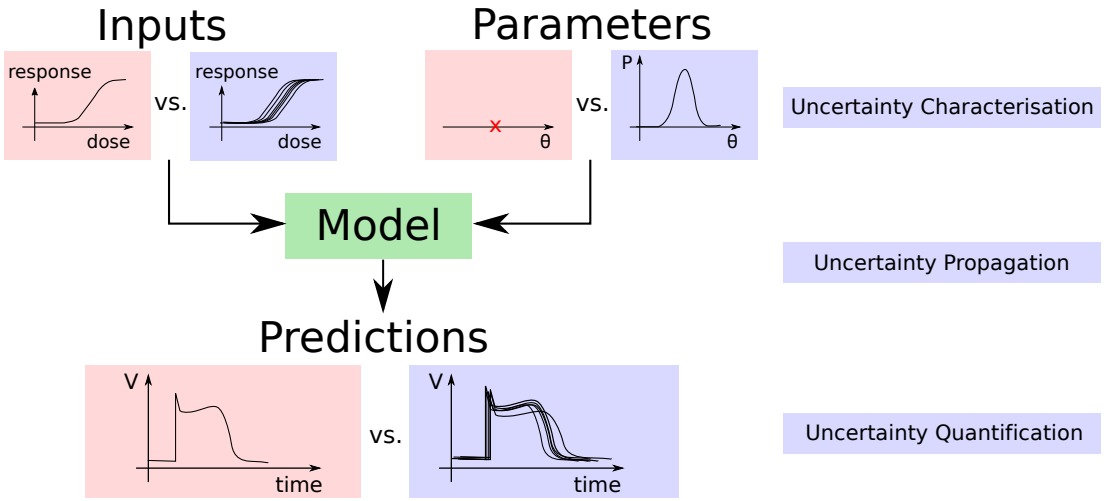

Figure 1.1: Comparison of approaches to making predictions of how a drug will affect whole-cell electrical behaviour. Red: single best-fit parameter values provide a single output. Blue: uncertainty in inputs and model parameters propagated into a range of possible outputs.

In Figure 1.1 we illustrate how uncertainty characterisation and propagation might look in the case of using electrophysiology models to predict how a drug will affect a cell's electrical behaviour. In the top-left of Figure 1.1, there are typical curves that represent how much of a particular ion channel is being blocked at different concentrations of a pharmaceutical compound. The simplest model is a single algebraic function that contains just two parameters which can be fitted to experimental data. The top-right of Figure 1.1 represents the parameterisation of the cell model when fitting to experimental data. The complexity of the cell models often allows for some correlation between parameters, and sometimes the experimental data do not contain enough information to strongly constrain certain parameters. Because of this parameter uncertainty, we might choose to parameterise models with probability distributions instead of single values. To propagate these uncertainties into model predictions, we would take samples from the ion channel screening fits and from the model parameter probability distributions, and make predictions using these samples. Each individual prediction is as likely as any

other one, but when plotted together, they will construct a distribution of possible outcomes, as illustrated at the bottom of Figure 1.1.

Single best-fit model parameterisations and predictions might provide the 'most likely' outcome, but do not take into account other very real possible outcomes. For example, using only single parameter value estimates, we might predict that a pharmaceutical compound at some concentration might have a 'safe' effect on a heart cell. However, characterising and propagating some of the uncertainty associated with such modelling, we might agree that the most likely outcome is 'safe', but rather that there is a 70% chance of a 'safe' outcome, and a 30% chance of a 'dangerous' outcome. This scenario is illustrated in Figure 1.2.

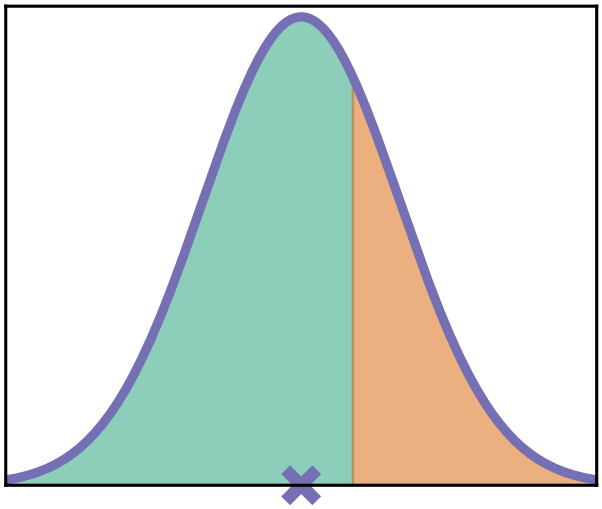

Figure 1.2: An illustration of predicted risk when characterising and propagating uncertainty into model predictions. The blue curve represents some predictor of risk, with the most likely point at its maximum, emphasised by the cross on the x-axis. This point would be the single output when not accounting for uncertainty. The green area (70%) represents a 'safe' outcome, and the red area (30%) represents a 'dangerous' outcome. While a dangerous outcome is less likely than a safe one, 30% is a non-negligible probability and so should be accounted for.

In this thesis, we propose a methodology to bring a level of uncertainty quantification and propagation into the cardiac drug safety process. We incorporate uncertainty into whole-cell electrical behaviour modelling, and into ion channel screening data, to predict a range of responses to a pharmaceutical compound at

different concentrations. This will allow for better-informed decision making when considering the cardiac risk of a newly-developed drug.

In Chapter 2, we provide an introduction to cardiac electrophysiology from the whole-heart scale down to the single-cell and ion channel level. We discuss the development of mathematical models for excitable cell electrical behaviour and for drug block. We then introduce the statistical inference approaches that will be used throughout the thesis.

In Chapter 3, we use some of the statistical inference methods introduced in Chapter 2 to characterise uncertainty in recently-published FDA ion channel screening data. We consider two different statistical models for dealing with the inherent variability of these biological systems. We demonstrate how these uncertainties can be propagated into whole-cell models to predict ranges of compound effects on whole-cell electrical behaviour. We also discuss a method for model selection that is consistent with our statistical framework.

In Chapter 4, we apply similar statistical methods to those used in Chapter 3 to synthetic noisy membrane voltage data to investigate how we can quantify parameter uncertainty in a range of whole-cell models of varying complexity. We also investigate how much information a membrane voltage dataset contains about each model parameter, and how much more information, if any, is provided by a more complex experimental protocol.

In Chapter 5, we apply these statistical inference methods to real membrane voltage recordings obtained from a canine heart cell. We explore possible causes of beat-to-beat variability by inferring parameter distributions for each beat separately, and we consider a hierarchical statistical model which attempts to relate these inter-experiment distributions.

In Chapter 6, we attempt to re-parameterise an existing whole-cell model by fitting parameter distributions to experimental data that I collected at Roche in Basel, Switzerland. We investigate which parameters are the most important in this model, and incorporate this parameter uncertainty with ion channel screening

uncertainty from Chapter 3 to demonstrate how predictions would be made for this cell type.

In Chapter 7, we discuss our overall findings and conclusions, as well as some of the limitations of the methods discussed throughout the thesis. We suggest extensions of the work presented, and some new directions for future research.

2

Background and motivation

We aim to use mathematical models of cardiac cells to evaluate and predict how a pharmaceutical compound will affect cardiac rhythm, and to assess the cardiac safety of the compound. Pharmaceutical compounds act at the molecular level, and the cumulative effect translates to the cellular level and through to the whole-organ scale. In this thesis, we look at novel methods for parameterising mathematical models of cardiac cell electrophysiology. We operate in a statistical framework that characterises some of the uncertainty associated with experimental data, which will allow us to make better-informed predictions about how a drug will interact with the heart.

In this chapter, we will provide the necessary background in cardiac electrophysiology and in mathematical modelling of single-cell cardiac electrophysiology. We also introduce the statistical methods that we will use for parameter inference and uncertainty characterisation. We introduce the physiology of the heart at the whole-organ scale, before describing cardiac electrophysiology at the cellular level. We discuss the evolution of mathematical models of single cardiac cells, how these models are typically constructed and parameterised, and how they are being more-widely adopted into the drug discovery and development process. We then consider some of the issues associated with typical methods for parameterising new cardiac cell models, as well as potential problems introduced by the natural

variability and uncertainty in biological experimental data used to parameterise these models, and how these problems can be mitigated. We then introduce the statistical framework and methods that we will use throughout the thesis to characterise and propagate this uncertainty into model predictions, which allows for better-informed decision-making when assessing the cardiac safety of a pharmaceutical compound.

2.1 Cardiac electrophysiology

2.1.1 Cardiac physiology

The heart is a muscle which is responsible for transporting oxygen and nutrients by pumping blood around the body. It is divided into four chambers: the lower and larger *ventricles* and the upper and smaller *atria*. The heart is divided into left and right sides, and each side consists of one atrium and one ventricle. A diagram of the heart is given in Figure 2.1. The right side of the heart pumps deoxygenated blood through the lungs for oxygenation and back to the heart. The left side of the heart then pumps the oxygenated blood around the rest of the body. The walls of these chambers are made up of layers of muscle cells called *cardiomyocytes* and other cells, *e.g.* fibroblasts and Purkinje fibres. The pumping is a result of the cardiomyocytes contracting. See Katz (2010), for example, for a more complete description.

Cardiomyocytes have many different transmembrane proteins, some of which are *ion channels*. These ion channels allow the passage of multiple or specific ions across the membrane, *e.g.* calcium, sodium, potassium, *etc.* Ions moving across the membrane carry electrical charges that result in a current. Cardiomyocyte membrane ion channels are generally voltage-sensitive; different voltages across the cell membrane can induce conformational changes in these membrane proteins, either allowing or preventing the flow of ions (Hille *et al.*, 2001).

When there is an influx of positively-charged ions, the intracellular potential is raised. Once the cell's membrane voltage reaches a threshold, sodium ion channels open, and so the fast sodium ion current can flow across the membrane, rapidly increasing the cell membrane voltage (depolarisation). At different membrane voltages, different ion channels open and close, and so different ion currents flow

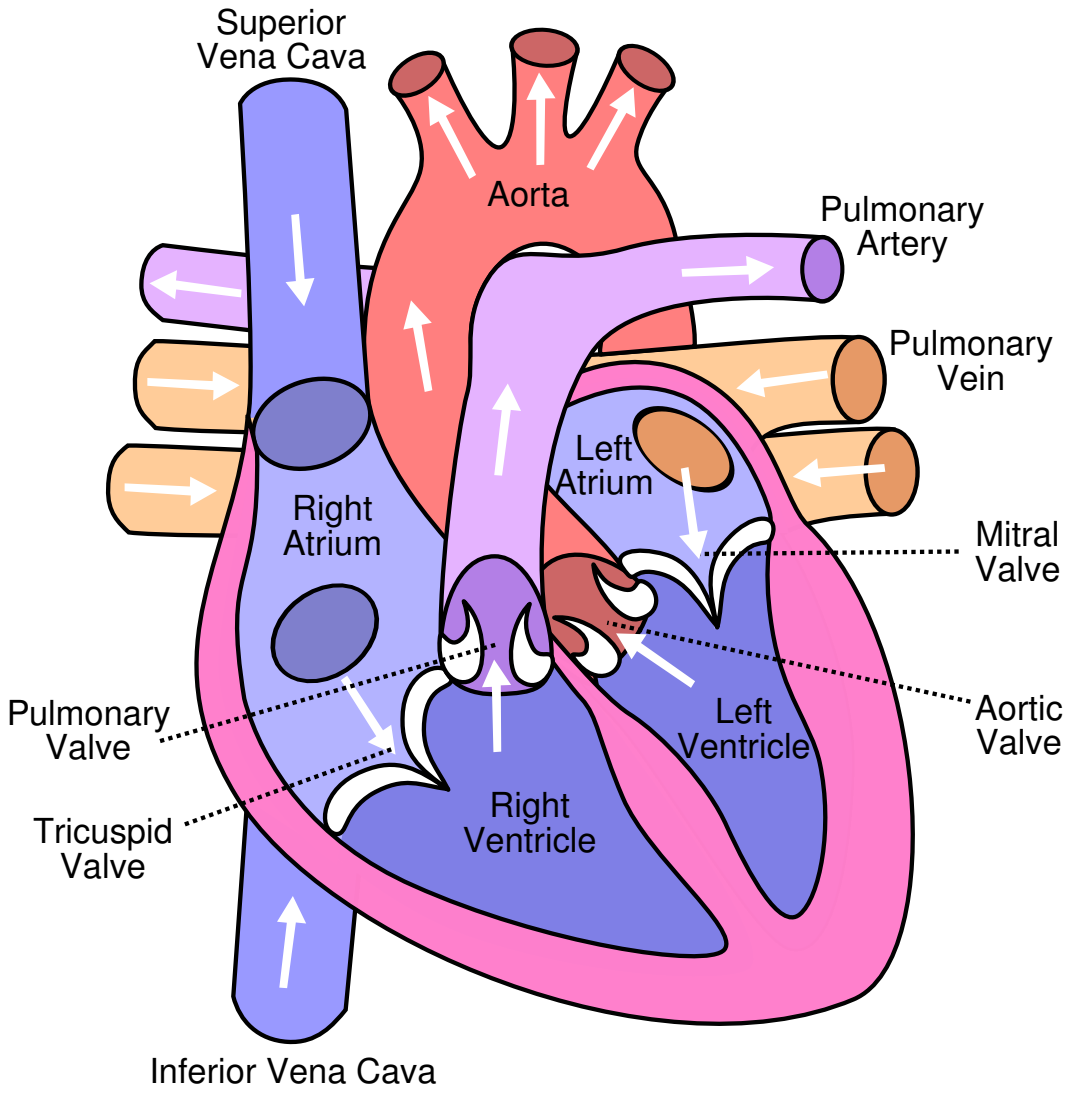

Figure 2.1: The primary components of the heart. Source: *Wikimedia Commons*.

in and out of the cell membrane, and the membrane voltage returns more slowly to its resting voltage (repolarisation). In the case of single-cell experiments, an externally-applied current ("stimulus current") is applied to increase the membrane voltage past the initial threshold. This excitation of the membrane voltage recorded over time is called an *action potential*, and cells with this property are described as *excitable*. Muscular contractions occur when calcium ions released within the cardiomyocytes bind to protein filaments in the cell. These intracellular calcium ions are released in response to an inward-travelling calcium ion current flowing across the cell membrane. See Noble (1979) for a more detailed description.

Since we are interested in how pharmaceutical compounds interact with the heart, we must consider how a compound affects the heart's electrophysiology at molecular (*e.g.* ion channel), cellular (*e.g.* cardiomyocyte), and whole-body (*e.g.* ECG) levels. These effects are then propagated through to the whole-heart scale by the mechanisms described above. We therefore begin by considering existing mathematical models of cardiomyocyte electrophysiology, which help us to understand the translation across these scales.

2.2 Mathematical modelling of cellular electrophysiology

The first mathematical description of an excitable cell was published by Hodgkin & Huxley in 1952. The authors studied a squid giant axon, not a human cell. An action potential recorded by Hodgkin & Huxley is given in Figure 2.2. Future mathematical models of cardiac cells would be based on their work, following mostly the same biophysical principles they laid down.

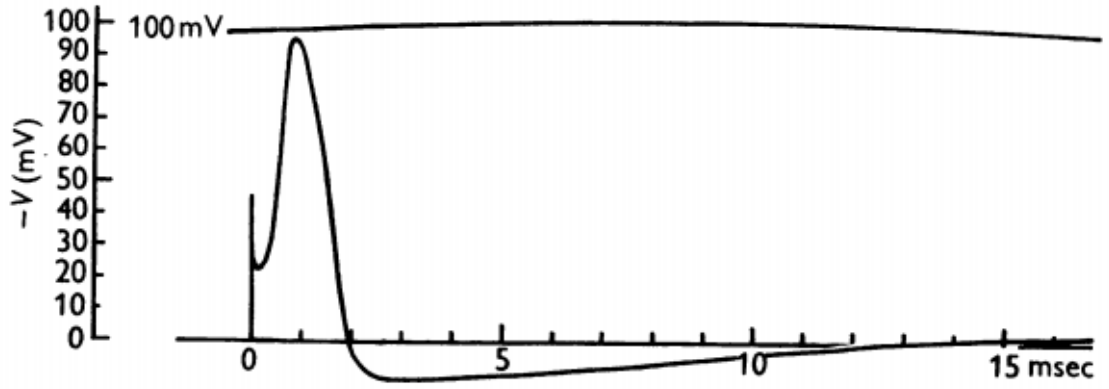

Figure 2.2: Action potential trace recorded from a squid giant axon by Hodgkin & Huxley.

2.2.1 The cell membrane as a circuit

Hodgkin & Huxley modelled the squid giant axon membrane as a simple electrical circuit, with a capacitance provided by the membrane, and three distinct ion currents, with variable resistances, flowing in parallel. The three ion currents

in the model were sodium, potassium, and a 'leak' current which accounted for the remaining, unidentified, ion currents. The circuit diagram from Hodgkin & Huxley (1952a) is given in Figure 2.3.

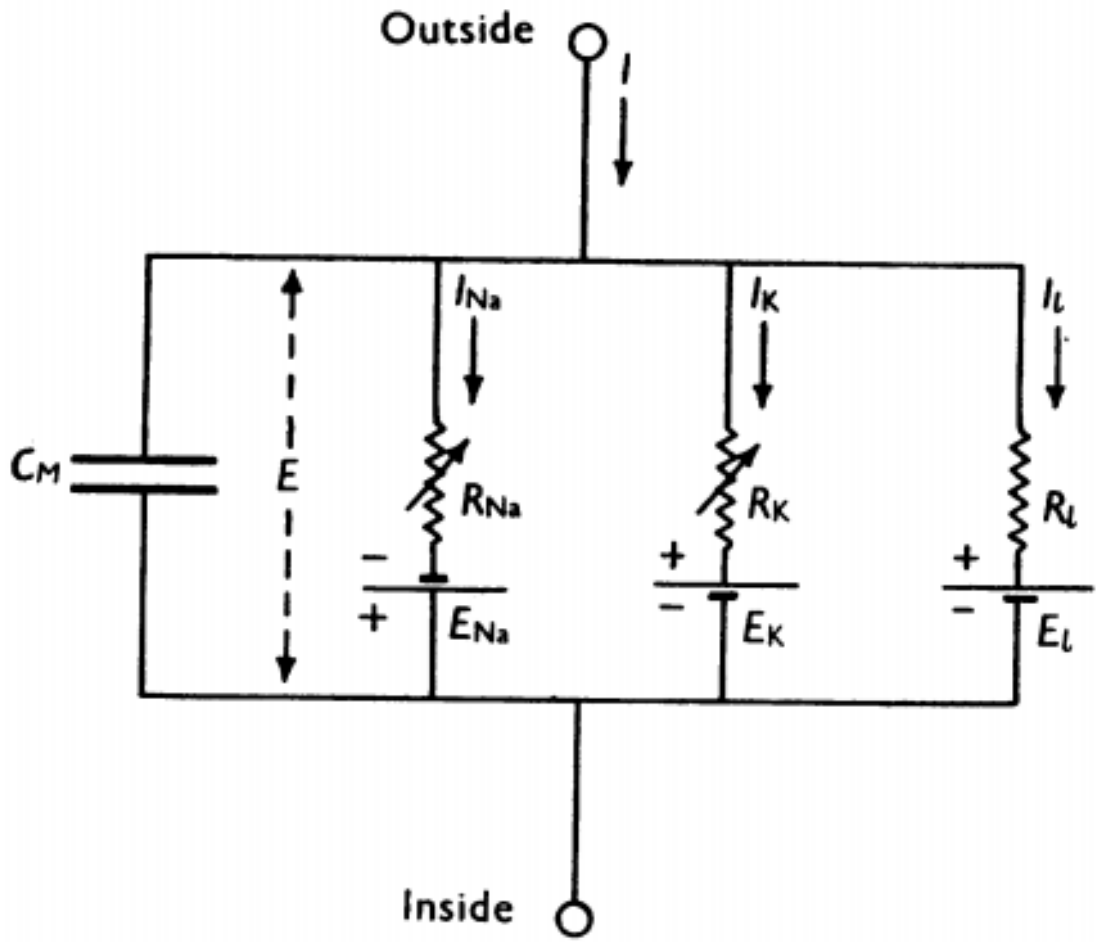

Figure 2.3: Circuit diagram from Hodgkin & Huxley (1952a), describing the squid giant axon membrane as a capacitive circuit with three distinct ion currents, with variable resistances, flowing in parallel.

Each ion current I_i had its own resistance R_i and reversal potential E_i, and was described by Ohm's law:

$$I_i = \frac{V - E_i}{R_i} = G_i(V - E_i), \tag{2.1}$$

where V is the voltage across the cell membrane, $G_i = 1/R_i$ is the *conductance* of the i ion current, and i can represent sodium, potassium, or leak. The conductance G_i corresponds to 'how much' current is able to flow, while the driving potential $(V - E_i)$ corresponds to 'how fast' the current is flowing. Pharmaceutical compounds

can block ion channels, reducing the amount of those ion currents that can flow. This corresponds to a reduction in the conductance, and so these conductance parameters are important when modelling drug action on excitable cells.

To model the cell's response to these currents, Hodgkin & Huxley used the capacitance equation differentiated with respect to time:

$$C_m \frac{\mathrm{d}V}{\mathrm{d}t} = I, \tag{2.2}$$

where C_m is the capacitance of the cell membrane, V is the membrane voltage, and I is the total current flowing across the membrane. When the ion currents I_i are considered separately, and the direction of current flowing across the cell membrane is defined appropriately, Equation (2.2) becomes

$$C_m \frac{\mathrm{d}V}{\mathrm{d}t} = -\sum_i I_i. \tag{2.3}$$

Equation (2.3) does not take the externally-applied current I_{stim} into account. With the presence of such a current, a general model describing the electrical behaviour of an excitable cell becomes

$$C_m \frac{\mathrm{d}V}{\mathrm{d}t} = -\left(\sum_i I_i + I_{stim}\right) \tag{2.4}$$

$$= -\left(\sum_i G_i(V - E_i) + I_{stim}\right), \tag{2.5}$$

We will refer to such models as *action potential models*.

2.2.2 Voltage-gated ion channels

Hodgkin & Huxley supposed that there were *gates* that determined whether an ion channel was open or closed, and assumed that each of these gates was independently open or closed. All of the gates for a particular ion channel would have to be open for the ion channel as a whole to be open and allow that ion current to flow. For example, consider a gating variable j which describes the proportion of those particular gates that are open. The opening and closing dynamics could be modelled as a reaction equation:

$$1 - j \underset{\beta_j}{\overset{\alpha_j}{\rightleftharpoons}} j, \tag{2.6}$$

where a closed gate opened with rate α_j, and an open gate closed with rate β_j. Following experimental observations, these rates were assumed to be voltage-dependent. This 'reaction' was described by the mass action equation

$$\frac{\mathrm{d}j}{\mathrm{d}t} = \alpha_j(1-j) - \beta_j j. \tag{2.7}$$

Assuming the voltage is being held constant, Equation (2.7) has solution

$$j = j_\infty + (j_0 - j_\infty)e^{-t/\tau_j}, \tag{2.8}$$

where j_0 is the initial value of j, $j_\infty = \alpha_j/(\alpha_j + \beta_j)$ is the steady-state value of j, and $\tau_j = 1/(\alpha_j + \beta_j)$ is the time constant.

Hodgkin & Huxley (1952b) used different ionic solutions to separate the sodium and potassium ion currents. They performed a series of experiments which consisted of holding the cell membrane at a particular voltage and then instantaneously increasing the voltage to a new constant value, and then instantaneously decreasing the voltage. This was designed to measure how the ion channels 'opened' and 'closed' over time as a function of voltage. From these recordings, Hodgkin & Huxley empirically formulated equations for gating variables of the form given in Equations (2.6)–(2.8).

For example, the Hodgkin & Huxley (1952a) formulation for the potassium ion current conductance, G_K, was

$$G_K = \hat{G}_K n^4, \tag{2.9}$$

where $\hat{G}_K$ is the *maximal ion current conductance*, and n is an activation gate representing the probability that an individual ion channel gate is open; all 4 gates must be open for potassium ions to flow, and n satisfies Equations (2.6)–(2.8). The potassium current then had four parameters to be determined: $\hat{G}_K$, n_0, α_n, and β_n. The last two were determined by fitting n_∞ and τ_n to experimental recordings of the cell membrane being instantaneously stepped to different voltages.

Similarly, the sodium ion current conductance was determined to take the form

$$G_{Na} = \hat{G}_{Na} m^3 h, \tag{2.10}$$

which had seven parameters to determine, since m and h both satisfy Equations (2.6)–(2.8). These parameters were all fitted to the experimental data heuristically.

We refer to the conductance divided by its maximal conductance as the *open probability*, which corresponds to the probability that a single ion channel is open, or the overall fraction of ion channels that are open in the cell membrane. The maximal ion current conductance parameters were fitted to current-voltage (I-V) curves recorded from the cell. For example, given recordings of the potassium ion current, the maximum value of $\hat{G}_K = I_K/(V-E_K)$ was recorded as the maximal conductance. This provides an approximation to $\hat{G}_K$ as it assumes that the open probability is 1, *i.e.* that all ion channels are open. The maximal ion current conductance can be thought of as proportional to the total number of that particular type of ion channel in the cell membrane. A more general description of an ion current is

$$I_i = \hat{G}_i O_i (V - E_i), \tag{2.11}$$

where $\hat{G}_i$ is the maximal ion current conductance, O_i is the open probability for this ion channel, V is the membrane voltage, and E_i is the reversal potential for ion i. The more complete version of Equation (2.5) is then

$$C_m \frac{\mathrm{d}V}{\mathrm{d}t} = -\left(\sum_i \hat{G}_i O_i (V - E_i) + I_{stim}\right). \tag{2.12}$$

Open probabilities of the form of Equations (2.9) and (2.10) are now referred to as *Hodgkin-Huxley-style* open probabilities.

Since pharmaceutical compounds affect ion channel dynamics, we can model a compound's effect as a change to the open probability dynamics, or more simply as a change to the maximal conductance value. The latter option has historically been a popular method for describing compound effects on cardiomyocytes (Brennan *et al.*, 2009), and so these maximal conductance parameters are often the parameters of most interest when fitting action potential models to experimental data.

2.2.3 Development of action potential models

Markov models of ion channels

Following advances in molecular biology, it was discovered that membrane proteins can have different types of conformation, such as open, closed, and inactivated in the case of ion channels. A class of mathematical models used to describe these more complex ion channel gating kinetics is *Markov models*, which take time and voltage dependence into account by allowing for an ion channel to move between different configurations. The simplest example is Equation (2.6), where each gating variable is described by a two-state Markov model. α_j and β_j are typically voltage-dependent.

More generally, Markov models can have an arbitrary number of different conformations. There exist many such models for different ion channels, of varying structural complexity. For example, the Clancy & Rudy (2001) Markov model of a particular potassium ion channel is

$$\begin{array}{ccccccccc} & & & & & & I & & \\ & & & & & \swarrow\!\!\!\swarrow & & \nearrow\!\!\!\nearrow & \\ & & & & & \alpha_2 & & \alpha_i & \\ & & & & & \beta_i & & & \\ C_3 & \underset{\beta_1}{\overset{\alpha_1}{\rightleftharpoons}} & C_2 & \underset{k_b}{\overset{k_f}{\rightleftharpoons}} & C_1 & \underset{\beta_2}{\overset{\alpha_2}{\rightleftharpoons}} & O & & \end{array} \tag{2.13}$$

where there are three closed states (C_1, C_2, C_3), one open state (O), and one inactivated state (I).

Markov models can capture more complex gating behaviour, but are more computationally expensive to model because of the increased number of mass action ordinary differential equations that must be solved numerically, as well as having a greater number of parameters to fit. In more-modern action potential models, some ion current open probabilities are formulated as Markov models, while others are formulated as Hodgkin-Huxley-style open probabilities.

GHK equation for ion currents

An alternative formulation for ion currents is based on the *Goldman-Hodgkin-Katz* (GHK) equation. It is a more complex model formulation than Equation (2.11), and is a function of intra- and extracellular ionic concentrations, voltage across the cell

membrane, and various physical constants (Hille *et al.*, 2001). It does not contain explicit reversal potential-dependence, but the reversal potential is implicitly defined by setting the GHK equation equal to 0. The GHK equation for the flow of an ion S is

$$\Phi_S = P_S z_S^2 \frac{VF^2}{RT} \frac{[S]_{int} - [S]_{ext} \exp(z_S VF/RT)}{1 - \exp(z_S VF/RT)}, \tag{2.14}$$

where P_S is the permeability of the membrane to the ion S, $[S]_{int}$ and $[S]_{ext}$ are the intra- and extra-cellular concentrations of S, respectively, z_s is the valence of the ions that are being carried, F is the Faraday constant, R is the gas constant, T is the absolute temperature, and V is the membrane voltage.

In modern action potential models, typically some ion currents are modelled with a GHK equation, while others are modelled as in Equation (2.11). Where ion currents are described as GHK equations, we treat their permeabilities P_S as parameters of interest, and include them when we refer to "maximal conductance parameters".

More-complex cell structures

The first action potential model of a cardiac cell was published by Noble in 1962, extending the work by Hodgkin & Huxley (1952a). Not many ion currents could be measured independently at the time, and so the model published by Noble only had three ion currents, and one leak current that effectively comprised the remaining unknown currents without modelling them explicitly. A schematic of the dynamically-modelled components of this cell is shown in Figure 2.4. While there is only one potassium channel depicted in this figure, the mathematical model contains two potassium currents with different properties.

As understanding of molecular biology has improved over the years, many more-complicated models have been presented, taking into account more understanding of ion channel kinetics, pumps, and exchangers in cardiomyocytes (Fink *et al.*, 2011). For example, O'Hara *et al.* published a model of the human ventricular cardiomyocyte in 2011 that included 15 ion currents flowing across the cell membrane, as well as intracellular dynamics of ion concentrations which also affect how the currents flow, and hence affect the action potential. O'Hara *et al.* chose to use

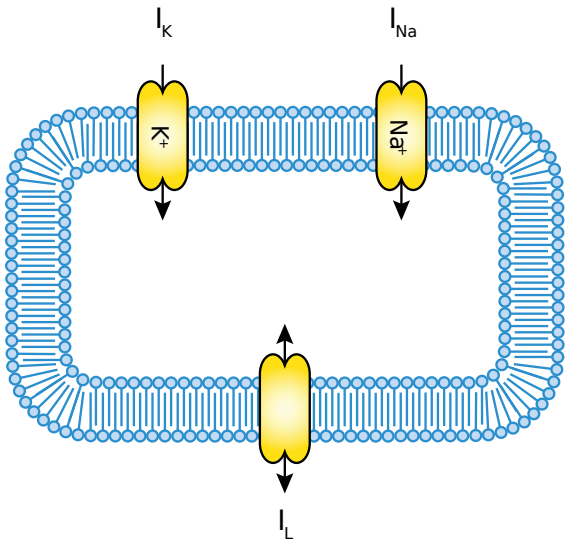

Figure 2.4: A schematic of a cardiac cell given by the Noble (1962) action potential model. The potassium current has two components which can be modelled as two separate potassium currents. Source: cellml.org.

Hodgkin-Huxley-style ion channel formulations, and they discuss how Markov models would increase computational complexity. There are 41 state variables in the basic O'Hara *et al.* (2011) model, which is already a large number resulting in a heavy computational cost. Due to the modular nature of action potential models, however, a modeller can easily substitute their own model component formulations, such as Markov models for ion channels of interest. A schematic of the dynamically-modelled components of the O'Hara *et al.* (2011) model is shown in Figure 2.5.

CellML repository

When a new model is published, it is important that another modeller can implement the same model to verify the results, or to use it for their own predictions. However, different modellers often use different software and programming languages, meaning such models are not easily shared. To remedy this situation, the *CellML repository* (Cuellar *et al.*, 2003) was started. CellML is based on the XML markup language, and provides a standard language for encoding models, which can easily be converted into different programming languages. CellML allows for the encoding of mathematical systems, and for the tagging of parameters and variables so that they can be individually analysed or varied. Not all published models are in the repository,

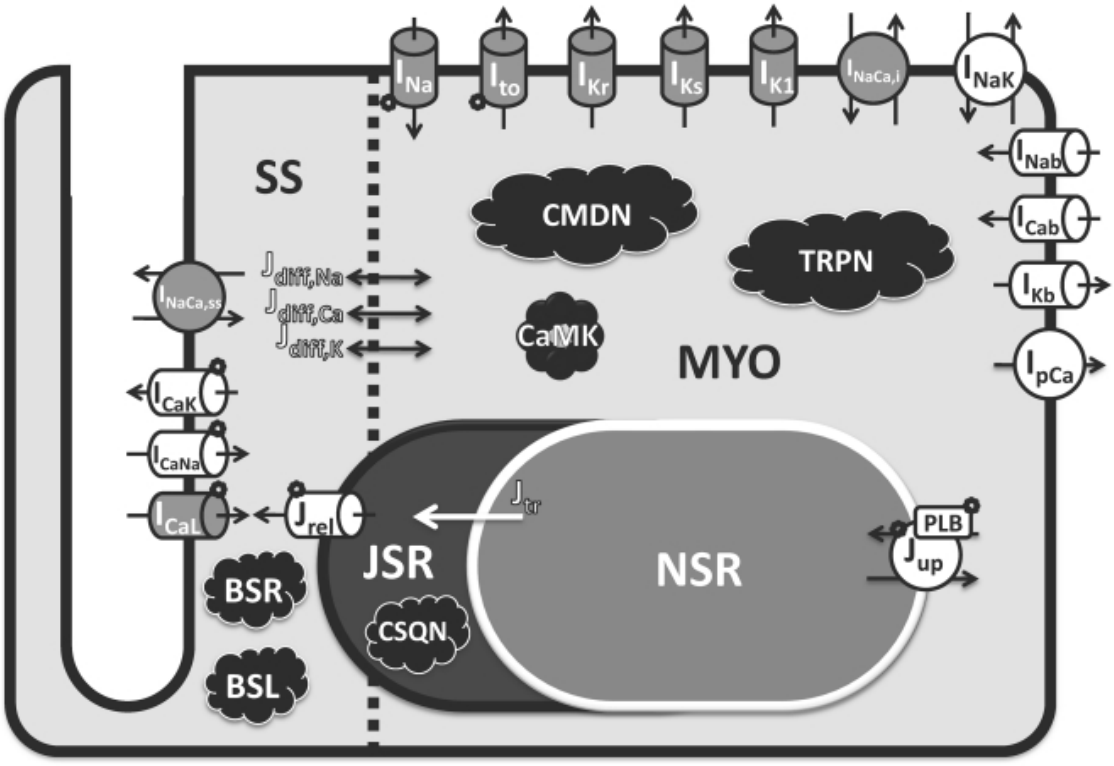

Figure 2.5: Figure 5 from O'Hara *et al.* (2011). A schematic of the ventricular cardiomyocyte described by the O'Hara *et al.* (2011) model.

however, but it does contain roughly 280 mathematical models of electrophysiological systems, and also many models from a range of topics. With such a large number of existing models, it would be time-consuming to search through the literature to find them, so a single repository is a helpful solution.

2.2.4 Parameterising action potential models

As discussed in Section 2.2.2, early action potential models were constructed according to the physiological understanding of the system at the time. For example, Hodgkin & Huxley were only able to experimentally isolate sodium and potassium ion currents, and so they grouped all remaining 'unknown' currents into a single "leak" current. Model components were empirically formulated on experimental data, and the model parameters were hand-tuned so that the model output matched the data. As more ion channel membrane proteins were discovered, modellers were able to explicitly include many ion currents in their models, instead

of grouping all 'unknown' currents into a single leak current. Newer models are typically based on earlier models, but with the inclusion of a newly-discovered ion current, or other cellular process. These new model components are usually formulated and parameterised according to experimental data either previously published or provided by an experimental collaborator. Sometimes, some of the parameters inherited from the earlier model will be adjusted to allow for a better fit to the new data, but very often the inherited parameter values are left unaltered. For example, in 1994, Luo & Rudy published a model of the mammalian ventricular action potential, fitted mostly to experimental data obtained from Guinea pig cells. The model included a formulation of the sodium–potassium pump current, I_{NaK}. In 1998, Jafri *et al.* published a model of the Guinea pig action potential, including the same model formulation for I_{NaK}, with only the maximal current density parameter changed — all other parameters in the I_{NaK} model remained the same. Similarly, in 1999, Winslow *et al.* published a model of the canine action potential, again using the same formulation for I_{NaK}, with the maximal density parameter changed again and all other parameters unchanged.

Of particular interest in this thesis are the maximal ion current conductance parameters. Generally, drug interactions with a cardiomyocyte are modelled as a drug molecule physically blocking an ion channel, thereby reducing the overall number of 'available' ion channels. We assume that drug molecules bind to open and closed channels with the same affinity, and so ion current blocking is independent of gating. This can be modelled as a reduction in the maximal ion current conductance parameter, and so we think of these parameters as the targets of pharmaceutical compounds. It is therefore important to know if we can have confidence in the conductance parameters in published action potential models, and for that it is important to understand where the published parameter values originally came from. Additionally, these parameter values are typically published as single fixed values, or sometimes as a mean value with standard errors when obtained by fitting to data. This does not account for possible uncertainty introduced when fitting the model to experimental data, and so we will not be able to account for these

uncertainties when making predictions using these action potential models. In this section, we discuss inheritance of model components and parameter values from earlier models, and in the next section we discuss how to account for uncertainty when fitting models to experimental data.

Niederer *et al.* (2009) performed a meta-analysis of the Iyer *et al.* (2004) and Ten Tusscher *et al.* (2004) models of human ventricular cardiomyocytes. The authors traced the origins of these models' parameter values to find the original sources of the reported values. The authors aimed to find the possibility that different models of the same system will produce different results.

Some inherited model components were originally fitted to different types of experimental data, and when these components are continually inherited and parameterised from preceding models, their origins become unclear and may not be as relevant to the situation being investigated. For example, Niederer *et al.* say that "both cell models are based on ∼50% human and ∼25% guinea-pig data, with the remaining ∼25% taken from a wide range of species". The authors also discuss how, due to the natural complexity of mechanistic biophysical models, there is often information passed between model components over the course of a simulation. When one component is compromised by having an uncertain history, this is propagated and compounded by further computations being misinformed, resulting in a potentially untrustworthy model output.

The sources of parameter values can mean that it is difficult for us to rely upon model outputs, especially in the safety-critical context of pharmaceutical compound predictions that is of primary interest here. We therefore want to re-parameterise as much of an action potential model as possible by fitting to data from as few sources as possible: a single cell, or cells from the same batch of cells, *etc.* Throughout the rest of this thesis, we will explore how much information is present in membrane voltage action potential recordings when attempting to re-parameterise action potential models.

In addition to the types of data that models are being fitted to, there is also natural biological variability present when performing experiments. Typically,

parameters fitted to experimental data are reported as single values, while it may be that a range of parameter values could plausibly fit the data similarly well. In the next section, we therefore discuss some of the formal techniques used in Engineering to assess uncertainty from experimental data, and other sources, and how we will apply them later in the thesis to cardiac modelling.

2.3 Verification, validation, and uncertainty quantification (VVUQ)

Models of the cardiac action potential are established, valuable, and important research tools because they integrate biophysical mechanisms quantitatively, and so have explanatory and predictive power. Contemporary models are sufficiently detailed to allow the effects of ion channel gene mutations, pharmaceuticals, and disease to be examined in mechanistic detail (Roberts *et al.*, 2012; Mirams *et al.*, 2012). However, while the present generation of models are powerful tools, model parameters are generally assigned a fixed value, which means that the models produce a fixed prediction, as discussed in Section 2.2.4. In contrast, the experimental action potentials recorded from real cardiac cells are variable, with changes from beat to beat in a single cell (termed *intrinsic* variability), and from one cell to another (*extrinsic* variability). Intrinsic variability may be caused by seemingly-random processes such as stochastic ion channel gating (it is possible that the process is truly deterministic, but so complex that a stochastic model is the best approximation available to us), or non-linear dynamics such as alternans of action potential duration (APD). Extrinsic variability is considered to be caused by quantities that genuinely vary from cell to cell, *e.g.* cell size or ion channel expression. In practice, it can be difficult to distinguish between these sources of variability. Additionally, variability can be compounded by measurement errors when data from experiments are used to generate parameters for use in action potential models.

In this section, we introduce the components of VVUQ in the general case, before discussing its application to action potential modelling. These methods will allow us later in the thesis to account for some of the uncertainties associated with fitting

action potential models to experimental data, and propagate these uncertainties into model predictions. This is important for preclinical drug development, by allowing better-informed decision-making about the drug's cardiac risk.

Oberkampf *et al.* define *verification* and *validation* as

- "*Verification*: The process of determining that a model implementation accurately represents the developer's conceptual description of the model and the solution to the model."

- "*Validation*: The process of determining the degree to which a model is an accurate representation of the real world from the perspective of the intended uses of the model."

Uncertainty quantification (UQ) determines how uncertainty in the model inputs (*e.g.* parameters and initial conditions) affects the model outputs. Interest in UQ has grown as part of a drive for rigorous and formal approaches to assess the credibility of computational models. The heavy use of computational models for safety-critical applications in the automotive, aerospace, nuclear and structural engineering industries in particular motivated the development of *'Verification, Validation, and Uncertainty Quantification'* (VVUQ), which forms a set of methodologies, frameworks and best practices for improved assessment of the reliability and robustness of model predictions (National Research Council, 2012; Oberkampf & Roy, 2010). In this context, *verification* is defined as the process of confirming that a computational model (software) correctly implements an underlying mathematical model, and *validation* compares a model's predictions with reality. Although UQ forms part of the overall VVUQ process, each of the stages are intertwined, and in particular UQ improves the ability to perform validation, since understanding the uncertainty in model predictions facilitates comparison with experimental results. Historically, VVUQ has not been performed often in cardiac electrophysiology modelling.

There are several potential sources of uncertainty in a computational model of a real system; these include, at least, the following (Vernon *et al.*, 2010):

- *Observational uncertainty* is uncertainty or measurement errors in experimental data. For example, uncertainty represented by error bars in measurements of the current-voltage profile for a particular ion channel used to assign model parameters, or error bars in measurements of APD restitution used to evaluate model performance.

- *Parameter uncertainty* refers to uncertainty in model parameters, which may be a consequence of observational uncertainty as well as variability, or simply lack of information. It may be advantageous to express a model parameter (such as a maximal conductance) as a random variable with a distribution, rather than a fixed value.

- *Condition uncertainty* describes our uncertainty about the initial conditions and boundary conditions. For a cardiac action potential model the initial conditions are typically set by running the model until it has reached a steady state, but this will not capture the constantly varying environment of temperature, ion concentrations, and metabolism in which a real cell operates.

- *Structural uncertainty* accounts for the differences between a model and the real system that it represents. For example, a model of an ion channel will not be an exact representation of the biophysical dynamics of a population of proteins in the membrane, and structural uncertainty aims to quantify this difference.

- *Simulator uncertainty* addresses the uncertainty introduced when using an approximation to the true solution of the equations of the mathematical model when we perform a simulation. This includes any uncertainty introduced by using discretisation in numerical methods (numerical error), or uncertainty when a fast-running *surrogate* model (*e.g.* an *emulator*) is used to approximate the outputs of a computational model that is expensive to solve.

When comparing computational models to experimental data, there is also the issue of parameter *identifiability*. There are two types of identifiability:

- *Structural identifiability* is when a model's parameters can be determined from measurements of its outputs. Depending on the model formulation, this is sometimes not possible. For example, two parameters might only ever appear in a model multiplied together, and so this product can be identified, but not each parameter separately. This can be an issue if the two parameters represent different physical properties which might be needed separately when making some kind of prediction.

- *Practical identifiability* refers to how many model parameters can be determined by comparing with experimental data. For example, once a best-fit set of parameter values is obtained, if one parameter value is varied but the fit to the data does not change, this parameter is practically non-identifiable locally.

Walch & Eisenberg (2016) investigated the identifiability of parameters of Hodgkin-Huxley-style ion channel open probabilities. They found that from the same type of data that Hodgkin & Huxley fit their model to, not all of the parameters were identifiable. Different combinations of parameters were identifiable, however, but not all separately. Making further assumptions about known initial conditions allowed Walch & Eisenberg to identify all parameters.

Techniques for UQ provide a means to deal with these different sources of uncertainty. There are two stages to UQ related to parameter uncertainty:

1. *Uncertainty characterisation* regards uncertainty in model *inputs*. In this stage uncertainty in parameters is characterised by assigning probability distributions to input parameters instead of single values, although sometimes simple statistics (*i.e.* means and variances) are used. If the input is a parameter that is directly measurable, this is a purely experimental task because the probability distribution is informed by the experimental observations. On the other hand, if the input is a parameter that is indirectly inferred from other data, statistical methods may be required to estimate the parameter uncertainty.

2. *Uncertainty propagation* (or *uncertainty analysis*) regards uncertainty in model *outputs*. Here the aim is to establish the uncertainty in model outputs due to the uncertainty in inputs, again as probability distributions or simple statistics. Generally, this stage is very computationally-demanding, since a large number of simulations are needed to generate outputs for the different combinations of inputs that are possible.

2.3.1 VVUQ in cardiac modelling

Parameter uncertainty suggests that it might be helpful to express a model parameter as a random variable with a probability distribution, rather than a single fixed value. As such, when fitting models to synthetic and experimental data later in this thesis, we will assign a likelihood to a set of parameter values, which will provide a measure of how good a fit they provide to the data, but will also provide a way to compare the relative likelihoods of different sets of parameters. Problems related to uncertainty and variability are not unique to cardiac electrophysiology, and new approaches are beginning to emerge from areas as diverse as models of the atmosphere (Lee *et al.*, 2013) and galaxy formation (Vernon *et al.*, 2010).

Until recently, VVUQ has not been a priority for cardiac modelling, because this type of model has not been widely used in high-risk or safety-critical applications. However, the present generation of cardiac action potential models are sufficiently detailed that there is the prospect that they could be used as both as part of clinical applications and also for drug safety assessment. Both of these applications are safety critical. For clinical applications, the model output could be guidance for ablation in clinical procedures, and the inputs would include personalised measures of tissue conductivity and anatomy (Trayanova *et al.*, 2012). For safety testing in drug development, the output could be a measure of action potential prolongation, and the inputs would include a quantification of the reduction of different ion currents as a function of compound concentration (Mirams *et al.*, 2014). In both types of application it will be important to express a measure of confidence in the model outputs, given uncertainties and errors in the inputs. As

a result, there has been growing interest and application of (VV)UQ in cardiac modelling (Geneser *et al.*, 2007; Elkins *et al.*, 2013a; Pathmanathan & Gray, 2013, 2014; Pathmanathan *et al.*, 2015a).

The first formal UQ of the steady-state inactivation of the fast sodium current, I_{Na}, was performed by Pathmanathan *et al.* (2015a). Inactivation dynamics of an ion current can be measured by holding a cell at a particular voltage, to open a fraction of the ion channels, and then stepping the voltage up and recording the resulting current (Bendahhou *et al.*, 1999). The maximum current measured after each of these voltage steps is normalised to the overall maximum, and these normalised currents are plotted against their respective voltages to give a graph of how inactivation gating is voltage-dependent. Generally, a sigmoid function is fitted to averages of multiple recordings, with just a set of 'best fit' parameters. This describes a 'population-averaged' approach, but may not necessarily describe the behaviour of an 'average' cell. Pathmanathan *et al.* applied an 'individual-based' statistical methodology to assess voltage clamp data: *nonlinear mixed effects* modelling was used to infer the behaviour of an 'average' cell, and the variation across cells. Pathmanathan *et al.* found that the population-averaged approach "can fail to capture the behaviour of any cell".

We will use VVUQ, in a particular statistical framework, throughout the thesis to quantify uncertainty in experimental data, and to assess parameter identifiability in ion channel screening data in Chapter 3, and in cardiomyocyte membrane voltage action potential data in Chapters 4–6. These uncertainties can then be propagated to make a range of predictions about how a particular drug will interact with a cardiomyocyte's electrophysiology, and these predictions will have quantified probabilities associated with them.

2.4 Pharmaceutical cardiac safety

In this section, we describe normal cardiac rhythm, and common effects of drugs on cardiac rhythm. We then discuss a recent initiative calling for greater implementation of computational cardiac models in the drug safety assessment process. Throughout

the thesis we will use VVUQ techniques discussed in Section 2.3 to allow the propagation of uncertainty into the *in silico* stage of the safety assessment process, allowing for better-informed decision-making about a drug's cardiac risk.

2.4.1 Cardiac rhythm

At the tissue level, electrical impulses are passed on to neighbouring cells. Each cell produces an action potential which contributes to the overall electrical activity of the heart. This whole-heart electrical activity can be measured in an *electrocardiogram* (ECG). An ECG is a recording of voltage against time measured by placing electrodes on the skin. The typical ECG shape of a healthy heart is shown at the top of Figure 2.6. An example of a real ECG from a relatively healthy heart is shown in Figure 2.7.

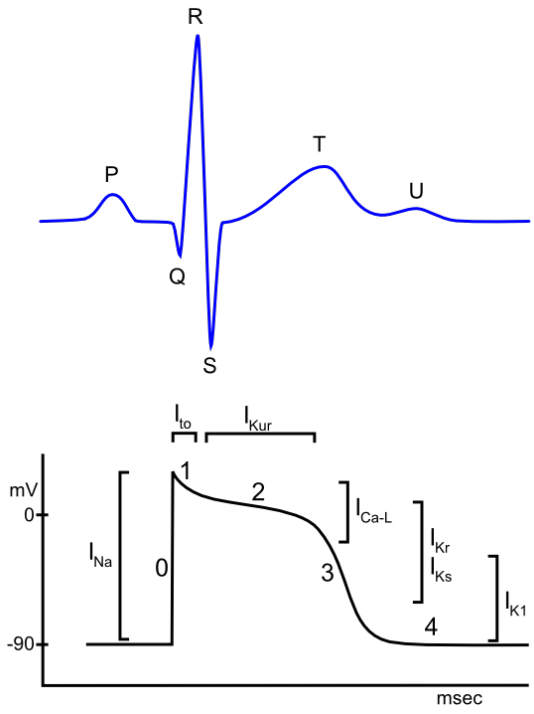

Figure 2.6: Top: a typical ECG trace. Bottom: a typical action potential trace with different ion currents labelled when they are active. Source: ecgpedia.org

There are many forms of irregular heartbeat, known as *arrhythmias*, that can have serious consequences if left untreated. An important type of cardiac arrhythmia is *Torsades de Pointes* (TdP), which is a ventricular tachycardia (high heart rate).

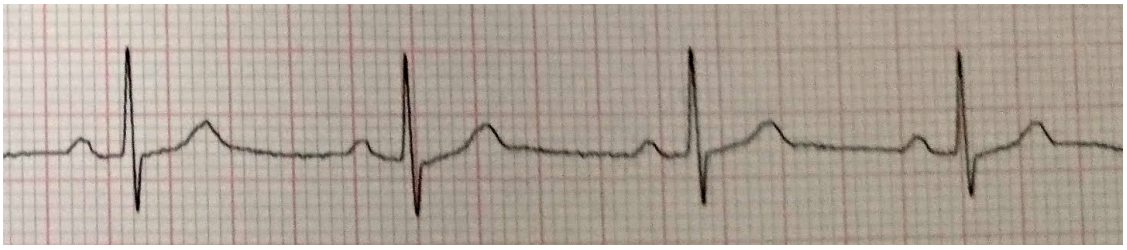

Figure 2.7: An ECG recorded from the mostly-healthy heart of a former rower.

It has a characteristic form on an ECG where it appears to be twisting around an imaginary axis. TdP can lead to sudden cardiac death, and is associated with Long-QT syndrome, a prolongation of the time between the points labelled Q and T in Figure 2.6 (Napolitano *et al.*, 1994).

2.4.2 Drug development

An important criterion in cardiac safety is to test whether or not a drug can induce TdP. There is a direct relation between action potential duration (APD), the time from depolarisation until repolarisation, and the QT interval (Belardinelli *et al.*, 2003). Because of this connection of TdP with Long-QT syndrome, drugs are tested to see if they prolong APD as a predictor of whether they will prolong the QT interval.

Antiarrhythmic drugs are designed to suppress abnormal cardiac rhythms, but drugs with other purposes can have side effects affecting cardiac rhythm. Not all drugs that prolong APD increase the risk of TdP, and not all drugs that increase the risk of TdP prolong APD (Belardinelli *et al.*, 2003). Between 1988 and 2001, at least 9 non-antiarrhythmic drugs were withdrawn from the market, due to their potential to increase TdP risk (Shah, 2005).

In 2005, the International Committee on Harmonization published the E14 clinical guidelines saying that QT interval prolongation should be considered carefully when developing new drugs. The E14 guidelines recommend the use of a "thorough QT" study to assess the proarrhythmic risk of a new drug. Membrane proteins comprise more than 60% of drug targets (Davey, 2004), and are commonly

affected as off-targets, and so modelling drugs' effects on membrane proteins is of great interest and importance.

A new initiative for safety pharmacology is the Comprehensive *in-vitro* Proarrhythmia Assay (CiPA, Cavero & Holzgrefe, 2014). CiPA has three main stages: the effects of a pharmaceutical compound on multiple ion channels are measured, these measurements are used in *in-silico* human ventricular cardiomyocyte models to predict potential proarrhythmic biomarkers, and finally these predictions are verified by measuring the effects of the compound on the electrophysiology of stem cell-derived cardiomyocytes. CiPA is designed to replace the thorough QT study with a focus on *in-silico* and *in vitro* approaches, potentially reducing the risks and costs of the drug discovery and development process. There is an additional stage of a human trial, looking at more informative biomarkers than the existing ones.

2.4.3 Stem cell-derived cardiomyocytes

The third stage of CiPA involves applying a pharmaceutical compound to a stem cell-derived cardiomyocyte, and comparing these measurements with the predictions obtained from ion channel screening data and action potential modelling. We therefore wish to consider their viability as a surrogate for human cardiomyocytes in the drug safety process.

Human induced pluripotent stem (iPS) cells (Yu *et al.*, 2007) are a promising cell type for many types of investigation, particularly for the CiPA initiative. They can be driven toward different cell types, and can be acquired more readily than embryonic stem cells, meaning there are fewer ethical concerns; this approach also admits the possibility of patient-specific modelling. There has been progress in driving iPS cells to differentiate into cardiomyocytes, and they display qualitatively similar electrophysiological behaviour to adult human cardiomyocytes. For example, Cellular Dynamics International (CDI) have developed a line of iPS cell-derived cardiomyocytes that they call "iCells" (Ma *et al.*, 2011). A batch of these cells is approximately 95% composed of cardiomyocyte-like cells. These cardiomyocyte-like cells can be further divided into ventricular-, atrial-, or nodal-like cells, *i.e.* cells

from different regions of the heart, just like native cardiomyocytes. Under ideal conditions, these iCells produce spontaneous action potentials, where the three different cell types have distinctive action potential morphologies; the cell types can be classified based on these morphologies.

However, there is *a priori* uncertainty of which cell types are present, which makes automated high-throughput screening difficult. In addition to the uncertainty surrounding which types of cell are present, stem cell-derived cardiomyocytes from different providers exhibit different behaviours, as do different batches of cells from the same provider. This variability suggests that we should attempt to construct cell-specific models, so that we can understand the output of stem cell assays more completely.

2.4.4 Uncertainty propagation

In this thesis, we will discuss various sources of uncertainty when parameterising action potential models from experimental data for use in drug safety predictions. We wish to characterise these uncertainties to allow us to propagate this uncertainty into model outputs. That is, we wish to say "given these sets of experimental data and this action potential model, what range of behaviour do we predict, together with associated probabilities?". This uncertainty characterisation and propagation will allow for better-informed decision-making at the pre-clinical level when considering the cardiac safety of a pharmaceutical compound.

In particular, we will characterise uncertainty in ion channel screening data: inter-experiment variability and observation noise make it difficult to know exactly how much of a particular ion channel a drug will block. By characterising this uncertainty, we can assign likelihoods to different levels of block, allowing us to assign probability densities to predictions about how the drug will affect whole-cell electrophysiology. We will also characterise uncertainty in action potential data: re-parameterising the maximal ion current conductance parameters in action potential models will allow us to explore identifiability issues in these complex models, and to assess how much information is present in the experimental data about these parameters. This can

inform experimental design by identifying experimental protocols which produce data containing more information about particular parameters of interest. Combining these uncertainties and propagating them into action potential model predictions, we will be able to assign probability densities to the effect of a concentration of a drug on action potential behaviour.

In the next section, we describe the statistical framework we adopt to allow us to characterise uncertainties and express them as probabilities. We also discuss the statistical methods we employ to infer such probabilities and propagate them into model outputs, so that model predictions can be expressed as probability distributions.

2.5 Methods

As discussed in Section 2.2, single-cell cardiac electrophysiology models have become very complex, often including many state variables and many more parameters. Many of these parameters have been inherited from earlier models, and it is often very difficult to find the original source of particular parameter values. This leads us to question how relevant these parameters, and hence models, can be when used for predictions in the human case. Where parameters have been fitted to experimental data, single best-fit parameter values are typically reported, and possible sources of uncertainty have not been considered. As discussed in Section 2.2.4, these single best-fit parameter values do not allow us to consider the uncertainties of a drug's effect on the action potential, and so we are not making use of all the information available to us.

In accounting for some sources of uncertainty when fitting models to noisy experimental data, we do not wish to obtain single best-fit sets of parameter values. Such point estimates do not allow for propagation of uncertainty into model predictions and so might not capture important variation in model outputs. For example, the effect of a pharmaceutical compound on action potential duration (APD): best-fit parameter values may predict that a particular compound will prolong APD by a 'safe' margin, but propagating uncertainty from the ion channel screening data could suggest that the APD will be prolonged by a 'safe' margin with

the greatest probability, but with a non-negligible probability that the APD will be prolonged by an 'unsafe' margin. These extreme (but non-negligible) cases are missed by best-fit parameter value estimates. This is illustrated in Figure 1.2 of Chapter 1, where in this example the x-axis is APD and the y-axis is probability density.

To apply some of the uncertainty analysis techniques discussed in Section 2.3, we will use *Bayesian* statistical methods. Instead of obtaining single values for parameters from experimental data, we treat model parameters as random variables, and wish to infer their likely probability distributions given the available experimental data. That is, given some experimental data, how likely is it that different combinations of parameter values can give rise to the observed data? We will then consider the effects of this uncertainty characterisation on the relevant stages of the drug safety process as discussed in Section 2.4.2.

2.5.1 Statistical model

In order to infer a probability distribution for the parameters contained within a particular mathematical model, we need to specify our statistical model, which describes how we believe our mathematical model relates to the observed experimental data. The simplest (non-trivial) statistical model is to assume that our mathematical model accurately represents the underlying system behaviour, and that the only source of error is observation noise when performing the experiment:

$$\text{Observed data} = \text{System behaviour} + \text{Noise model}. \tag{2.15}$$

We will denote the parameters of the mathematical model by $\boldsymbol{\theta}$, and a new parameter σ which will define the spread of the noise model in Equation (2.15). Therefore the complete set of mechanistic and statistical model parameters is $\{\boldsymbol{\theta}, \sigma\}$.

A simple noise model is that the observed data points are Normally distributed around the model output values with some standard deviation σ. We can either set σ to some fixed value if, for example, we know with certainty how our experimental equipment behaves, or we can treat σ the same as our mathematical model parameters, and attempt to infer a probability distribution for σ, along with

the mathematical model parameters, from the experimental data. We will use the latter option, and attempt to infer a joint distribution for the model parameters $\boldsymbol{\theta}$ and the noise standard deviation parameter σ. A schematic representing this simple statistical model is given in Figure 2.8.

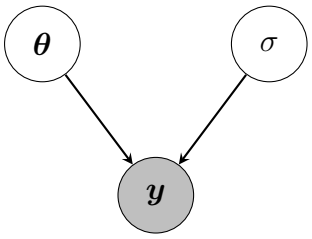

Figure 2.8: Parameter interaction and dependence under our simple statistical model for a single dataset $\boldsymbol{y}$. $\boldsymbol{\theta}$ is the set of mathematical model parameters. σ is the noise model standard deviation. All non-shaded variables are parameters for which we wish to infer probability distributions. $\boldsymbol{\theta}$ and σ have no inward-pointing arrows and therefore require prior distributions to be specified.

Suppose our mathematical model output is described by a function of its parameters, $\boldsymbol{f}(\boldsymbol{\theta})$, a vector of equal length to the vector of experimental data. In general, $\boldsymbol{f}$ is a vector containing the output of any mathematical model, which could be a system of differential equations, or a simple analytic expression, for example. $\boldsymbol{f}$ could be a time series, such as action potential data, or just the solution to the mathematical model at discrete points, such as ion channel screening data. As in the ion channel screening data case, $\boldsymbol{f}$ might also be a function of experimentally-controlled variables, separate from the mathematical parameters. For now, we will not explicitly denote this dependence, as we are focusing on the mathematical model parameters. In all cases considered in this thesis, the statistical model is that each data point y_i is independently Normally distributed according to

$$y_i \sim \mathcal{N}\left(f_i(\boldsymbol{\theta}), \sigma^2\right), \tag{2.16}$$

where f_i represents the i^{th} entry of the vector $\boldsymbol{f}$.

We can write down the probability density function of a Normal distribution, so we know that the likelihood of $\{\boldsymbol{\theta}, \sigma\}$ given a single data point y_i is given by

$$p(y_i|\boldsymbol{\theta}, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_i - f_i(\boldsymbol{\theta}))^2}{2\sigma^2}\right). \tag{2.17}$$

The likelihood of $\{\boldsymbol{\theta}, \sigma\}$ given the entire experimental dataset $\boldsymbol{y}$ is the product of Equation (2.17) over i, since we treat them as conditionally independent given the parameters:

$$L(\{\boldsymbol{\theta}, \sigma\}) := p(data|\boldsymbol{\theta}, \sigma) \tag{2.18}$$

$$= \prod_{i=1}^{N} p(y_i|\boldsymbol{\theta}, \sigma) \tag{2.19}$$

$$= (2\pi)^{-N/2}\sigma^{-N} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{N}(y_i - f_i(\boldsymbol{\theta}))^2\right). \tag{2.20}$$

We often find that with large N, we are multiplying lots of very small numbers together, which can sometimes result in a loss of precision. We therefore take the natural log of Equation (2.20) so that we instead deal with sums instead of products:

$$l(\{\boldsymbol{\theta}, \sigma\}) := \log(L(\{\boldsymbol{\theta}, \sigma\})) \tag{2.21}$$

$$= -\frac{N}{2}\log(2\pi) - N\log(\sigma) - \frac{1}{2\sigma^2}\sum_{i=1}^{N}(y_i - f_i(\boldsymbol{\theta}))^2 \tag{2.22}$$

$$\propto_+ -N\log(\sigma) - \frac{1}{2\sigma^2}\sum_{i=1}^{N}(y_i - f_i(\boldsymbol{\theta}))^2, \tag{2.23}$$

where $\propto_+$ means "equal up to the addition of a constant". For notational simplicity, in the future we will take equality in Equation (2.23). Throughout this thesis, whenever we are dealing with a likelihood, we will only ever be required to evaluate it up to a constant factor.

2.5.2 Bayes' rule

As discussed in Section 2.2.4, action potential models are typically published with single best-fit parameter values reported. This does not allow for the propagation of any uncertainties in the model inputs, and so there will be a single-value model prediction. This does not allow for as well-informed decision-making as possible in the drug safety decision-making process. In addressing the issue of uncertainty in model inputs and parameters, we will adopt a *Bayesian* approach. This means that we will treat the maximal ion current conductance parameters as random variables and infer probability distributions for them. We are only treating the

maximal conductance parameters as random variables because they are the most important when considering the effects of a pharmaceutical compound on cardiac electrophysiology. In theory, we could treat all model parameters in this way, but we believe that the conductance parameters are the most important for capturing action potential behaviour. We would not expect to gain much information from fitting time constant parameters to action potential traces, for example. Once we have inferred probability distributions for these parameters, *i.e.* likely sets of parameter values given the experimental data available to us, we will be able to propagate this uncertainty into predictions by sampling new parameter values from these distributions and computing the model output for each of these samples. This will construct a probability distribution for possible model outputs.

The Bayesian approach is not the only statistical methodology to account for uncertainty in predictions. A more classical approach is the *frequentist* approach where each observation is a sample from some population. It relies on repeatability of the experiments so that observations more accurately represent samples from a population. Bayesian, unlike frequentist, inference includes subjective belief in how the model parameters behave, and is typically better-suited to where there is a relatively small number of experimental realisations. See Robert (2007) for a more complete discussion.

We infer the joint probability distribution for the model parameters $\{\boldsymbol{\theta}, \sigma\}$ using *Bayes' rule*, where we absorb σ into $\boldsymbol{\theta}$ for notational simplicity:

$$\pi(\boldsymbol{\theta}) = p(\boldsymbol{\theta}|data) = \frac{p(data|\boldsymbol{\theta})p(\boldsymbol{\theta})}{\int_{\boldsymbol{\theta}} p(data|\boldsymbol{\theta})p(\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{\theta}}, \tag{2.24}$$

where $\pi(\boldsymbol{\theta})$ is the *posterior* (or "target") distribution for the model parameters $\boldsymbol{\theta}$, $p(data|\boldsymbol{\theta})$ is the *likelihood* of the parameter values in $\boldsymbol{\theta}$ given the observed data, and $p(\boldsymbol{\theta})$ is the *prior* distribution of $\boldsymbol{\theta}$. In practice, the prior distribution contains any existing knowledge or beliefs we have in the model parameters before observing the experimental data.

With many parameters, we cannot directly compute $\pi(\boldsymbol{\theta})$, and more generally we are unable to evaluate integrals with respect to $\pi(\boldsymbol{\theta})$, while such integrals are

the answer to any question one might ask a Bayesian statistician — see Robert (2007). We therefore use Markov chain Monte Carlo (MCMC, Robert & Casella, 2013) methods, which, in the simplest case, only require being able to evaluate $p(data|\boldsymbol{\theta})$ at any given parameter set $\boldsymbol{\theta}$. More generally, MCMC methods only require that we can evaluate $\pi(\boldsymbol{\theta})$ up to a factor of a constant. MCMC methods output a sequence/chain of parameter sets that form approximate independent samples from our target probability distribution $\pi(\boldsymbol{\theta})$.

Inferred probability distributions for model parameters incorporate some level of identifiability analysis. For example, if a marginal posterior distribution for a single parameter is very similar to the marginal prior distribution for that parameter, this suggests that there is very little information in the data about that particular parameter. There have been some studies relating inferred distributions to identifiability (Raue *et al.*, 2009; Siekmann *et al.*, 2012; Daly *et al.*, 2015).

2.5.3 Adaptive Metropolis-Hastings algorithm

In general, the posterior distribution defined in Equation (2.24) cannot be expressed analytically. The normalising integral in the denominator is also computationally intractable when considering more than a small number of parameters. In this section, we describe the sampling method we will use throughout the thesis to construct sequences of parameter values that will approximate independent samples from $\pi(\boldsymbol{\theta})$. We will use this sampling method to approximate $\pi(\boldsymbol{\theta})$ for model parameters in dose-response models of ion channel screening data in Chapter 3 and in action potential models in Chapters 4–6.

In 1953, Metropolis *et al.* published an algorithm to compute energies of interacting molecules, based on research performed during the development of the hydrogen bomb. Particles were arranged randomly, and the energy of the system computed. After a small uniform perturbation to a particles position, the new energy of the system was computed, and the new arrangement of particles was accepted with some probability based on the difference in energies. In 1970, Hastings generalised the Metropolis *et al.* (1953) method, describing a more general

acceptance probability, allowing for asymmetric perturbations from the current state, *i.e.* proposal distributions. In 2001, Haario *et al.* published an "adaptive Metropolis-Hastings" algorithm, which updates how new sets of parameter values are proposed. These proposals skew in size according to how many newly proposed states have been accepted, and in space according to where the most accepted states are. This method allows a faster convergence to the target distribution, since the proposal distribution should match with the target distribution after some time, so there are fewer 'wasted' proposals.

We use this adaptive Metropolis-Hastings algorithm (Haario *et al.*, 2001) to construct a sequence of sets of parameter values which approximates independent samples from the target distribution $\pi(\boldsymbol{\theta})$. Given the most-recently accepted set of parameter values $\boldsymbol{\theta}_{[t]}$, the algorithm proposes a set of parameter values $\boldsymbol{\theta}^*$ by sampling from a Normal distribution centred on $\boldsymbol{\theta}_{[t]}$, and accepts this proposed set into the sequence with probability α:

$$\alpha = \min\left\{\frac{\pi(\boldsymbol{\theta}^*)}{\pi(\boldsymbol{\theta}_{[t]})}, 1\right\} \tag{2.25}$$

$$= \min\left\{\frac{p(data|\boldsymbol{\theta}^*)p(\boldsymbol{\theta}^*)}{p(data|\boldsymbol{\theta}_{[t]})p(\boldsymbol{\theta}_{[t]})}, 1\right\}, \tag{2.26}$$

which follows from the definition of $\pi(\boldsymbol{\theta})$ given in Equation (2.24). We therefore do not have to be able to evaluate $\pi(\boldsymbol{\theta})$ pointwise, but rather we must be able to evaluate it up to a factor of a constant. We can do this if we can evaluate the likelihood and prior distributions at any particular set of parameter values. To be able to evaluate the likelihood, we need to specify our *statistical model*, which describes how we believe the experimental data is related to the mathematical model.

The likelihood is often a product of very small values, and so is prone to numerical difficulties. Because of this, we work with the natural log of the likelihood (and prior distribution) when computing the acceptance probability α as in Equation (2.26). We define the *log-likelihood* of the parameters $\boldsymbol{\theta}$:

$$l(\boldsymbol{\theta}) := \log(p(data|\boldsymbol{\theta})). \tag{2.27}$$

Equation (2.26) then becomes

$$\log(\alpha) = \min\left\{l(\boldsymbol{\theta}^*) + \log(p(\boldsymbol{\theta}^*)) - l(\boldsymbol{\theta}_{[t]}) - \log(p(\boldsymbol{\theta}_{[t]})), 0\right\}. \tag{2.28}$$

If there are any added constants in $\log(p(data|\boldsymbol{\theta}))$, we can ignore them in our definition of $l(\boldsymbol{\theta})$ because they will be cancelled out when taking differences in Equation (2.28).

The definition of the acceptance probability α given in Equation (2.26) means that a proposed set of parameter values will always be accepted into the chain if it has a greater (scaled) posterior density. If the proposed parameter values have a lesser posterior density, they may still be accepted into the chain, but with some probability of less than 1. This means that the algorithm will tend to move in directions of increasing posterior density values, and will usually find local maxima. In a complex parameter space, however, it is possible that there might be multiple local maxima with a global maximum far away in parameter space. The algorithm can get trapped in these local maxima without proposing parameter values far enough away to 'jump' to the global maximum. We therefore choose to perform an optimisation before starting our MCMC algorithm for exploring parameter space — we discuss this in greater detail in Chapter 4. While not theoretically necessary, it can save us a lot of computation time by sampling parameter space where there will be the most posterior probability mass, and to make sure that we have identified local maxima before running the MCMC algorithm so that we can check that the chain visits all of them often enough.

We generally choose the objective function to be the negative scaled posterior density function, or simply a sum of square differences function. This should give us a point estimate for the greatest log-target value. An MCMC algorithm with a Gaussian proposal distribution will work best if the mass is distributed only around this point (*i.e.* there are no other modes elsewhere). Since we cannot know for certain where the global maximum of $\pi(\boldsymbol{\theta})$ is, we run multiple optimisations from different starting positions. If several independent optimisation runs find only one point with the same best value, we expect that the target distribution will be uni-modal.

We use a minimisation algorithm called Covariance Matrix Adaptation – Evolution Strategy (CMA-ES, Hansen, 2006). The CMA-ES minimiser samples n points from a multivariate Normal distribution centred on some point. The likelihoods of these n points are computed, and a weighted average of the best few points is taken to be the new mean point. The covariance matrix is also updated to skew in this favourable direction. We treat CMA-ES as a 'black box' optimiser, using default values for most options. In our most recent work, we perform 32 initial CMA-ES optimisations from different starting positions. Although more initial optimisations provide stronger evidence that a particular point is indeed the global optimum, we choose 32 to make use of the 16 nodes on a single cluster of Oxford's Advanced Research Computing facilities. So far, we have found 32 to be large enough to provide good fits, and we have not encountered two distinct modes with similarly good fits.

Once we have obtained our estimate for the global maximum of the (scaled) target distribution $\pi(\boldsymbol{\theta})$, we run our adaptive Metropolis-Hastings algorithm, as described in Algorithm 1. If $\boldsymbol{\theta}$ is d-dimensional, we do not adapt the algorithm until after $1000d$ iterations, to allow a coarse search of the parameter space, just in case there is some better region of parameter space, before narrowing down the search to look more locally. However, the need for this is somewhat mitigated by the initial optimisation, since we do not expect there to be distinct regions of parameter space worth exploring. Adaptation is of the form of skewing the proposal Normal distribution's covariance matrix to align with directions of better (scaled) target distribution values.

We want our MCMC algorithm to have good mixing properties, *i.e.* we want it to efficiently explore the log-target surface around the maximum, rather than getting stuck. This efficiency is related to the proportion of states accepted by the chain with respect to all proposed states. Given Gaussian target and proposal distributions, 25% is considered a good average acceptance rate to ensure good mixing of the MCMC algorithm (Roberts *et al.*, 1997). Since there is no general result, however, we use 25% as a guideline for a good acceptance rate. This is done by adjusting how large the proposal distribution is, *i.e.* how far away it is likely to propose new points.

If too many proposed points are being rejected, the MCMC algorithm will explore the space closer and closer to the current point by shrinking the proposal distribution's covariance matrix. As the proposed points become closer and closer to the current point, the log-likelihood surface becomes approximately flat, and so more of these points are accepted. Conversely, if too many points are being accepted, the MCMC algorithm will propose points further away. This is achieved by scaling our proposal distribution by a factor, $a_{[t]}$, which is automatically adjusted after each iteration to ensure that the acceptance rate tends to 25%.

Algorithm 1 Metropolis-Hastings algorithm with adaptive proposal covariance matrix.

$\log(a_{[0]}) \leftarrow 0$
$\boldsymbol{\theta_{[0]}} \leftarrow$ initial optimisation
$\Sigma_{[0]} \leftarrow D$, where D is a diagonal matrix
$t \leftarrow 0$
if $t > 1000d$ **then**
  $\gamma_{[t]} \leftarrow (t - 1000d + 1)^{-0.6}$
else
  $\gamma_{[t]} \leftarrow 0$
end if
loop
  Given the current parameter state $\boldsymbol{\theta_t}$, sample $\boldsymbol{\theta^*} \sim \mathcal{N}(\cdot|\boldsymbol{\theta_{[t]}}, a_t\Sigma_{[t]})$
  if $p(\boldsymbol{\theta^*}) \neq 0$ **then**
    Compute $\log(\alpha) = \min\left\{l(\boldsymbol{\theta^*}) + \log(p(\boldsymbol{\theta^*})) - l(\boldsymbol{\theta_{[t]}}) - \log(p(\boldsymbol{\theta_{[t]}})), 0\right\}$
    Sample $u \sim \mathcal{U}(0, 1)$
    if $u < \alpha$ **then**
      $\boldsymbol{\theta_{[t+1]}} \leftarrow \boldsymbol{\theta^*}$
      $accepted \leftarrow 1$
    else
      $\boldsymbol{\theta_{[t+1]}} \leftarrow \boldsymbol{\theta_{[t]}}$
      $accepted \leftarrow 0$
    end if
  else
    $\boldsymbol{\theta_{[t+1]}} \leftarrow \boldsymbol{\theta_{[t]}}$
    $accepted \leftarrow 0$
  end if
  $\Sigma_{[t+1]} \leftarrow (1 - \gamma_{[t]}) \times \Sigma_{[t]} + \gamma_{[t]} \times (\boldsymbol{\theta_{[t+1]}} - \boldsymbol{\mu_{[t]}})^T(\boldsymbol{\theta_{[t+1]}} - \boldsymbol{\mu_{[t]}})$
  $\boldsymbol{\mu_{[t+1]}} \leftarrow (1 - \gamma_{[t]}) \times \boldsymbol{\mu_{[t]}} + \gamma_{[t]} \times \boldsymbol{\theta_{[t+1]}}$
  $\log(a_{[t+1]}) \leftarrow \log(a_{[t]}) + \gamma_{[t]} \times (accepted - 0.25)$
  t++
end loop

In practice, we perform *thinning* on the chain after having run the MCMC algorithm. This means saving iterations at regular intervals, for example every 5^{th}, and discarding the rest. This helps to reduce autocorrelation of the samples and better approximates independent samples from the target probability distribution. As a practical consideration, this also greatly reduces the file size of the output of the MCMC algorithm, which we save as a text file. We also discard a number of iterations from the start of the chain as a *burn-in*, as the algorithm will still be adapting as the acceptance fraction tends to 0.25 and the proposal covariance matrix adapts. During this adaptation period, the chain has not settled into its stationary distribution and so will not yet be approximating the target distribution well. The adaptation scaling factor $\gamma_{[t]}$ is 0 initially so that there is no adaptivity. When the adaptive steps begin, $\gamma_{[t]}$ is set to $2^{-0.6}$ and then tends to 0. We discard a burn-in large enough to "hide" the adaptivity; we want $\gamma_{[t]}$ to be close to 0 for all saved iterations. When dealing with a large number of parameters, we typically discard the burn-in before saving the remaining samples. However, for more manageable numbers of parameters, it can be useful to save the burn-in for analysis, as it can be easily discarded later.

2.5.4 Hierarchical statistical model

Biological systems have natural variability, and so repeated experiments with the same setup can produce different results. It is important to capture this variability as it can provide more information about how the system might behave in the future, instead of simply predicting 'average' behaviour. In addition to considering the effects of experimental noise within a single experiment, we will also consider inter-experiment variability while continuing to operate within a Bayesian framework. In Chapter 3, we will consider inter-experiment variability in ion channel screening data, where repeated doses of a pharmaceutical compound at the same concentration can be recorded as blocking an ion channel at varying levels. In Chapters 4–6, we will consider beat-to-beat variability in action potential

recordings, where action potentials are induced at a steady rate, but slightly different behaviour is recorded at each pace.

Given multiple datasets, we wish to infer probability distributions for the mathematical model parameters that might give rise to each dataset as well as inferring probability distributions for new parameters that govern how the mathematical model parameters vary between experiments. To achieve this, we use a *hierarchical* statistical model. The mathematical model parameters that give rise to the experimental data are now 'mid-level' parameters, which are conditioned on 'top-level' parameters above them in the statistical model. A schematic of this hierarchical statistical model is given in Figure 2.9. We will use hierarchical statistical models in Chapter 3 to account for inter-experiment variability in ion channel screening data, and in Chapters 4–5 to explore possible causes of beat-to-beat variability in action potential recordings.

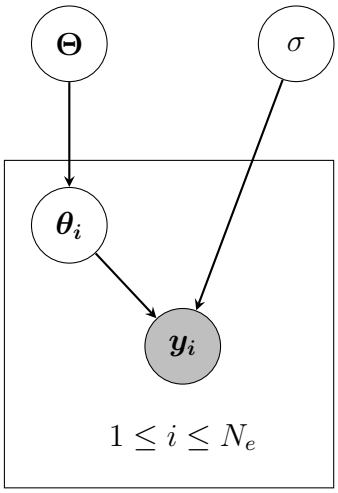

Figure 2.9: Parameter interaction and dependence under our hierarchical statistical model for multiple recordings from repeats of the same experiment. i indexes the individual experiments/datasets. $\boldsymbol{\theta_i}$ is the set of parameters which is used to generate $\boldsymbol{y_i}$, the dataset from experiment i. $\boldsymbol{\Theta}$ is the set of parameters that determines how the $\boldsymbol{\theta_i}$ are generated. All non-shaded variables are parameters for which we wish to infer probability distributions. Parameters with no inward-pointing arrows require prior distributions to be specified.

If the total number of parameters in the statistical model is not too large, we can use the same adaptive Metropolis-Hastings algorithm described in Section 2.5.3 to infer a joint probability distribution for all parameters given the experimental data. However, if we have a complex mathematical model such as an action potential

model with 15 maximal conductance parameters, the total number of parameters in the statistical model can become too large for the adaptive Metropolis-Hastings algorithm to comfortably handle. For example, with 3 datasets of action potential data, this would mean our statistical model has 46 + (dimension of $\boldsymbol{\Theta}$) parameters for which to infer probability distributions. We have observed that our adaptive Metropolis-Hastings algorithm struggles to mix well when using high-dimensional action potential models. This is likely due to the complex target and numerical issues, both introduced by solving the action potential models.

After inferring a joint distribution for a large number of parameters, we wish to predict what might happen in a future experiment, based on our observations from previous experiments. That is, we would like to generate $\boldsymbol{\theta}_{\boldsymbol{N_e+1}}$ based on whatever rule governs how the previous $\boldsymbol{\theta}_{\boldsymbol{i}}$ were generated. To do this, we construct a *posterior predictive* distribution which takes contributions from the sequence of parameter values obtained by the MCMC algorithm.

Posterior predictive distribution

We wish to use our inferred probability distributions to make predictions about how the system would behave if a future experiment were performed. To do this, we can consider the graphical model in Figure 2.9, but include an additional $\boldsymbol{\theta}_{\boldsymbol{N_e+1}}$ node that is dependent on $\boldsymbol{\Theta}$ but does not itself influence a dataset. Such an updated model is shown in Figure 2.10.

To construct the posterior predictive distribution, we use the model structure shown in Figure 2.10 to write

$$p(\boldsymbol{\theta}_{\boldsymbol{N_e+1}}|\text{rest}) = \int_{\boldsymbol{\Theta}} p(\boldsymbol{\theta}_{\boldsymbol{N_e+1}}|\boldsymbol{\Theta})p(\boldsymbol{\Theta}|\text{rest})\text{d}\boldsymbol{\Theta}. \tag{2.29}$$

We can construct an approximation to $p(\boldsymbol{\theta}_{\boldsymbol{N_e+1}}|\text{rest})$ in Equation (2.29) by summing the probability density functions of every sample of $\boldsymbol{\Theta}$ from the MCMC output, and then normalising so that it is indeed a valid probability density function. Supposing there are T samples in the chain after thinning and discarding

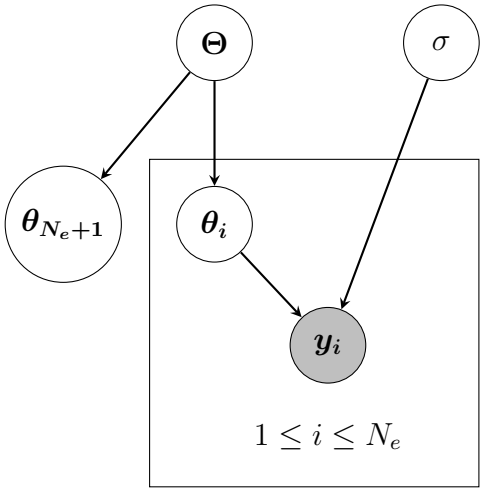

Figure 2.10: The statistical model shown in Figure 2.9 updated to include an additional $\boldsymbol{\theta}_{N_e+1}$ node. The posterior distribution for $\boldsymbol{\theta}_{N_e+1}$ conditioned on everything else provides a probability distribution describing how a future experiment will behave.

a burn-in, we have

$$p(\boldsymbol{\theta}_{N_e+1}|data) \approx \frac{1}{T}\sum_{t=1}^{T} f\left(\boldsymbol{\theta}_{N_e+1}|\boldsymbol{\Theta}_{[t]}\right), \tag{2.30}$$

where $f(\boldsymbol{\theta}_{N_e+1}|\boldsymbol{\Theta}_{[t]})$ is the probability distribution of $\boldsymbol{\theta}_{N_e+1}$ given $\boldsymbol{\Theta}_{[t]}$. In general, we cannot sample from this probability distribution directly. However, in one dimension, we can use the inverse-cumulative distribution function (CDF) method by performing the same addition and normalisation as when constructing posterior predictive distribution, but adding CDFs instead of PDFs. Using the inverse-CDF method, we can then approximately sample from $p(\boldsymbol{\theta}_{N_e+1}|data)$ to predict parameter values that will govern how the system behaves in a future experiment.

The inverse-CDF method is not possible in greater-than-one dimensions. Alternatively, we could include the $\boldsymbol{\theta}_{N_e+1}$ node when running our MCMC algorithm and infer a distribution for $\boldsymbol{\theta}_{N_e+1}$ along with the other parameters shown in Figure 2.10.

We can take multiple samples from $p(\boldsymbol{\theta}_{N_e+1}|data)$ and use these as inputs into the mathematical model. These inputs will produce a range of outputs, which correspond to the possible outcomes of future experiments, where these predictions have been informed by previous experiments.

2.5.5 Model selection with Bayes factors

In statistical modelling, model selection is important. That is, we wish to compare the evidence for one statistical model against the evidence for another. For example, two models of a system may be presented, one of which is more mechanistically complex. However, if there are issues of non-identifiability in the more complex model, there may not be any advantage in using it instead of the simpler model. We will apply the Bayesian method of model selection to ion channel screening data to compare two simple models and discuss whether a particular simplification is justified.

In a Bayesian framework, *Bayes factors* are used to compare two different models. A Bayes factor is a single number that compares the relative evidence for two different models. They favour models that provide good fits to the data but also penalise overly complex models, *i.e.* models with more parameters. We will apply Bayes factors to model selection when fitting to ion channel screening data in Chapter 3.

We wish to consider the posterior probability of a model $\mathcal{M}_i$ given some observed data $\boldsymbol{y}$. Equations (2.31) – (2.34) are standard definitions, see Kass & Raftery (1995),

$$p(\mathcal{M}_i|\boldsymbol{y}) = \frac{p(\boldsymbol{y}|\mathcal{M}_i)p(\mathcal{M}_i)}{\sum_i p(\boldsymbol{y}|\mathcal{M}_i)p(\mathcal{M}_i)}. \tag{2.31}$$

To compare two models, we compute the *posterior odds* as a ratio of the two models' posterior probabilities:

$$\frac{p(\mathcal{M}_2|\boldsymbol{y})}{p(\mathcal{M}_1|\boldsymbol{y})} = \frac{p(\boldsymbol{y}|\mathcal{M}_2)p(\mathcal{M}_2)}{p(\boldsymbol{y}|\mathcal{M}_1)p(\mathcal{M}_1)}, \tag{2.32}$$

where we have the Bayes factor

$$B_{21} = \frac{p(\boldsymbol{y}|\mathcal{M}_2)}{p(\boldsymbol{y}|\mathcal{M}_1)}, \tag{2.33}$$

which compares what the data tells us about the models, and the *prior odds*

$$\frac{p(\mathcal{M}_2)}{p(\mathcal{M}_1)}, \tag{2.34}$$

which incorporates our prior beliefs of the models.

We drop the $\mathcal{M}_i$-dependence from the notation for simplicity. We wish to approximate $p(\boldsymbol{y})$ from the numerator in Equation (2.31), also known as the *marginal likelihood*, for two models to compute their Bayes factor.

Kass & Raftery summarised how to interpret the Bayes factor B_{21} in Table 2.1.

B_{21}	Evidence against $\mathcal{M}_1$
1 – 3	Not worth more than a bare mention
3 – 10	Substantial
10 – 100	Strong
> 100	Decisive

Table 2.1: How to interpret the Bayes factor B_{21} in terms of evidence in favour of $\mathcal{M}_2$ against $\mathcal{M}_1$. Taken from Kass & Raftery (1995).

We can use our existing MCMC code to approximate $p(\boldsymbol{y})$ using *thermodynamic integration* (Calderhead & Girolami, 2009). This requires us to run our MCMC algorithm multiple times with slightly different algorithm parameter values, and so the overall computational cost is much greater than that of inferring a posterior distribution for model parameters. For more details on computing $p(\boldsymbol{y})$, see Appendix A.

Having now described the background biology and established the statistical framework that will be used in the rest of this thesis, in the next chapter we describe our first application of these methods in ion channel screening dose-response data. In the subsequent chapters, we discuss the application of these methods to action potential recordings.

3

Bayesian inference for ion channel screening dose-response data

The majority of this chapter is based on our published paper, Johnstone *et al.* (2016a), with some updates to Section 3.3, where we have used a slightly different likelihood model to account for some data being manually capped at 0 and 100, and we have used the same prior distribution for σ as we do in Section 3.5. We have also included some discussion on model selection in the Bayesian framework in Section 3.4, which was not in the published paper.

3.1 Introduction

As discussed in Chapter 2, the *Comprehensive in vitro Proarrhythmia Assay* initiative (CiPA, Sager *et al.*, 2014) proposes that pharmaceutical compounds be tested on up to seven ion channels that both strongly influence ventricular repolarisation and are frequently blocked by pharmaceutical compounds. This process is known as "ion channel screening". A particular ion current is measured by over-expressing the gene that encodes the ion channel of interest in an immortal cell line such as human embryonic kidney cells. Effectively, the cell membrane then only has these ion channels and any current measured can be attributed to the ion current of interest. The proposal is for ion channel screening data to be passed into an *in-silico* human

ventricular action potential model to see if the compound produces proarrhythmic behaviour, or indicators of such behaviour, at the whole-cell level.

Levels of block are measured by first measuring the ion current of interest in a control setting with no drug applied. This current is the baseline against which future experiments will be compared. After adding some concentration of a drug, the ion current is measured again, and the level of block is the difference between the new current and the baseline current as a percentage of the baseline current. It is possible to measure levels of block below 0% if, for example, the baseline current was measured to be less than its true value. If a drug that has no effect on this ion current is then applied and the current measured again, the true baseline value might be recorded. This gives the impression that the current has increased after application of the drug, and hence the drug appears to cause a negative level of block. Alternatively, drugs that increase an ion current are known as *agonists*, however none of the drugs considered here are thought to be agonists.

"Dose-response" (or "concentration-effect") curve-fitting is generally performed to describe how increasing compound concentration provokes a response in a process. "Dose-response" generally relates to *in-vivo* experiments where a drug dose is administered but the concentration at the relevant site is not precisely known, whereas 'concentration-effect' generally relates to *in-vitro* experiments where the concentration is accurately applied. We will refer to both as 'dose-response' in this chapter for simplicity. Here, the 'response' is binding and blocking of ion channels, measured via inhibition of ion currents.

In this chapter, we discuss a methodology to infer distributions of possible dose-response curves from experimental data using the methods introduced in Section 2.5. When making predictions of block at a given concentration, these distributions of possible dose-response curves provide us with a probability distribution for the level of block at that concentration. To illustrate the consequences of this, we use these distributions of block (for multiple ion channels) inferred from data provided in Crumb *et al.* (2016) as inputs into an *in-silico* action potential model. We then run forward simulations to predict a distribution of outputs — in this case action

potential durations given application of a compound at a particular concentration. We show that given the limited number of repeats of ion-channel experiments, there are wide ranges of predicted action potentials, with overlapping results for different compounds with different associated proarrhythmic risks.

While all biological measurements vary across experiments, there are sometimes intra-experiment trends that suggest that each experiment must be treated separately, rather than treating all experimental measurements the same. To account for this, we consider a more complex hierarchical statistical model to account for inter-experiment variability in ion channel screening data. We discuss how this approach can be used to make predictions about how a drug would interact with an ion channel in a future experiment, and how this approach can untangle some of the inter-experiment behaviour and give a better idea of the 'true' underlying behaviour.

Finally, we use the standard dose-response model, and a simplified version, as a test case for model selection in a Bayesian framework. We compute Bayes factors, comparing these two models against each other given the available data, and discuss when we can use a simplified model. This is important to consider as a more-complex model might not provide a sufficient improvement in the quality of the fit to justify the inclusion of additional parameters.

3.1.1 Dose-response curves

Dose-response curves are summarised by two parameters: an Inhibitory Concentration 50% (IC_{50}) value — that is the concentration of the compound that gives 50% of the maximum effect; and a Hill coefficient which sets the 'steepness' of the curve as it passes the IC_{50}. Examples of dose-response data and a fitted curve are given in Figure 3.1.

The equation for a dose-response curve was proposed by Hill (1910), and consequently another name for the curve is a *Hill curve* (Weiss, 1997). Suppose we have an ion channel-blocking compound C, and (unblocked) receptor membrane proteins R. If n molecules of C are required to block one unit of R, they form a

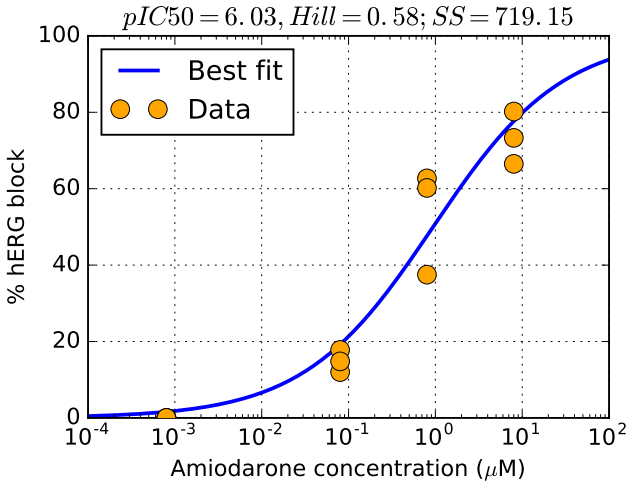

Figure 3.1: A dose-response curve fitted to experimental data for hERG block by Amiodarone from the Crumb *et al.* (2016) dataset. A dose-response curve is shown fitted to all data points at once. The sum of square differences was minimised. The final parameter and objective function values are shown at the top of the graph for easier comparison when plotting the results from multiple datasets.

complex S, equivalent to being a blocked channel. Suppose they bind at rate k_1 and dissociate at rate k_2. The rate equation describing this binding is

$$nC + R \underset{k_2}{\overset{k_1}{\rightleftharpoons}} S. \tag{3.1}$$

We use the law of mass action to write down the rate of change of S with respect to time:

$$\frac{\mathrm{d}S}{\mathrm{d}t} = -k_2 S + k_1 C^n R \tag{3.2}$$

$$= -k_2 S + k_1 C^n (R_0 - S), \tag{3.3}$$

noting that $R + S = R_0$, a constant. At steady state ($\mathrm{d}S/\mathrm{d}t = 0$) we have

$$S = \frac{k_1 C^n R_0}{k_2 + k_1 C^n} \tag{3.4}$$

$$= \frac{R_0}{1 + \left(\frac{K}{C}\right)^n} \tag{3.5}$$

where $K = (k_2/k_1)^{1/n}$. If we let s be the proportion of membrane proteins that are blocked, *i.e.* $s = S/R_0$, we have

$$s = \frac{1}{1 + \left(\frac{K}{C}\right)^n}. \tag{3.6}$$

We then have that n is the Hill coefficient, and K is the IC_{50}. From this derivation, we expect that n should be an integer, but non-integer values often provide the best fit to experimental data. We might also expect non-integer exponents from a more general form of the equation,

$$nC + mR \underset{k_2}{\overset{k_1}{\rightleftharpoons}} mS, \tag{3.7}$$

which could happen if channels co-localise and a molecule of C could affect more than one channel at once. We therefore allow the Hill coefficient to take any positive value.

If we let x be the concentration of a compound, we describe the effect of the compound by

$$\text{response} = f_{DR}(x; IC_{50}, Hill) := \frac{100}{1 + \left(\frac{IC_{50}}{x}\right)^{Hill}}, \tag{3.8}$$

where IC_{50} and $Hill$ are parameters that take positive values. In our motivating example this response will be "% block" of a particular type of ion channel.

3.1.2 Standard fitting procedure

A Hill curve is often fitted to all data points simultaneously, to obtain 'average' IC_{50} values and Hill coefficients for a particular curve. This gives the most likely set of parameter values. For example, Crumb *et al.* (2016) recently published dose-response screening data for 30 compounds on 7 different ion channels, along with best-fit IC_{50} values and Hill coefficients. As discussed in Chapter 2, taking this approach means that there is no associated probability given to these IC_{50} and Hill values; different possible ranges for these parameter values are not considered. The usual fitting procedure can also give rise to models which differ in behaviour from each individual experiment, as shown by Pathmanathan *et al.* (2015b) in the case of inactivation of the fast sodium current in action potential models, and as we will show in case studies below.

3.2 Bayesian statistical modelling approach

To explore and characterise the uncertainty in the dose-response measurements published by Crumb *et al.*, and to propagate these uncertainties into model predictions, we use a Bayesian statistical framework as introduced in Section 2.5.1 of Chapter 2. This Bayesian framework allows us to consider different possible dose-response curves that may have given rise to these data. Instead of computing point-estimates for the IC_{50} value and Hill coefficient, we infer probability distributions of these parameters ($\boldsymbol{\theta}$ in Figure 2.8), as well as a distribution for the possible observational noise (σ in Figure 2.8). This provides us with a method for propagating uncertainty in experimental data into simulations by drawing parameters from these inferred distributions (Equation (2.24)), and using these samples as simulation inputs.

We consider two different types of Bayesian statistical models: one where all data points are treated equally, as though they were obtained from the same experiment (described in Section 2.5.1 of Chapter 2), and another where we believe that each repeat of an experiment has distinct properties (through some source of inter-experiment variability) and therefore its own set of parameters to infer (described in Section 2.5.4 of Chapter 2). The first case we will refer to as "single-level", and the second case as "hierarchical". The single-level case does not consider extrinsic variability, defined in Section 2.3, since we are assuming that all data points are generated by the same behaviour with only observational noise — see Figure 2.8. The hierarchical case does consider extrinsic variability, which we model by assuming that each experimental dataset was generated according to its own IC_{50} value and Hill coefficient, which may vary across experiments, as well as a consistent level of observational noise across experiments — see Figure 2.9.

Instead of working with IC_{50}, we work with pIC_{50}, where

$$pIC_{50}\ [\log(\text{Molar})] = 6 - \log_{10}(IC_{50}\ [\mu\text{Molar}]), \tag{3.9}$$

with square brackets indicating units. This transformation makes it much easier for fitting algorithms to explore the parameter space, as linear variation in IC_{50}s does

not result in linear changes to dose-response curves, which are commonly plotted on log scales. The dose-response model we therefore work with in practice is

$$\% \text{ channel block} = f(x; pIC_{50}, Hill) := f_{DR}\left(x; 10^{(6-pIC_{50})}, Hill\right). \tag{3.10}$$

We assume that the underlying behaviour is described by the dose-response model, f, given by Equation (3.10). We also assume that an experimental observation is Normally distributed around some underlying behaviour with some standard deviation, σ (that has the same units as the measured response). That is, given an applied compound concentration, x, our statistical model is that the response, y, is a Normally-distributed random variable with mean $f(x; pIC_{50}, Hill)$ and standard deviation σ, that is:

$$y \sim \mathcal{N}\left(f(x; pIC_{50}, Hill), \sigma^2\right). \tag{3.11}$$

When we have noisy data, different sets of parameters might allow us to fit the equation to the experimental data equally well, *i.e.* the parameters are not practically identifiable. In our Bayesian framework, we treat these model parameters as random variables, in part due to the uncertainty introduced through observational error and any parameter identifiability problems. We therefore want to infer a probability distribution, instead of point-estimates, for the parameters $pIC_{50}, Hill$, and σ. Given the experimental data, we wish to infer the posterior distribution $\pi(pIC_{50}, Hill, \sigma)$, as defined in Equation (2.24) in Chapter 2.

3.3 Single-level model

In the example shown in Figure 3.1, each experiment consisted of applying one or more concentrations of a compound to a cell and measuring the degree of block of an ion current. There were multiple recordings, leading to multiple response data points at each concentration.

3.3.1 Methods

In this statistical model, we will assume that there is no inter-experiment variability, so all data points are considered to be from one experiment, and all the data points are generated using the same set of parameter values. Under this model, the data for hERG block by Amiodarone, for example, is generated by the process shown schematically in Figure 3.2, adapted from Figure 2.8.

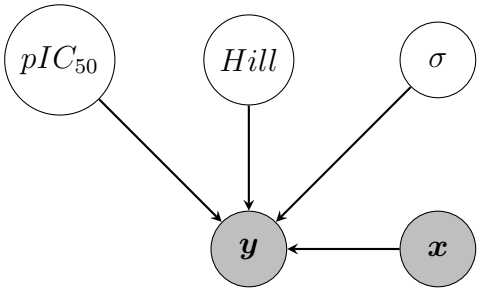

Figure 3.2: The statistical model for the generation of dose-response data. All non-shaded variables are parameters for which we wish to infer probability distributions. $\boldsymbol{x} = \{x^{(1)}, \ldots, x^{(K)}\}$ is the vector of all experimentally-applied drug concentrations, and $\boldsymbol{y} = \{y^{(1)}, \ldots, y^{(K)}\}$ is the vector of all experimentally-recorded response data points.

Under this statistical model, all data points $y^{(j)}$ are independently distributed according to

$$y^{(j)} \sim \mathcal{N}\left(f(x^{(j)}; pIC_{50}, Hill), \sigma^2\right), \tag{3.12}$$

where $x^{(j)}$ is the applied compound concentration, and $j = 1, \ldots, K$, where K is the total number of data points. The likelihood of the set of parameters $pIC_{50}, Hill$, and σ, given a single data point $y^{(j)}$ is then

$$p(y^{(j)}|pIC_{50}, Hill, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\left(y^{(j)} - f(x^{(j)}; pIC_{50}, Hill)\right)^2}{2\sigma^2}\right). \tag{3.13}$$

In practice, we work with the log of the target distribution, and therefore the log of the likelihood given in Equation (3.13).

We have not considered possible uncertainty in the drug concentrations $\boldsymbol{x}$. This could be addressed by letting $\boldsymbol{x} = \hat{\boldsymbol{x}} + \boldsymbol{\delta}$, where $\hat{\boldsymbol{x}}$ are the desired experimentally-applied concentrations and $\boldsymbol{\delta}$ are random variables representing possible errors in

the concentrations. This would mean that now $\boldsymbol{x}$ are random variables for which we would infer probability distributions by inferring probability distributions for $\boldsymbol{\delta}$. However, we expect that these input concentrations are much easier to measure and control than the resulting levels of current block, and so we treat such input uncertainty as negligible and use fixed values of $\boldsymbol{x}$.

In the data published by Crumb *et al.*, any observations (with one exception) below 0% or above 100% were capped to 0% or 100%, respectively (personal communication), because these extreme values are assumed to be due to observational error. It would be better not to filter the data in this way, since repeated zero entries can lead to the erroneous conclusion that there is no noise σ on the data. A single negative response (a level of block recorded at less than 0%) was given in the published data, so in the following work we set this value to 0, and 'forget' the reported value, so that all reported data points can be treated the same. In this framework, if we encounter a 0 or a 100 in the data, the true recorded value could have been any value less than 0 or greater than 100, respectively. To take these extreme possibilities into account, we truncate the Normal distribution in Equation (3.12) at 0 and 100, and assign the excluded probability mass to their respective endpoints. We therefore have the likelihood of $\{pIC_{50}, Hill, \sigma\}$ given a single data point y_i:

$$p(y_i|pIC_{50}, Hill, \sigma) = \begin{cases} G_{\mathcal{N}}\left(0; f(x_i; pIC_{50}, Hill), \sigma^2\right), & y_i = 0 \\ g_{\mathcal{N}}\left(y_i; f(x_i; pIC_{50}, Hill), \sigma^2\right), & 0 < y_i < 100 \\ 1 - G_{\mathcal{N}}\left(100; f(x_i; pIC_{50}, Hill), \sigma^2\right), & y_i = 100 \end{cases} \tag{3.14}$$

where $g_{\mathcal{N}}(\,\cdot\,; \mu, \omega^2)$ is the Normal probability density function with mean μ and standard deviation ω, $G_{\mathcal{N}}(\,\cdot\,; \mu, \omega^2)$ is the corresponding Normal cumulative distribution function, and f is the dose-response function defined in Equation (3.10).

Since $pIC_{50}, Hill$, and σ are parameters that we infer, we need to specify a prior distribution across them, corresponding to $p(\boldsymbol{\theta})$ in Equation (2.24). We choose independent prior distributions for each parameter: a uniform distribution for the Hill coefficient, and a Gamma distribution the observation noise standard deviation σ, and a negatively-shifted exponential distribution for the pIC_{50}, as illustrated

in Figure 3.4. Where available, but we could choose a more informative prior based on previous ion channel screening data.

We allow $Hill$ to take values in (0,10). The Hill coefficient must be positive and describes the steepness of the dose-response curve, so after a certain point, increasing the Hill coefficient does not make a noticeable difference to the curve, so we choose 10 as a generous upper bound, above any biologically-plausible drug-binding we are aware of. Similarly, a compound that has no measurable effect could be thought of as having a very large IC_{50}, and it makes no difference practically to model it as having an even larger IC_{50}. This corresponds to a negative pIC_{50} value, and so we choose to allow pIC_{50} to be greater than -3 as values outside this interval will not have much effect (see Figure 3.3).

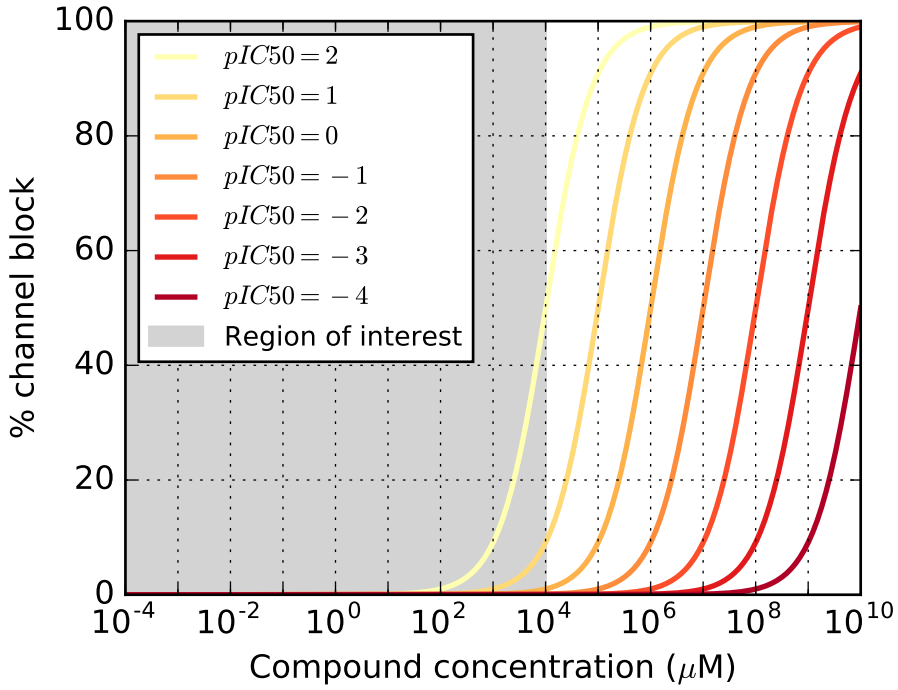

Figure 3.3: Large IC_{50} values are indistinguishable when they are orders of magnitude above the relevant concentration range. Here we show the effect of decreasing pIC_{50} (increasing IC_{50}), while maintaining $Hill = 1$. The shaded "region of interest" covers the minimum and maximum concentrations in the data published by Crumb *et al.* As the pIC_{50} value decreases, there is no significant change to the dose-response curve across the relevant range of concentrations — all predictions are close to zero response.

σ is a standard deviation parameter and therefore also positive. We therefore

choose a Gamma distribution, which puts probability mass only on positive values, and is also shifted so that it can take values only greater than 10^{-3}. We give σ a prior lower bound of 10^{-3} because we have run into numerical problems when all measured responses were 0, and because we do not believe that we have perfect measuring capabilities, and so there will never be no observational noise. We choose the hyperparameters governing the Gamma distribution so that the mode is on 6, and so that most of the probability mass is around 3–10%, a reasonable range of observation error that we might expect. We choose an exponential prior distribution for pIC_{50} starting from -3 so that there has to be sufficient evidence in the data that the drug is indeed having an effect, otherwise posterior probability mass will stack up against this lower bound. These prior distributions are plotted in Figure 3.4.

As described in Section 2.5.3, we first perform a Covariance Matrix Adaptation Evolution Strategy optimisation (CMA-ES Hansen, 2016) to find an optimal starting point for exploring possible parameter sets. We then use an adaptive Metropolis-Hastings MCMC algorithm (Haario *et al.*, 2001) to infer an approximation to $\pi(pIC_{50}, Hill, \sigma)$. After constructing the sequence of samples, we plot normalised marginal histograms to approximate the marginal posterior distributions.

The posterior distribution in the single-level model is given by (up to a factor of a constant)

$$p(pIC_{50}, Hill, \sigma|\boldsymbol{y}) \propto \left(\prod_{j=1}^{K} p(y^{(j)}|pIC_{50}, Hill, \sigma) \right) p(pIC_{50})p(Hill)p(\sigma), \quad (3.15)$$

where the first term on the right-hand-side is given by Equation (3.13).

3.3.2 Results

In Figure 3.5 we plot normalised marginal and pairwise histograms for the values of the parameters for each sample of the MCMC algorithm output, in the case of Amiodarone applied to the hERG channel. These histograms are approximate projections of the posterior distribution across these parameters. The spread in each distribution corresponds to the uncertainty in that parameter; if a parameter's

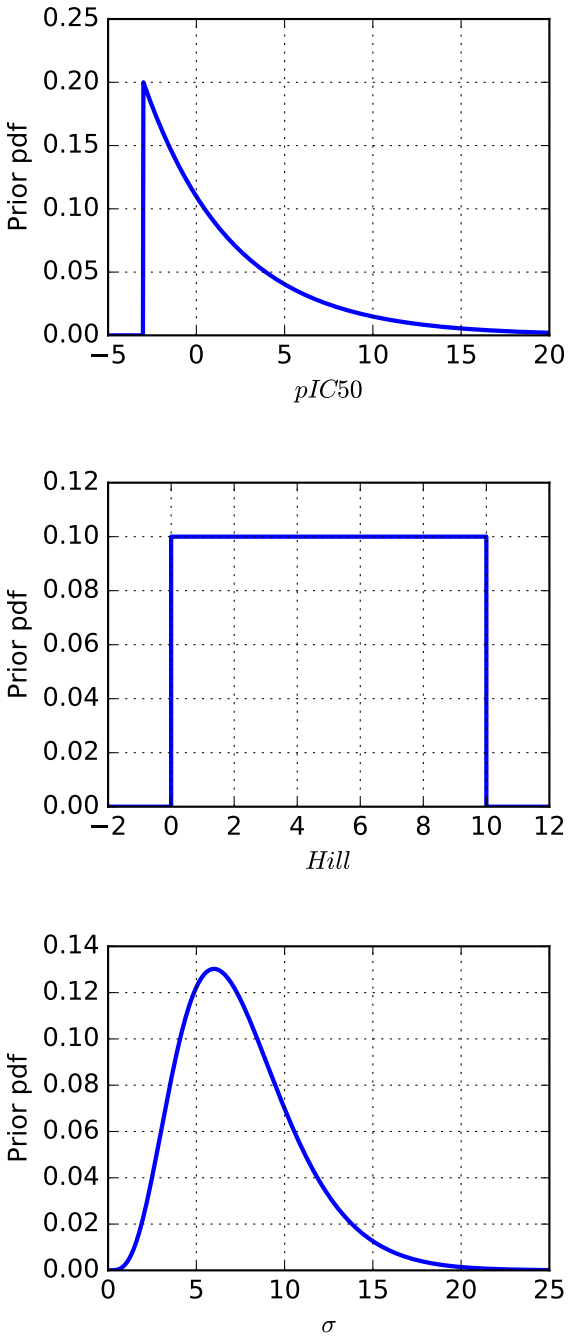

Figure 3.4: Prior probability distributions. Top: exponential shifted down to -3 with rate parameter 0.2. Middle: uniform between 0 and 10. Bottom: Gamma distribution shifted positively by 10^{-3}, with shape parameter 5 and scale parameter approximately 1.5 so that the mode is at 6

marginal posterior distribution is narrower, we are more certain about its value from the observed data.

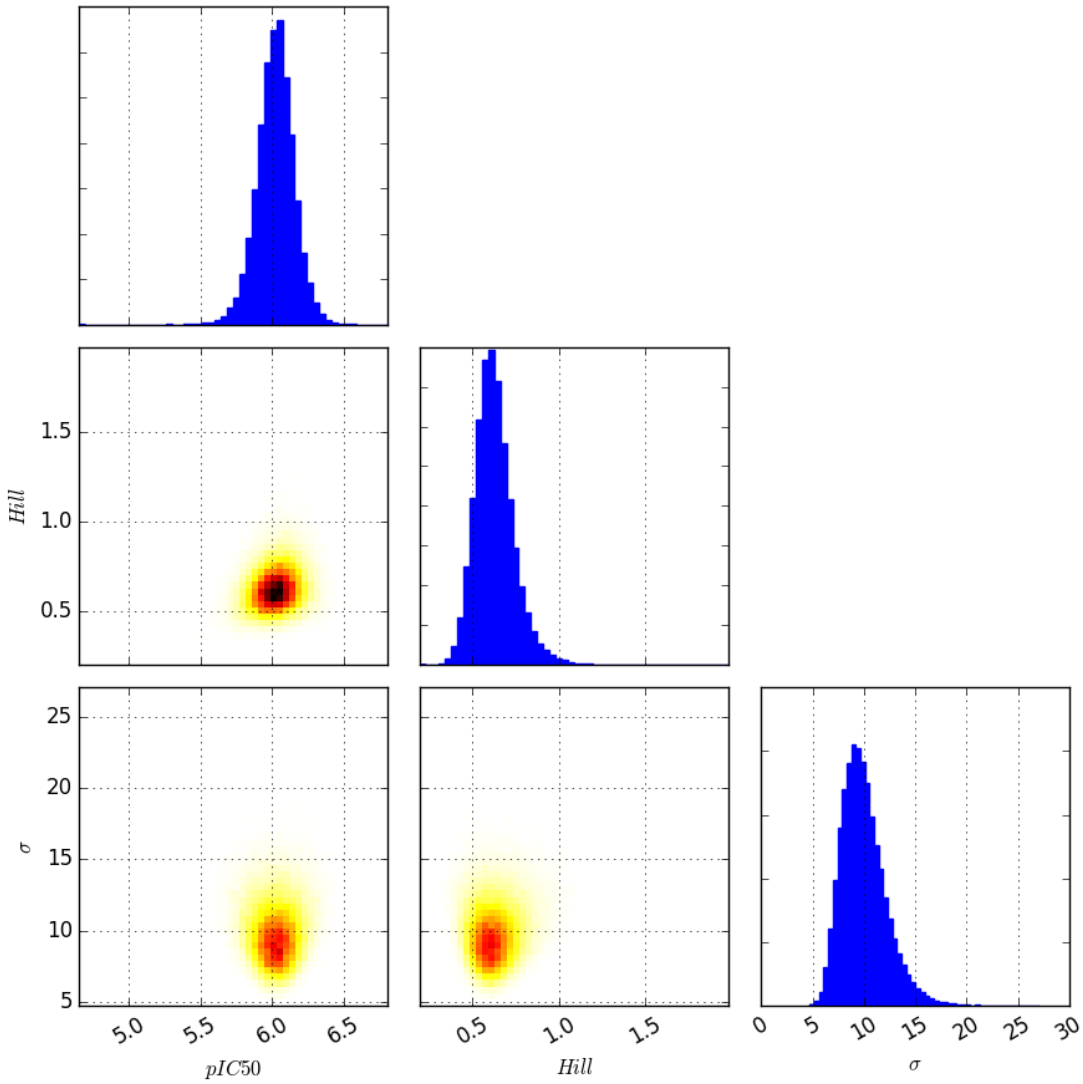

Figure 3.5: Matrix plot of normalised marginal and pairwise marginal histograms of the MCMC algorithm output samples for each parameter, in the Amiodarone applied to hERG example.

Before propagating these uncertainties, we first draw $(pIC_{50}, Hill)$ samples from the MCMC output, and plot dose-response curves with these parameter values. Examples are given in Figure 3.6, where Amiodarone has a measurable blocking effect on hERG, but no measurable effect on Kir2.1. As we take more samples, the plotted curves build up a distribution of possible dose-response curves given the experimental data. For each compound concentration, we then have a range

of possible responses with their relative probability densities being given by the density of dose-response curves at that concentration.

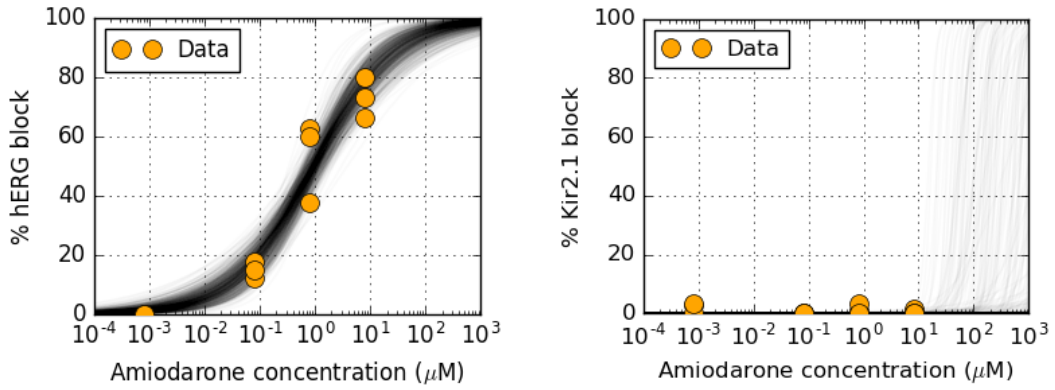

Figure 3.6: Inferred dose-response curves from MCMC $(pIC_{50}, Hill)$ samples, each plotted with near-transparency. Individual curves are equally likely, but regions with a greater density of curves appear darker and correspond to a greater probability density. Left: Amiodarone with measurable effect on hERG. Right: Amiodarone with no measurable effect on Kir2.1. Outside of the range of measured concentrations, the MCMC algorithm was unable to find narrow ranges for possible parameter values, because the experimental data does not contain enough information.

We can compare the marginal histograms of Amiodarone applied to hERG (Figure 3.5) with those of Amiodarone applied to Kir2.1 (Figure 3.7). There is no information about pIC_{50} or $Hill$ in the Kir2.1 case; all parameter values are equally likely for any curves that start increasing after the last data point. This is why the curves in the right-hand plot of Figure 3.6 appear zero where the data points are, and uniform after. There is a lower and smaller range for σ, which is not surprising since noise is only introduced around 0, where we are confident there is no response happening.

3.4 Model selection using Bayes factors

Model selection is important when modelling experimental data, as some models will be able to fit to the data better, but might be unphysical or unjustified in their formulation. For example, a polynomial with arbitrarily high degree will be able to fit to some data points, but this model would have no physical justification.

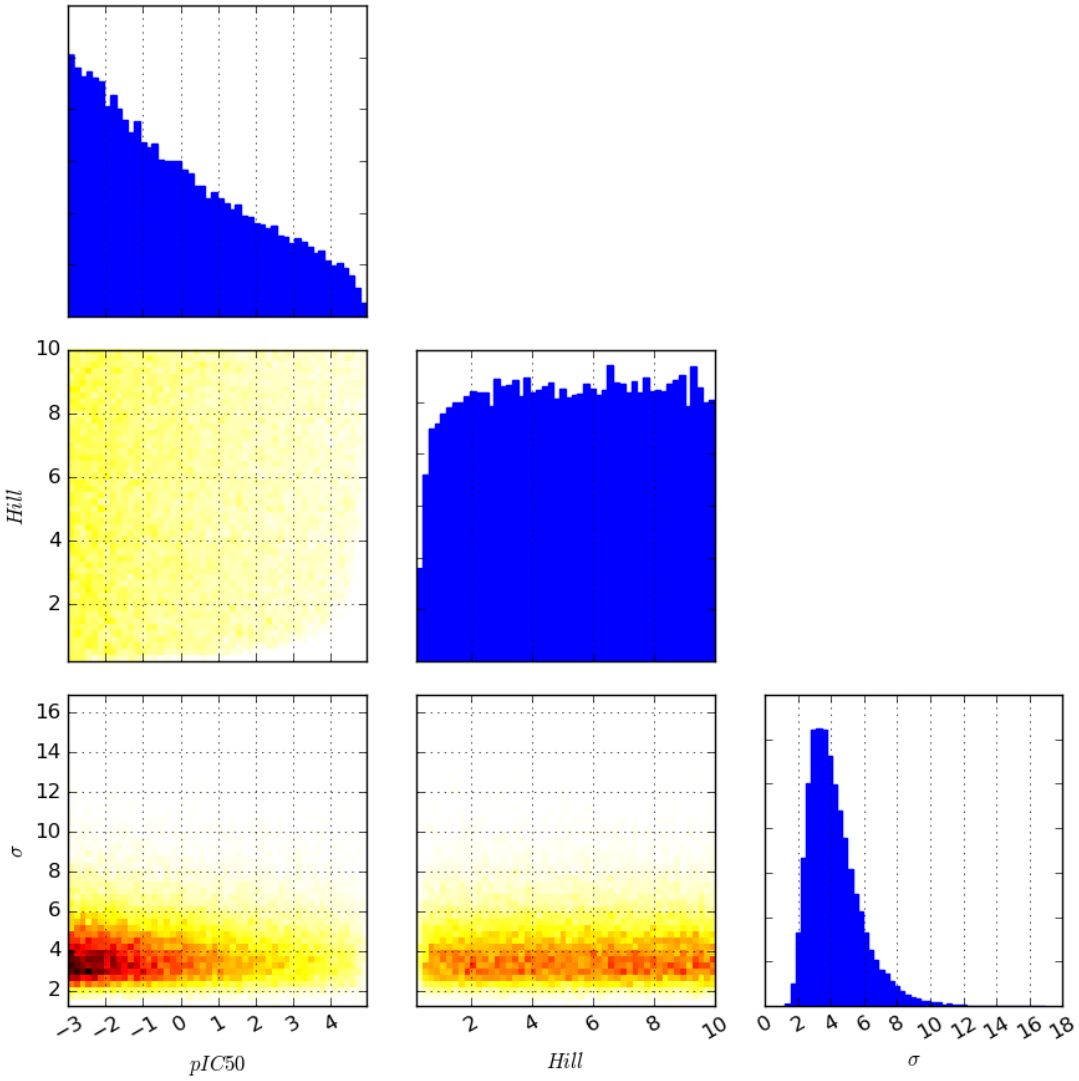

Figure 3.7: Matrix plot of normalised marginal and pairwise marginal histograms of the MCMC algorithm output samples for each parameter, in the Amiodarone applied to Kir2.1 example. There is no measurable response to the drug, and so there is very little information about pIC_{50} and $Hill$ to inform the inferred posterior distributions, and so they resemble their respective prior distributions.

In general, we might expect a model with more parameters, so more degrees of freedom, might be able to fit to experimental data better. But the increase in goodness of fit might not be enough to justify an increased model complexity. Increased model complexity can lead to over-fitting to the data which will result in poor predictions in the general case.

In the case of ion channel screening data, the simplest model would be to fix the Hill coefficient equal to 1 in Equation (3.10). This is not unreasonable as it corresponds to a single drug molecule blocking a single ion channel, *i.e.* $n = 1$ in Equation (3.1). $Hill$ would no longer be a parameter to fit, or infer a distribution for, and we would only have pIC_{50} remaining, as well as the noise standard deviation σ. In general, the best-fit parameter values will not give a Hill coefficient of 1, so we need a method to decide whether it is worth using the simpler model ($Hill = 1$) or a more complex model ($Hill$ can vary). In this section, we use Bayes factors, described in Section 2.5.5, to perform such model selection in a Bayesian framework. We compare two single-level models, of the form introduced in Section 3.3.

We model % channel block using $f(x; pIC_{50}, Hill)$ as defined in Equation (3.10). Often, when fitting ion channel screening data, $Hill$ is set to 1, and only IC_{50} (or pIC_{50}) is fitted to the data. We will consider these two models: "Model 1" — fixing $Hill = 1$ and allowing pIC_{50} to vary, and "Model 2" — allowing pIC_{50} and $Hill$ to vary.

As in Equation (3.14), we assume that the measured levels of block at concentration x are Normally distributed around $f(x; pIC_{50}, Hill)$ (but capped at 0 or 100) with standard deviation σ, which we also treat as a random variable to infer with our MCMC algorithm. We keep the same prior distributions for pIC_{50} and σ in Model 1, and additionally $Hill$ in Model 2. In Model 1, we fix $Hill = 1$. These prior distributions are plotted in Figure 3.4.

After running our adaptive Metropolis-Hastings MCMC algorithm (Algorithm 1 in Chapter 2), we take a subset of samples from the chains and plot the dose-response curves with these parameter values. These curves build up a distribution of possible levels of block that are actually occurring when we observe the experimental data.

3.4.1 Results

We have computed the Bayes factors, introduced in Section 2.5.5, comparing Models 1 and 2 for all drugs and ion channels in Crumb *et al.* (2016). Given n and c as defined in Appendix A, we take $n = 40$ and $c = 3$ to define temperatures as in Equation (A.15). We therefore have to run our MCMC algorithm (Algorithm 1) 41×2 times per drug and ion channel combination. These can be run in parallel.

Here we discuss a few particular cases to illustrate the main scenarios:

1. Model 2 is selected over Model 1,
2. Model 1 is selected over Model 2,
3. Neither model is selected — there is no measurable drug effect,
4. Neither model is selected — there is some measurable drug effect.

For each drug and ion channel combination, we take samples from the MCMC algorithm output and plot dose-response curves generated from these parameter values, labelled "Samples". We do this separately for Model 1 (we take pIC_{50} samples and set $Hill = 1$) and for Model 2 (we take $(pIC_{50}, Hill)$ samples). Each sample is plotted with a faint line, and each line has equal probability density. However, when multiple lines lie on top of each other, they will appear darker. These darker regions correspond to greater probability density. We also plot and highlight the curve generated by the parameters that resulted in the greatest log-target value when running the MCMC algorithm. This curve corresponds to the point/line of maximum posterior probability density, and is labelled "Max PD".

1. Quinine + Nav1.5-late

We obtain the Bayes factors

$$B_{12} = 0.000295, \tag{3.16}$$

$$B_{21} = 3388. \tag{3.17}$$

According to Table 2.1, we have "decisive" evidence in favour of Model 2 over Model 1.

The sampled dose-response curves for both models are plotted in Figure 3.8. Model 2 provides better fits through the data when evaluating the target function, primarily because the data points tend to be closer (vertically) to the dose-response curves provided by Model 2. Our conclusion from the Bayes factor value is that the improvements in fits from Model 1 to Model 2 justify the inclusion of an additional parameter in the statistical model.

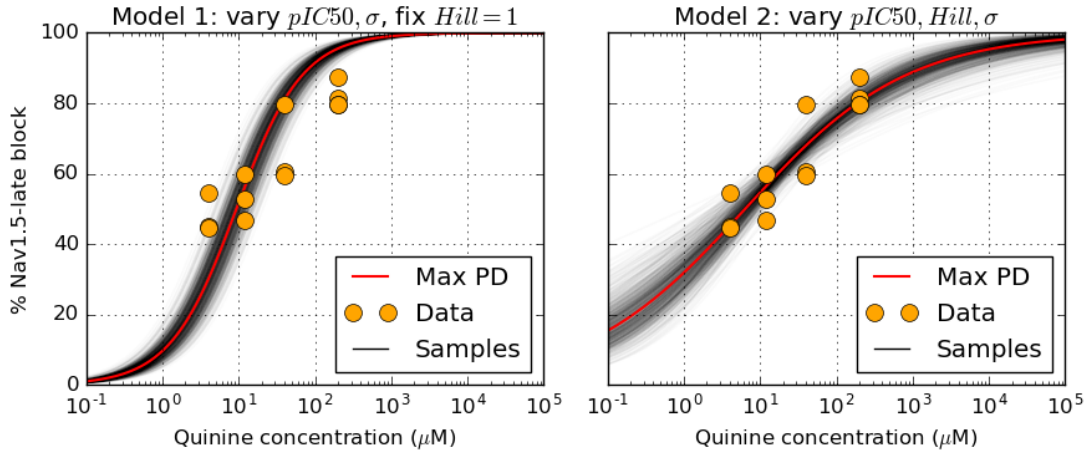

Figure 3.8: Dose-response curves generated from samples from the output of the MCMC algorithm for Quinine and Nav1.5-late. Left: the Hill coefficient is fixed to 1. Right: the Hill coefficient is allowed to vary as an additional parameter.

2. Quinine + Kv4.3

We obtain the Bayes factors

$$B_{12} = 34.65, \tag{3.18}$$

$$B_{21} = 0.0289. \tag{3.19}$$

According to Table 2.1, we have "strong" evidence in favour of Model 1 over Model 2.

The sampled dose-response curves for both models are plotted in Figure 3.9. Although not clear from the figure, Model 2 does allow for better fits to the data, as we would expect from a model with more parameters. However, from the Bayes

factor, we conclude that the improvements in fits from Model 1 to Model 2 do not justify the addition of an additional parameter in the statistical model.

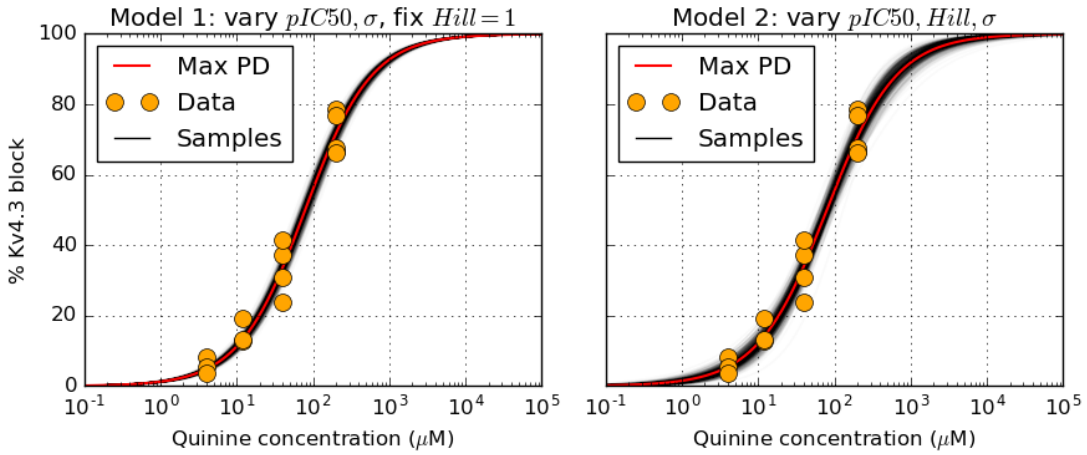

Figure 3.9: Dose-response curves generated from samples from the output of the MCMC algorithm for Quinine and Kv4.3. Left: the Hill coefficient is fixed to 1. Right: the Hill coefficient is allowed to vary as an additional parameter.

3. Amiodarone + KvLQT1/mink

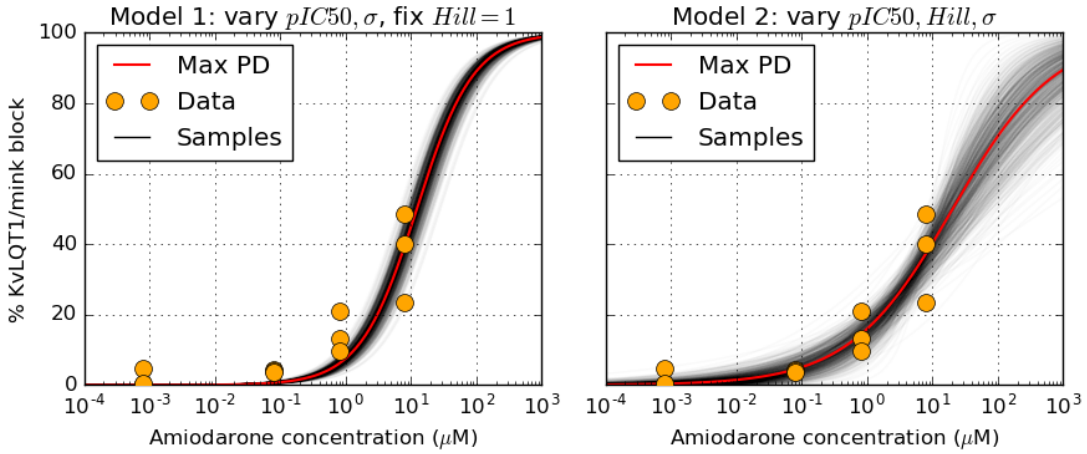

Figure 3.10: Dose-response curves generated from samples from the output of the MCMC algorithm for Amiodarone applied to KvLQT1/mink. Left: the Hill coefficient is fixed to 1. Right: the Hill coefficient is allowed to vary as an additional parameter.

For this drug/channel interaction we obtain the Bayes factors

$$B_{12} = 1.007, \tag{3.20}$$

$$B_{21} = 0.993. \tag{3.21}$$

According to Table 2.1, there is not any evidence in favour of either model over the other.

The sampled dose-response curves for both models are plotted in Figure 3.10. Model 2 does allow for better fits to the data, but it is not a significant improvement over Model 1. In particular, a slightly-increased σ can account for the discrepancies introduced by Model 1, and we observe this in the marginal posterior distributions for σ, shown in Figure 3.11.

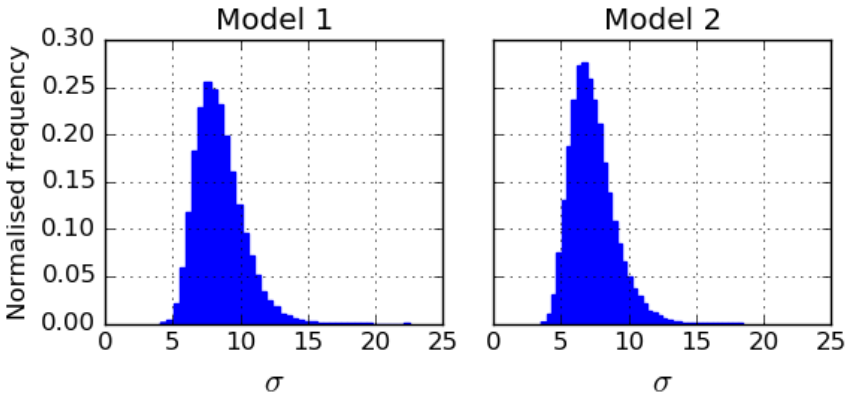

Figure 3.11: Normalised marginal histograms of σ for Models 1 and 2. The distribution obtained from Model 1 has a higher mode than that of Model 2. Model 1 also produces a wider distribution, with more probability mass on larger values of σ.

4. Amiodarone + Kir2.1

We obtain the Bayes factors

$$B_{12} = 0.9444, \tag{3.22}$$

$$B_{21} = 1.059. \tag{3.23}$$

According to Table 2.1, there is not any evidence in favour of either model over the other.

The sampled dose-response curves for both models are plotted in Figure 3.12. There is no measurable effect, beyond observation noise, of Amiodarone on Kir2.1. Both models can fit a flat line through zero, with no knowledge of what might happen at greater concentrations. This is illustrated in the fairly uniform distribution of dose-response curves past the highest experimental concentration. The marginal posterior distributions for pIC_{50} and $Hill$ resemble their prior distributions, as there is not enough information in the data to identify these parameters.

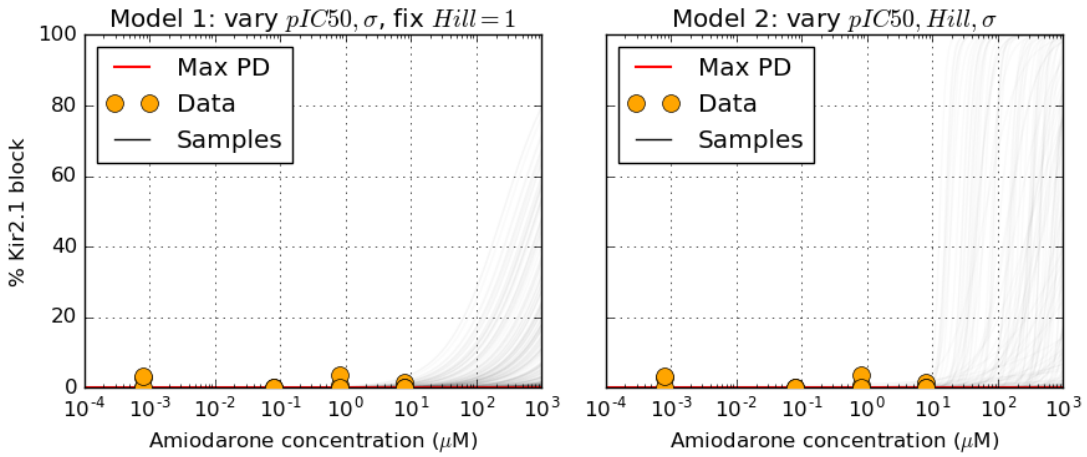

Figure 3.12: Dose-response curves generated from samples from the output of the MCMC algorithm for Amiodarone applied to Kir2.1. Left: the Hill coefficient is fixed to 1. Right: the Hill coefficient is allowed to vary as an additional parameter.

All Bayes factors

Since the difference between Models 1 and 2 is defined by fixing, or allowing to vary, the Hill coefficient, we expect that this parameter has the greatest effect on the Bayes factor comparing the two models. We compute B_{12} (the Bayes factor used to evaluate evidence in favour of Model 1 over Model 2) for all drug and ion channel combinations. We then plot these B_{12} values against the $Hill$ value that gave rise to the greatest posterior density value when running the MCMC algorithm under Model 2. We also plot the standard deviations of all of the $Hill$ samples obtained while running the MCMC algorithm under Model 2. These standard deviations give us an idea of how tightly-constrained the marginal distributions for

the *Hill* parameter are. For example, we expect that if the marginal distribution for the *Hill* parameter has a large standard deviation, then many different values of *Hill* could plausibly fit the data, and therefore the inclusion of this additional parameter does not provide a sufficiently-improved fit to the data. We also expect that if the *Hill* value giving rise to the greatest posterior density is close to 1, and that marginal posterior distribution has a small standard deviation, then Model 1 will be preferred over Model 2.

Before plotting, we take the $\log_{10}$ of all values to get a clearer idea of the variation, since lots of points are grouped in similar places. This 3-dimensional plot is shown in Figure 3.13 from two different angles. We see a line of points corresponding to $\log_{10}(B_{12}) \approx 0$ and relatively large $\log_{10}$ values of the *Hill* standard deviations. These correspond to the cases where the drug had no measurable effect on the ion channel, and so the marginal posterior distributions resembled their respective prior distributions. In the right-hand plot, we see a 'parabola' of points, where the apex corresponds to $Hill = 1$ and the greatest values of B_{12}. As the maximum posterior density value of *Hill* moves further away from 1, $B12$ decreases until eventually there is more evidence for Model 2 over Model 1.

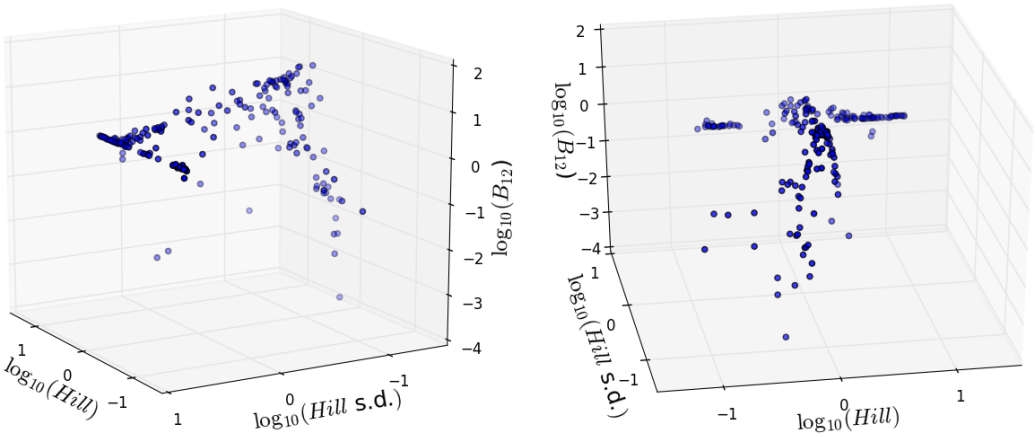

Figure 3.13: Scatter plot of all drug and ion channel combinations. The same plot is shown twice from different angles. x-axis: $\log_{10}$ of *Hill* giving the maximum posterior density in Model 2. y-axis: $\log_{10}$ of the standard deviation of all *Hill* values from Model 2. z-axis: $\log_{10}$ of B_{12}.

Of the 210 total drug and ion channel combinations, Model 1 (vary pIC_{50}, fix $Hill = 1$) was selected in 65 cases, and Model 2 (vary pIC_{50} and *Hill*) was selected

in 23 cases. The remaining 122 cases had $1/3 < B_{12} < 3$, which provides no substantial evidence in favour of either model, according to Table 2.1.

Evidence	Model 1	Model 2
Substantial	30	6
Strong	35	11
Decisive	0	6
	65	23

Table 3.1: Number of drug and ion channel combinations whose Bayes factors provide different levels of evidence, as defined in Table 2.1, in favour of each model. Model 1 was selected in 65 cases, Model 2 was selected in 23 cases, and there was no real evidence in favour of either model in the remaining 122 cases. Bayes factor values: 3–10 provides "substantial" evidence, 10–100 provides "strong" evidence, and >100 provides "decisive" evidence.

Since the prior distributions have to be specified by the modeller, and can be chosen somewhat arbitrarily to encode prior knowledge of beliefs about the parameters, Bayes factors can be sensitive to choice of prior distribution over the model parameters (Kass & Raftery, 1995). Typically it is useful to perform a prior sensitivity analysis on Bayes factors. For example, varying the rate parameter in the exponential distribution for pIC_{50} shown in Figure 3.4, and computing the Bayes factor. We would then have a range of Bayes factors effectively as a function of the pIC_{50} prior distribution rate parameter. Of course, this requires computing the Bayes factor for each parameter value, which multiplies the already-intensive computation time by however many parameter values we wish to sample. These are reasons why Kass & Raftery suggest the arbitrary-seeming range of 10–100 for "strong" evidence in favour of one model. A small change in a small Bayes factor could tip the favour to the other model, whereas with a much larger Bayes factor, a small change would have a much smaller relative effect.

3.5 Hierarchical statistical model

When we plot the ion channel screening data and group the data points according to their respective experimental repeats, instead of treating them as data points from one experiment as in Section 3.3, we see that data points from the same

experiment generally keep their relative positions. That is, we often see that the highest value at each concentration was observed during the same experiment, as shown for the Amiodarone applied to hERG case in Figure 3.14.

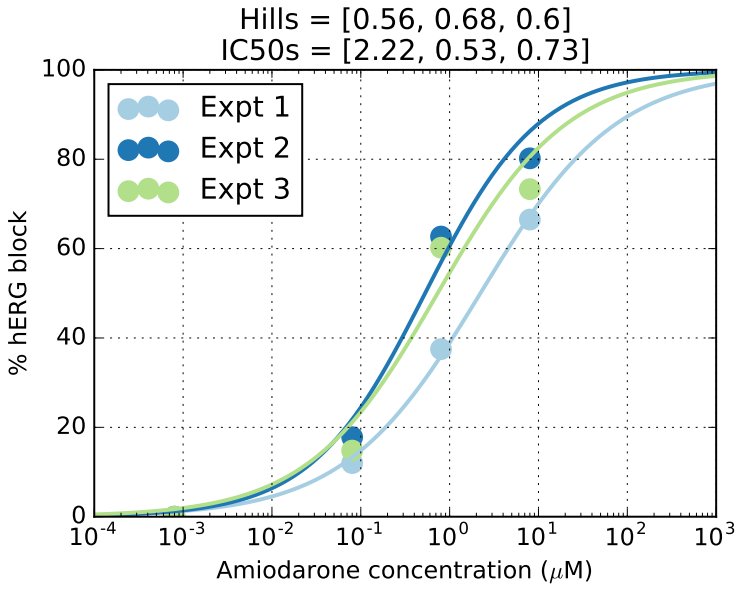

Figure 3.14: Data for hERG block by Amiodarone suggests inter-experiment variability. Different whole-cell patch-clamp experiments are plotted with different colours. In this case, the responses are consistently in the same ordering relative to the other experiments at each compound concentration (*e.g.* Experiment 2 always shows the largest response, and Experiment 1 always the smallest). This suggests inter-experiment variability that is distinct from observational error σ.

3.5.1 Methods

The intra-experiment correlation seen in Figure 3.14 suggests that each experiment has its own distinct properties. We model this as inter-experiment variability in the pIC_{50} value and Hill coefficient. That is, we treat each experiment as having its own pIC_{50} and $Hill$ parameters for which we will attempt to infer distributions. We let N_e be the number of experiments performed. The vector of data points obtained from experiment i is $\boldsymbol{y_i}$, where $i = 1, \ldots, N_e$. We take a hierarchical approach, as discussed in Section 2.5.4, and assume that there is some 'higher-level' distribution that governs how these parameters vary across experiments.

Hill coefficient and pIC_{50} distributions for ion channel screening datasets of up to $N_e > 12,000$ repeats were published in Elkins *et al.* (2013b), there they were found to fit independent log-logistic and logistic distributions, respectively. We assume the same type of distributions would occur here (if the experiments were repeated enough): that is, each experiment's $Hill_i$ is drawn from a log-logistic distribution with parameters α and β; and each experiment's $pIC_{50,i}$ is drawn from a logistic distribution with parameters μ and s. We have assumed that the observational errors are drawn from the same Normal distribution across all N_e repeats, so we infer just a single noise standard deviation parameter σ. A schematic of this hierarchical statistical model is given in Figure 3.15, adapted from Figure 2.9, with the 'mid-level' parameters and 'bottom-level' data points being independently distributed according to

$$Hill_i \sim \text{log-logistic}(\alpha, \beta), \tag{3.24}$$

$$pIC_{50,i} \sim \text{logistic}(\mu, s), \tag{3.25}$$

$$y_i^{(j)} \sim \mathcal{N}\left(f(x_i^{(j)}; pIC_{50,i}, Hill_i), \sigma^2\right), \tag{3.26}$$

where $y_i^{(j)}$ is the j^{th} entry in experiment i's responses $\boldsymbol{y_i}$, and $x_i^{(j)}$ is the j^{th} entry in experiment i's applied concentrations $\boldsymbol{x_i}$. We suppose that every experiment i has K_i data points (to generalise to cases where different experiments tests different numbers of concentrations), so $j = 1, \ldots, K_i$ and $i = 1, \ldots, N_e$.

We now need to specify prior distributions over the 'top-level' parameters (see Figure 3.15): μ, s, α, β, and σ. Prior distributions are chosen to contain any prior information or beliefs we have about the parameters before observing the data. We can therefore inform our choice of prior distributions by considering previously-published ion channel screening data. We use Gamma distributions for all of these, since Gamma distributions, in general, put probability only on positive values, and because these parameters are all positive, with the exception of μ. μ, however, can take any value and represents the centre of the logistic distribution. For μ, we used a Gamma distribution which is shifted along the x-axis down to -4, so there is little probability mass below -2. We choose this because as the pIC_{50} value becomes

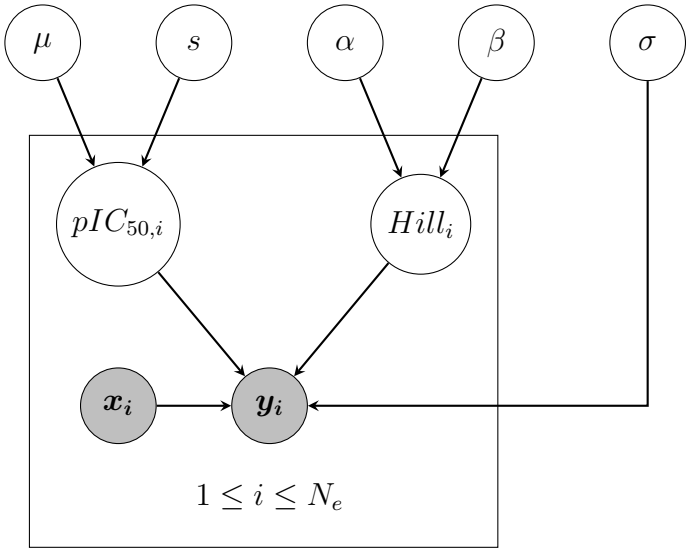

Figure 3.15: Hierarchical statistical model for dose-response data. i indexes the individual experiments. All non-shaded variables are parameters for which we wish to infer probability distributions.

lower and lower, the IC_{50} becomes larger, and eventually any possible compound effects occur well above the experimental concentrations. This was illustrated above in Figure 3.3, where dose-response curves (with $Hill = 1$) are plotted for varying values of pIC_{50}. The shaded "region of interest" covers the minimum and maximum concentrations in the data published by Crumb *et al.* (2016). Ion channel screening is generally performed at concentrations that range to well-above therapeutic concentrations, and so we do not want to infer how a compound will behave at even higher concentrations. These Gamma prior distributions were tuned to cover values provided by Elkins *et al.* (2013b), but also allow more room for variation. Plots of the prior distributions for α, β, μ, s, and σ are given in Figure 3.16.

In addition to covering the values published by Elkins *et al.*, we restricted β to be greater than 2, so that all log-logistic distributions generated would have no probability mass at 0, and also the gradient of the probability density function would be zero. This prevents Hill coefficients equal to 0 from being sampled by the MCMC algorithm. We also enforce that σ be greater than 10^{-3}, since we believe there is always the possibility of observation error, and hence there must be a positive standard deviation, we also run into division-by-zero numerical problems with the

evaluation of the target distribution if we sample $\sigma = 0$ (see Equation (3.13)). We therefore use the same Gamma prior as in Section 3.3 for σ.

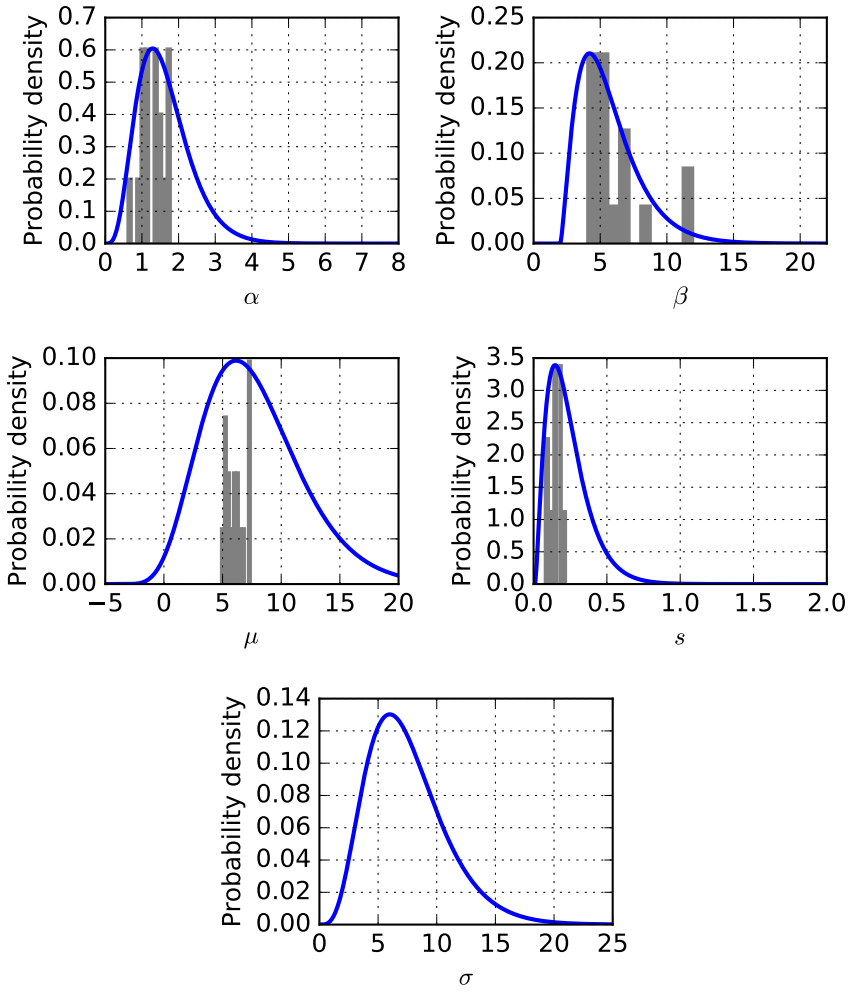

Figure 3.16: Prior distributions. Blue: Gamma distributions used as the prior distributions over α, β, μ, s, and σ. Grey: histograms of parameter estimates for different strongly-blocking control compounds, with large numbers of repeats, as previously published in Elkins *et al.* (2013b). However, we want to be able to fit to data obtained from any compound, not just the compounds published by Elkins *et al.*, and including ineffective compounds that elicit no response, hence the increased width of the prior distributions on certain parameters (μ particularly).

The choice of prior distribution will have an effect on the posterior distributions (via Equation (2.24)). However, the more information that is contained in our data, the less effect we expect our prior distribution to have on our posterior distribution. For example, Figure 3.17 shows how the marginal posterior distribution

for the 'top-level' parameters correspond to their respective prior distributions in the case of synthetic data, where we fit to different numbers of experimental datasets. This synthetic data were generated by sampling pIC_{50} values from a logistic distribution with $\mu = 6$ and $s = 0.1$, and Hill coefficients were sampled from a log-logistic distribution with $\alpha = 1$ and $\beta = 5$. Normally-distributed observation noise with standard deviation $\sigma = 1$ was added to the dose-response model at every concentration.

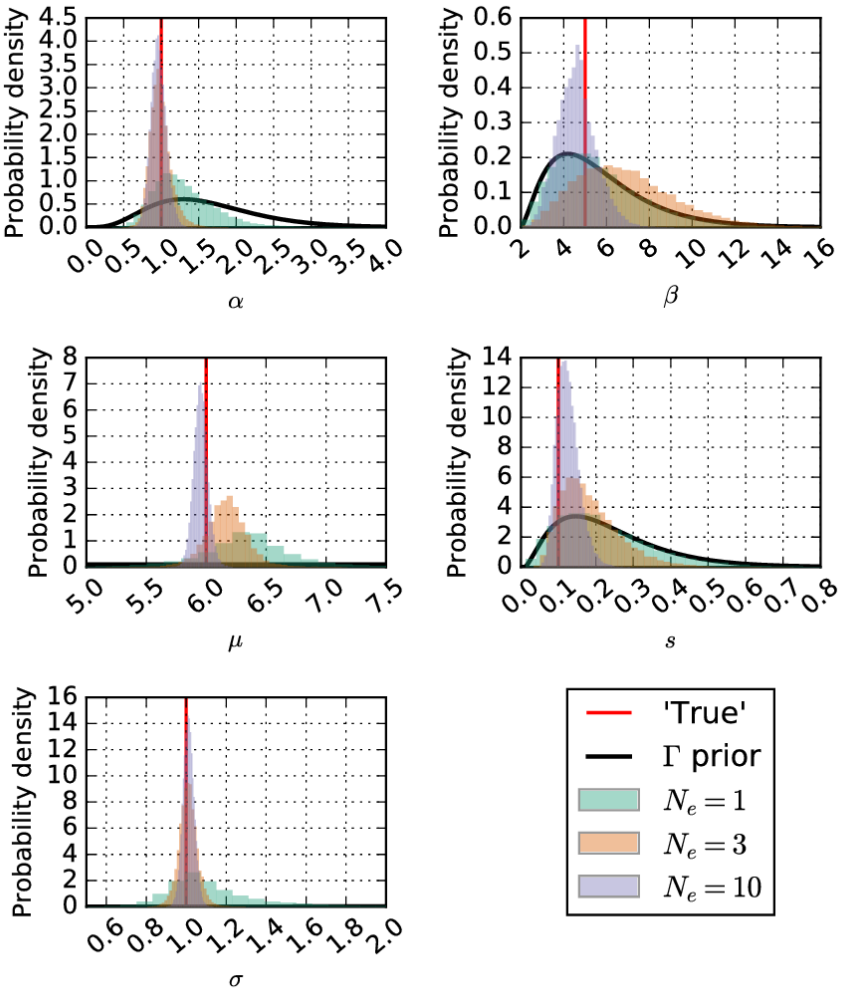

Figure 3.17: A comparison between marginal posterior distributions for 'top-level' parameters in the hierarchical model, with their respective Gamma (Γ) prior distributions. The number of (synthetic) experimental datasets, N_e, being fitted was increased. As we fit to more experiments, the prior distributions have a smaller effect on the posterior distributions. Where the black line for the prior distributions looks as if it lies along the x-axis (for μ and σ, the prior distribution was much wider than the marginal posterior distribution (there is a lot of information on these parameters with just one experiment).

We want to infer the posterior probability distribution for $\mu, s, \alpha, \beta, \sigma, Hill_i$, and $pIC_{50,i}$, for $i = 1, \ldots, N_e$, giving a total of $5 + 2N_e$ parameters. Using Bayes' Theorem, the posterior distribution in Equation (2.24) is now given by

$$\begin{aligned} &p(\mu, s, \alpha, \beta, \sigma, \{Hill_i\}_{i=1}^{N_e}, \{pIC_{50,i}\}_{i=1}^{N_e} | \{\boldsymbol{y_i}\}_{i=1}^{N_e}) \\ &\propto \left(\prod_{i=1}^{N_e} p(\boldsymbol{y_i} | \sigma, Hill_i, pIC_{50,i}) p(Hill_i | \alpha, \beta) p(pIC_{50,i} | \mu, s) \right) p(\mu) p(s) p(\alpha) p(\beta) p(\sigma). \end{aligned} \tag{3.27}$$

We use the same adaptive Metropolis-Hastings MCMC algorithm as in Section 3.3.1 — the adaptive Metropolis-Hastings algorithm described in Algorithm 1 — to infer a posterior distribution from the experimental data. Since we have many more parameters than in the non-hierarchical case, we expect to have to run our MCMC algorithm for more iterations to adequately approximate the posterior distribution. For most cases in this dataset, we have $N_e = 3$ or $N_e = 4$, which is not too demanding since our mathematical model is a simple analytic expression, and does not require solving differential equations as in the case of action potential models. However, if N_e became very large, we may have to use alternative MCMC techniques, such as a Metropolis-within-Gibbs algorithm which we discuss in Chapter 4.

3.6 Data and software availability

We used Python to perform all the inference and analysis in this chapter, and have published this code under the name *PyHillFit*, the latest version of which is available at `https://github.com/mirams/PyHillFit`. The code contains the experimental input data required to reproduce the examples shown here. Installation instructions can be found in the `README` file, in the main folder at the above links.

3.6.1 Results

As before, we plot normalised marginal histograms to approximate the marginal posterior distributions for each parameter. Such histograms are plotted for μ, s, α, β, and σ in Figure 3.18 for the Amiodarone applied to hERG case. We can compare these to the prior distributions shown in Figure 3.16, and we see that in most cases

we have much narrower marginal posterior distributions than prior distributions. This tells us that the data contains enough information about those parameters to constrain them to narrower intervals.

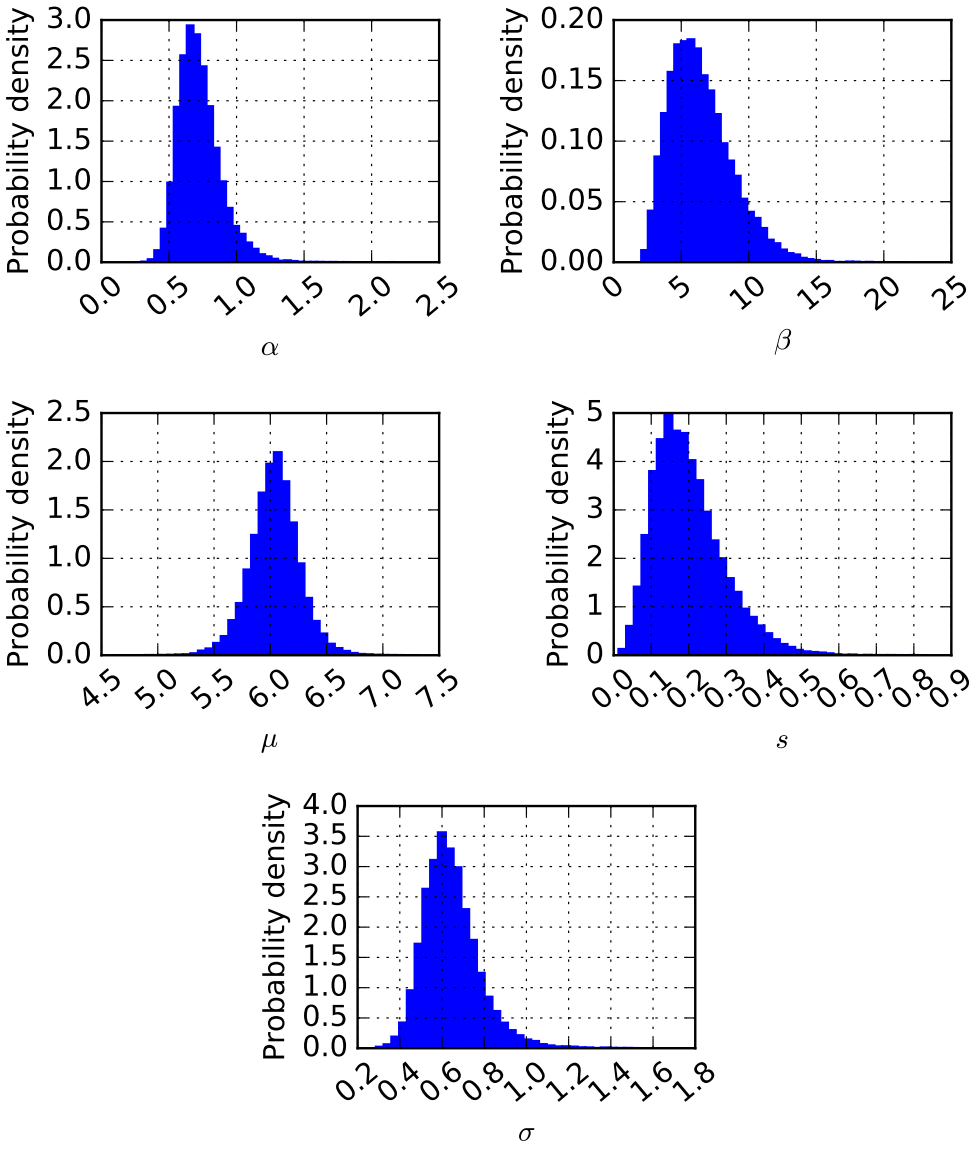

Figure 3.18: Normalised marginal histograms for the 'top-level' parameters, α, β, μ, s, and σ after running the fitting the hierarchical model to the Amiodarone applied to hERG dataset using the MCMC algorithm. Most of these distributions are narrower than their respective prior distributions in Figure 3.16, with the exception of β. We therefore conclude that the experimental data does not contain much information about β, in line with the synthetic data study shown in Figure 3.17.

We can also superimpose the normalised histograms for each $Hill_i$ and $pIC_{50,i}$

to give us an idea of how much inter-experiment variability is present in these parameters. These superimposed histograms are plotted in Figure 3.19 for the Amiodarone applied to hERG case.

To make predictions about how a particular compound and channel will interact if we perform another experiment, we consider the posterior predictive distribution, described in Section 2.5.4, for $Hill_i$ and $pIC_{50,i}$. That is, what is $p(Hill_{N_e+1}, pIC_{50,N_e+1}|\text{data})$? $Hill_i$ and $pIC_{50,i}$ are modelled as being drawn from log-logistic and logistic distributions, respectively. We therefore take the product of a log-logistic distribution and a logistic distribution that are generated from the 'top-level' parameters at a subset of T iterations from our hierarchical MCMC output. We then sum these T products and normalise by dividing the sum by T to obtain a new probability distribution:

$$p(Hill_{N_e+1}, pIC_{50,N_e+1}|\text{data}) \approx \frac{1}{T}\sum_{t=1}^{T} \text{log-logistic}(Hill_{N_e+1}; \alpha_t, \beta_t) \times \text{logistic}(pIC_{50,N_e+1}; \mu_t, s_t), \quad (3.28)$$

where t indexes the samples in our Markov chain, after having discarded a number of initial samples as a burn-in. The posterior predictive distribution takes the form of a product of distributions in Equation (3.28) because of possible dependence between the $Hill_i$ and $pIC_{50,i}$ parameters. If the parameters were independent then Equation (3.28) would decouple into two distinct normalised sums.

Since the posterior predictive distribution in Equation (3.28) is two-dimensional, we cannot use the inverse-CDF method mentioned in Section 2.5.4 of Chapter 2. Instead, we sample $(\alpha_t, \beta_t, \mu_t, s_t)$ from the MCMC output, and then sample $pIC_{50} \sim \text{logistic}(\mu_t, s)$ and $Hill \sim \text{log-logistic}(\alpha_t, \beta_t)$. We then plot the dose-response curve defined by these newly-sampled $pIC50$ and $Hill$ parameter values. This process is repeated many times, so that we construct a distribution of possible dose-response curves without directly constructing an approximation to the posterior prediction distribution defined in Equation (3.28).

A plot of predicted dose-response curves for a future experiment, following the hierarchical model MCMC, is given in plot A of Figure 3.20. To make a prediction

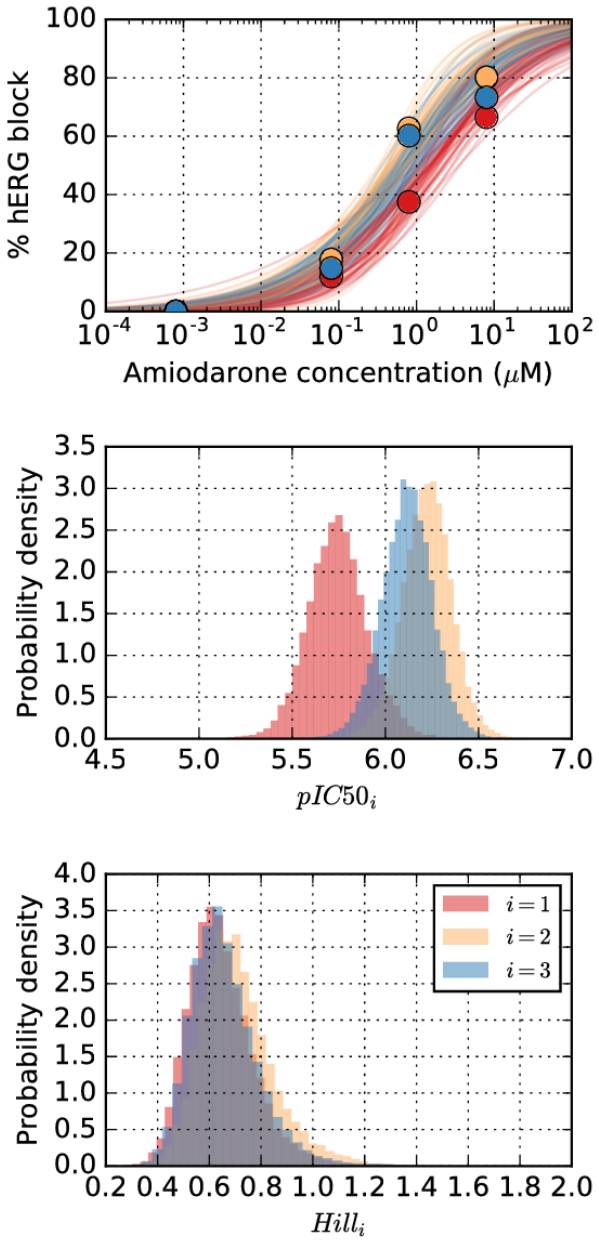

Figure 3.19: Inferred parameters for individual experiments. Top: Dose-response curves plotted using the 'mid-level' $pIC_{50,i}$ and $Hill_i$ samples from our MCMC algorithm output from the Amiodarone applied to hERG dataset. Middle & bottom: Superimposed normalised histograms for $pIC_{50,i}$ and $Hill_i$, after fitting our hierarchical model to the Amiodarone applied to hERG dataset using the MCMC algorithm. We find that the Hill coefficient does not vary much between experiments, however there is variability within the pIC_{50} value.

of what % block will be induced by that compound at a particular concentration, we take a vertical cross-section through these dose-response curves and plot a normalised histogram of these levels of block to approximate a probability distribution, as shown in plot B of Figure 3.20.

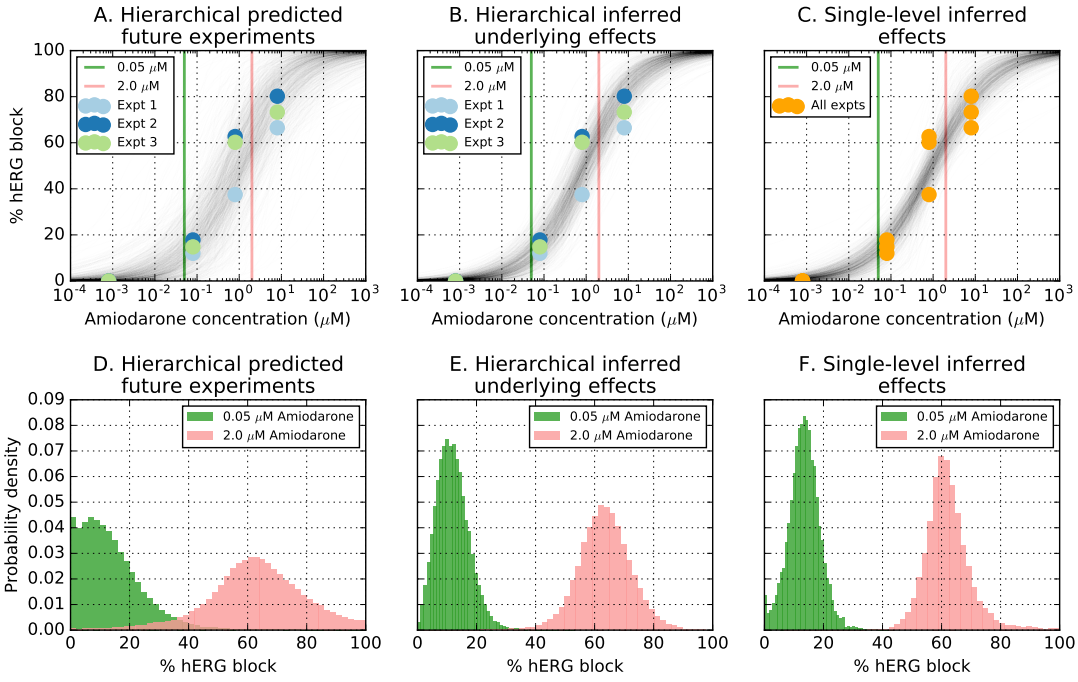

Figure 3.20: Predicted dose-response curves and associated probability distributions for levels of block at example concentrations when fitting to the Amiodarone applied to hERG dataset. A: Hierarchical model — predictions for how a future experiment will behave, with samples taken from the posterior predictive distributions. B: Hierarchical model — inferred distribution for the underlying behaviour of the system, plotted by using μ and α samples from the MCMC algorithm output directly as values for pIC_{50} and $Hill$, respectively. C: Single-level — inferred distribution, plotted by taking samples from the MCMC algorithm output. D,E,F: Histograms of the intersections between the vertical lines and dose-response curves in A,B,C, respectively, at two different concentrations of Amiodarone.

Note that the hierarchical model allows us to make two sets of predictions. Firstly, we can use the posterior predictive distribution given by Equation (3.28) as shown in Figure 3.20 (panel A). This distribution includes inter-experiment variability, and can therefore be considered a distribution that predicts where data points from future experiments may lie. Secondly, we can examine the variability in the underlying properties of the compound; the 'average' effect, before it is

altered by inter-experiment variability (panel B). We generated this plot by taking samples of α and μ to use as Hill and pIC_{50} values. We would expect panel B to be more directly comparable with the single-level approach (which fits 'average' data points), which is shown in panel C for comparison.

Which of the two distributions (illustrated in panels A or B in Figure 3.20) one may wish to use for predictions is subtle. If we consider that the source of variability between experiments is also present in the system that we are making predictions for, then the first case (panel A) would be the best to use. If however we consider that there is a single underlying effect, and the act of measuring it introduced inter-experiment variability that is not present in the real system, then the second distribution (panel B) would be more appropriate. Most biological experiments implicitly assume the second case is true — that by taking repeated measurements and then taking the average, a more accurate assessment of the underlying system is made.

3.7 A comparison of single-level and hierarchical models

There are advantages and disadvantages to choosing either the single-level statistical model, or the hierarchical statistical model. The main benefit of the single-level model is that we are only fitting three parameters, meaning that the parameter space of interest is relatively easy to explore. This means that we need run our MCMC algorithm for fewer iterations to obtain an acceptable approximation of the posterior distribution than if we had a larger number of parameters, reducing overall computation time. However, an algebraic model such as the one we have here (Equation (3.10)) can be solved very quickly, and so computation time is generally not a problem for even the hierarchical version of the model, for this application. There is also little sensitivity to the prior distributions, as there is a lot of information about all three parameters in even one experiment.

In Figure 3.20 we can compare the results of the two types of inference for real data on Amiodarone block of the hERG current. In the single-level case, by

fitting to all data points at once, the inference can misinterpret inter-experiment variability and assign it to the 'wrong' parameter(s). To demonstrate this, we generated two sets of synthetic data corresponding to these fictitious compounds, with fictitious inter-experiment variability properties:

1. For Shamiodarone, $Hill$ was fixed as 1 across all experiments, and the $pIC_{50,i}$ were drawn from a logistic distribution, with $\mu = 6$ and $s = 0.2$.

2. For Shamitriptyline, pIC_{50} was fixed as 6 across all experiments, and the $Hill_i$ were drawn from a log-logistic distribution with $\alpha = 1$ and $\beta = 2.5$.

In both cases, we simulated 5 experiments where each experiment consists of measuring % channel block at 4 different compound concentrations. We added Normal observation noise with standard deviation $\sigma = 0.5$ to each point.

Case 1: fixed Hill and varying pIC$_{50}$

A sample of the inferred curves for this case are given in Figure 3.21. The plots in panels A & D represent pIC_{50} and $Hill$ parameter values being drawn from their respective posterior predictive distributions, this gives predictions of how we believe the observations from a future experiment would behave. The plots in B & E are based on α and μ samples — what we believe to be the underlying 'average' behaviour of the compound interacting with an ion channel, when experimentally-introduced variability is discounted. The hierarchical model was able to identify consistency within the $Hill_i$, and the MCMC algorithm generally only infers that $pIC_{50,i}$ varied between experiments.

The histograms in panels D–F are a cross-section of the dose-response curves at different concentrations, and represent the probability density of % block at that compound concentration. Note that each curve in panel B has approximately the same slope, corresponding to a consistent Hill coefficient, whereas in panel C we see that there is a greater range of slopes, corresponding to (slightly more) variability in the Hill coefficient. Comparing plot E with plot F, we see that at 0.05 μM concentration, the non-hierarchical model is less certain about the % channel block

than the hierarchical model, because the former has incorrectly inferred there is more variation in the Hill coefficient. So the single-level model tends to compensate for the varying parameter values by fitting curves that fit through an 'average' of the points. The algorithm does this by varying both $Hill$ and pIC_{50} to obtain curves that could fit the data reasonably well, even when the synthetic data was generated by holding one parameter fixed and varying the other.

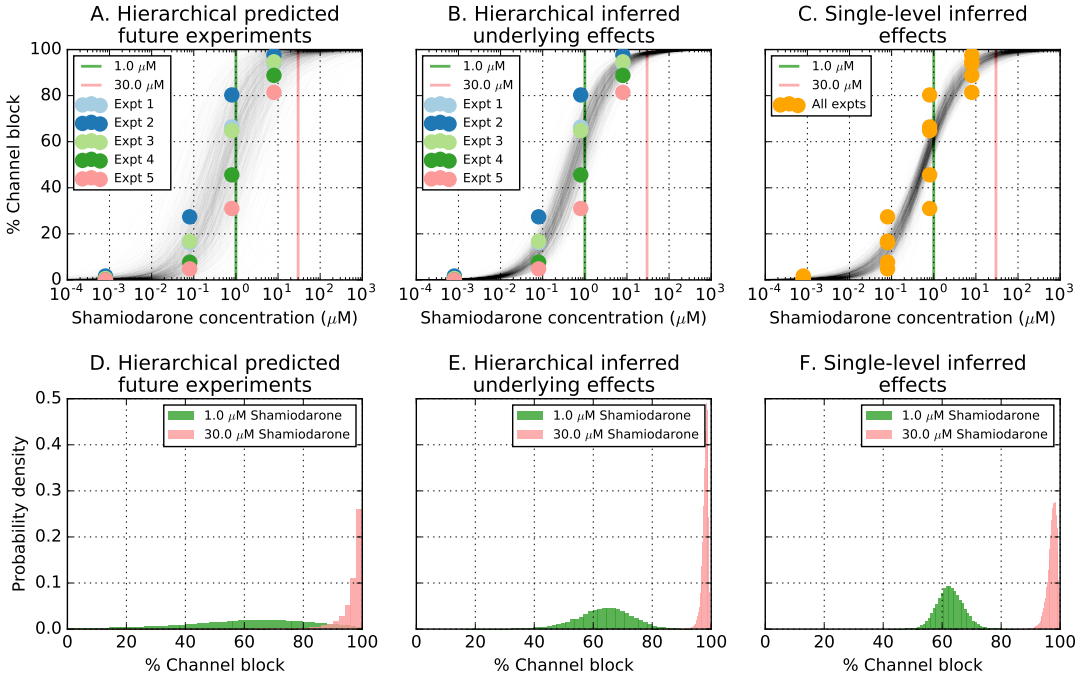

Figure 3.21: Inference on synthetic data generated by fixing $Hill = 1$ and varying pIC_{50}. A: Predicted dose-response curves, with pIC_{50} and $Hill$ sampled from their respective posterior predictive distributions, taking inter-experiment variability into account. B: Inferred underlying behaviour of the compound-ion channel interaction, with inter-experiment variability discounted. C: Inferred dose-response curves from single-level inference. D–F: Normalised histograms of the cross sections plotted with vertical lines in plots A–C. These histograms represent probability density functions of % block at a particular concentration, given the (synthetic) experimental data.

Case 2: fixed pIC_{50} and varying Hill

A sample of the inferred curves for this case are given in Figure 3.22. The single-level model in panel C does show small variability in the Hill coefficient, as well as small variability in the IC_{50} (and hence pIC_{50}). This leads to a reasonably spread prediction of ion channel block at both a concentration near pIC_{50} and at a higher

concentration (panel F). But we know that the underlying data had the same pIC_{50}, and so variability near the IC_{50} should be minimal, and indeed the spread of predictions at a higher concentration should be larger. The hierarchical model captures the desired underlying variability better (compare panels E and F). The Hill coefficient varies (panel B) while also capturing the low variability in the pIC_{50} value (there is still some variability in the inferred pIC_{50} distribution due to observational noise and a low number of repeat experiments). As a result, the predicted percentage blocks in panels E and F are different. As for Case 1, either too-much or too-little variability is predicted by the single-level inference, depending on the concentration.

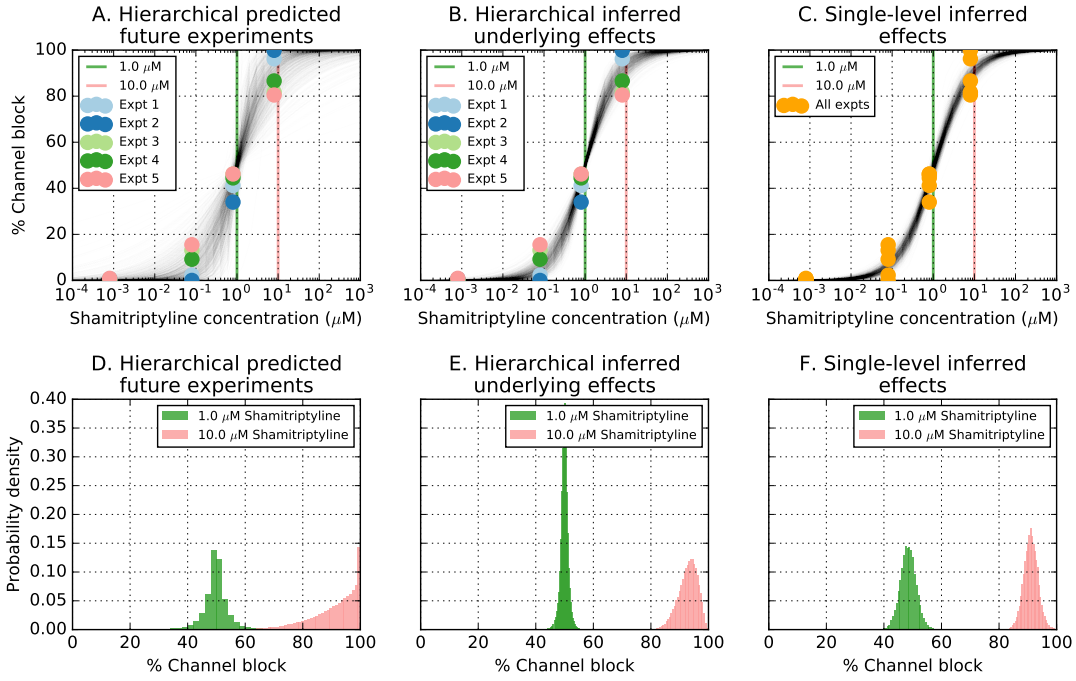

Figure 3.22: Inference on synthetic data generated by fixing pIC_{50} and varying $Hill$. A: Predicted dose-response curves, with pIC_{50} and $Hill$ sampled from their respective posterior predictive distributions, taking inter-experiment variability into account. B: Inferred underlying behaviour of the compound-ion channel interaction, with inter-experiment variability discounted. C: Inferred dose-response curves from single-level inference. D–F: Normalised histograms of the cross sections plotted with vertical lines in plots A–C. These histograms represent probability density functions of % block at a particular concentration, given the (synthetic) experimental data.

3.8 Propagating dose-response uncertainty

As discussed in Section 2.4.2, the proposed Comprehensive in-vitro Proarrhythmia Assay (CiPA) recommends the use of computational action potential models in the drug safety process. Ion channel screening will be performed, and the IC_{50} values and Hill coefficients obtained from these experiments are to be used in action potential models to predict whether or not a compound is likely to be proarrhythmic. One simple proposed measure of proarrhythmia is action potential duration prolongation (Mirams *et al.*, 2011), directly related to prolongation of the QT-interval, which can be a precursor to potentially fatal arrhythmias such as Torsade de Pointes.

Using best-fit IC_{50} values and Hill coefficients obtained from ion channel screening data, we can compute a predicted level of block of each of the ion currents in an action potential model, at a particular compound concentration. We then simulate an action potential and measure the action potential duration prolongation relative to the control case (see Beattie *et al.* (2013) for an example of this approach). However, when using best-fit IC_{50} values and Hill coefficients, we obtain a single predicted action potential after simulating a particular compound concentration.

If instead, we use Bayesian-statistical frameworks to infer probability distributions for each experiment's IC_{50} value and Hill coefficient, we can take samples from these distributions and compute an action potential duration for each sample. This allows us to construct a probability distribution for predicted action potential durations based on the original ion channel screening data (uncertainty propagation).

To illustrate this approach, we use our methodology to fit hierarchical dose-response parameters to thirty drug compounds for seven ion currents each using the Crumb *et al.* (2016) dataset. We then take samples, based on the 'underlying effects' curves from hierarchical fits (see Figures 3.20, 3.21 & 3.22), and simulate action potential durations after applying blocks to seven ion channels using the approach outlined in Mirams *et al.* (2011). We plot the resulting predicted distributions of action potential durations as violin plots in Figure 3.23. The violin plot for each of the thirty compounds is coloured according to the Torsade/QT risk categorisation in the CredibleMeds database (`https://crediblemeds.org/`). We also plot the

control action potential duration in the O'Hara model along with a 10% action potential duration prolongation.

We see that the vast majority of the compounds have overlapping probability distributions for predicted APD at maximum free therapeutic plasma concentration, suggesting that at least this previously-proposed measure will be insufficient to distinguish compounds in terms of risk based on data such as these.

For a more direct comparison between the drugs, we can take a random sample from these distributions of APD_{90} for each drug, and then order these samples from longest APD_{90} to shortest. This is then repeated many times, and we can plot histograms of the resulting rankings for each drug. We performed this 10,000 times, and the histograms are shown in Figure 3.24. Strong overlap between histograms causes problems for drug safety because it is difficult to assess relative risk.

This suggests that either action potential prolongation is not a good enough marker of proarrhythmia, or that there was too much uncertainty associated with the experimental data to constrain these distributions to narrow distinct ranges. To counter the latter point, we would suggest that more experimental repeats be performed. In either case, it is imperative to realise that the data that are being used have a level of uncertainty which means it is not possible to rank the majority of these drugs in terms of their predicted APD.

3.9 Discussion

A Bayesian framework is a useful tool to address uncertainty characterisation in ion channel screening data. When no uncertainty characterisation is performed, one can obtain best-fit parameter values for the data presented, but there is no associated probability in terms of the behaviour that generated the data, or in predictions informed by the data. The single-level Bayesian inference model discussed in Section 3.3 can provide ranges of possible dose-response curves (and underlying parameters) that fit ion channel screening data. However, parameter-specific inter-experiment variability can be missed when using a single-level statistical model, since the MCMC algorithm treats all points equally and so varies the parameters without

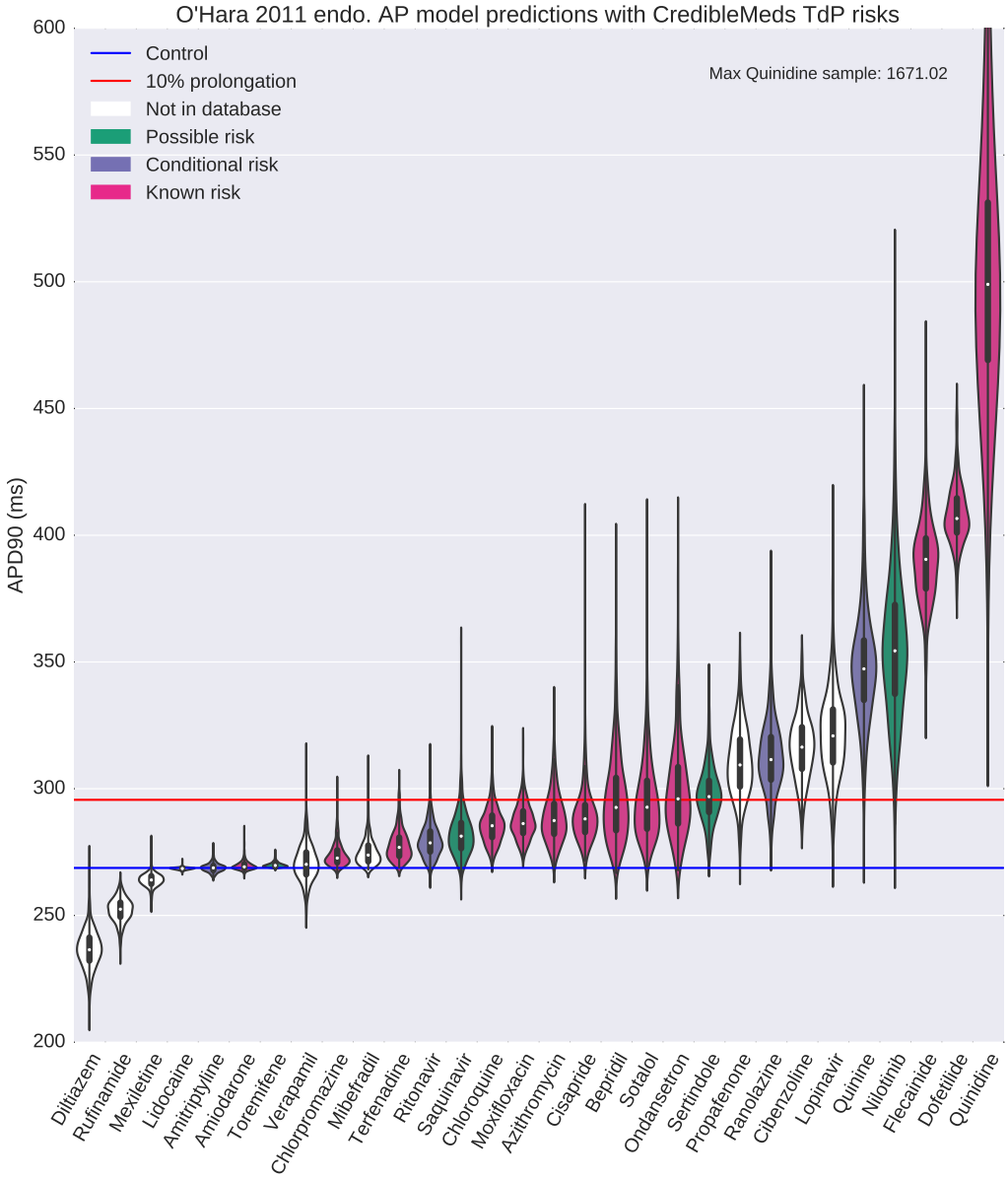

Figure 3.23: Violin plots of predicted action potential duration (APD_{90}, time taken for the cell to return to 90% repolarised after depolarisation) using samples from the iterations of our MCMC algorithm under the hierarchical statistical model. Simulations were run using the O'Hara *et al.* (2011) human ventricular cardiomyocyte action potential model, and APD_{90}s were computed. We used 'AP-predict' (Williams & Mirams, 2015), a bolt-on project for the Chaste open-source computational biology C++ library (Mirams *et al.*, 2013). Histograms are plotted for the 30 compounds discussed by Crumb *et al.*. In one case, two samples led to very long action potential durations and so are omitted for clarity.

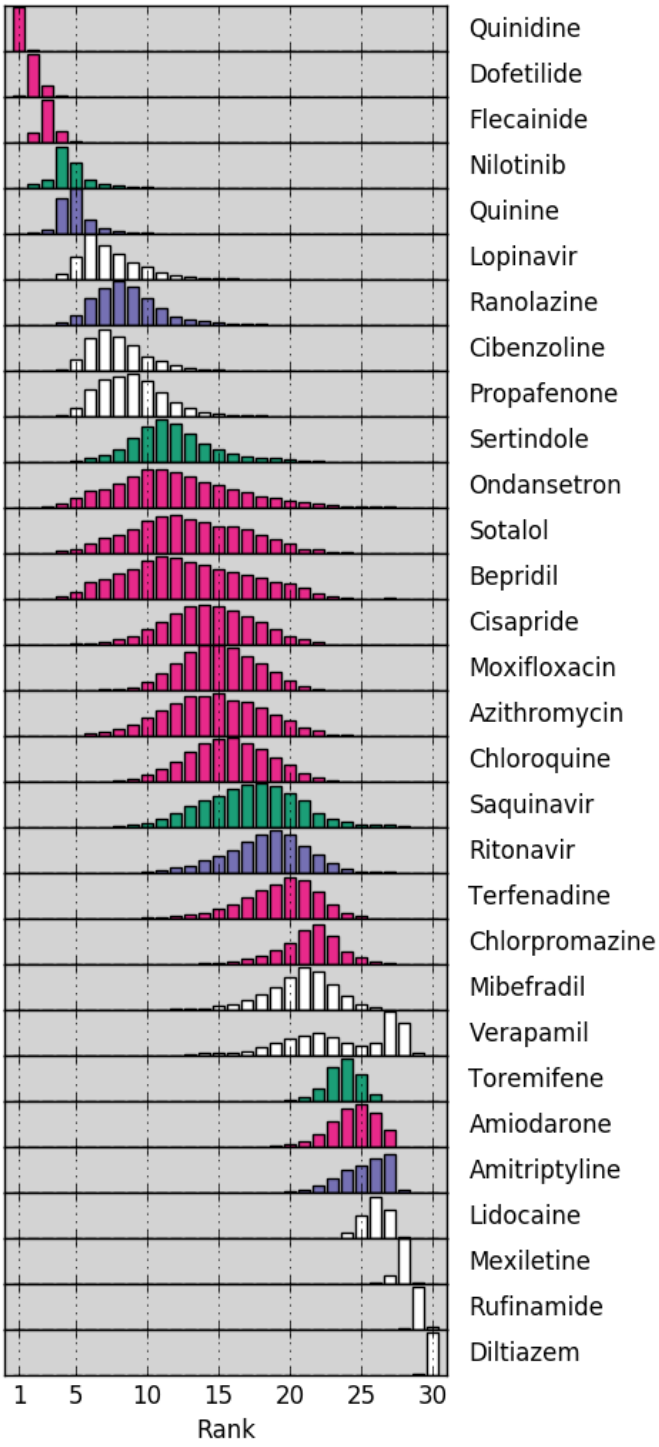

Figure 3.24: Histograms of drug rankings after taking random APD_{90} samples and ordering from highest to lowest, and repeating 10,000 times. The colours correspond to the same risk categories as in Figure 3.23. The y-axes are not on the same scale, because the spread for Bepridil is much greater than for Quinidine, for example, and it would be difficult to visualise all drugs if using the same y-scale if normalising the histograms to represent probability distributions.

considering the inter-experiment correlations. This leads to an 'averaging' effect, where the dose-response model is fitting to an average of the experimental data points, but may not reflect the behaviour of any individual experiment. However, this single-level inference is quick to run as it requires only the fitting of three parameters, and provides a better approximation of probability distributions than a single best fit.

A hierarchical statistical[1] model can capture inter-experiment variability within dose-response parameters, as demonstrated in the synthetic cases discussed in Section 3.7. The hierarchical model can therefore be used to infer inter-experiment behaviour, and hence predict how a future experiment might behave. By taking samples for the 'top-level' parameters from our MCMC output, we can build distributions of how we believe the compound is interacting with the ion channel. At a given compound concentration, we can then build a probability distribution for possible levels of ion channel block. The hierarchical model is able to determine what variability is being introduced at the experimental level, and allows us to make probabilistic statements about the underlying behaviour.

Our hierarchical model is similar to *nonlinear mixed effects* (NLME) modelling, but we operate in a Bayesian framework. NLME assumes a similar structure to that shown in Figure 3.15, but infers best-fit values for the 'top-level' parameters, and a distribution from which the 'mid-level' parameters are sampled. While it does capture inter-experiment variability and would allow us to make predictions about how a future experiment might behave, it only provides a point-estimate for underlying behaviour, rather than different possibilities with relative probabilities.

There are different possible interpretations of what leads to inter-experiment variability. One possibility is that a pIC_{50} value, say, is specific to the cell being measured, and that we expect the fitted values to vary from cell to cell. Alternatively, given that a drug molecule has a particular structure, and ion channel membrane proteins also have particular structures, we might expect a drug to have a particular pIC_{50} value when acting upon an ion channel, regardless of which cell is being measured, due to their respective physical properties. In

[1]Cryptic crossword clue: "Statistical framework covered by Wahba?" "Yes", I answered (8)

this latter case, the observed inter-experiment variability might be caused by any number of experimental variables.

In the first case, we would consider each cell to have its own properties when interacting with a drug, and therefore we would take samples from the 'possible future experiment' distribution when propagating uncertainty. However, in the second case, we believe there is some true underlying behaviour with some unknown causes of the observed inter-experiment variation. We would therefore take samples from the distributions that we believe describe the 'true' parameters for uncertainty propagation.

A possible limitation of the hierarchical model is that computation time increases with the number of experimental datasets being fit to at once. This is not a problem for up to 4 or 5 experiments, but the number of parameters quickly becomes intractable for the adaptive-Metropolis MCMC algorithm (Haario *et al.*, 2001) that we have been using. In general, with MCMC methods, we want to run our algorithm for as long as possible, to best approximate independent samples from the posterior distribution. There is therefore no upper limit for how long this method takes, although for these examples we have run our algorithm for 500,000 iterations, which takes approximately 12 minutes for the Amiodarone-hERG case which has 3 experimental datasets of 4 concentrations each.

Another possible limitation of the hierarchical model is the dependence on the prior distributions for the 'top-level' parameters. As shown in Figure 3.17, when there is not much data, the posterior is heavily influenced by the prior. However, we chose our priors based on data published by Elkins *et al.* which was based on 12,000 ion channel screening experiments, and we therefore have some confidence in their shapes (Figure 3.16). In a Bayesian framework, should new data become available, we can compute new posterior distributions for the parameters according to Equation (2.24), by using a previous posterior distribution as the new prior distribution.

A benefit of both inference techniques is that we introduce the observation noise as a parameter to be fitted, along with the pIC_{50} values and Hill coefficients. Instead

of estimating the observation noise and then fitting dose-response curves based on our estimate, we allow the MCMC algorithm to find likely levels of noise, while also quantifying the uncertainty in those estimates. Since all of these parameters are being fit at the same time by the MCMC algorithm, we can extract how much noise on dose-response parameters is introduced by inter-experiment variability, and how much noise is due to observation error.

Single best-fit parameter estimates from ion channel screening data can give a *most likely* set of dose-response curve parameters. However, this approach does not provide us with a measure of uncertainty around these parameters. The hierarchical Bayesian approach we present can quantify some of the uncertainty associated with dose-response curves, with its prior distributions set for ion channel screening data, for the purposes of propagating this uncertainty into further quantitative studies. This hierarchical Bayesian statistical model provides a method for taking inter-experiment variability into account. It is more able to take consistent changes in parameters across experiments and 'unpick' them from observation noise.

It is very useful to perform uncertainty characterisation when fitting dose-response models to ion channel screening data. This is evident in Figure 3.23, where the risk rankings of various compounds have changed when using the O'Hara *et al.* model to predict APD_{90}s. When presented with experimental data with clear variability, and hence uncertainty, it is important to characterise and propagate this uncertainty to have predictions that take as much of the information present as possible into account.

Bayes factors can help with model selection when working in a Bayesian framework, if used carefully. However, they will not always provide a definitive answer in favour of one model over another. Additionally, they are very computationally expensive, as they require multiple runs of our MCMC algorithm. As is often the case, the more MCMC runs the better, and running the algorithm for as long as possible is also preferable.

Possible continuations of this work could include using different mathematical models to describe ion channel block. For example, the addition of a saturation

level might more accurately describe the behaviour of some drugs; a drug will not necessarily block a channel 100% even at an arbitrarily high concentration. Alternatively, we might consider a dose-response model that allows for the activation of an ion channel by a drug. These different model formulations can be directly compared using Bayes factors, if there are reasonable prior distributions across their parameters.

In the next chapter, we will perform more uncertainty characterisation, but on model parameters of cardiac action potential models.

4

Inferring action potential model conductance parameters from synthetic voltage traces

The results from the first half of this chapter were published in Johnstone *et al.* (2016b) (single-level statistical model as described in Section 2.5.1 of Chapter 2), and adapted from Johnstone *et al.* (2016a) (hierarchical statistical model as described in Section 2.5.4 of Chapter 2).

4.1 Introduction

In this chapter we infer conductances for action potential models from synthetic voltage versus time data traces, which we generate using the action potential models as published with the addition of random noise. We expected that (i) the complexity of the model, and (ii) the richness of the experiment would determine the conductances that were constrained by data. For example, for a simple model with a small number ($\leqslant 6$) of ion currents, we might expect that only a certain combination of conductance values produces the observed action potential, and therefore we can infer all of them. For a more complicated model with tens of conductances, there may be many options for 'balancing the currents' to give the same action potential, and so we would not be able to establish unique estimates for

each conductance (at least in the presence of observation noise). Similarly, a simple experiment such as a single action potential recording may contain less information than a more complex protocol with, for instance, multiple pacing frequencies.

There are a number of experimentally plausible protocols that can be used to generate action potential data, including steady pacing at different rates, changing ion concentrations of the extracellular solution, or applying a pharmaceutical intervention to block a specific ion current. We repeated the following study with a range of mathematical action potential models, and a range of protocols, to examine how each affects the proportion of conductances that we can infer.

We present our progress and methods when investigating the application of MCMC methods to action potential model parameter inference, starting from a more basic MCMC algorithm than the adaptive Metropolis-Hastings algorithm introduced in Chapter 2.5.3, and some of the problems that led us to investigate more efficient algorithms.

4.1.1 Methods

Bayesian inference

We wish to infer probability distributions for action potential models, given some experimental data. We will therefore work in the same kind of Bayesian framework introduced in Chapter 2.5.1. A similar approach was taken recently to estimate single ion channel kinetics (Siekmann *et al.*, 2011, 2012).

Differences between simulated voltage trace and the observed (synthetic) data trace are assumed to be due to noise, which we assumed to follow a Normal distribution with mean zero and an unknown standard deviation σ. (While we take a specific value of σ to generate the synthetic data, in general this would not be known, and so we attempted to infer σ alongside the conductances.) A schematic of parameter interaction and dependence under this statistical model is given in Figure 4.1, which is the same as Figure 2.8, but with $\boldsymbol{G}$ used to reflect that we are interested in conductance parameters. $\boldsymbol{G}$ and σ require prior distributions to be specified.

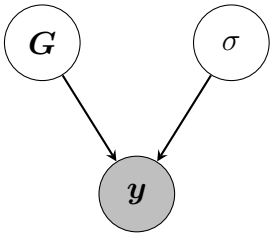

Figure 4.1: Parameter interaction and dependence under the statistical model for a single action potential recording. $\boldsymbol{G}$ is the set of maximal ion current conductance parameters which are used to generate $\boldsymbol{y}$, the dataset. All non-shaded variables are parameters for which we wish to infer probability distributions. Parameters with no inward-pointing arrows require prior distributions to be specified.

Given our data and some action potential model, we compute the likelihood of observing this data given a certain set of parameter values. We make the following assumptions:

- The only source of error in our measurements is that of experimental noise. In other words, that the measured data is Normally distributed around some true value, with a standard deviation of σ mV.
- We treat σ as an unknown parameter to be inferred, along with the maximal conductance parameters.
- The values of the noise around the true solution at each time point are independently distributed.
- We have a rough idea of what the parameters are; prior knowledge can be accurately represented by a uniform distribution across large intervals (which we choose to be from 0.1 $\times$ (the true parameter value) to 10 $\times$ (the true parameter value)). Their joint prior distribution, $p(\boldsymbol{G}, \sigma)$, is given by

$$p(\boldsymbol{G}, \sigma) = \begin{cases} c & \{\boldsymbol{G}, \sigma\} \text{ in the specified intervals,} \\ 0 & \text{otherwise,} \end{cases} \tag{4.1}$$

where c is a non-zero, finite normalising constant.

In this case, we supposed that we had very little prior information on the parameter values, and took a uniform prior over intervals ranging from 0.1$\times$ to 10$\times$ the original model value (*i.e.* $p(\boldsymbol{G})$ is equal to a positive constant within these

intervals and zero otherwise). Note that current-voltage (I-V) curves may provide more useful prior information for some ion currents where available.

Numerical considerations

We define the "log-likelihood surface" as the set of log-likelihood values given any set of conductance parameter values. Since this log-likelihood surface is defined by a solution to a system of ordinary differential equations (ODEs), we expect it to depend on the solver, as well as the parameters and equations themselves.

To illustrate this, we generate synthetic data from the O'Hara *et al.* (2011) model by solving the system of ODEs to obtain a time series of voltage points, and then adding Normally-distributed observation noise to every voltage value in this time series. Increasing the ODE solver's sampling frequency will slow down each MCMC iteration by slowing down each ODE solve. Since we want to run many MCMC iterations, this can add up to a large effect. Alternatively, decreasing the sampling frequency will speed up each solve, but the coarser time series is more likely to miss the peak of the action potential upstroke, which contains important information about the sodium current conductance. In practice, we choose the ODE solver's sampling frequency to be equal to the sampling frequency of the experimentally-recorded data we are analysing so that we can appropriately compute log-target values.

We run an MCMC algorithm to infer distributions for the conductance parameters given this synthetic dataset. We then take the set of parameter values that give the best log-likelihood value, and compare it to the set of parameter values used to generate the synthetic data. The uncertainty introduced by the observation noise means that these two sets of parameter values can be distinct points in parameter space. We then take a straight line connecting these two points in parameter space, and compute the log-likelihood values along this line. We repeat this for a number of different tolerances in the ODE solver. As shown in Figure 4.2, the (1-dimensional) log-likelihood surface becomes smoother as the solver tolerances are tightened. With laxer solver tolerances, the log-likelihood surface has many

local maxima, so the MCMC algorithm could get stuck more easily. With relative and absolute tolerances of 10^{-7} and 10^{-9}, respectively, the log-likelihood surface appears reasonably smooth, and so we use these values for future analysis.

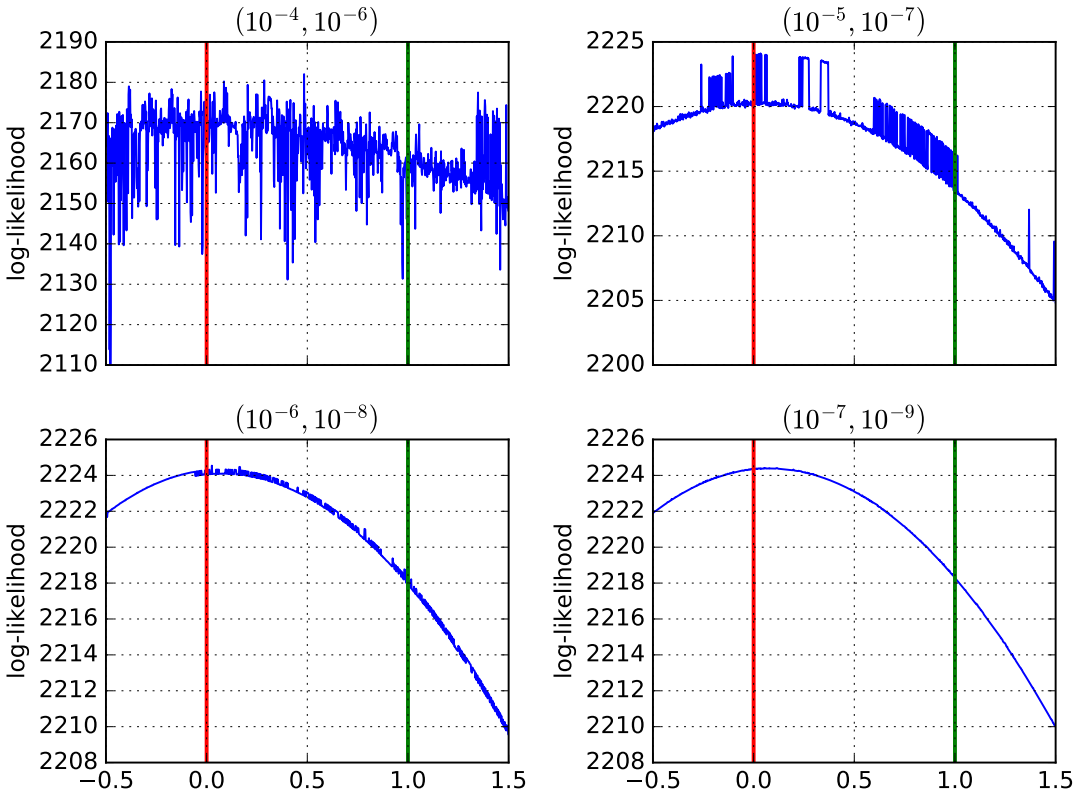

Figure 4.2: Log-likelihood values along the straight line connecting the best-fit parameter values found by the MCMC algorithm (red vertical line, left) and the parameter values used the generate the synthetic data (green vertical line, right). Log-likelihood values are given for different ODE solver tolerances of the form (relative, absolute). The (1-dimensional) log-likelihood surface becomes smoother as the solver tolerances are tightened.

Initial MCMC

For our proposal distribution, we choose an independent multivariate Normal distribution centred at the current accepted point (set of parameter values), which is equivalent to the products of univariate Normal distributions for each parameter, centred at the current parameter value. The initial variance of each parameter is proportional to the width of the prior interval for that parameter, and is scaled by the auto-adjusting scaling factor a. The pseudocode is given in Algorithm 2.

Algorithm 2 Metropolis-Hastings algorithm with adaptive proposal scale.

$\log(a_{[0]}) \leftarrow 0$
$\boldsymbol{G_{[0]}}$ sampled from independent uniform random distributions
$\Sigma \leftarrow D$, where D is a diagonal matrix
$t \leftarrow 0$
if $t > 1000d$ **then**
  $\gamma_{[t]} \leftarrow (t - 1000d + 1)^{-0.6}$
else
  $\gamma_{[t]} \leftarrow 0$
end if
loop
  Given the current parameter state $\boldsymbol{G_{[t]}}$, sample $\boldsymbol{G^*} \sim \mathcal{N}(\cdot|\boldsymbol{G_{[t]}}, a_t\Sigma)$
  if $p(\boldsymbol{G^*}) \neq 0$ **then**
    Compute $\log(\alpha) = l(\boldsymbol{G^*}) - l(\boldsymbol{G_{[t]}})$
    Sample $u \sim \mathcal{U}(0, 1)$.
    if $u < \alpha$ **then**
      $\boldsymbol{G_{[t+1]}} \leftarrow \boldsymbol{G^*}$
      $accepted \leftarrow 1$
    else
      $\boldsymbol{\theta_{t+1}} \leftarrow \boldsymbol{G_{[t]}}$
      $accepted \leftarrow 0$
    end if
  else
    $\boldsymbol{G_{[t+1]}} \leftarrow \boldsymbol{G_{[t]}}$
    $accepted \leftarrow 0$
  end if
  $\log(a_{[t+1]}) \leftarrow \log(a_{[t]}) + \gamma_{[t]} \times (accepted - 0.25)$.
  t++
end loop

As a first test case, we simulated some (synthetic) data from the Hodgkin & Huxley (1952a) model (discussed in Section 2.2 of Chapter 2) in the form of a single action potential, generated using the original model's maximal current conductance parameter values. At each point of this synthetic membrane voltage trace we added random noise, independently drawn from a Normal distribution with zero mean and standard deviation $\sigma = 0.25\,\text{mV}$. This value was chosen by computing the standard deviation of a 'flat' section of a real action potential recording provided to us by collaborators in Utrecht — we study this data in Chapter 5. We then attempted to infer probability distributions for them, *i.e.* likely sets of parameter values which could give rise to this (synthetic) data, ignoring the knowledge we already had about

the parameter values. In addition to inferring distributions for the conductance parameters, we also 'forgot' the level of noise we added to the voltage trace, and try to infer a distribution for σ too. In this first case, we are therefore inferring a joint posterior distribution for 4 parameters: the maximal ion current conductance parameters G_{Na}, G_K, and G_l, and the noise standard deviation σ.

Figure 4.3 shows a normalised histogram of G_{Na} samples and the path taken through parameter space during the course of the MCMC algorithm run. The normalised histogram approximates the marginal probability distribution for that parameter. It was run for 2,000,000 million iterations in total, with thinning performed by saving every 5$^{\text{th}}$ iteration, and discarding the first half of all samples as a burn-in.

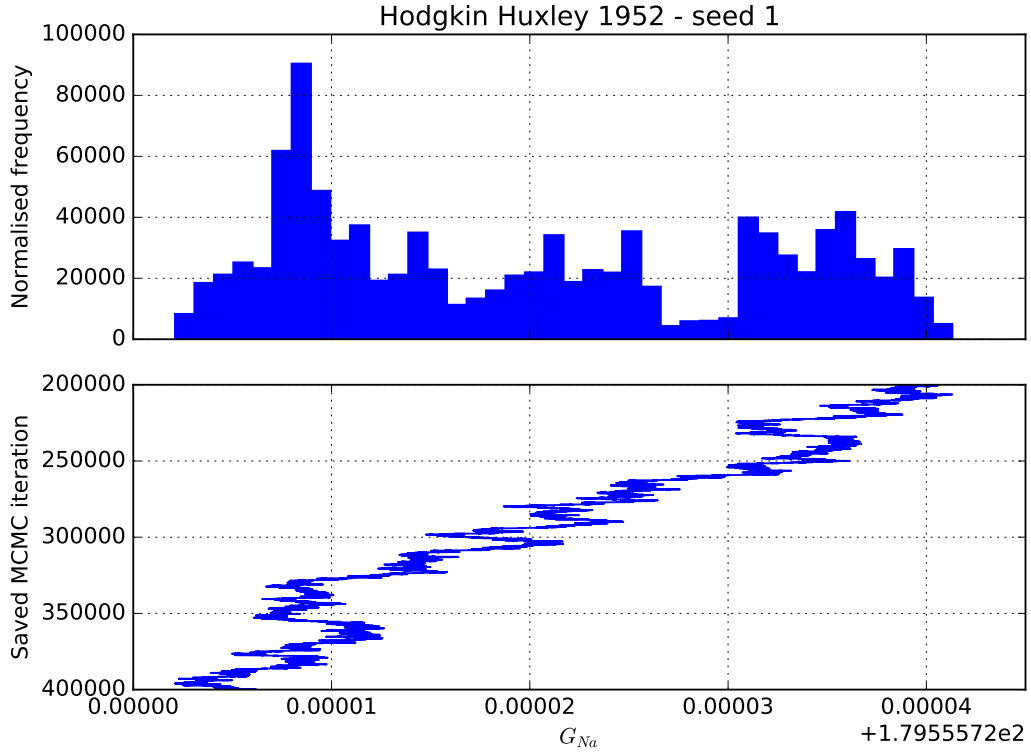

Figure 4.3: Marginal outputs of G_{Na} after running Algorithm 2 for 2,000,000 iterations, saving every 5$^{\text{th}}$ sample, and discarding the first half of all samples as a burn-in. The initial position was sampled randomly. Top: normalised histogram of the saved samples. Bottom: the path taken through parameter space by the chain.

We also plot a 3-dimensional scatter plot of a random subset of the same Markov chain as shown in Figure 4.3, representing random samples drawn from the posterior

distribution. This scatter plot is given, from two different angles, in Figure 4.4. We can see that samples tend to lie along a particular direction, suggesting co-dependence between the parameters, *i.e.* there is good identifiability of G_{Na}, G_K, and G_l relative to each other, but the combination can vary together. This means that the Markov chain can only take a path of similar log-target values through parameter space if all conductance parameter values are varying along the same vector together. Proposals that are approximately orthogonal to this direction are more likely to be rejected, and so there will be many 'wasted' proposed samples.

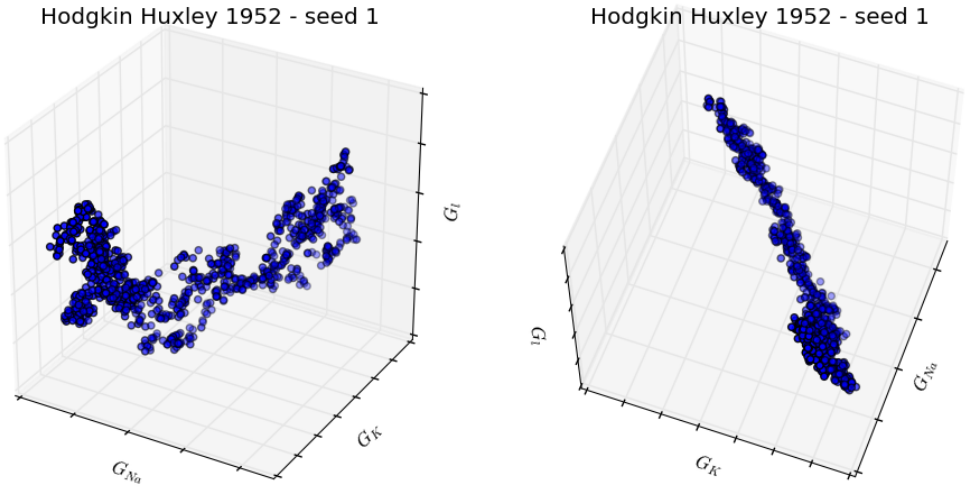

Figure 4.4: Scatter plots of 1000 samples from the chain shown in Figure 4.3, after discarding the first half of all samples as a burn-in. The two plots are the same samples, just viewed from different angles. The path suggests a proposal distribution in a certain direction will have more success.

It does not seem as though the Markov chain has reached an equilibrium distribution. However, due to our tuning of the parameter a, the final acceptance rate of this chain was 0.244 (3 d.p.), quite close to our target of 0.25. Looking more closely at a, we found that $\log(a) \approx -36.06$, giving a very small value to a, meaning that proposed states were very close, if not identical, to the currently-accepted states in the chain. We conclude that the chain had become stuck in some local optimum in parameter space, and that it was not effectively exploring the regions of interest.

We ran the same MCMC algorithm using different random seeds in Python, thus giving us different starting positions and different proposed points, but still

according to Algorithm 2. We found that in the majority of cases, the chain would get stuck in some similar kind of local optimum, without finding the region containing the 'true' parameter values, which is where we expect the best log-target values to be. Looking at the parameter values that gave rise to the best log-target scores in these chains, we found that they were often approximately some constant multiplied by the 'true' values. This is summarised in Table 4.1, and the relatively low standard deviation in "Best/true", in most cases, shows that each parameter is being multiplied by the same constant. This suggests that the log-target is generally greater when near to the line in parameter space defined by constant multiples of the 'true' parameter values.

Seed	Best	Best/true	Mean(best/true)	Std(best/true)
1	(179.56, 53.62, 0.44)	(1.5, 1.49, 1.48)	1.488	0.007
2	(120.06, 35.96, 0.3)	(1.0, 1.0, 1.0)	1.001	0.001
3	(712.69, 216.16, 1.77)	(5.94, 6.0, 5.89)	5.946	0.045
4	(1132.04, 317.16, 2.48)	(9.43, 8.81, 8.27)	8.837	0.476
5	(357.64, 111.66, 0.87)	(2.98, 3.1, 2.91)	2.998	0.079
6	(936.68, 272.92, 2.23)	(7.81, 7.58, 7.44)	7.609	0.151
7	(185.95, 53.83, 0.43)	(1.55, 1.5, 1.45)	1.497	0.042
8	(1099.26, 309.34, 2.47)	(9.16, 8.59, 8.22)	8.658	0.386
9	(77.37, 24.2, 0.22)	(0.64, 0.67, 0.73)	0.684	0.038
10	(119.94, 35.93, 0.3)	(1.0, 1.0, 1.0)	0.999	0.001
11	(115.71, 34.55, 0.29)	(0.96, 0.96, 0.97)	0.965	0.004
12	(266.52, 79.49, 0.62)	(2.22, 2.21, 2.07)	2.168	0.066
13	(805.74, 240.38, 1.96)	(6.71, 6.68, 6.52)	6.637	0.084
14	(657.63, 203.12, 1.65)	(5.48, 5.64, 5.5)	5.542	0.072
15	(714.42, 215.62, 1.73)	(5.95, 5.99, 5.77)	5.905	0.095
16	(308.55, 93.02, 0.74)	(2.57, 2.58, 2.47)	2.542	0.051
17	(391.36, 124.6, 1.02)	(3.26, 3.46, 3.39)	3.37	0.083
18	(809.64, 242.89, 1.95)	(6.75, 6.75, 6.51)	6.667	0.113

Table 4.1: 18 runs of Algorithm 2, with different Python random seeds. "Best" refers to the sample with the best log-target score in the entire chain. "Best/true" refers to these best conductance parameter values divided their respective 'true' values. The highlighted rows are where the chain found the 'true' parameter values.

This suggests two ideas that might increase efficiency and accuracy of our MCMC algorithm:

1. allow the chain to search along particular directions which have good log-target values, and

2. choose a better-informed starting position.

Adaptive covariance matrix Metropolis-Hastings

In order to allow the MCMC algorithm to find directions in parameter space that have good log-target values, we allow the proposal distribution's covariance matrix to adapt, in addition to its scaling factor. That is, when new points are accepted into the chain, the proposal covariance matrix skews itself in that direction, so it is more likely to keep travelling in that direction.

We ran this MCMC algorithm, detailed in Algorithm 1 in Chapter 2, from the same 18 starting positions as in Section 4.1.1 with the same respective Python random seeds. Eight, as opposed to the original two, of the chains were able to find the 'true' parameter values and then efficiently explore this region of parameter space, with the remaining chains getting stuck in local optima. A summary of the results of these MCMC algorithm runs is given in Table 4.2.

Seed	Best	Best/true	Mean(best/true)	Std(best/true)
1	(177.33, 52.46, 0.43)	(1.48, 1.46, 1.44)	1.457	0.017
2	(120.11, 35.97, 0.3)	(1.0, 1.0, 1.0)	1.001	0.001
3	(744.55, 224.82, 1.82)	(6.2, 6.24, 6.05)	6.168	0.082
4	(1502.85, 398.84, 3.18)	(12.52, 11.08, 10.59)	11.398	0.821
5	(120.17, 35.99, 0.3)	(1.0, 1.0, 1.0)	1.002	0.002
6	(1466.12, 391.16, 3.12)	(12.22, 10.87, 10.4)	11.159	0.772
7	(126.39, 38.0, 0.31)	(1.05, 1.06, 1.03)	1.048	0.009
8	(1094.55, 308.41, 2.44)	(9.12, 8.57, 8.12)	8.604	0.408
9	(120.05, 35.95, 0.3)	(1.0, 1.0, 1.0)	1.0	0.001
10	(120.11, 35.98, 0.3)	(1.0, 1.0, 1.0)	1.001	0.001
11	(120.22, 36.0, 0.3)	(1.0, 1.0, 1.0)	1.002	0.001
12	(258.73, 78.07, 0.61)	(2.16, 2.17, 2.04)	2.121	0.059
13	(831.57, 247.0, 1.96)	(6.93, 6.86, 6.53)	6.775	0.173
14	(677.35, 205.11, 1.64)	(5.64, 5.7, 5.47)	5.603	0.099
15	(1219.71, 336.17, 2.68)	(10.16, 9.34, 8.95)	9.483	0.507
16	(120.12, 35.97, 0.3)	(1.0, 1.0, 1.0)	1.001	0.001
17	(120.09, 35.97, 0.3)	(1.0, 1.0, 1.0)	1.001	0.001
18	(1486.21, 395.04, 3.15)	(12.39, 10.97, 10.49)	11.284	0.802

Table 4.2: 18 runs of Algorithm 1, with different Python random seeds, and the same respective starting positions as in Table 4.1. "Best" refers to the sample with the best log-target score in the entire chain. "Best/true" refers to these best conductance parameter values divided their respective 'true' values. The highlighted rows are where the chain found the 'true' parameter values.

Initial log-target optimisation

As discussed in Chapter 2.5.3, we perform a CMA-ES optimisation, described in Section 2.5.3 of Chapter 2, before starting the adaptive Metropolis-Hastings algorithm. We do this so that the MCMC algorithm starts in a region of good log-target space, and so that we do not waste time while the algorithm searches for better regions of log-target space. Since we are using Uniform priors, we just optimise the log-likelihood, as any values allowed by the prior just add a constant to the log-target value, while disallowed values add $-\infty$ and will therefore never be accepted into the chain, or accepted as an improved set of parameter values during the initial optimisation. This is equivalent to minimising the sum of square differences between the data trace and model output trace, varying only the maximal conductance parameters. Given the form of Equation (2.23), the optimal value of σ can then be analytically computed. Snapshots of this initial optimisation are given in Figure 4.5, where the fits obtained by the CMA-ES optimiser are plotted at different iterations, along with the current parameter values as percentages of their 'true' values.

This optimisation will hopefully give us a point estimate for the greatest log-target value. We have only ever found one mode when modelling action potential data; there have been identifiability issues with some parameters, meaning that the mode is not a maximum along the these parameters' axes, but there have been no distinct modes other than the ones returned by CMA-ES.

Figure 4.6 shows an example of the log-likelihood of the parameter set during a CMA-ES run, followed by each iteration of the MCMC algorithm, using the Luo & Rudy (1991) model because it has only 6 conductance parameters and so both the CMA-ES and MCMC do not take as long to converge as more complex models. Sets of parameters at three different points have been selected and their corresponding action potential traces plotted over the experimental trace to show the convergence of the minimisation and subsequent local search by MCMC. All of the simulations were performed using the Chaste C++ library (Mirams *et al.*, 2013; Cooper *et al.*, 2014). We combined these methods to perform inference on synthetic

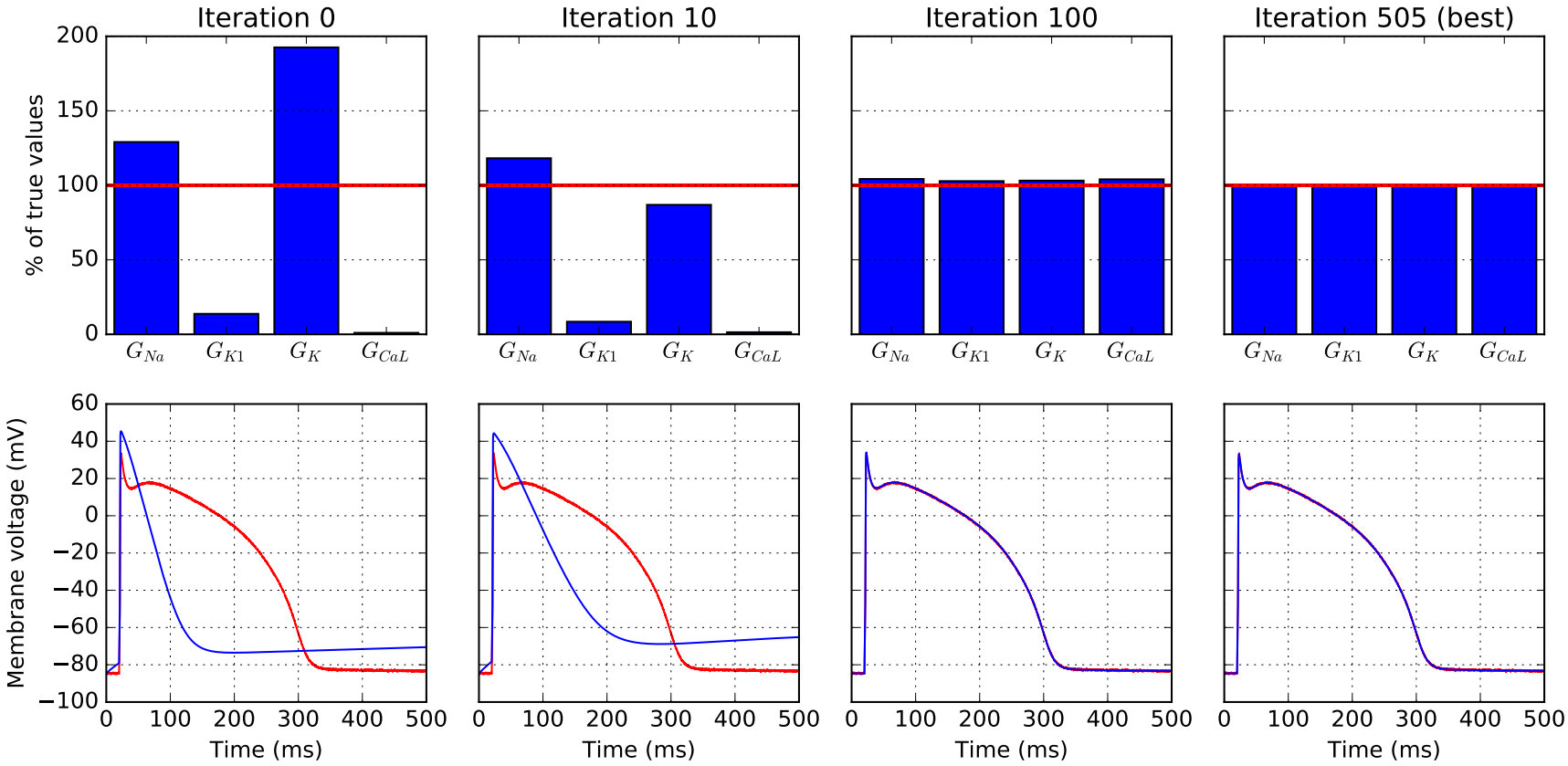

Figure 4.5: Sum-of-squares minimisation by CMA-ES of the Beeler & Reuter model, with synthetic data generated using the original published conductance parameter values. Top: the current iteration's parameter values as a percentage of the values used to generate the synthetic data. Bottom: the action potential trace generated using the current iteration's parameter values (blue), and the synthetic data trace (red).

data, generated as described in Section 4.1.1, from a range of action potential models of increasing complexity, *i.e.* increasing number of maximal conductance parameters.

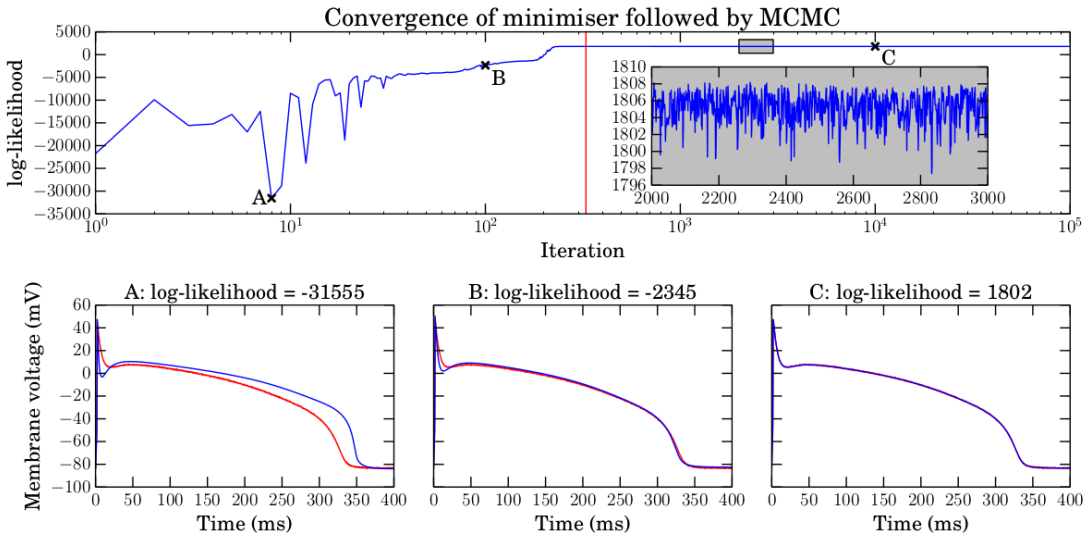

Figure 4.6: Top: synthetic data example of optimisation and MCMC for the Luo & Rudy (1991) model. The Luo & Rudy (1991) model only contains 6 conductance parameters and so the CMA-ES and MCMC do not take as long to run as with the more complex action potential models. The negative log-likelihood is minimised to find a starting point for the MCMC algorithm followed by log-likelihoods obtained during the MCMC algorithm run. The MCMC begins at the vertical red line, note the log scale on the x–axis — minimisation is a much smaller proportion of the total number of iterations than the MCMC. Points A, B and C are at iterations 8, 100 and 2000, respectively. Inset is a zoomed in view of the highlighted box, showing the MCMC algorithm exploring the log-likelihood between iterations 2000 and 3000. Panels A–C: the blue traces are the APs for the proposed parameters at iterations marked as A, B and C. Red traces are the same synthetic data in all 3 plots.

Using noisy synthetic data as our voltage recordings provides a best-case scenario of the recordings we would observe if the action potential model was 'correct' (eliminating the need to consider structural uncertainty). In practice, this means that in the case of identifiable parameters, we can typically recover the parameter values used to generate the synthetic data to a level of accuracy that depends on the level of observation noise. Increasing the level of observation noise typically widens the inferred marginal distributions for the identifiable parameters, but does not typically change the point of maximum posterior density drastically.

4.1.2 Results

In this section we examine how much information on individual ion current conductances is present in synthetic data from a range of protocols, in terms of the certainty we gain in possible conductance values after performing the inference described above.

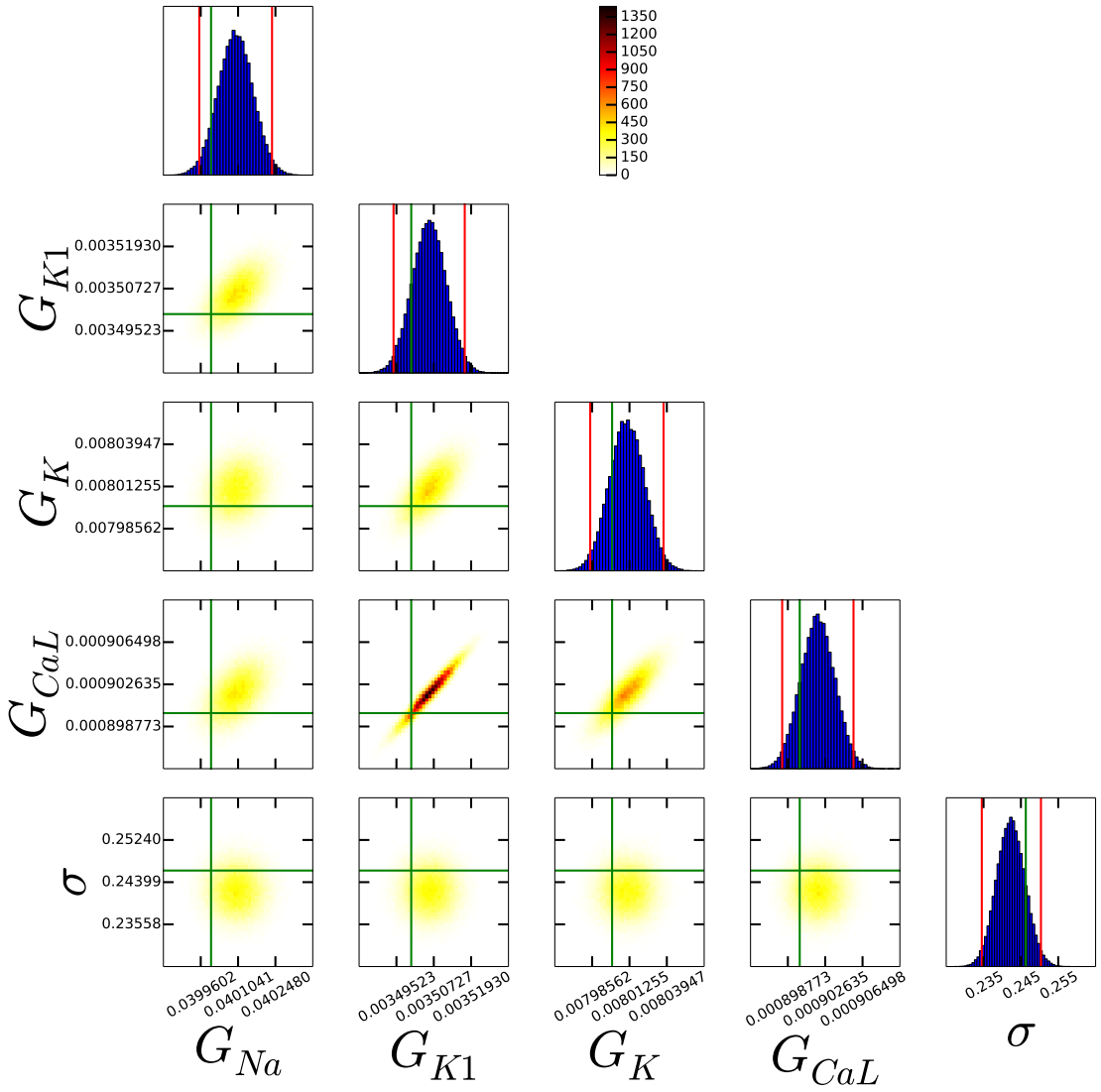

Figure 4.7: Marginal and 2-d normalised histograms representing marginal and pairwise probability distributions of each parameter, using the Beeler & Reuter model and single action potential protocol. All parameters have been successfully inferred from the data (see main text). The vertical red (outside) lines give the 95% credible interval. The vertical green (inside) line is the parameter value used to generate the synthetic data.

The first experimental protocol we tested was 1 Hz pacing, recording the

membrane voltage for sufficiently long to capture a single action potential (as shown in Figure 4.6 A–C). As an example, Figure 4.7 shows an array of scatter plots showing parameter distributions that were inferred for the Beeler & Reuter (1977) model. In this plot the main diagonal contains histograms of the values each parameter takes through the MCMC run, and the off-diagonal entries contain heat maps showing the proportion of the run spent in that area for each pair of parameters. Also plotted are: the original parameter values, as vertical and horizontal green lines across the plots; and 95% credible intervals for each parameter as vertical red lines on the histograms (the intervals within which 95% of the values lie — constructed by 'discarding' the outermost 2.5% of the samples from the histogram, on each side).

In Figure 4.7 all of the probability distributions are very narrow, in that the credible intervals (red lines) are very close, in absolute terms, to the original parameter values (green lines). We conclude that all 4 conductance parameters are successfully inferred from a single action potential trace, and the noise parameter σ has also been inferred well. From this matrix-style plot, we can also observe pairwise dependence between parameters. For example, G_{CaL} and G_{K1} appear to be strongly positively correlated. This suggests that when using a drug that blocks I_{CaL}, say, we might expect I_{K1} to also decrease. We can therefore make predictions about possible side effects of drugs with particular targets.

To quantify this, we require a measure of how successful the inference process has been for any model and protocol. For each parameter θ_i, we define its *mean absolute percentage error* (M_i) across the MCMC run (excluding the burn-in period) as:

$$M_i = \frac{1}{N} \sum_{t=1}^{N} \left| \frac{\hat{\theta}_i - \theta_{t,i}}{\hat{\theta}_i} \right|, \tag{4.2}$$

where $\hat{\theta}_i$ is the actual value of the parameter, *i.e.* the value used when creating the synthetic data, $\theta_{t,i}$ is the value of parameter i at iteration t of the MCMC run, and N is the total number of iterations in the MCMC run. We say we have successfully inferred a parameter if $M_i < 0.05$ (on average, the parameter value in the chain is within 5% of the actual value of the parameter).

A summary of the number of maximal conductance parameters successfully inferred for each model under the protocols is given in Table 4.3. The noise parameter σ is always successfully inferred, and therefore omitted from the table. A more detailed table of which conductance parameters were successfully inferred from the ten Tusscher *et al.* (2004), O'Hara *et al.* (2011) & Davies *et al.* (2011) models is given in Table 4.4. "Success" refers to a parameter having a MAPE less than 0.05.

For simple models (with $\leqslant 6$ conductance parameters), a single 1 Hz action potential provides enough information for us to successfully infer all conductance parameters. However, a single action potential did not provide enough information for us to successfully infer all parameters from the more complex models. For example, Figure 4.8 shows a subset of the scatter plot array from the ten Tusscher *et al.* model. Here, G_{Na} is inferred successfully, and we see narrow histograms for this parameter. However, the background currents G_{bCa} and G_{bNa} are not successfully inferred, and their histograms extend to the edge of the prior range we considered ($0.1\times$ the original value). There is also a strong interdependence between the two parameters visible in their pairwise plot; it seems that a high value for G_{bCa} and low value for G_{bNa} fit the data just as well as a low value for G_{bCa} and high value for G_{bNa}. This dependence is an indication that there is nonidentifiability of parameters — multiple parameter sets fit the data equally well, and the individual conductances cannot be identified (using this protocol). P_{NaK} appears to have formed a well-defined distribution, but it is too wide to meet our criterion of success.

Other protocols shown in Table 4.3 include halving and doubling the extracellular potassium concentration, $[K^+]_o$, when measuring a single AP. The latter reduced the number of conductance parameters we could infer in the more complex models. In particular, using the ten Tusscher *et al.* (2004) model and doubling $[K^+]_o$, the histograms for G_{K1}, G_{Kr}, and G_{Ks} are shifted to lower values (shown in Figure 4.9). As such, the values moved away from the true parameters and so their M_i values increased. In an attempt to obtain a more informative voltage trace, we investigated longer and more complex protocols. We used an S1-S2 pacing protocol, 1 Hz followed by 2 Hz, and measured for a total of 2000 ms, giving a trace with three APs. In

Model	1 Hz 1 AP	$0.5[K^+]_o$ 1 AP	$2[K^+]_o$ 1 AP	1 & 2 Hz 3 APs	1 Hz for 2000 ms, $G_x = 0$ at 1000 ms					1 Hz 10 s	1 Hz 20 s
					G_{CaL}	G_{K1}	G_{Kr}	G_{Ks}	G_{to}		
Hodgkin & Huxley (1952a)	3/3	n/a	n/a	3/3	n/a	n/a	n/a	n/a	n/a	3/3	3/3
Noble (1962)	4/4	n/a	n/a	4/4	n/a	4/4	n/a	n/a	n/a	4/4	4/4
Beeler & Reuter (1977)	4/4	n/a	n/a	4/4	4/4	4/4	n/a	n/a	n/a	4/4	4/4
Luo & Rudy (1991)	6/6	6/6	6/6	6/6	6/6	6/6	n/a	n/a	n/a	6/6	6/6
ten Tusscher *et al.* (2004) epi.	7/12	7/12	5/12	9/12	10/12	11/12	10/12	8/12	10/12	11/12	12/12
O'Hara *et al.* (2011) endo.	5/13	6/13	3/13	9/13	10/13	10/13	10/13	10/13	9/13	10/13	11/13
Davies *et al.* (2011)	5/14	6/14	5/14	10/14	9/14	11/14	11/14	12/14	11/14	12/14	12/14

Table 4.3: Number of conductance parameters that were constrained sufficiently given a particular model and simulated protocol. Our definition of success is that $M < 0.05$ for a parameter for the length of the MCMC chain (having removed the first quarter as a burn-in period). 'n/a' stands for 'not applicable', and indicates that the models do not include the components necessary to perform these protocols. N.B. G_{to1} appears in Davies *et al.* (2011), but G_{to} is in all other models in which the current is modelled.

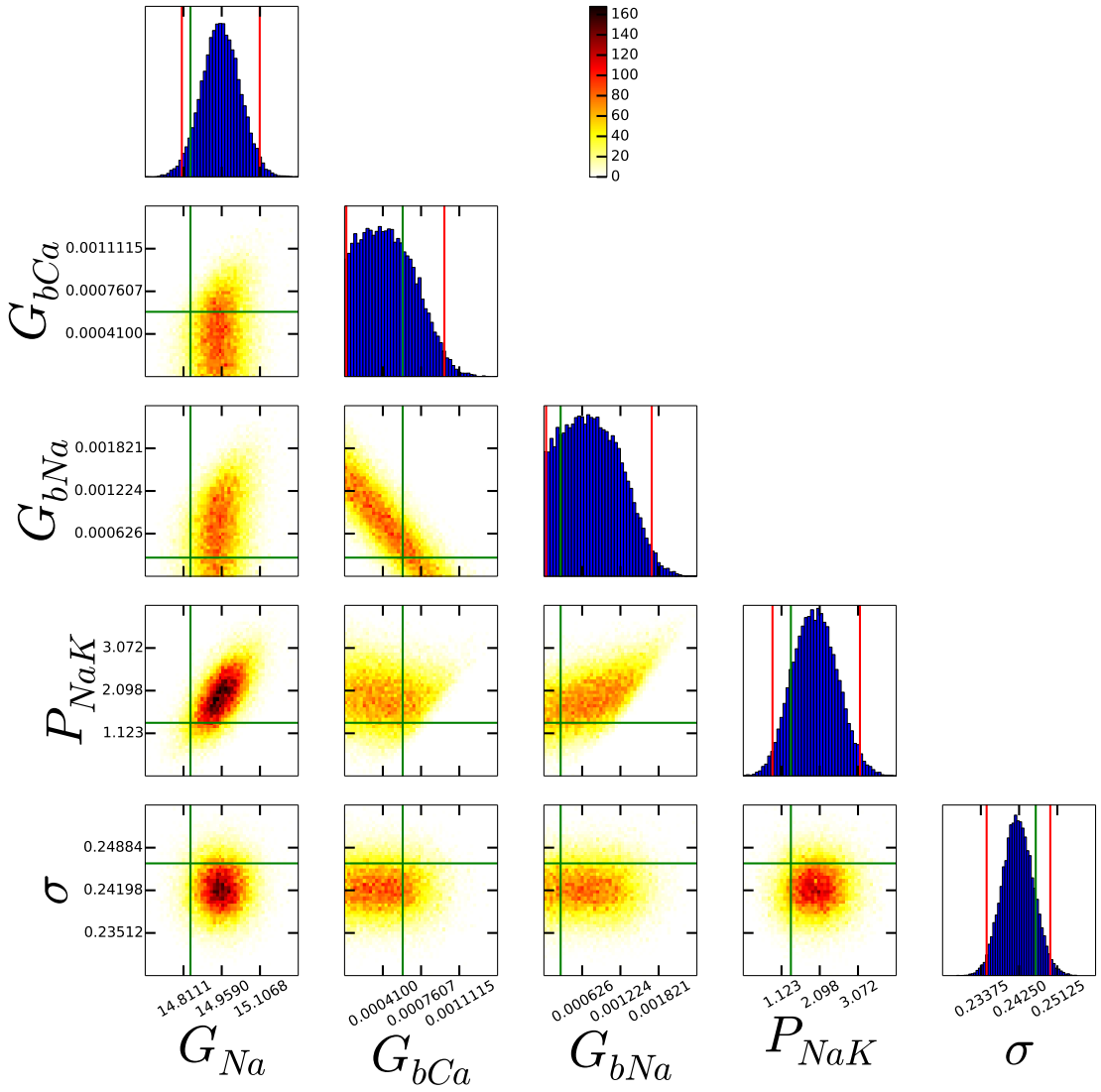

Figure 4.8: Subset of the histograms and pairwise probability distributions of each parameter using the ten Tusscher *et al.* model and single action potential protocol for synthetic data. Not all parameters have been successfully inferred from the data, *e.g.* G_{bCa}, G_{bNa} and P_{NaK}. P_{NaK} has the shape of a typically converged distribution, but fails our test of success by being too wide. The vertical red (outside) lines give the 95% credible interval. The green (inside) lines are the parameter values used to generate the synthetic data.

the case of the two most complex models, this increased the overall number of successfully inferred conductances, presumably as different currents give different contributions at the two rates, but we still could not infer all conductances.

We then tested a series of protocols in which we simulated the effects of pharmaceutical block of particular ion currents. With a 1 Hz pacing rate, we measured for a total of 2000 ms and therefore obtained 2 APs. At 1000 ms we set one maximal conductance parameter to zero, knocking out this current from the cell. This represents an 'ideal' case where the block is 100% effective and specific, and works instantaneously. Intuitively, one would expect such a protocol to give good information on the conductance of the current that was blocked, since it directly shows the effect on the action potential of removing a particular ion current. The conductances we set to zero can always be inferred with these protocols, and in many cases the alteration in behaviour allows many additional conductances to be inferred as well (see Table 4.4 for the full results).

The last protocols that were tested were 1 Hz pacing for 10,000 ms or 20,000 ms. These longer protocols allow slower time-scale dynamics to reveal their effects, and therefore for information about the conductances that govern these changes to be inferred more readily than from a single AP, for example the smaller background and pump currents. Somewhat remarkably, the inference process was able to successfully recover all 12 conductances in the ten Tusscher *et al.* (2004) model, 11/13 in the O'Hara *et al.* (2011) model, and 12/14 in Davies *et al.* (2011) using just twenty 1 Hz APs. The two conductances in the O'Hara *et al.* (2011) model which were not successfully inferred from the 20 s protocol are G_{bNa} and G_{pCa}, and in the Davies *et al.* (2011) model G_{bCl} and G_{pCa}.

Superimposed normalised histograms of inferred distributions for G_{K1}, G_{Kr} and G_{Ks} using the ten Tusscher *et al.* model under two protocols are given in Figure 4.9. Note that all inferred distributions in the $2[\mathrm{K}^+]_o$ protocol are shifted to the left, *i.e.* the inferred conductance values are generally smaller than when $\mathrm{K}^+]_o$ is left unchanged.

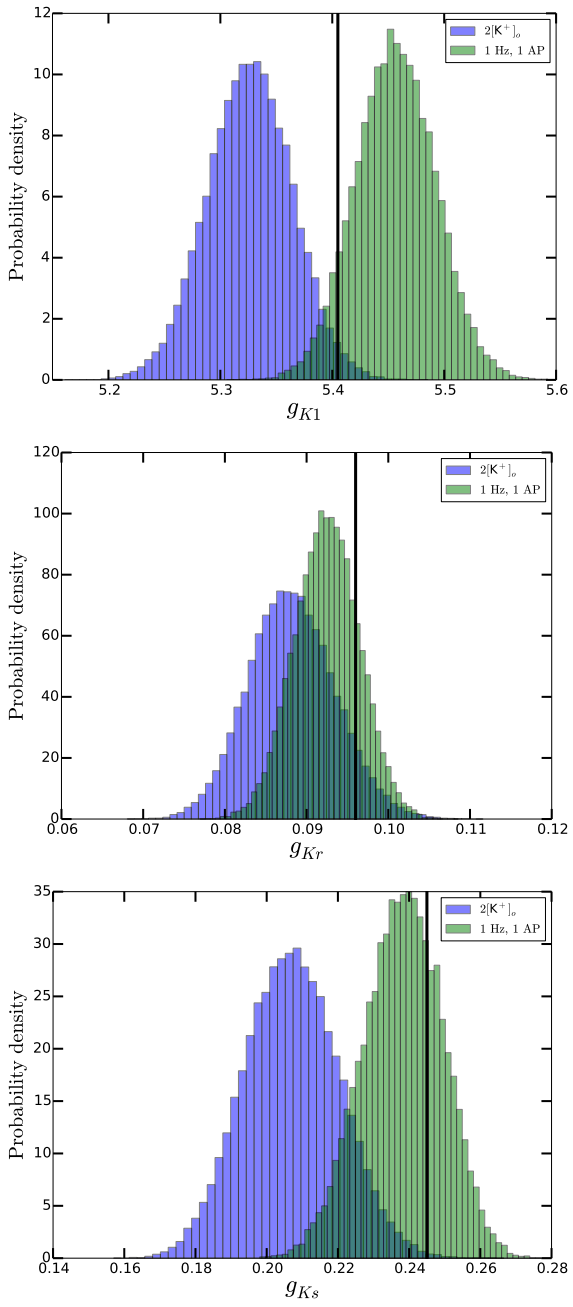

Figure 4.9: Superimposed normalised histograms for G_{K1}, G_{Kr} and G_{Ks} for two protocols using the ten Tusscher *et al.* model. The blue histograms are from the $2[K^+]_o$ protocol, the green are from the 1 Hz single action potential protocol. The vertical black lines are the original values given in the model.

A summary of which conductances were successfully inferred from the ten Tusscher *et al.* (2004), O'Hara *et al.* (2011) and Davies *et al.* (2011) models is given in Table 4.4. All conductances were successfully inferred for the four simpler models.

4.2 Modelling action potential variability in a Bayesian framework

In this section, we will use Bayesian statistical models to explore variability between action potentials. This is observed for many reasons, for example when recording single action potentials from different cells, or recording multiple action potentials from a single cell. We are exploring variability in maximal ion current conductances, which is often thought of as being due to differences in ion channel gene expression. Between cells, there is expected variability in gene expression, giving rise to different action potential behaviour. However, in the case of observed beat-to-beat variability within a single cell, it is likely that maximal conductances are varying for different reasons. A possible explanation is the stochastic nature of ion channel opening, where voltage-dependence describes the probability, or average openness, of an ion channel at a particular voltage. This stochasticity is always present, but we consider a 'mean field approximation' of conductances throughout a single pace. Under this assumption, we would suppose that a cell's maximal conductances are constant within a single pace, but can differ between paces.

In this section, we will use synthetic data to explore 'general' variability in action potential behaviour — not necessarily beat-to-beat variability as described above. We will compare a similar statistical model to the one presented in Section 4.1.1 (that we will refer to as a "single-level" statistical model) with a "hierarchical" statistical model. In the single-level model, each dataset (we will also refer to a dataset as an "experiment") is considered separately, and probability distributions for the model parameters are inferred just from that one experiment. In the hierarchical model, we suppose that each experiment has its own set of action potential model parameter values, but that these are related to each other by some inter-experiment probability distribution. We will attempt to infer a (joint) probability distribution

Parameter	Protocol										
	1 Hz 1 AP	$0.5[K^+]_o$ 1 AP	$2[K^+]_o$ 1 AP	1 & 2 Hz 3 APs	1 Hz for 2000 ms, $G_x = 0$ at 1000 ms					1 Hz 10 s	1 Hz 20 s
					G_{CaL}	G_{K1}	G_{Kr}	G_{Ks}	G_{to}		
ten Tusscher *et al.* (2004)											
G_{Na}	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
G_{CaL}	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
G_{K1}	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
G_{Kr}	✓	✓	✗	✓	✓	✓	✓	✓	✓	✓	✓
G_{Ks}	✓	✓	✗	✓	✓	✓	✓	✓	✓	✓	✓
k_{NaCa}	✗	✗	✗	✓	✓	✓	✓	✗	✓	✓	✓
G_{to}	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
G_{bCa}	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	✓
G_{bNa}	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓
G_{pCa}	✗	✗	✗	✓	✓	✓	✗	✗	✓	✓	✓
G_{pK}	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
P_{NaK}	✗	✗	✗	✗	✗	✓	✗	✗	✗	✓	✓
O'Hara *et al.* (2011)											
G_{Na}	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
G_{CaL}	✗	✓	✗	✓	✓	✓	✓	✓	✓	✓	✓
G_{bK}	✗	✗	✗	✓	✓	✓	✓	✓	✗	✓	✓
G_{K1}	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
G_{Kr}	✓	✓	✗	✓	✓	✓	✓	✓	✓	✓	✓
G_{Ks}	✗	✗	✗	✓	✓	✓	✓	✓	✗	✓	✓
k_{NaCa}	✗	✗	✗	✓	✓	✓	✓	✓	✓	✓	✓
P_{NaK}	✗	✗	✗	✗	✓	✗	✓	✗	✗	✓	✓
G_{to}	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
G_{bCa}	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓
G_{bNa}	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗
G_{pCa}	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗
G_{NaL}	✓	✓	✗	✓	✓	✓	✓	✓	✓	✓	✓
Davies *et al.* (2011)											
G_{Na}	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
G_{CaL}	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
G_{K1}	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
G_{pK}	✗	✓	✗	✓	✓	✓	✓	✓	✓	✓	✓
G_{Ks}	✗	✗	✗	✗	✗	✗	✗	✓	✓	✓	✓
G_{Kr}	✗	✗	✗	✓	✓	✓	✓	✓	✓	✓	✓
G_{pCa}	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗
G_{bCa}	✗	✗	✗	✗	✗	✓	✓	✓	✗	✓	✓
k_{NaCa}	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
P_{NaK}	✗	✗	✗	✓	✗	✓	✓	✓	✓	✓	✓
G_{to1}	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
G_{to2}	✓	✗	✗	✓	✓	✓	✓	✓	✓	✓	✓
G_{bCl}	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗
G_{NaL}	✗	✗	✗	✓	✓	✓	✓	✓	✓	✓	✓

Table 4.4: Inference success for each conductance in the ten Tusscher *et al.* (2004), O'Hara *et al.* (2011) and Davies *et al.* (2011) models. "Success" (✓) is defined by MAPE < 0.05, defined in Equation (4.2). The other four models had every parameter successfully inferred for every protocol.

for these experiment-level parameters, as well as the parameters governing the inter-experiment variability. We will compare the predictive power of each of these statistical models when considering future action potential behaviour.

4.2.1 Model parameterisation

The number of membrane proteins is governed by the level of expression of the particular gene that encodes these proteins. As such, a maximal ion current conductance parameter can be thought of as a measure of a particular gene expression. Gene expression is the product of many different chemical processes. Beal (2017) showed that extrinsic variability in gene expression levels in biological systems can be well-approximated by log-Normal distributions as a result of many different probability distributions — representing stochastic chemical processes — being multiplied together.

We therefore use a new Bayesian statistical model to describe maximal ion current conductances, where we work with the natural log of the conductance parameter instead of the conductance parameter itself. This is because if G is log-Normally distributed, then $\log(G)$ is Normally distributed. Working with $\log(G)$ naturally enforces the constraint of positive conductances. Using the same statistical model as before also means that we can effectively reuse our existing models and code for performing parameter inference, even though we are inferring distributions for a transformed variable.

4.2.2 Synthetic data

As before, we first test our inference method and statistical model on synthetic data that we generate, so that we know the 'correct answer' and can verify our results.

We define Normal distributions from which log(conductance) parameter values are drawn. The means of these distributions were chosen to be the natural log of the action potential model conductance parameter values, since we know these parameter values produce plausible action potential behaviour, and so we expect other action potential-producing parameter values might be near to these model values. The

standard deviations were manually adjusted to provide a realistic range of action potential morphologies. Since we are working with log(conductance) and not simply conductance, the standard deviation can be the same size for all parameters. We therefore chose the same standard deviation for each parameter, which was 0.01. All experimental parameter values were sampled independently from each other.

To speed up computation time when solving the action potential model ODEs, we sampled the voltage trace every 1 ms, starting the model from its original initial conditions (1 Hz pacing steady state). Normally-distributed observation noise with standard deviation 0.5 mV was then added to every point of the trace. This value was chosen by measuring the level of noise on real action potential traces that I recorded from a stem cell-derived cardiomyocyte at Roche in Basel, which is discussed further in Chapter 6. These measured levels of noise varied between approximately 0.3 and 0.6 mV, with the majority being closer to 0.4 mV. We chose 0.5 mV to allow a greater number of possible solutions to be fitted to the data. The greater level of noise used here than in Section 4.1.1 will allow for possibly-wider posterior distributions.

32 synthetic datasets were generated for the Beeler & Reuter (1977) and O'Hara *et al.* (2011) models. These synthetic action potential traces are given in Figure 4.10. All randomly-sampled sets of parameter values produced viable action potentials, and so we did not have to further constrain the synthetically-generated experimental parameter values.

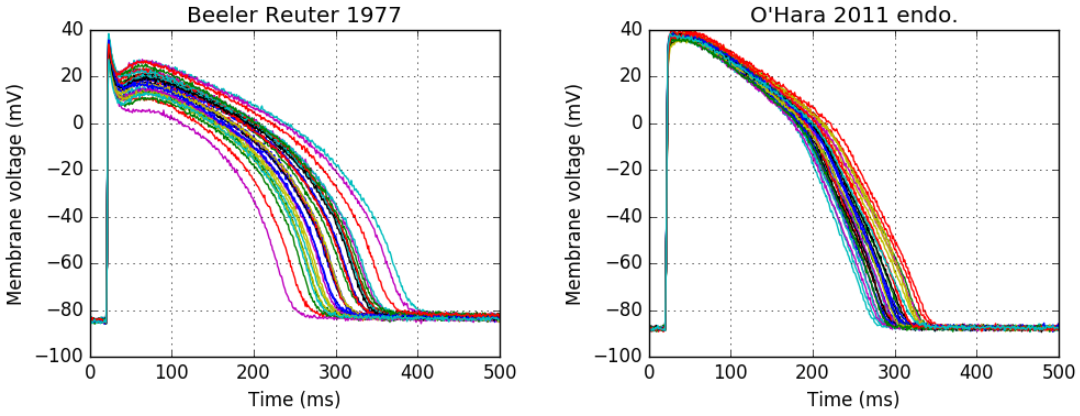

Figure 4.10: 32 synthetic action potential traces generated by sampling log(conductance) parameter values from Normal distributions.

4.2.3 Single-level statistical model

We adapt the statistical model shown in Figure 4.1, using $\log(G)$ instead of G as discussed above. This new statistical model is given in Figure 4.11.

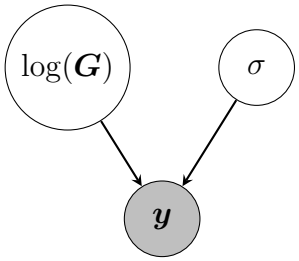

Figure 4.11: Parameter interaction and dependence under the statistical model for a single action potential recording. $\log(\boldsymbol{G})$ is the set of the natural logs of the maximal ion current conductance parameters which are used to generate $\boldsymbol{y}$, the dataset. All non-shaded variables are parameters for which we wish to infer probability distributions. Parameters with no inward-pointing arrows require prior distributions to be specified.

We will use different priors for the log(conductance) parameters than the uniform distributions used in Section 4.1.1. We choose a Normal prior distribution for each log(conductance) parameter to reflect the log-Normal distributions of gene expression levels discussed above. We choose the mean of the Normal prior distribution to be the natural log of the model conductance parameter value, $\log(\hat{G})$. Similar to the choice of prior distribution in Section 4.1.1, we want there to still be reasonable probability mass between $0.1\times$ and $10\times$ the model conductance parameter value. We therefore choose a standard deviation ω so that this distribution is wide enough to capture these extreme cases. That is, we choose ω such that $2\omega = \log(10\hat{G}) - \log(\hat{G}) = \log(10)$. The prior distribution for a general log(conductance) parameter $\log(G)$ is given in Figure 4.12.

As in Section 4.1.1, we choose a Uniform prior distribution for the observation noise standard deviation parameter σ. We expect to never have perfect recording equipment, and so we choose a lower bound of $10^{-3}\,\text{mV}$, and we choose a generous upper bound of $25\,mV$.

We then have the (non-normalised) posterior distribution for our statistical

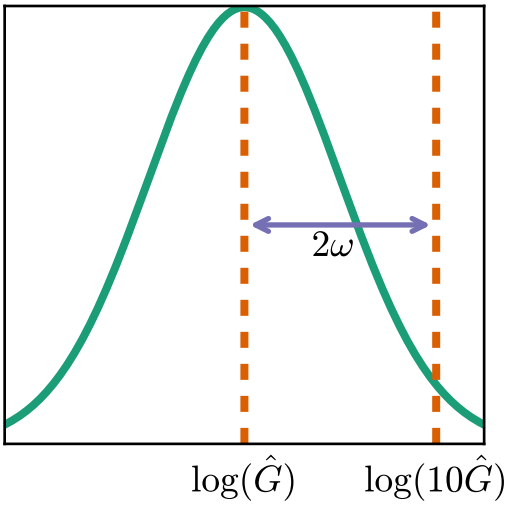

Figure 4.12: Normal prior distribution for $\log(G)$, where G is any conductance parameter. It is centred on $\log(\hat{G})$, where $\hat{G}$ is the value of G given in the action potential model of interest. It has standard deviation $\omega = \log(10)/2$.

model parameters $\log(\boldsymbol{G})$ and σ:

$$\begin{aligned}\pi(\log(\boldsymbol{G}), \sigma) \propto & \left(\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_i - f_i(\boldsymbol{G}))^2}{2\sigma^2}\right)\right) \\ & \times \left(\prod_{j=1}^{d} \frac{1}{\sqrt{2\pi\omega^2}} \exp\left(-\frac{(\log(G^{(j)}) - \log(\hat{G}^{(j)}))^2}{2\omega^2}\right)\right) \\ & \times I(10^{-3} < \sigma < 25), \end{aligned} \tag{4.3}$$

where the first term on the right-hand side is the likelihood of $\{\log(\boldsymbol{G}), \sigma\}$ given the data, the second term is the prior density of $\log(\boldsymbol{G})$, and the final term is the prior density of σ, where I is an indicator function providing a finite, positive constant value if the condition inside the parentheses is satisfied, and 0 otherwise. d is the number of conductance parameters in the action potential model, and j indexes these conductance parameters.

We will use the same adaptive Metropolis-Hastings algorithm described in Section 2.5.3 of Chapter 2. As usual, we work with the natural log of the (non-

normalised) target distribution, *i.e.*

$$\log(\pi(\log(\boldsymbol{G}),\sigma)) \propto_+ l(\log(\boldsymbol{G}),\sigma) + \log(p(\log(\boldsymbol{G}),\sigma)) \tag{4.4}$$

$$\propto_+ l(\log(\boldsymbol{G}),\sigma) + \log(p(\log(\boldsymbol{G})) + \log(p(\sigma)) \tag{4.5}$$

$$\propto_+ -n\log(\sigma) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - f_i(\boldsymbol{G}))^2 - \frac{1}{2\omega^2}\sum_{j=1}^{d}\left(\log(G^{(j)}) - \log(\hat{G}^{(j)})\right)^2, \tag{4.6}$$

where $\propto_+$ means "up to the addition of a constant".

4.2.4 Hierarchical statistical model

In all biological systems, there is intrinsic and extrinsic variability, as discussed in Section 2.3 of Chapter 2. This means that when an experiment is performed, such as current-clamp experiments for measuring membrane voltage action potentials, the recorded data will vary between experiments, even if the (controllable) experimental conditions have not changed.

Considering the synthetic data generated in Section 4.2.2, we will refer to each dataset as being obtained from different experiments, although in practice they could be a train of action potentials, say, cut up into many single action potential recordings. We then wish to see if we can infer probability distributions for each experiment's model parameters, and infer the probability distribution that govern how the different experiments' model parameter values were generated. The aim is to establish whether, given multiple datasets from a single cell, can we make a more informed prediction about how that cell will behave in a future experiment by taking into account how its past behaviour varied across experiments. This might then be generalisable to predicting how a new cell from a batch of cells will behave if we can fit a model to datasets from other cells from the same batch.

4.2.5 Methods

Given multiple datasets of action potential recordings, we wish to construct a statistical model that reflects how we believe the parameters governing each dataset are inter-connected. The model we choose is that the experimental dataset $\boldsymbol{y_i}$ (where

i indexes the experiment) has entries that are Normally distributed around the solution to the action potential model with parameters $\boldsymbol{G_i}$, with standard deviation σ mV. We then also suppose that the $\log(\boldsymbol{G_i})$ are independently drawn from a Normal distribution with mean $\boldsymbol{m}$ and variance $\boldsymbol{s^2}$. $\boldsymbol{m}$ and $\boldsymbol{s^2}$ are both vectors, with each entry corresponding to a conductance parameter in $\log(\boldsymbol{G_i})$. That is,

$$\log\left(G_i^{(j)}\right) \sim \mathcal{N}\left(m^{(j)}, s^{(j)2}\right) \tag{4.7}$$

$$y_{i,k} \sim \mathcal{N}\left(f_k(\boldsymbol{G_i}), \sigma^2\right), \tag{4.8}$$

where $f_k(\boldsymbol{G_i})$ is the k^{th} entry of the voltage trace returned by solving the action potential model with maximal conductance parameters $\boldsymbol{G_i}$. A schematic of parameter interaction and dependence under this statistical model is given in Figure 4.13, adapted from Figure 2.9 in Chapter 2. The 'top-level' parameters $\boldsymbol{m}, \boldsymbol{s^2}$, and σ require prior distributions to be specified.

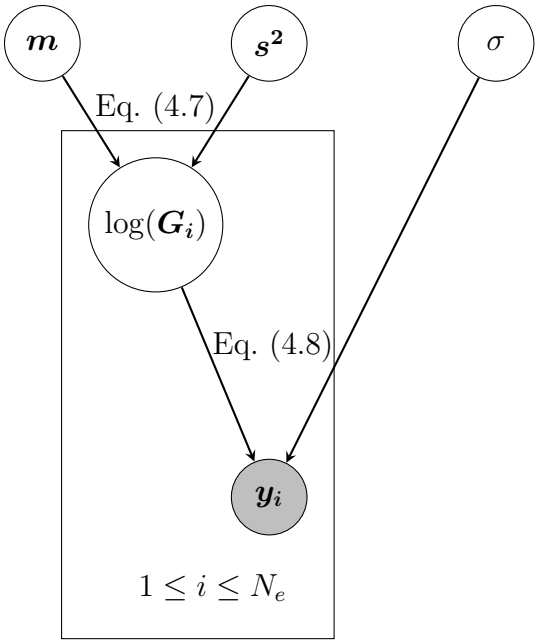

Figure 4.13: Parameter interaction and dependence under our hierarchical statistical model for multiple action potential recordings. i indexes the individual experiments (or datasets). $\boldsymbol{G_i}$ is the set of maximal ion current conductance parameters which is used to generate $\boldsymbol{y_i}$, the dataset from experiment i. $(\boldsymbol{m}, \boldsymbol{s^2})$ are the mean and variance, respectively, parameter values that give rise to the $\log(\boldsymbol{G_i})$. All non-shaded variables are parameters for which we wish to infer probability distributions. Parameters with no inward-pointing arrows require prior distributions to be specified.

If the action potential model that we are fitting to the data has d maximal conductance parameter values, then this statistical model has $(N_e + 2)d + 1$ parameters for which we wish to infer probability distributions at the same time. Even for a relatively simple action potential model, this can quickly become a difficult computational problem with the number of parameters involved. For example, when using the Beeler & Reuter (1977) action potential model, we are fitting 4 maximal conductance parameters. Given just four action potential datasets, we will be inferring distributions for 25 parameters. This is a much greater number of parameters than we were dealing with in the ion channel screening data case. Due to the large number of parameters involved in action potential models, we will use a different MCMC algorithm called *Gibbs Sampling* (Geman & Geman, 1984).

If we can sample directly from the parameters' conditional posterior distributions, then Gibbs Sampling allows us to sample from the full posterior distribution by sampling from each parameter's conditional posterior distribution in turn. Given a join probability distribution for random variables X and Y, $f_{X,Y}(x, y)$, the conditional distribution for X is $f_X(x) = \int_y f_{X,Y}(x, y)\mathrm{d}y$. Unlike in the Metropolis-Hastings algorithm, these samples are always accepted into the chain, and so there are no 'wasted' proposals that are rejected by the algorithm.

In fact, the only parameters whose conditional posterior distribution we can sample from directly are $(\boldsymbol{m}, \boldsymbol{s^2})$. To be able to do so, we make use of *conjugate priors* (Schlaifer & Raiffa, 1961). A conjugate prior is a prior distribution that, when multiplied by a likelihood function as in Equation (2.24), the product is in the same family of probability distributions as the prior. For example, a Normal-Inverse-Gamma distribution multiplied by a Normal distribution gives a new (scaled by a constant) Normal-Inverse-Gamma distribution with different parameters. We choose a Normal-Inverse-Gamma distribution as a (joint) prior distribution for $(\boldsymbol{m}, \boldsymbol{s^2})$ because the likelihood is a product of Normal distributions. Note that these Normal distributions all collapse in the conjugate prior product and result in a new Normal-Inverse-Gamma distribution. The conditional posterior distribution (treating all

other parameters as constants) for $(\boldsymbol{m}, \boldsymbol{s^2})$ is therefore a Normal-Inverse-Gamma distribution, from which we can directly sample.

There are no such closed forms for any other parameters' conditional posterior distributions, and so we cannot sample from them directly. We therefore cannot use full Gibbs Sampling to cycle through all parameters, sampling from their conditional posterior distributions. Instead of full Gibbs sampling, therefore, we perform a single step of the adaptive Metropolis-Hastings algorithm, where the target distribution is the conditional posterior distribution for the remaining parameters. This process is known as *Metropolis-within-Gibbs* (Gilks *et al.*, 1995).

Target distribution

Following the parameter dependences shown in Figure 4.13, the full posterior, or target, distribution for all parameters is

$$\pi\left(\boldsymbol{m}, \boldsymbol{s^2}, \{\log(\boldsymbol{G_i})\}_{i=1}^{N_e}, \sigma\right) \propto \left(\prod_{i=1}^{N_e} p(\boldsymbol{y_i}|\log(\boldsymbol{G_i}), \sigma) p(\log(\boldsymbol{G_i})|\boldsymbol{m}, \boldsymbol{s^2})\right) p(\boldsymbol{m}, \boldsymbol{s^2}) p(\sigma), \tag{4.9}$$

where $p(\boldsymbol{y_i}|\log(\boldsymbol{G_i}), \sigma)$ is the likelihood of $\{\log(\boldsymbol{G_i}), \sigma\}$

We choose independent Normal-Inverse-Gamma distributions as the (joint) prior distributions for each pair of parameters in $(\boldsymbol{m}, \boldsymbol{s^2})$. Normal-Inverse-Gamma distributions themselves have 4 parameters, which become hyperparameters in our statistical model. [1] We therefore have to choose a level at which to stop adding prior distributions on parameters. We therefore have

$$p(\boldsymbol{m}, \boldsymbol{s^2}) = \prod_{j=1}^{d} \mathcal{N}\text{-}\Gamma^{-1}\left(m^{(j)}, s^{(j)2}|\boldsymbol{\eta_j}\right), \tag{4.10}$$

where $(m^{(j)}, s^{(j)2})$ is the j^{th} entry of $(\boldsymbol{m}, \boldsymbol{s^2})$, and $\boldsymbol{\eta_j}$ is the set of hyperparameters governing the relevant Normal-Inverse-Gamma distribution.

[1] In theory, we could infer distributions for these hyperparameters too, but would need to specify prior distributions over them, clearly increasing the computational cost of our parameter inference. One could then put prior distributions on parameters, then put prior distributions on the parameters of the previous prior distribution, and this could be repeated *ad infinitum*. At some point, we would expect the additional prior distributions to add no benefit to the model, and to just increase computational cost.

We have the same formulation for $p(\boldsymbol{y_i}|\log(\boldsymbol{G_i}), \sigma)$ as in the non-hierarchical case:

$$p(\boldsymbol{y_i}|\log(\boldsymbol{G_i}), \sigma) = \prod_{k=1}^{K} \mathcal{N}\left(y_{i,k}|f_k(\boldsymbol{G_i}), \sigma^2\right) \tag{4.11}$$

$$\propto \sigma^{-K} \exp\left(-\frac{1}{2\sigma^2}\sum_{k=1}^{K}\left(y_{i,k} - f_k(\boldsymbol{G_i})\right)^2\right), \tag{4.12}$$

where $y_{i,k}$ is the k^{th} entry in dataset $\boldsymbol{y_i}$, which has K points in total, and $f_k(\boldsymbol{G_i})$ is the k^{th} entry of the voltage trace returned by solving the action potential model with maximal conductance parameters $\boldsymbol{G_i}$.

Similarly, the probability density of $\log(\boldsymbol{G_i})$ conditioned on $(\boldsymbol{m}, \boldsymbol{s^2})$ is

$$p(\log(\boldsymbol{G_i})|\boldsymbol{m}, \boldsymbol{s^2}) = \prod_{j=1}^{d} \mathcal{N}\left(\log({G_i}^{(j)})|m^{(j)}, s^{(j)2}\right) \tag{4.13}$$

$$\propto \left(\prod_{j=1}^{d} s^{(j)2}\right)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\sum_{j=1}^{d}\frac{\left(\log(G_i^{(j)}) - m^{(j)}\right)^2}{s^{(j)2}}\right). \tag{4.14}$$

We use a uniform prior distribution for the observation noise standard deviation σ and let it take values between 10^{-3} and 25:

$$p(\sigma) \propto I(10^{-3} < \sigma < 25), \tag{4.15}$$

where I is the same indicator function, and hence uniform prior, as in Equation (4.3).

4.2.6 Hyperparameters

Since we have chosen Normal-inverse-Gamma distributions as prior distributions for the top-level parameters $(\boldsymbol{m}, \boldsymbol{s^2})$, we need to specify the parameters that define these prior distributions. A single Normal-inverse-Gamma distribution has four parameters: μ, ν, α, and β. First, a random variable s^2 is sampled from an inverse-Gamma distribution defined by α and β, which are shape and scale parameters, respectively. Next, a random variable m is sampled from a Normal distribution with mean μ and variance s^2/ν.

To decide how we choose these four parameters (per conductance parameter), we consider the *prior predictive* distribution. This is the distribution from which we would sample log(conductance) parameter values if we had no experimental data to

inform our predictions. That is, we integrate the distribution for $\log(G)$ given (m, s^2) with respect to the prior distribution $p(m, s^2)$. This can be approximated by taking many samples (m, s^2) from the Normal-inverse-Gamma distribution, computing the probability density functions of the corresponding Normal distributions, summing these probability density functions together, and then normalising by dividing by the total number of samples taken.

To be comparable with the single-level method described in Section 4.2.3, we choose the hyperparameters so that the prior predictive distributions resemble the Normal prior distributions from Section 4.2.3. To approximate this, we consider the modes of the different probability distributions and propagate these to obtain the 'most likely' Normal distributions.

The mode of a Normal distribution with mean μ is also μ, and we want the prior predictive distribution to be centred on the original model parameter values, as we did for the prior distributions in Section 4.2.3. We therefore choose $\mu = \log(\hat{G})$, where $\hat{G}$ is the original model conductance parameter value. The standard deviation of this Normal distribution is $\sqrt{s^2/\nu}$, which we want to be (approximately) equal to $\log(10)/2$, for the same reasons described in Section 4.2.3. We therefore want $\nu \approx 4s^2/\log(10)^2$. The mode of an inverse-Gamma distribution with parameters α and β is $\beta/(\alpha+1)$, and so we take s^2 to be approximately this value. Therefore $\nu \approx 4\beta/((\alpha+1)\log(10)^2)$.

We choose α and β such that the mode of the inverse-Gamma (*i.e.* the most likely value of s^2) to be equal to $4\times$ the variance used to generate the synthetic data, to make greater values of s^2 more likely, thus requiring enough evidence in the data to infer values closer to the true value. We therefore choose α and β such that $\beta/(\alpha+1) = 0.04$, where 0.01 is the value given in Section 4.2.2.

This gives us the hyperparameters:

$$\mu = \log(\hat{G}), \tag{4.16}$$

$$\beta = 0.04(\alpha+1), \tag{4.17}$$

$$\nu = \frac{4\beta}{(\alpha+1)\log(10)^2} = \frac{0.16}{\log(10)^2}, \tag{4.18}$$

and α is still a free variable. From empirical observations, α was chosen to equal 4 as the resulting inverse-Gamma provided a good range of possible values for s^2.

Parameter inference

To perform Gibbs sampling, we need to write down the conditional posterior distributions for the (sets of) parameters we wish to sample in sequence. To do this, we can simply 'ignore' the parameters in Equation (4.9) that have no influence on the parameters of interest by treating them as if they were constants and dropping them from proportional equations. We will use the same notation π to indicate conditional posterior distributions, with the arguments defining which distribution we are using.

Since we have defined independent Normal-Inverse-Gamma prior distributions for each pair in $(\boldsymbol{m}, \boldsymbol{s^2})$, with hyperparameters $\boldsymbol{\eta_j}$, and a Normal likelihood for $\log(\boldsymbol{G_i})$ given $(\boldsymbol{m}, \boldsymbol{s^2})$, we know that the conditional posterior for $(\boldsymbol{m}, \boldsymbol{s^2})$ will also be a product of independent Normal-Inverse-Gamma distributions:

$$\pi(\boldsymbol{m}, \boldsymbol{s^2}) \propto \prod_{j=1}^{d} \mathcal{N}\text{-}\Gamma^{-1}\left(m^{(j)}, s^{(j)2} | \tilde{\boldsymbol{\eta}}_{\boldsymbol{j}}\right). \tag{4.19}$$

These new $\tilde{\boldsymbol{\eta}}_{\boldsymbol{j}}$ are functions of the old $\boldsymbol{\eta_j}$ and the $\log(G_i^{(j)})$, and their new forms are described by Murphy (2007). We can therefore sample directly from this conditional posterior distribution after we update the hyperparameters based on the current values of $\log(G_i^{(j)})$. The remaining conditional posterior distributions do not have such closed-form distributions, and so we cannot sample from them directly.

We now consider the conditional posterior distribution for each $\log(\boldsymbol{G_i})$ in turn:

$$\pi(\log(\boldsymbol{G_i})) \propto p(\boldsymbol{y_i} | \log(\boldsymbol{G_i}), \sigma) p(\log(\boldsymbol{G_i}) | \boldsymbol{m}, \boldsymbol{s^2}) \tag{4.20}$$

$$\propto \exp\left(-\frac{1}{2\sigma^2} \sum_{k=1}^{K} \left(y_{i,k} - f_k(\boldsymbol{G_i})\right)^2\right) \exp\left(-\frac{1}{2} \sum_{j=1}^{d} \frac{\left(G_i^{(j)} - \hat{G}^{(j)}\right)^2}{\sigma^{(j)2}}\right), \tag{4.21}$$

where we have omitted the σ and $\sigma^{(j)2}$ terms from Equations (4.12) and (4.14), respectively, since we are treating the non-$\log(\boldsymbol{G_i})$ parameters as constants when considering conditional distributions.

Finally, we have the conditional posterior distribution for the observation noise standard deviation σ:

$$\pi(\sigma) \propto \left(\prod_{i=1}^{N_e} p(\boldsymbol{y_i}|\log(\boldsymbol{G_i}), \sigma)\right) p(\sigma) \tag{4.22}$$

$$\propto \left(\prod_{i=1}^{N_e} \sigma^{-K} \exp\left(-\frac{1}{2\sigma^2}\sum_{k=1}^{K}\left(y_{i,k} - f_k(\boldsymbol{G_i})\right)^2\right)\right) p(\sigma) \tag{4.23}$$

$$\propto \sigma^{-N_e K} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{N_e}\sum_{k=1}^{K}\left(y_{i,k} - f_k(\boldsymbol{G_i})\right)^2\right) I(10^{-3} < \sigma < 25), \tag{4.24}$$

where $I(\cdot)$ is an indicator function with value 1 if its condition is satisfied, and 0 otherwise.

When performing Gibbs sampling, we sample from the conditional posterior distribution for each parameter in turn. To begin with, we do this for the parameters $(\boldsymbol{m}, \boldsymbol{s^2})$. As discussed above, we cannot do this for the remaining parameters. We therefore, in turn, approximate a sample from each experiment's conditional posterior distribution for $\log(\boldsymbol{G_i})$ with a single Metropolis-Hastings step. Finally, we perform a Metropolis-Hastings step for σ. The pseudocode for this Metropolis-within-Gibbs algorithm is given in Algorithm 3. As shown in line 8 of Algorithm 3, the Metropolis-Hastings step is not adaptive for the first $100 \times$ (number of parameters) iterations. This reduction to 100 from 1000 in the single-level case is just to reduce the time taken before the adaptation begins — we have found that very few, if any, of the proposed parameter values during this pre-adaptation period are accepted into the chain.

4.2.7 Results

As in Figure 4.7 in Section 4.1.2, we can plot normalised marginal histograms of the experiment-level log(conductance) parameter samples to approximate those parameters' marginal posterior distributions. Similar to the single-level model case, these represent probability densities for log(conductance) parameters given that experimental dataset. However, because of the structure of the hierarchical model, these experiment-level distributions can influence each other through the top-level distribution. This is because, as we sample from the conditional posterior

Algorithm 3 Metropolis-within-Gibbs for inferring probability distributions for the parameters shown in Figure 4.13. $[t]$ indexes the iteration of the algorithm. Vectors are treated as row vectors such that we obtain square matrices in line 23.

1: $\log(a_{i:N_e,[0]}), \log(a_{\sigma,[0]}) \leftarrow 0$.
2: $\boldsymbol{m}_{[0]}, \boldsymbol{s}^{\boldsymbol{2}}_{[0]}, \log(\boldsymbol{G}_{\boldsymbol{1:N_e}})_{[0]}, \sigma_{[0]} \leftarrow$ initial optimisation.
3: $\Sigma_{1:N_e,[0]} \leftarrow$ diagonal matrices.
4: $\Sigma_\sigma \leftarrow$ constant.
5: $\gamma \leftarrow 0$.
6: $t \leftarrow 0$.
7: **loop**
8: **if** $t > 100d$ **then**
9: $\gamma \leftarrow (t - 100d + 1)^{-0.6}$
10: **end if**
11: Sample $\left(\boldsymbol{m}_{[t+1]}, \boldsymbol{s}^{\boldsymbol{2}}_{[t+1]}\right) \sim \pi\left(\boldsymbol{m}, \boldsymbol{s}^{\boldsymbol{2}} | \log(\boldsymbol{G}_{\boldsymbol{1:N_e}})_{[t]}\right)$
12: **for** i in $\{1, 2, \ldots, N_e\}$ **do**
13: Propose $\log(\boldsymbol{G}^*_{\boldsymbol{i}}) \sim \mathcal{N}(\cdot | \log(\boldsymbol{G}_{\boldsymbol{i}})_{[t]}, a_{i,[t]}\Sigma_{i,[t]})$.
14: Sample $u \sim \mathcal{U}(0, 1)$.
15: **if** $\log(u) < \log\left(\pi\left(\log(\boldsymbol{G}^*_{\boldsymbol{i}}) | \boldsymbol{m}_{[t+1]}, \boldsymbol{s}^{\boldsymbol{2}}_{[t+1]}, \sigma_{[t]}\right)\right)$
16: $- \log\left(\pi\left(\log(\boldsymbol{G}_{\boldsymbol{i}})_{[t]} | \boldsymbol{m}_{[t+1]}, \boldsymbol{s}^{\boldsymbol{2}}_{[t+1]}, \sigma_{[t]}\right)\right)$ **then**
17: $\log(\boldsymbol{G}_{\boldsymbol{i}})_{[t+1]} \leftarrow \log(\boldsymbol{G}^*_{\boldsymbol{i}})$
18: $accepted \leftarrow 1$
19: **else**
20: $\log(\boldsymbol{G}_{\boldsymbol{i}})_{[t+1]} \leftarrow \log(\boldsymbol{G}_{\boldsymbol{i}})_{[t]}$
21: $accepted \leftarrow 0$
22: **end if**
23: $\Sigma_{i,[t+1]} \leftarrow (1 - \gamma) \times \Sigma_{i,[t]} + \gamma \times (\log(\boldsymbol{G}_{\boldsymbol{i}})_{[t+1]} - \boldsymbol{\mu}_{\boldsymbol{i},[t]})^T (\log(\boldsymbol{G}_{\boldsymbol{i}})_{[t+1]} - \boldsymbol{\mu}_{\boldsymbol{i},[t]})$
24: $\boldsymbol{\mu}_{\boldsymbol{i},[\boldsymbol{t+1}]} \leftarrow (1 - \gamma) \times \boldsymbol{\mu}_{\boldsymbol{i},[t]} + \gamma \times \log(\boldsymbol{G}_{\boldsymbol{i}})_{[t+1]}$
25: $\log(a_{i,[t+1]}) \leftarrow \log(a_{i,[t]}) + \gamma \times (accepted - 0.25)$
26: **end for**
27: Propose $\sigma^* \sim \mathcal{N}(\cdot | \sigma_{[t]}, a_{\sigma,[t]}\Sigma_\sigma)$.
28: Sample $u \sim \mathcal{U}(0, 1)$.
29: **if** $\log(u) < \log\left(\pi\left(\sigma^* | \log(\boldsymbol{G}_{\boldsymbol{1:N_e}})_{[t+1]}\right)\right) - \log\left(\pi\left(\sigma_{[t]} | \log(\boldsymbol{G}_{\boldsymbol{1:N_e}})_{[t+1]}\right)\right)$ **then**
30: $\sigma_{[t+1]} \leftarrow \sigma^*$
31: $accepted \leftarrow 1$
32: **else**
33: $\sigma_{[t+1]} \leftarrow \sigma_{[t]}$
34: $accepted \leftarrow 0$
35: **end if**
36: $\log(a_{\sigma,[t+1]}) \leftarrow \log(a_{\sigma,[t]}) + \gamma \times (accepted - 0.25)$
37: t++
38: **end loop**

distributions in turn with our Metropolis-within-Gibbs algorithm, the experiment-level parameters' log-target value depends on the top-level parameters, and then the top-level parameters' log-target value depends on the experiment-level parameters. Because of this, the same log(conductance) parameters across experiments are not entirely independent.

We can also plot the posterior predictive distribution, which is the probability distribution that describes what values we would expect to observe for a log(conductance) parameter in a future experiment. We compute it by integrating the probability density of the log(conductance) parameters with respect to the probability densities defined by the top-level parameters, as discussed in Section 2.5.4 of Chapter 2. We can approximate this integral by summing the probability densities defined by the top-level parameter samples and then normalising.

In the case of synthetic data, we have defined the probability distribution that generates the true experimental parameter values. Therefore, we would hope that the posterior predictive distribution would approximate this true distribution, at least when inferring parameter distributions from enough experiments.

When inferring distributions from synthetic data generated using the Beeler & Reuter (1977) model, we find that we can obtain well-defined tight marginal distributions for the experiment-level log(conductance) parameters. This is expected, as we were able to do so in Section 4.1.2. These marginal histograms, along with the posterior predictive and prior predictive distributions obtained from fitting to 4 experiments, and the true distributions are shown in Figure 4.14. The prior predictive distribution appears flat, and therefore uniform, in these plots, because of how relatively tight the posterior predictive and true probability distributions are.

We can also compare the prior distributions for m and s^2 with their respective marginal posterior distributions. These are given in Figure 4.15 for varying N_e. As N_e increases and more datasets are being fitted to at once, the marginal posterior distributions become narrower and centred on the true values used to generate the synthetic data, as expected.

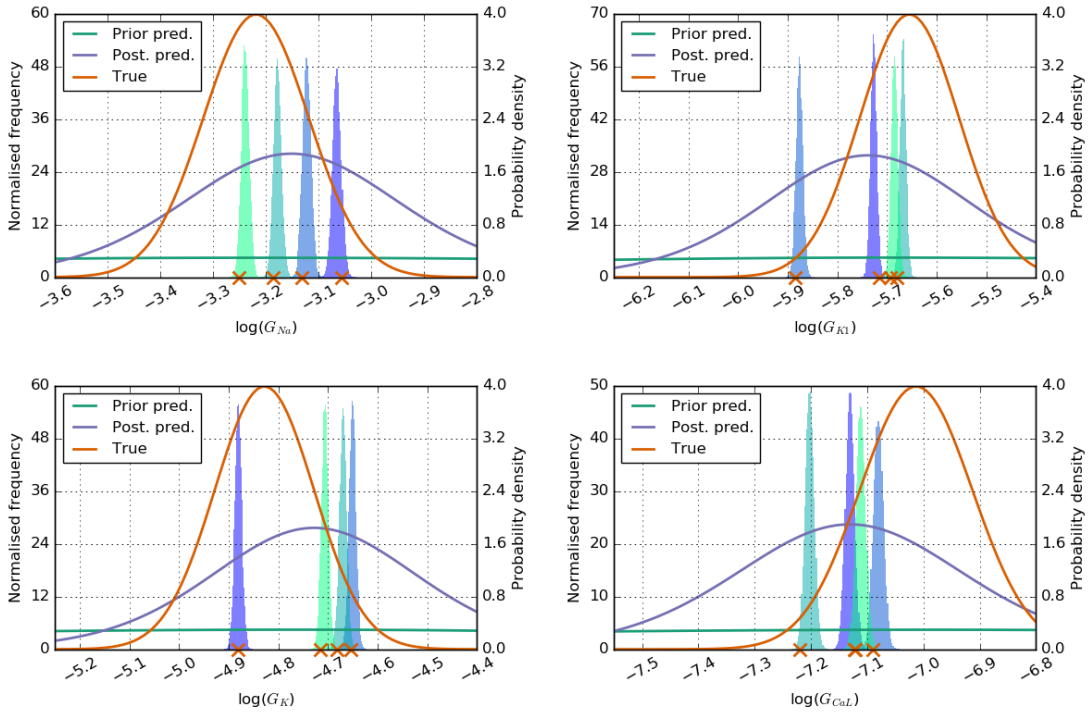

Figure 4.14: Parameter inference using the hierarchical model on synthetic data generated from the Beeler & Reuter (1977) model. y-axis: normalised marginal histograms of experiment-level log(conductance) parameters. Right y-axis: prior and posterior predictive distributions — approximations for the probability density of values of $\log(G)$ for a future experiment before and after observing the experimental data, respectively — and true probability density function used the generate the parameters. The crosses are the log(conductance) parameter values that were used to generate each dataset.

We want to see how closely we can approximate the true parameter-generating probability distribution using the posterior predictive distribution as we increase the number of datasets being fitted to. We would expect that having more datasets provides more information and will therefore allow for a better approximation. However, increasing the number of datasets being fitted to increases the overall computational cost of the parameter inference. The marginal posterior predictive distributions for the log(conductance) parameters in the Beeler & Reuter (1977) model are shown in Figure 4.16, plotted for different values of N_e, the number of datasets being fitted to at once. Since the prior predictive distributions are relatively wide, lots of information is required to obtain an accurate prediction. While the convergence to the 'true' distribution could be accelerated by hyperparameters that result in narrower prior predictives, we are assuming little prior knowledge

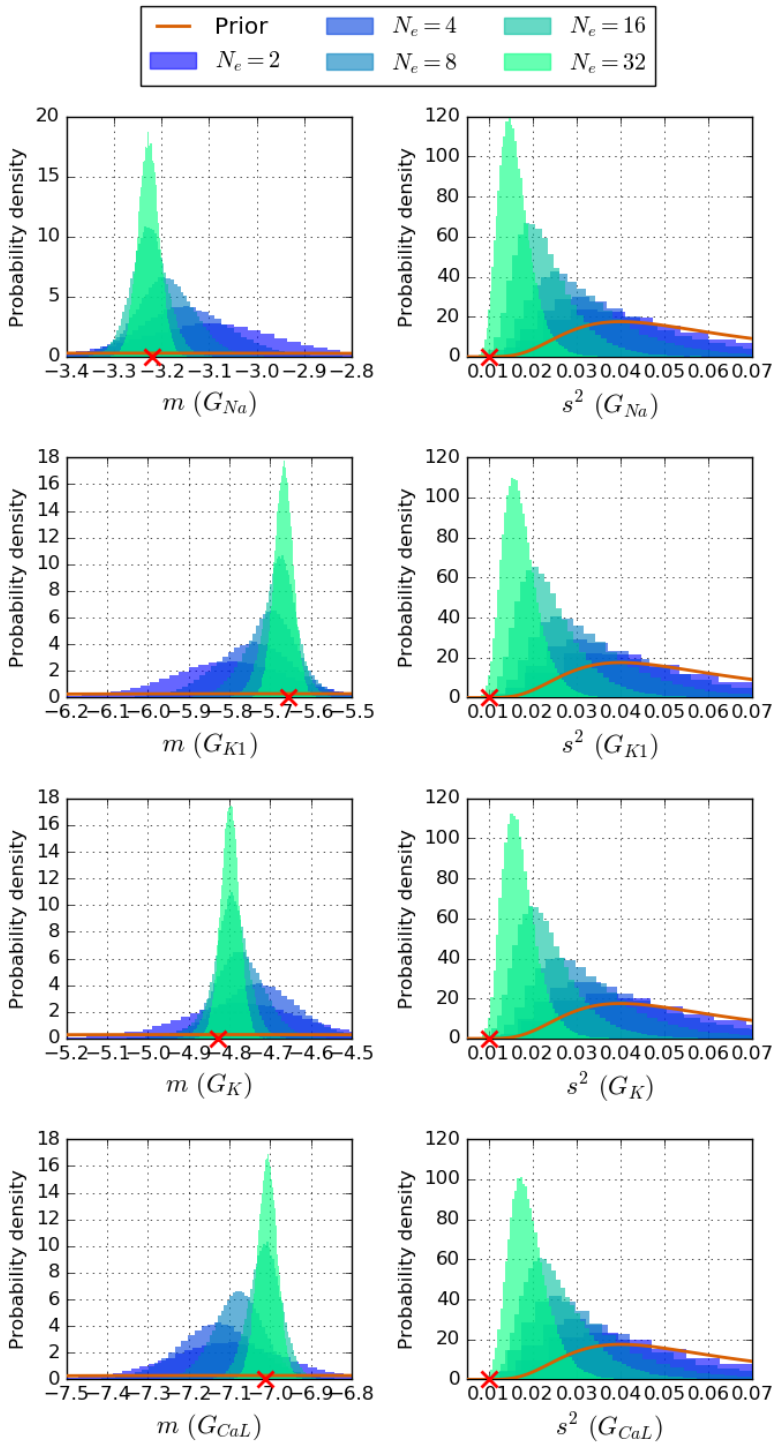

Figure 4.15: Prior and marginal posterior distributions for m and s^2 for each log(conductance) parameter in the Beeler & Reuter (1977) model, inferred from synthetic data using the hierarchical statistical model.

in the spread of these parameters, and so we believe that the prior predictive distributions are suitably wide.

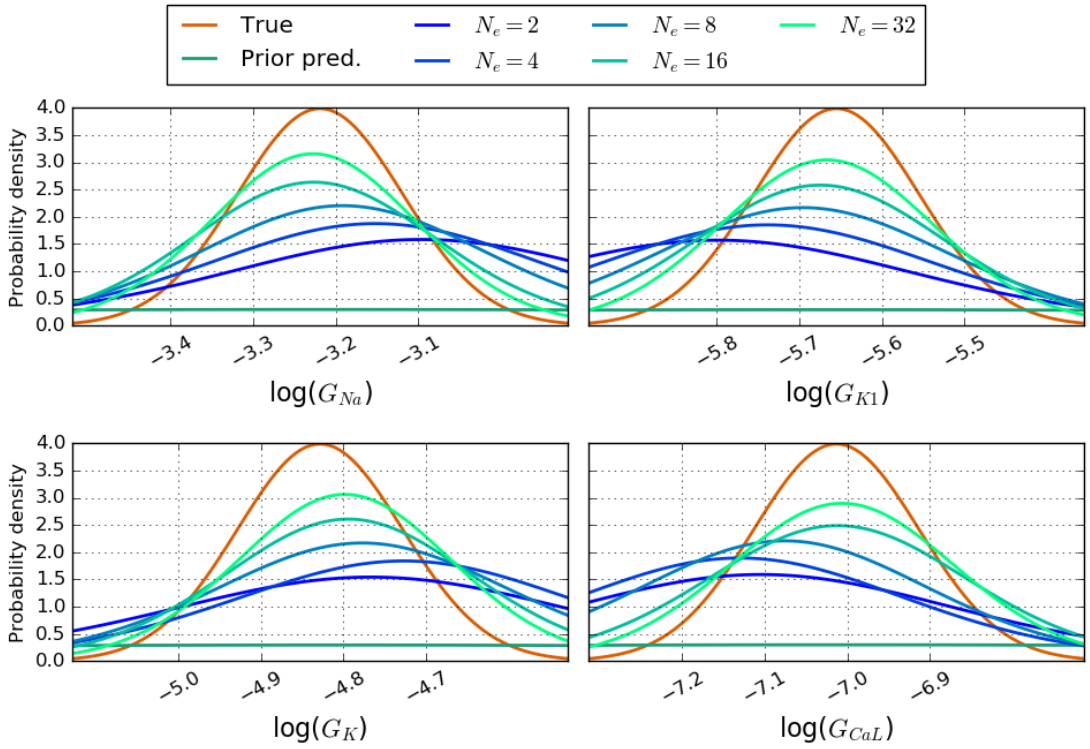

Figure 4.16: Marginal posterior predictive distributions for each log(conductance) parameter inferred from synthetic data generated from the Beeler & Reuter (1977) model. Parameter distributions were inferred under the hierarchical model, fitting to N_e experiments at once. As N_e increases, the posterior predictive distributions tend to the true distribution used to generate the experimental parameter values.

The posterior predictive distributions are tending toward the true distribution used to generate the experimental parameters. This is expected, but in practice the computational cost of fitting to so many datasets at once is very great. If running the MCMC algorithm in series, then the time taken scales linearly with the number of datasets being fitted to. This time can be reduced by performing the log($\boldsymbol{G_i}$) Metropolis-Hastings steps in parallel, but we were limited to 16 processes running in parallel. Additionally, real action potential recordings are typically sampled at a much higher frequency than the synthetic data we are using here, which increases the time taken to solve the system of differential equations in the action potential model. Downsampling the action potential traces is an option to reduce computational

cost, but naïve downsampling can lead to 'missing' the peak of the upstroke by discarding the maximum voltage point. It would seem like there was a smaller upstroke than there really was, leading to smaller values of the sodium current conductance being inferred. We could more-intelligently downsample by starting from the maximum recorded voltage and then outwardly sampling every n^{th} point.

We also note that the normalised marginal histograms for the experiment-level log(conductance) parameters match very closely with the normalised marginal histograms of the log(conductance) parameters for the output of the single-level model. For example, the normalised marginal histograms for the first 8 synthetic action potential traces from the Beeler & Reuter (1977) model are shown in Figure 4.17. The large number of data points in a single action potential trace constrains the range of possible action potentials that might fit the data, meaning that there is more information about the experiment-level parameters. We therefore expect that the single-level model outputs contain enough information to allow us to infer an approximate predictive distribution. This approximate predictive distribution should provide an alternative to the posterior predictive distribution obtained from the hierarchical model.

To approximate the posterior predictive distribution for a general $\log(G)$ given N_e datasets, we take a single $\log(G)$ sample from each of the N_e single-level MCMC runs. We then fit these samples to a Normal distribution, obtaining a mean and a standard deviation. We then compute the probability density function defined by this mean and standard deviation. This is repeated many (T, say) times, each time summing the newly-computed probability density functions. This sum is then normalised by dividing by T. These single-level-derived approximate predictive distributions are shown in Figure 4.18 for different values of N_e, next to their equivalent hierarchical-derived posterior predictive distributions for synthetic data from the Beeler & Reuter (1977) model. This method supposes that samples from each single-level chain represent independent samples from some overarching Normal distributions. We could have chosen a simpler method, such as fitting the single-level means to a Normal distribution. However, if all single-level distributions

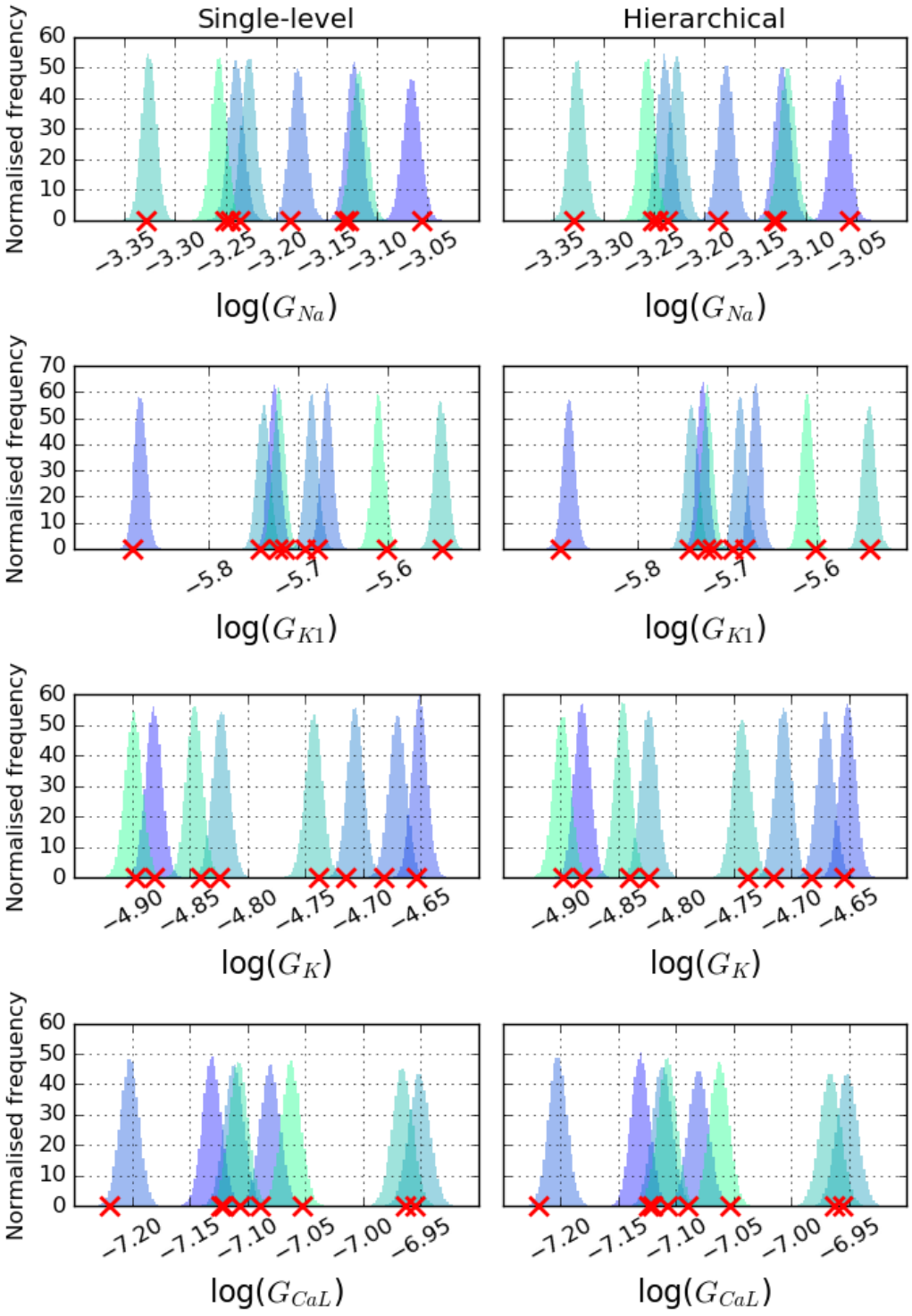

Figure 4.17: Normalised marginal histograms of the log(conductance) parameters in the Beeler & Reuter (1977) model. Left: single-level model. Right: hierarchical model, experiment-level. The red crosses are the true parameter values used to generate the synthetic data.

had the same mean, even if they were very wide, the predictive distribution would be close to a delta function.

We can see from Figure 4.18 that while the posterior predictive distributions tend to the true distribution as the number of experiments increases in a more 'uniform' manner, the single-level approximation provides reasonable approximations after $N_e = 16$, with not much improvement when $N_e = 32$.

We also construct these predictive distributions for the case of synthetic data generated from the O'Hara *et al.* (2011) model. These distributions are given in Figures 4.19–4.20. The distributions that appear relatively flat correspond well to those parameters in Table 4.4 for which we were not able to infer a tight distribution from a single (synthetic) action potential recording. This tells us that we are not 'accidentally' inferring extra information that may not be in the data about these uncertain parameters.

4.3 Discussion

In this chapter, we have used some of the Bayesian inference methods used in Chapter 3, as well as some more complex methods, to investigate uncertainty in computational modelling of synthetic cardiac action potentials. Using synthetic data is a useful first step in assessing the performance of these inference methods, as it gives us a 'best-case' scenario where we know that our model is an accurate representation of reality. In the case of trying to model noisy real-world experimental data, we will never have the 'correct' model, and so it is important to choose the best model for the application at hand, along with giving a well-reasoned justification for selecting such a model. We discuss this further in Chapters 5–6, where we are fitting action potential models to recordings obtained from canine cardiomyocytes and stem cell-derived cardiomyocytes, respectively.

We have seen how in the synthetic data case, we can change how much information is present in the data by varying the experimental protocol used to obtain the data. It is worth considering this approach when designing experiments to

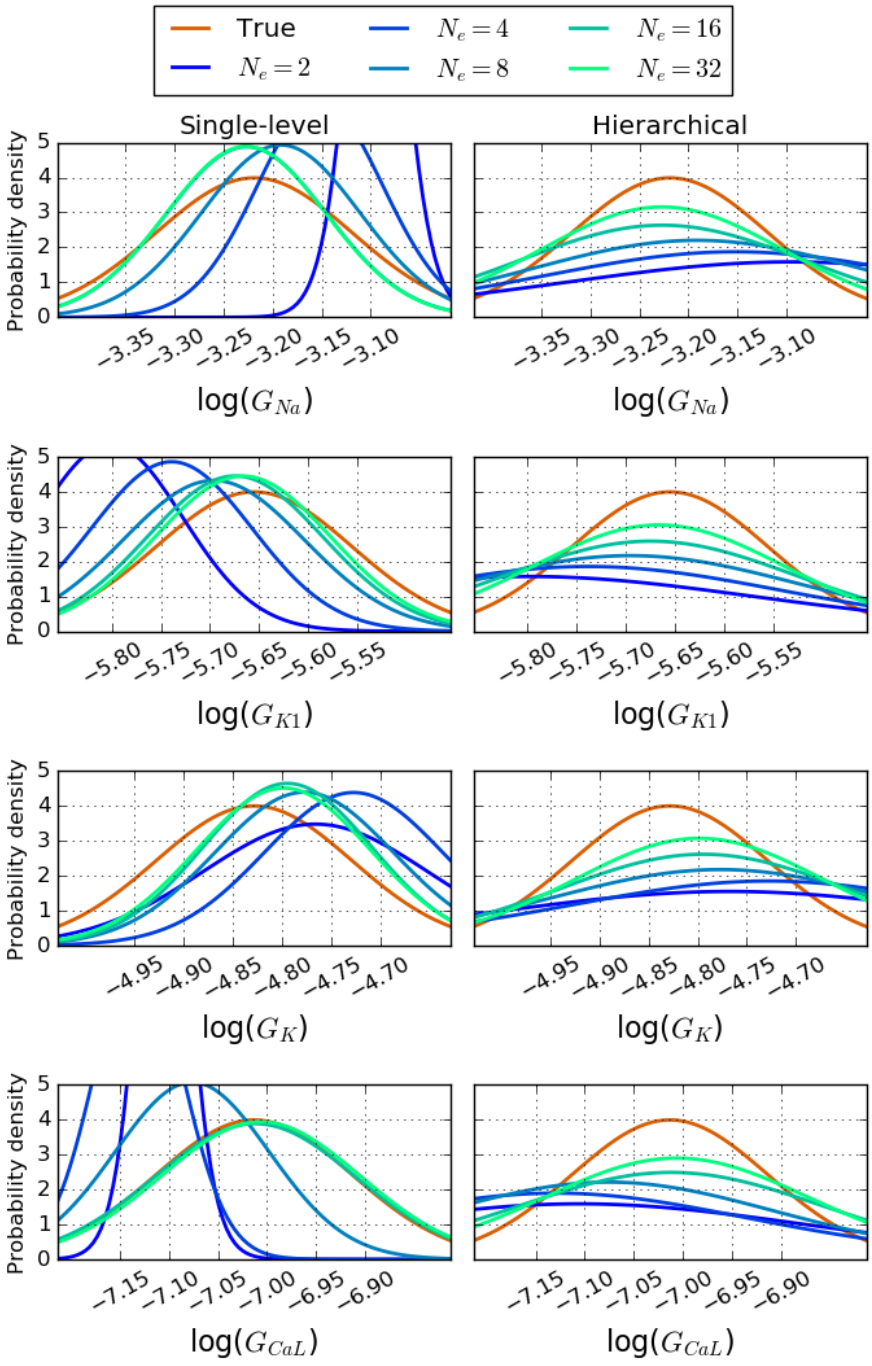

Figure 4.18: Predictive distributions from inferred log(conductance) parameter distributions from synthetic data from the Beeler & Reuter (1977) model. Left: single-level model. Right: hierarchical model. The y-axes have been capped at 5 so that the differences between distributions for higher values of N_e are more visible.

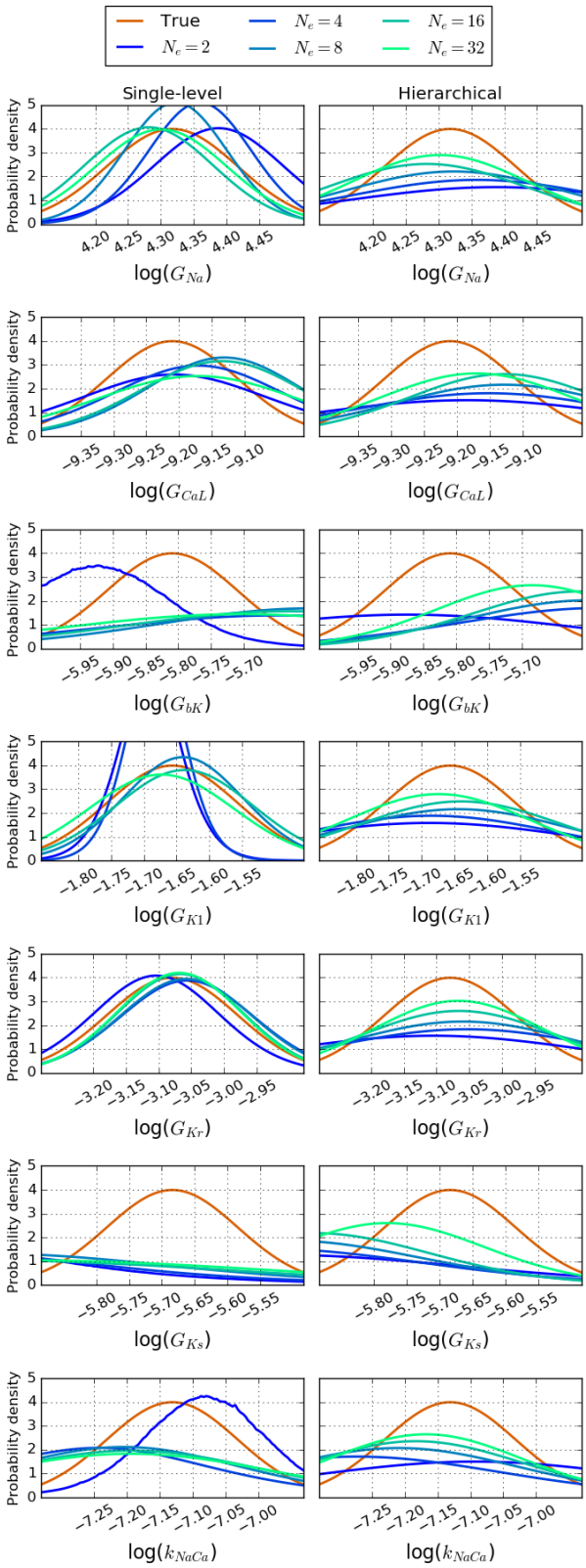

Figure 4.19: Left: Single-level predictive distributions. Right: hierarchical posterior predictive distributions. The y-axes have been capped at 5 so that the differences between distributions for higher values of N_e are more visible. Continued in Figure 4.20.

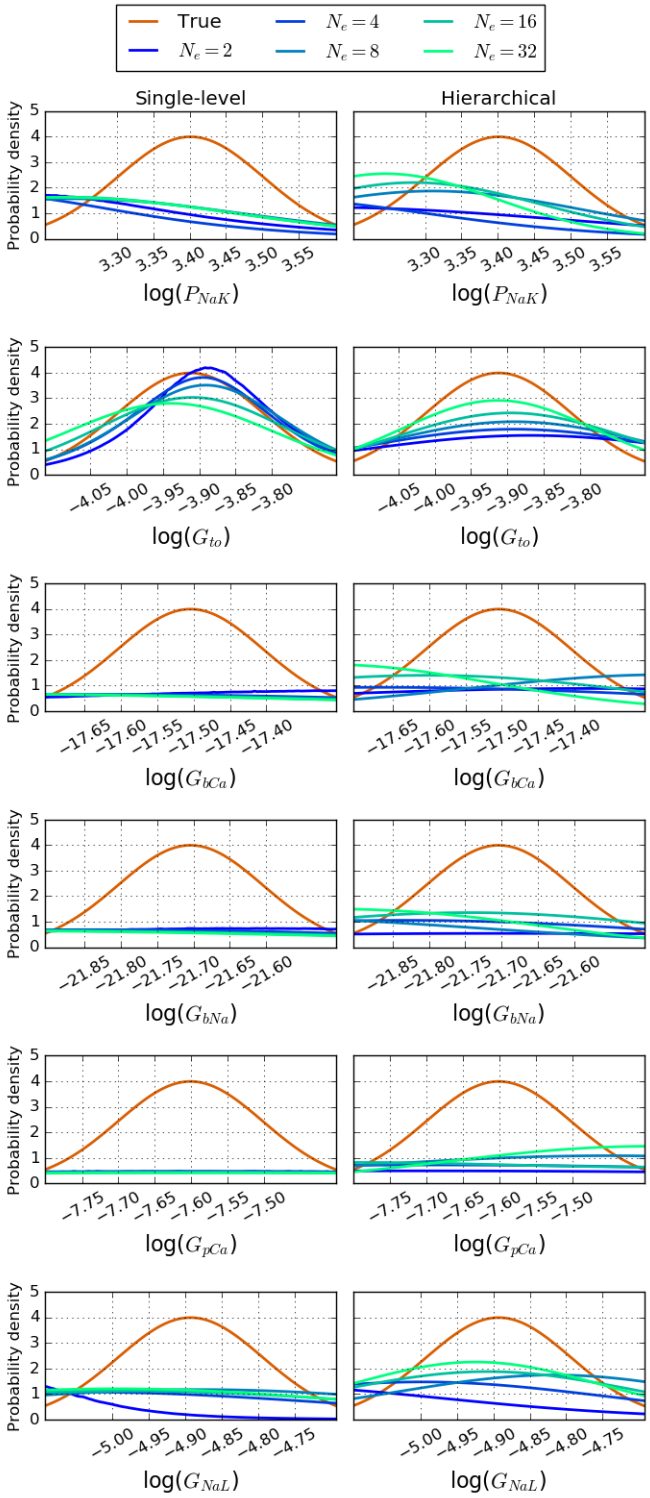

Figure 4.20: Continued from Figure 4.19. Left: Single-level predictive distributions. Right: hierarchical posterior predictive distributions.

investigate particular ion currents, for example, through action potential behaviour. These parameter inference methods also assess a parameter's identifiability.

As discussed in Chapter 3, a hierarchical statistical model can be used to quantify inter-experiment variability, and to predict what might happen in a future experiment. However, in this chapter we have been using much more complex models than in Chapter 3, and so using the same inference methods would have been very computationally expensive. The Metropolis-within-Gibbs algorithm we employed to perform parameter inference mitigates some of the computational cost, but it is still an expensive method. Instead, we performed parameter inference on the datasets separately using a single-level statistical model and an adaptive Metropolis-Hastings algorithm. We were able to use these single-level chains to construct approximations of the posterior predictive distributions — the distributions we would use to predict future experiment behaviour.

In the next chapter, we apply the methods from this chapter to infer probability distributions for log(conductance) parameters from canine cardiac action potential data. Considering the computational cost of inferring probability distributions for the parameters in the hierarchical statistical model as well as the approximations obtained from the single-level model, we will use the single-level statistical model and adaptive Metropolis-Hastings algorithm to infer parameter distributions from canine action potential data and to construct predictive distributions.

5

Fitting an action potential model to experimental data

In this chapter, we will apply some of the techniques discussed in Chapter 4 on experimentally-measured membrane voltage action potential recordings obtained from a canine ventricular cardiomyocyte. In particular, we use the methods discussed in Section 4.2 to investigate the beat-to-beat variability observed in contiguously-recorded action potentials, provided to us by collaborators in Utrecht.

The work in this chapter is based on work published in Johnstone *et al.* (2016b), but the version presented here includes more recent methods. For example, in Johnstone *et al.* (2016b), our MCMC algorithm sampled 'conductance space', whereas, in this chapter, we have used the statistical model described in Section 4.2.3 of Chapter 4, whereby our MCMC algorithm samples 'log(conductance) space'. This is because gene expression levels can be described by log-Normal distributions, as discussed in Section 4.2.1 of Chapter 4. Throughout the chapter, the words "trace" and "pace" are used interchangeably, since we only work with single action potential recordings.

5.1 Experimental procedure

The following experiments were performed by our collaborators at University Medical Centre, Utrecht.

Canine cardiomyocytes were isolated as described by Nalos *et al.* (2012). A drop of cell suspension was placed in a recording chamber, which was kept at 37°C. After the cells adhered, perfusion with a normal modified Tyrode's solution (Tyrode, 1910, a roughly isotonic solution) was started. Cardiomyocytes were patched with patch pipettes with a resistance of 3–4 MΩ, and action potentials were recorded in whole-cell configuration. After obtaining access and adjusting series resistance compensation ($> 70\%$), continuous 1 Hz pacing with a 3 nA, 2 ms stimulus was started. Time resolution of the recordings was 250 μs (4 kHz), while voltage resolution was 31.2 μV. Recordings in the dataset cover $\approx$ 800 ms of each pace This is because the processing time required to continuously log data is non-zero, and so recording for a full second would offset the 1 Hz pacing slightly after each pace.

The bath solution was composed of (concentrations in mM) NaCl 130, KCl 5.4, $CaCl_2$ 1.8, $MgCl_2$ 1.2, $NaHCO_3$ 18, HEPES 10 and glucose 10, and pH was set at 7.4 with NaOH. Pipettes were filled with (in mM) NaCl 10, KCl 130, $MgCl_2$ 0.5, HEPES 10, MgATP 5, and pH was set at 7.2 with KOH. After 200 s of pacing 10 μM of Moxifloxacin was introduced to the bath, and at 240 s the bath concentration of KCl was decreased to 4 mM. The voltage traces recorded during the first 200 s of pacing, in the control solution, are shown in Figure 5.1. The traces are recorded continuously from a single cell and are coloured sequentially, according to the colour bar to the right of the plot.

There is a clear trend of the earlier paces having longer action potential durations (APDs) than the later paces, which shows that the myocyte is not in a steady state at the start of the recording. We were interested in modelling a myocyte in an approximate steady state, so we plot the APDs of each trace to find a series of similar action potentials. Ideally, we would have a 'flat' region in the graph of APD against pace which would indicate no (or very minor) beat-to-beat variation, and we would select these traces as our experimental datasets for model fitting. The

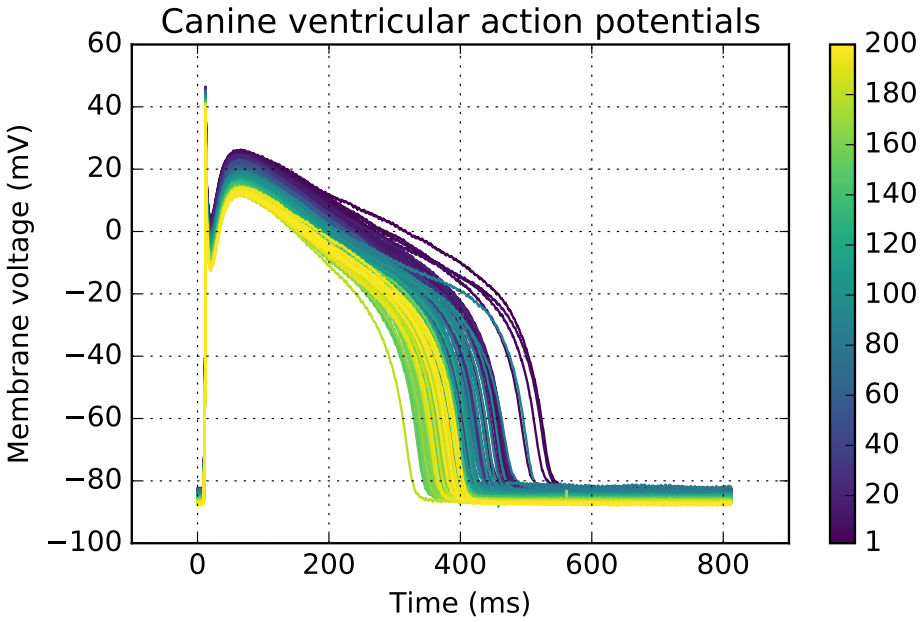

Figure 5.1: 200 experimental recordings of action potentials from a canine ventricular myocyte in control solution. The trace numbers are colour-coded according to the colour bar on the right. Trace #1 is the first recording.

APD of each experimental pace is shown in Figure 5.2, along with an indication of the paces that were selected as approximately steady-state for training and validation purposes. Paces #150-179 were chosen to model in the control setting because there was relatively little change in APD throughout that region, whereas before then it looks like the cell is adapting to the pacing. When considering beat-to-beat variability, we want as little as possible variation to be caused directly by the experimental stimuli so that the variability is because of the cell's intrinsic behaviour. Paces #400-429 were chosen to model after the addition of Moxifloxacin and the change in extracellular potassium concentration for similar reasons: the cell is adapting to its new environment, and the selected region appears relatively steady. When modelling these paces, we chose to model the first 16 from each region. This choice was a practical consideration where we had access to multiples of 16 cores from Oxford's "Advanced Research Computing" services.

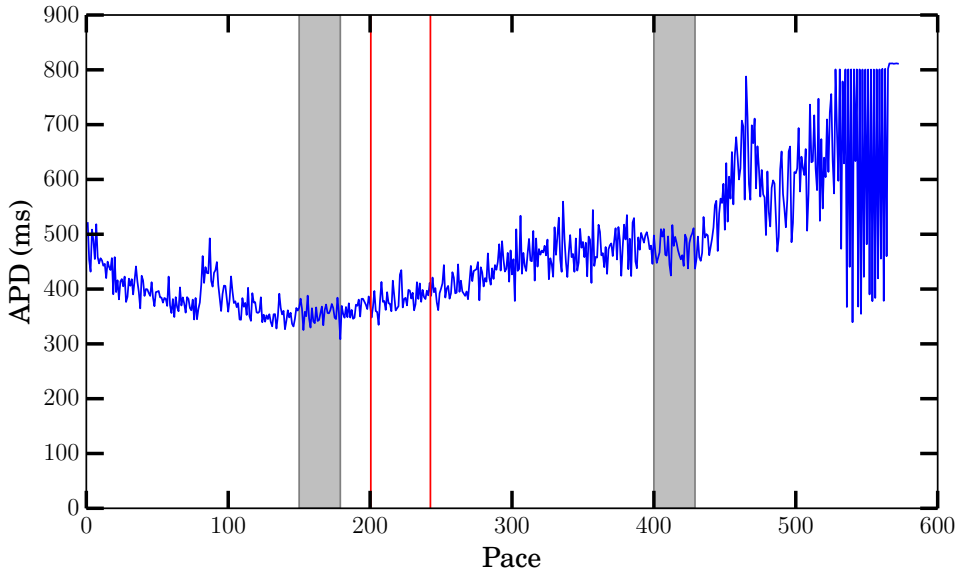

Figure 5.2: Summary of the whole patch clamp experiment. Action Potential Durations (APD_{90}) are shown for every pace in the experiment. The first grey shaded region, at paces #150–179, shows the training data, at control. The first red line shows when 10μM Moxifloxacin was added, the second when the bath KCl concentration was decreased to 4 mM (lowering the extracellular potassium concentration). The second shaded region, at paces #400–429, highlights the validation data paces. The shaded regions were chosen to be representative of the patch-clamped cell having settled to roughly steady behaviour in control and altered conditions. After pace $\sim$ #450 the patch-clamp becomes unstable and action potentials lengthen until they are over 1 s long. The values plotted after approximately #530 are therefore unreliable.

5.2 Modelling the canine action potential

We apply the methodology from Chapter 4 to the case of noisy real-world experimental data. That is, we will attempt to infer an approximation to the posterior distribution of action potential model parameters, informed by experimental recordings. This will allow us to quantify some of the uncertainties associated with modelling a single action potential recording. We also analyse the beat-to-beat variability of the myocyte. That is, we characterise how and why the action potentials change as pacing is performed. Beat-to-beat variability introduces more uncertainty, as we want to predict how the cell will respond to a drug in a future experiment, so we have to consider its different possible behaviours in the control setting. If we based our predictions on a single action potential recording, then

we limit the range of predictions we can make. Analysis of beat-to-beat variability can provide better estimates of proarrhythmic risk (Heijman *et al.*, 2013), however in this work we are just considering the additional uncertainty it introduces. We will attempt to identify which ion currents are the most sensitive to beat-to-beat variability, either because of complicated chaotic dynamics that the models do not yet capture (Glass, 2009), or because of stochasticity in how much the ion channel gates open at different voltages (Fox, 1997).

One approach to explore beat-to-beat variability is to treat different paces as independent experimental recordings, and perform the same inference on each of these separately. We will then have posterior distributions for the model parameters for each of these experimental traces. Marginal distributions for each parameter, for example, can then be compared directly. A limitation of this method is the choice of initial conditions. Since the action potentials were actually recorded in sequence, the initial conditions for one pace depend on the conditions at the end of the previous pace. As a first approximation, however, we reset the action potential model's state variables to their initial values provided in the CellML files — 1 Hz steady-state values. We will discuss possible alternatives to this in Section 5.5.

Figure 5.2 shows the periods of the experiment which were taken for this study, and provides an impression of the beat-to-beat variability of one aspect of the action potentials, namely their duration.

5.2.1 Data-clamp

We encountered problems using the built-in stimulus functionality in Chaste, since different action potential models deal with membrane capacitances in different ways and are coded in different units. To overcome this and accurately represent the effects of the stimulus current on the membrane voltage, we adapted a method for parameterising action potential models proposed by Dokos & Lovell (2004). Instead of fitting directly to the membrane voltage trace, the authors fitted to the total ionic current, I_{tot}, required to give rise to the experimental voltage trace, where $I_{tot} = \sum_{ion} I_{ion} + I_{stim} = -C_m \frac{\mathrm{d}V}{\mathrm{d}t}$. To ensure the model follows the experimental

voltage trace, Dokos & Lovell introduced a new current that acts as a restorative force, 'pushing' the model back onto the experimental data at a rate proportional to how far away the model solution is. This method is called a "data-clamp", and the data-clamp current takes the form

$$I_{data} = G_{data}\left(V - V_{data}\right), \tag{5.1}$$

where G_{data}, which can be thought of as the data-clamp current conductance, determines how quickly the computed voltage trace is pushed back onto the data trace, and V_{data} is the membrane voltage value at that time. This data-clamp current is added into the ODE model, so the general model is

$$C_m \frac{\mathrm{d}V}{\mathrm{d}t} = -\left(\sum_{ion} I_{ion} + I_{stim} + I_{data}\right), \tag{5.2}$$

$$= -\left(I_{tot} + I_{data}\right). \tag{5.3}$$

This data-clamp current could be thought of as a model discrepancy term, insofar as it provides the additional current required to ensure that V is equal to V_{data} at every time point along the recording. Since I_{tot} in general also depends on V, and hence $\frac{\mathrm{d}V}{\mathrm{d}t}$ depends on V, we cannot compute a discrepancy term/current by subtracting the computed model voltage trace from the experimental trace. This is because of the complex voltage-feedback system defined by the general action potential model. If we were to re-add this 'difference voltage' into the ODE model, we would not reproduce the desired voltage trace. If we therefore record I_{data} at every time point through the simulation, we would obtain an approximation for the sum of all of the unknown currents active in the myocyte.

We adapted this data-clamp method to clamp the generated action potential to the experimental voltage trace during just that part of the upstroke attributable to the stimulus current. We therefore did not use I_{data} as a possible model discrepancy term; we simply used the data-clamp method to ensure that we are representing the effect of the stimulus current exactly, and then let the action potential evolve as normal. To enable the CVODE solver to recover V_{data} at any time (it is taking adaptive time steps), linear interpolation is performed. This method was coded

into Chaste, and can now be added to any model generated from CellML. We switched on the data-clamp by setting its conductance G_{clamp} to 300 mS/μF during the time period shown in Figure 5.3, and zero otherwise.

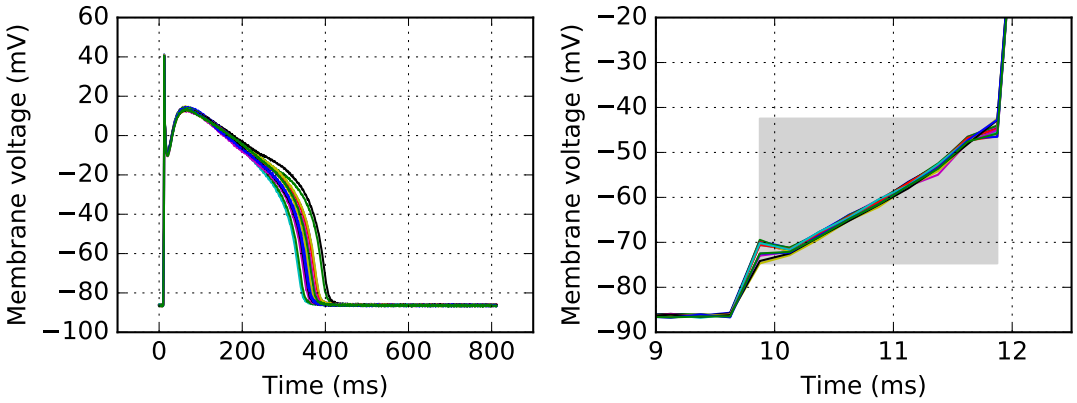

Figure 5.3: Left: Experimental action potential recordings of the first 16 paces from the first shaded region in Figure 5.2. Right: zoomed-in view to show the effect of the experimentally-applied stimulus on membrane voltage. The data-clamp current is applied during the time covered by the grey box.

5.2.2 Solving the action potential model

Using the experimental solutions described in Section 5.1 and the data-clamp method described in Section 5.2.1, we solve the (as yet unselected, we will discuss this in Section 5.2.3) action potential model according to the following procedure:

- Reset all state variables to the initial conditions provided in the model — 1 Hz pacing steady state.
- If the model contains them as parameters: set the intra- and extra-cellular potassium and sodium concentrations to their experimental values.
- Set the initial voltage value to the initial voltage value in the action potential trace.
- Solve the action potential model until the data-clamp start time.
- Apply the data-clamp current for a short time during the relevant part of the upstroke (see Figure 5.3).

- Turn off the data-clamp and allow the model to continue for as long as the experimental recording.

5.2.3 Choice of action potential model

There are many existing action potential models, and so it is important to choose an appropriate model for a particular application. Since we wish to model action potentials from a canine ventricular cardiomyocyte, we will compare two canine models from the CellML repository (Lloyd *et al.*, 2008): Decker *et al.* (2009) and Davies *et al.* (2011) (a modified version of Hund & Rudy (2004)). If both models can provide good fits to the experimental data, we want a well-defined reason to choose one model over the other before running our MCMC algorithm to perform parameter inference. We will use the *Bayesian Information Criterion* (BIC, Schwarz *et al.*, 1978) to decide between the two canine action potential models.

The BIC is defined as

$$\text{BIC} = \log(n)k - 2\log(\hat{L}), \tag{5.4}$$

where n is the number of data points, k is the number of parameters being estimated, and $\hat{L}$ is the maximum value of the likelihood function.

Equation (5.4) is derived from maximising the posterior probability, $p(\mathcal{M}_i|data)$, for a particular model — see Equation (2.31). During the derivation of the BIC, $-2\log(p(\mathcal{M}_i|data))$ is minimised, and a Taylor expansion approximation around the maximum likelihood point is used to disregard high-order terms, reducing to Equation (5.4). See Neath & Cavanaugh (2012) for a more thorough derivation of the BIC. The model with the lowest BIC is the preferred model among a set of candidates. Therefore, the BIC punishes model complexity (number of parameters) and rewards high likelihood (goodness-of-fit).

We wish to use the BIC to choose between the Davies *et al.* (2011) and Decker *et al.* (2009) canine action potential models. We therefore count the number of parameters we wish to estimate, and compute their maximum likelihoods. Both action potential models have 14 conductance (we use this term to describe all

maximal ion current densities) parameters that we can vary in order to fit to the experimental data.

Maximum likelihood

For the likelihood calculation, we again assume that each data point is Normally distributed around the trace generated by the action potential model with some standard deviation σ mV as in Equation (2.23), where $\boldsymbol{\theta}$ is now $\log(\boldsymbol{G})$, the vector of log(conductance) parameters. We therefore have 14 log(conductance) parameters and one noise standard deviation, σ, so the total number of parameters, k, is 15 for both models. The number of time points, n, is a property of the data that is also the same for both models. Since both n and k are the same, the first term on the right-hand side of Equation (5.4) will appear in both BIC calculations, and so selecting the model with the lowest BIC is equivalent to selecting the model with the greatest log-likelihood. Note that this is not true in the general case where models may have different numbers of conductance parameters.

We will use a CMA-ES optimiser to maximise $l(\log(\boldsymbol{G}), \sigma)$ as discussed in Section 4.1.1 of Chapter (4). As in Section 4.1.1, $\log(\boldsymbol{G})$ and σ decouple, and we can obtain a best-fit set of parameters by varying the log(conductance) parameters to minimise a sum of square errors, and then compute the best-fit value of σ analytically. Reducing the dimension of the parameter space being explored by the CMA-ES optimiser like this speeds up the overall computation.

CMA-ES optimisation

Unlike in the synthetic data case in Chapter 4, we no longer have a 'correct' model to fit to the data. There is therefore no ultimate 'true' solution that we know we are looking for. This means that it is more important here than in the synthetic data case to run as many initial CMA-ES optimisations as possible to increase the probability that we find the optimal point/region in parameter space.

We note that multiplying/dividing conductance parameters by constants has a more 'symmetric' effect on the action potential than adding/subtracting a proportion of the parameter values. This is illustrated in Figure 5.4. Here, we compare action

potentials generated by scaling the conductance parameter values by a constant and its reciprocal with action potentials generated by adding and subtracting a proportion of the parameter values. The case where we have multiplied and divided by a constant appears to result in a more symmetric spread of action potentials.

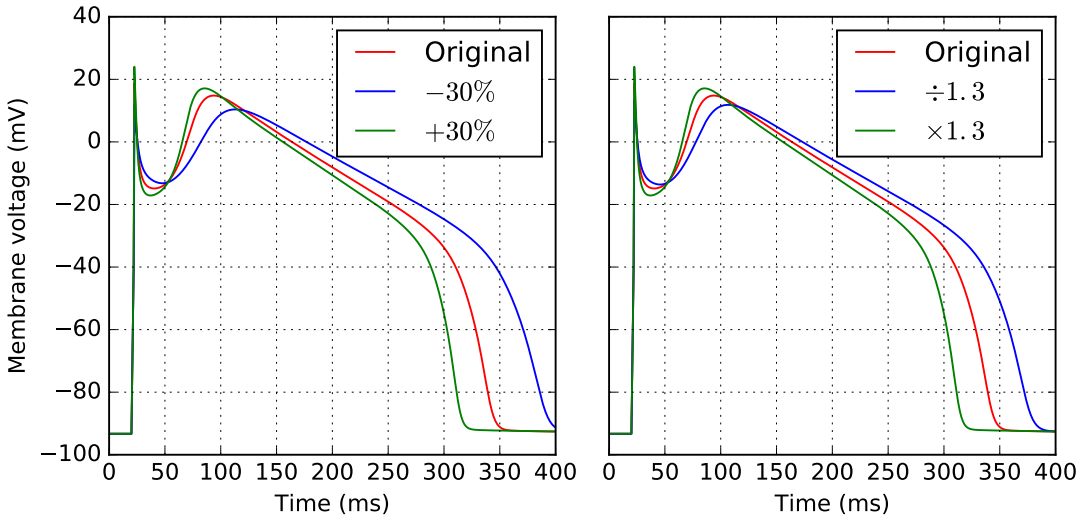

Figure 5.4: Action potentials generated by the Davies *et al.* (2011) canine model. Left: all conductance parameters scaled by ±30%. Right: all conductance parameters scaled by ×1.3 or ÷1.3.

When performing a CMA-ES minimisation, new points are proposed by a Normal distribution. In 1-d, a Normal distribution is symmetric about its mean μ, so $p(\mu + x) = p(\mu - x)$ for any x, where p is the probability density function of this Normal distribution. Figure 5.4 suggests that we actually want proposals such that the current parameter value $\times k$ and $\div k$ are equally likely. That is, roughly speaking, we want some transformation that allows $p(k\mu) = p(\mu/k)$. If x is the current CMA-ES iteration mean value, we want a scaling $G = f(x)$ such that if $f(x+a) = kG$, then $f(x-a) = G/k$, where a is the Normal perturbation to x, and G is the conductance parameter value that will be passed to the action potential model. This is further justification for considering log(conductance) parameters, since this property is already enforced: we take $f(x) = e^x$, since $f(x \pm a) = e^{x \pm a} = e^x e^{\pm a}$. As mentioned before, this scaling also automatically satisfies that condition that the conductance parameters be non-negative, and therefore there are fewer 'wasted' proposals.

Since some sets of parameter values fail to elicit an action potential, even when stimulated using Chaste's stimulus current or using the data-clamp method, we begin our CMA-ES optimisation near to the original model parameter values, as in Chapter 4. We know these parameter values produce an action potential, so it seems reasonable that the optimal set of parameter values to fit to the data could be relatively near to this point in parameter space. We therefore add different small Normally-distributed perturbations to these parameter values for every CMA-ES run we wish to perform. As with all optimisers, CMA-ES can theoretically get stuck in local minima optima and fail to find the global optimum. This is why we start the CMA-ES optimiser from multiple different points in parameter space.

Results

Two examples, for each model, of best log-likelihood fits found by the CMA-ES minimiser are plotted in Figure 5.5. The minimiser was run on each trace from 32 different starting points, and the best answers from all of those runs were saved. In the case of the Decker *et al.* (2009) model for the first trace, after selecting the best log-likelihood found by CMA-ES, 21 of the 31 remaining CMA-ES runs obtained log-likelihoods less than 0.5 less than the best log-likelihood, corresponding to likelihoods at least $\exp(-0.5) \approx 0.6$ times as likely as the best. The remaining 10 CMA-ES runs all found log-likelihoods less than 1.7×10^{-7} times as likely as the best log-likelihood. In the case of the Davies *et al.* (2011) model for the first trace, after selecting the best log-likelihood found by CMA-ES, 24 of the 31 remaining CMA-ES runs obtained log-likelihoods corresponding to likelihoods at least 0.8 as likely as the best. Two of the remaining 7 were between 0.2 and 0.6, and the remaining 5 failed to converge. While these fits seem to capture overall behaviour, such as action potential duration, quite well, there are clear points along the traces where there are noticeable differences, in particular at the turning points and the upstroke magnitude, more noticeably with the Decker *et al.* (2009) model. It is possible that these are indeed the respective global optima, but it is also possible that the minimiser is getting stuck in local optima.

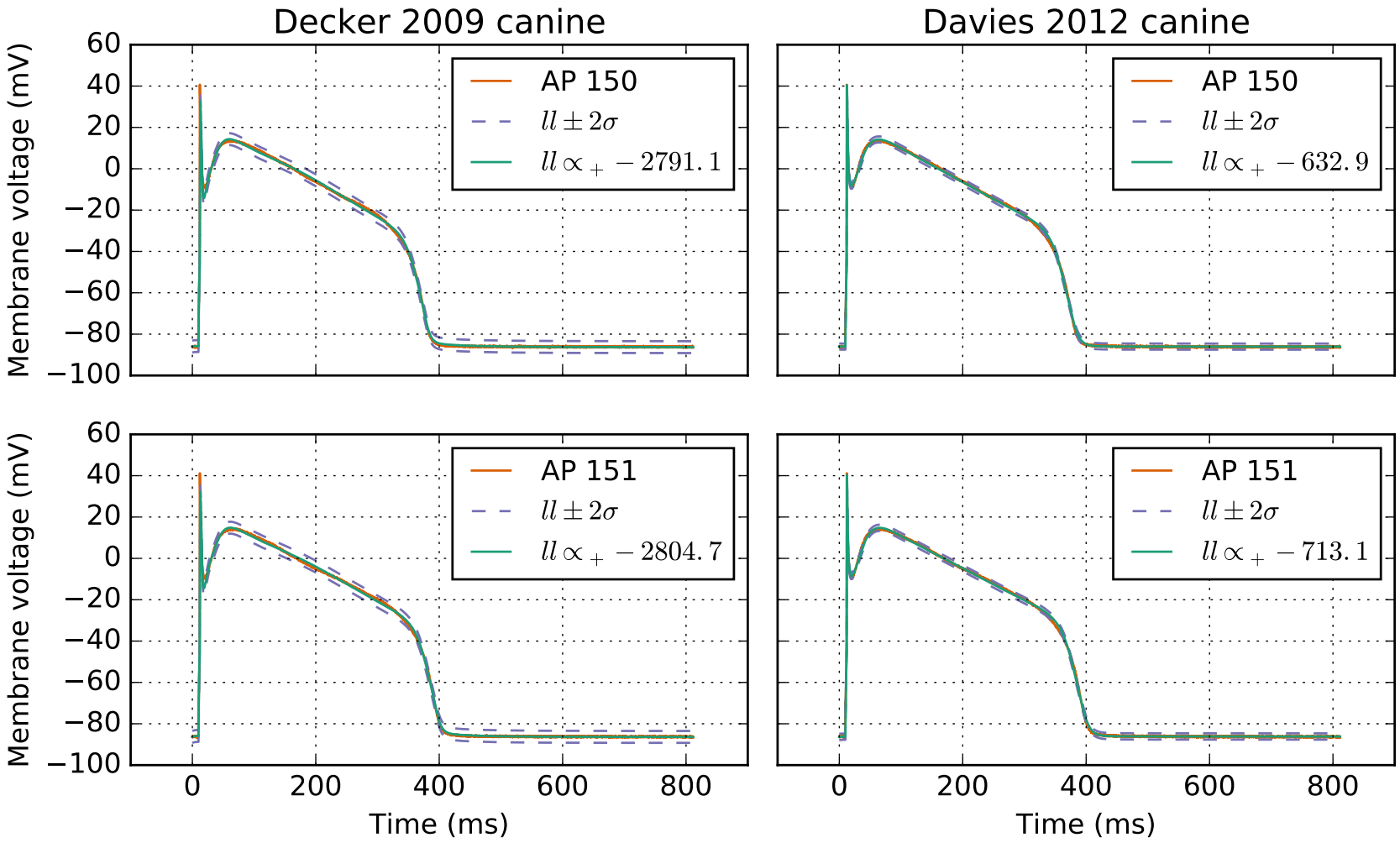

Figure 5.5: Maximum log-likelihood fits of the Davies *et al.* (2011) and Decker *et al.* (2009) models obtained from 32 CMA-ES optimisations.

For both action potential traces, the Davies *et al.* (2011) model gives a much greater maximum log-likelihood. In both cases, therefore, the Davies *et al.* (2011) model has a smaller BIC than the Decker *et al.* (2009) model. We have only performed this analysis on two action potential traces, but it suggests that the Davies *et al.* (2011) model may provide better fits to the present experimental data than the Decker *et al.* (2009) model. We only performed this BIC analysis for two of the action potential traces, but we have previously empirically observed better performance from the Davies *et al.* (2011) model. We will therefore continue to use the Davies *et al.* (2011) model when performing parameter inference from the present canine action potential data.

5.3 Parameter inference using MCMC

We will apply the Bayesian inference methods discussed in Section 4.2 of Chapter 4 to perform parameter inference and to explore variability between action potential traces. The experimental data provided to us were recorded contiguously from a single cardiomyocyte. The variability we will be considering is therefore beat-to-beat variability.

We concluded in Section 4.2.7 of Chapter 4 that the single-level model provides an adequate approximation to the hierarchical model relative to the computational cost. We will therefore run our adaptive Metropolis-Hastings (single-level model) algorithm to infer a probability distribution for the log(conductance) and observation noise standard deviation parameters for each action potential trace separately.

Before running our adaptive Metropolis-Hastings algorithm, we perform CMA-ES optimisations of the full log-target distribution. We perform 32 CMA-ES runs for each action potential recording (#150–165), as described in Chapter 4.

We consider a recording from a single canine cardiomyocyte, and infer possible conductances for a single pace (in the control setting). This process is repeated for a train of subsequent action potentials to uncover ionic currents that could be responsible for beat-to-beat variability within a single cell. While we do not believe that a sudden change of conductance values every second is a good model

for the emergence of beat-to-beat variability in a single cell, it may be adequate to characterise the variability and generate a predictive model for it.

In Figure 5.6, we present the log(conductance) parameters that were inferred for all membrane ion currents in the Davies *et al.* (2011) model, and the observation noise standard deviation parameter. Note that we do not expect all conductances to be fitted with narrow distributions to a single action potential, as there was insufficient information in the synthetic data situation to recover them all perfectly for this model with its original conductances (see Table 4.3 in Chapter 4), so we do not expect to recover background, pump or exchanger conductances. In Figure 5.6, we also plot the prior distribution for each parameter, but on a different y-scale from the normalised histograms. The top of the y-axis for these prior distribution plots is approximately equal to the maximum value of the prior probability density functions. We also plot our approximated predictive distribution, discussed in Section 4.2.7 of Chapter 4 on the same y-scale as the normalised marginal histograms.

The marginal histograms for most of the background, pump, and exchanger current log(conductances) ($\log(G_{pCa}), \log(G_{bCa}), \log(k_{NaCa})$), and the transient outward chloride current ($\log(G_{to2})$) are all very wide and nearly entirely overlapping between experiments. This suggests that there is not much information present in the individual datasets to constrain these parameters, and they could plausibly be removed from the model. However, we might expect these currents to have longer-term effects than over the course of a single pace. This is one drawback of fitting the action potential model to each pace separately — we lose information about longer-timescale effects.

We also note that the inferred distributions for the observation noise standard deviation σ roughly take values in the range 0.8–1 mV. This contrasts with the value of 0.25 mV we measured when considering a flat section of some of the recordings, as mentioned in Section 4.1.1 of Chapter 4. The inferred distributions for σ are wider because it is also taking overall model error into account, and not just the flat sections at the end. That is, where the model does not match the data, the statistical model treats this difference as observation noise and so increases σ to compensate.

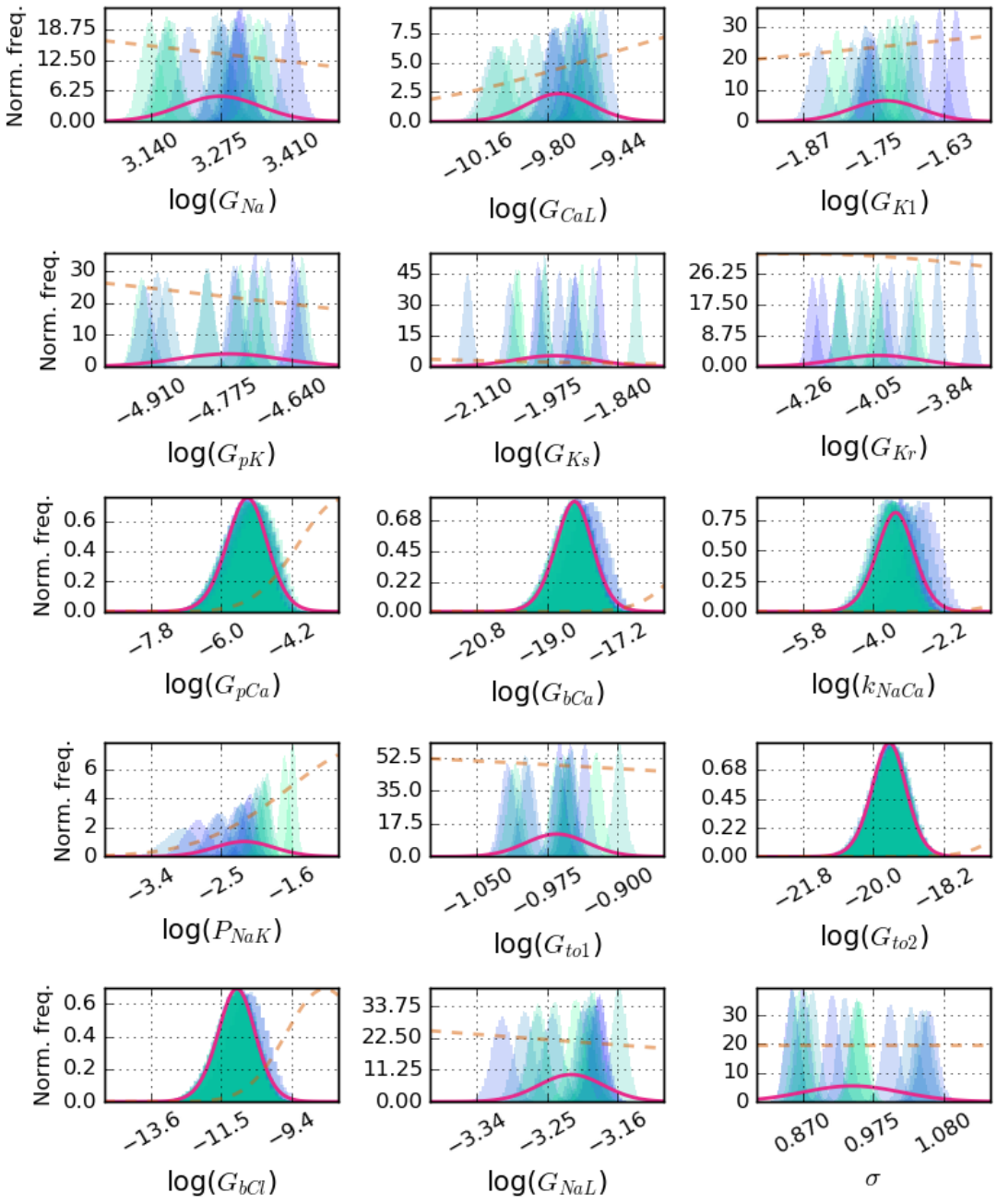

Figure 5.6: Normalised marginal histograms of log(conductance) parameters and the noise parameter σ, inferred from a series of 16 canine cardiomyocyte action potentials. The colours transition from semi-transparent blue to semi-transparent green through time. That is, ▭ represents the first pace, and ▭ represents the final pace. The dashed orange lines are the prior distributions, plotted on a different y-scale from the normalised histograms. Each log(conductance) prior y-axis goes from 0 to 0.35, which is approximately the maximum value of the priors. The σ prior y-axis goes from 0 to 2× the uniform constant value. The solid fuchsia lines are the approximated predictive distributions.

To compare inter-experiment variability, we can take the maximum posterior density points from each experiment and observe how these vary. We take the means and standard deviations for these maximum posterior density points across experiments. These values are presented, in decreasing order of standard deviation, in Table 5.1. Some parameters' marginal histograms overlap almost entirely, and thus we expect that these particular ion currents do not contribute to a single action potential, not that they are highly conserved across beats. In Table 5.1, we have highlighted the parameters whose maximum posterior density points have the greatest standard deviation between beats, but we have excluded the overlapping ones mentioned previously. We believe that these highlighted parameters have the biggest impact on the beat-to-beat variability observed in the canine action potential data.

Parameter	Mean	Standard deviation
$\log(P_{NaK})$	-2.189	0.333
$\log(k_{NaCa})$	-3.395	0.325
$\log(G_{bCl})$	-10.971	0.202
$\log(G_{pCa})$	-5.172	0.184
$\log(G_{CaL})$	-9.73	0.159
$\log(G_{to2})$	-19.548	0.154
$\log(G_{bCa})$	-18.225	0.144
$\log(G_{Kr})$	-4.037	0.127
$\log(G_{pK})$	-4.76	0.099
$\log(G_{Ks})$	-1.963	0.076
$\log(G_{Na})$	3.271	0.075
$\log(G_{K1})$	-1.728	0.061
$\log(G_{NaL})$	-3.221	0.041
$\log(G_{to1})$	-0.966	0.034

Table 5.1: Means and standard deviations of the maximum posterior density points inferred from 16 action potential traces. Parameters are ordered by decreasing standard deviation. The parameters that we believe contribute the most to the observed beat-to-beat variability are highlighted in green.

5.4 Prediction of beat-to-beat variability in a canine cardiomyocyte

We used the inferred log(conductance) parameters from Section 5.3 to predict the beat-to-beat variability that would be observed after Moxifloxacin addition and the change in bath concentrations. The action of 10 μM Moxifloxacin was modelled as conductance block with IC_{50} values of: 29 μM for I_{Kr} (Alexandrou *et al.*, 2006); 206.7 μM for I_{Na} (Harmer *et al.*, 2011); and 158 μM for I_{Ks} and I_{CaL} (both have pIC_{50} values of 3.8 in Mirams *et al.*, 2014). In this example, we take the Hill coefficient to be 1, and are using these single-value estimates for the IC50 values, and so are not taking uncertainty in ion channel screening into account, as discussed in Chapter 3. We predicted the correct trends in the data with an increased beat-to-beat variability in APD after the intervention, although the APDs themselves were generally shorter than those recorded experimentally, and the predicted resting membrane voltages were lower than those recorded experimentally.

We sample from the constructed approximate predictive distributions (fuchsia lines in Figure 5.6) 2000 times. We use these 2000 sets of log(conductance) parameters to simulate action potentials in the control environment. These are plotted as blue traces in the right-hand plot of Figure 5.7. We took 2000 more samples for each log(conductance) parameter, scaled the relevant parameters as described in the previous paragraph under the addition of 10 μM Moxifloxacin and 4 mM KCl rather than 5.4 mM extracellular concentration, and simulate new action potentials. These new action potentials are plotted as red traces in the right-hand plot of Figure 5.7. The model consistently predicts a lower resting potential and a longer APD than control, as did the experiment. But the model overestimates the reduction in resting potential and underestimates the APD prolongation. This could be due to the experiment not having reached a steady-state response, the resting potential was still dropping at paces #400–429. The original model parameterisation behaves the same way, so whatever the cause, this is not purely a product of our beat-to-beat characterisation procedure.

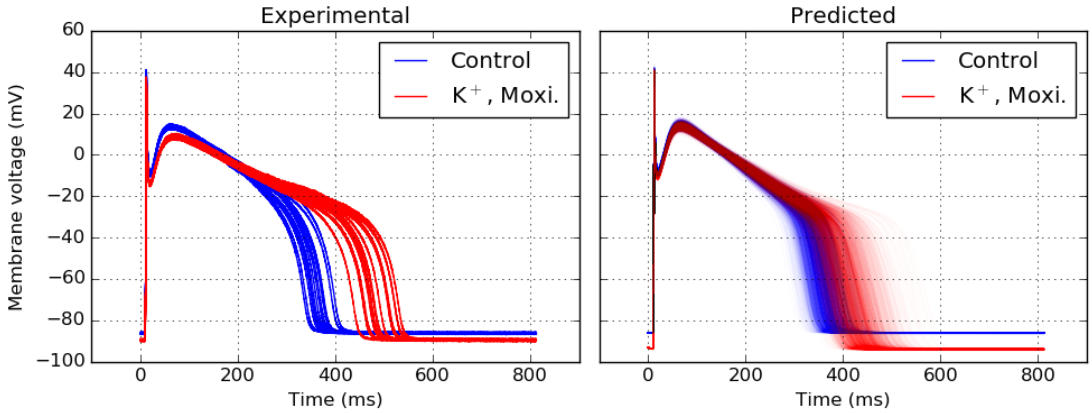

Figure 5.7: Action potentials generated from 2000 samples from the constructed predictive distributions. Blue: control. Red: Moxifloxacin.

In Figure 5.8 we plot violin plots of the APD_{90}s of the predicted action potentials plotted in Figure 5.7. We also plot the experimentally-observed APD_{90}s from both the control and Moxifloxacin settings. As expected, the predicted APD_{90}s for the control setting match well with the observed APD_{90}s. This is expected as the predictive distributions were constructed based on parameter inference performed on these experimental traces. However, the predicted APD_{90}s for the Moxifloxacin setting tend to be lower than the experimentally-observed values, although they are still contained within the tails of the distribution. The slight underestimate of variability in the simulation predictions is consistent with the observation that longer action potentials lend themselves to higher variability (Heijman *et al.*, 2013), and so we would expect that a model which reproduced the baseline response more closely may also improve the beat-to-beat variability predictions.

The fact that the model predictions are not very good for the absolute APD and resting potential indicates that Structural Uncertainty may be playing a large role. This may be caused by differences between the experimental preparation and its approximation in our simulations, rather than simply the model being inadequate/untested for simulating altered extracellular potassium and block of ion channels.

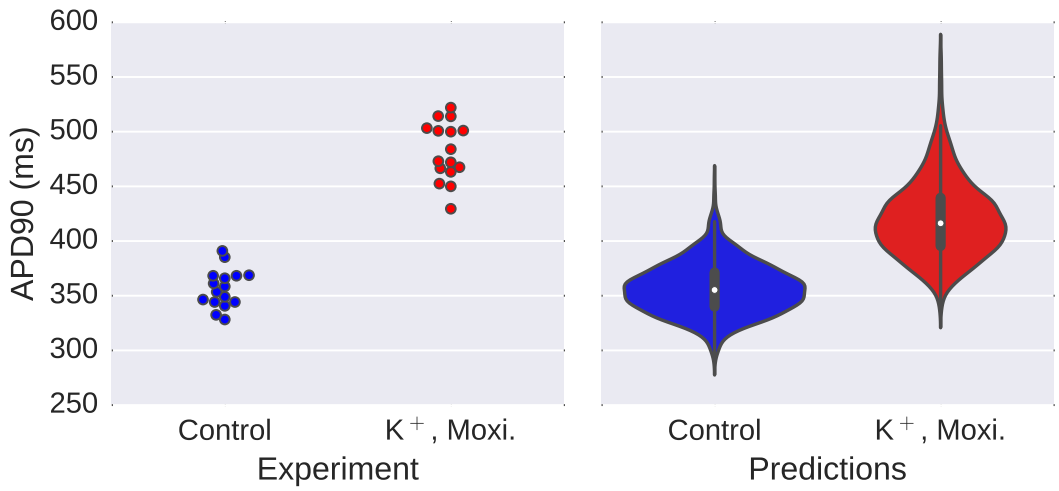

Figure 5.8: APD_{90}s computed from the action potentials shown in Figure 5.7. Left: experimental action potential recordings. Right: action potentials computed from samples drawn from predictive distributions.

5.5 Discussion

Models of the cardiac action potential are powerful tools that have provided mechanistic insight into complex biophysical systems. However, cardiac cells and tissue are adaptive and heterogeneous living systems, where there is variability in the shape and size of individual cells, and also in the density of ion channels, pumps, and exchangers in the cell membrane. Beat-to-beat fluctuations in ion concentrations and ion channel gating introduce further variability into the action potential. Although the present generation of cardiac action potential models can provide a mechanistic description of action potential generation, they neither capture these types of variability, nor is it always easy to determine how the model outputs depend on the parameters.

Parameter distributions were inferred from each individual voltage trace for a given pace, using action potential recordings obtained from a canine ventricular cardiomyocyte. Plotting the distributions for a single parameter from multiple different action potential traces together shows the beat-to-beat variability in maximal ion channel conductances. This is a new approach to investigating which currents underlie beat-to-beat variability, providing novel insight into normal cardiac

function. In particular, $\log(P_{NaK}), \log(G_{CaL})$, and $\log(G_{Kr})$ have among the widest spread of distinct marginal posterior probability distributions, suggesting that I_{CaL} and I_{Kr} are two of the main currents responsible for beat-to-beat variability.

Heijman *et al.* (2013) took an alternative approach, where they "developed stochastic implementations of 13 major ionic currents and fluxes in a model of canine ventricular-myocyte electrophysiology". They varied the ion channel density in this computational cell model, while re-scaling maximal ion current densities so that the total average current remained constant. Heijman *et al.* found that the currents I_{Na} and I_{Kr} had the greatest effect on beat-to-beat variability, while I_{Ks} also had a noticeable effect on beat-to-beat variability while generally not affecting the cell's action potential duration. The authors also found that "pumps and exchangers ... contributed little". Our method provides a different conclusion about the sodium-potassium pump current, whose marginal distributions in Figure 5.6 appear to be well-defined and distinct, suggesting that they vary between beats, and Table 5.1 shows that its maximum posterior density points vary the most. Table 5.1 agrees that I_{Na}, I_{Kr}, and I_{Ks} are important currents when considering beat-to-beat variability, however we also note that the potassium pump current appears to have a measurable effect between beats.

Lemay *et al.* (2011) also investigated effects of stochastic ion channel opening on beat-to-beat variability in cardiac action potentials. The authors' findings agree with Heijman *et al.* that I_{Na} and I_{Ks} play an important rôle in beat-to-beat variability. However, Lemay *et al.* also noted the importance of I_{CaL} in causing beat-to-beat variability. Our method agrees that I_{CaL} could be important in beat-to-beat variability — see Table 5.1. We have not considered the stochastic effects investigated by Heijman *et al.* and Lemay *et al.*, but stochastic model formulations could be incorporated into the methodology presented in this chapter for further investigation.

The variation in maximal ion current conductance represents the varying number of active ion channels in the membrane throughout each pace. The spread of distributions for G_{CaL} suggests that this number of active channels could vary by up to a factor of 2. Similarly, the spread of distributions for G_{Na}

suggests that the number of active sodium channels may vary by a factor of 1.4. Whether these fluctuations are due to stochastic opening of ion channels, or other regulation/kinetics operating over multiple paces, will require further investigation. Another possibility is that these model parameters are more sensitive to noise on the trace, rather than the maximal ion current density significantly varying between beats.

A major limitation of the approach discussed in this chapter is the choice of initial conditions for the action potential model. An action potential model consists of many state variables which are derived quantities based on the values of different variables and parameters at any time. For example, some intracellular ion concentrations have their dynamics described by differential equations, and so are not just fixed parameter values. In this case, different initial values of some ion concentration could affect the rate of change of membrane voltage, even if the initial membrane voltage value is unchanged. It is therefore important to consider initial values of all variables in the action potential model, and not just that of membrane voltage.

As a very simplistic approach, we used the 1 Hz pacing steady state with the original model parameters as the initial conditions for every solve of the action potential model, during both the CMA-ES and the MCMC. However, we are testing different conductance values in each iteration, and so these initial conditions do not necessarily reflect the steady state of the model given these parameter values. We have also assumed that all action potential recordings are independent. It is clearly not reasonable to assume that all action potential traces are independent of each other, since the recordings were obtained in sequence.Therefore, the initial conditions of one pace should be equal to the final state during the previous pace.

To find approximate initial conditions satisfying this requirement, we could find a best-fit set of parameter values for the first pace, solving the action potential model to steady state, and then using the final state variables as the initial conditions for the subsequent pace. It would not be reasonable to solve every pace to steady state, so from the second pace on we might use the final state variables as the initial conditions for the subsequent pace. This increases the computational cost of

finding the best-fit parameter values for the first pace, but not for the rest. However, this becomes more complicated when we wish to run our MCMC algorithm to infer parameter distributions. For subsequent action potentials, we might take the maximum posterior density parameters from the output of the MCMC algorithm from the first pace and use the final state variables from the resulting action potential, and repeat for subsequent paces. In this case, the inferences must be performed in series, increasing the overall time taken.

Another option is to use the data-clamp method. Instead of just applying the data-clamp to the portion of the recording where we believe only the stimulus current is acting, we could apply it to an entire preceding pace (or multiple preceding paces), so that the state variables are forced to be as close to the recording as possible. However, this method would not inform intracellular ion concentrations which would depend on conductance values chosen in previous paces. In this case, we would probably have to, as mentioned above, use the maximum posterior density set of parameter values obtained by the MCMC algorithm and its final state variables for the subsequent pace.

Instead, we could perform parameter inference by fitting to all action potential traces as one long recording, where we change the conductance parameter values at the beginning of each pace. This would allow the state variables to be passed along to subsequent paces. However, this would require fitting (number of conductance parameters) $\times$ (number of paces) parameters, which would be very computationally expensive.

In the next chapter, we discuss patch-clamp experiments that I performed at Roche in Basel, Switzerland, where I was recording action potentials from stem cell-derived cardiomyocytes. We apply the same parameter inference techniques discussed in this chapter, but with more knowledge of experimental parameters, and thus greater control over the action potential simulation process.

6

Stem-cell derived cardiomyocyte experiments

In Chapters 4–5, we discussed the use of Bayesian statistical models in quantifying and propagating uncertainty in action potential modelling. We now discuss applying these techniques in a setting related to drug safety testing: membrane voltage action potentials recorded from a stem cell-derived cardiomyocyte (SC-CM).

In this chapter, we will discuss the experimental methods for collecting action potential data, as well as the application of methods discussed in previous chapters to new experimental recordings obtained from SC-CMs. I spent two separate months at Roche in Basel, Switzerland, performing experiments to collect action potential recordings from SC-CMs. During the first visit, I was trained on how to perform patch-clamp experiments on Chinese hamster ovary (CHO) cells. CHO cells come from a very stable cell line and can be patched relatively easily. CHO cells are not excitable, however, and so I was not recording membrane voltage action potentials, but just learning the methods of physically patching a cell. I moved on to performing patch-clamp experiments on SC-CMs in the second visit, which I will discuss in Section 6.1.

As discussed in Chapter 2, SC-CMs are of increasing importance to the pharmaceutical industry as a surrogate for adult human cardiomyocytes, of which sources

are limited, in particular following the introduction of the CiPA initiative. The final stage of the CiPA initiative involves validating *in-silico* model predictions against experimental recordings taken from SC-CMs. It is therefore important that we are able to apply the methods discussed in previous chapters to this source of data. If we are able to do so, there can be better uncertainty quantification throughout the CiPA process, ultimately allowing for better-informed decision making in the drug discovery and safety process.

I performed all experiments to collect the data presented in this chapter. Some of the results from these experiments have been published in Lei *et al.* (2017).

6.1 Experimental methods

6.1.1 Whole-cell patch-clamp

Membrane voltage action potentials can be recorded from excitable cells by performing *patch-clamp* experiments. The cells are prepared in a bath containing "external solution", chosen to keep the cells alive and stable. A glass micropipette is filled with "internal solution", chosen to replicate the cytosol of the cells as closely as possible. The micropipette also contains a recording electrode. A ground electrode is placed in the bath. Both of these electrodes are connected to an amplifier which holds either the voltage ("voltage-clamp") or current ("current-clamp") at a specified level, and measures the resulting current or voltage, respectively.

The experiments I performed at Roche were "whole-cell" current-clamp experiments. The tip of the micropipette is placed against the cell membrane, and a negative pressure is applied to encourage the cell to form a seal around the tip of the micropipette. When the micropipette resistance is in the giga-ohm range (known as a *giga-seal*), a greater negative pressure is applied in the micropipette to break open the cell membrane. The volume of the micropipette is much greater than the volume of a cell, and so the cell's original contents become negligible as a fraction of the combined contents. We can therefore model the chemical composition of the intracellular region as being the same as that of the micropipette. If the

giga-seal has remained intact, the cell membrane now acts as a variable resistor in the electrical circuit. A simplified diagram of this setup is given in Figure 6.1.

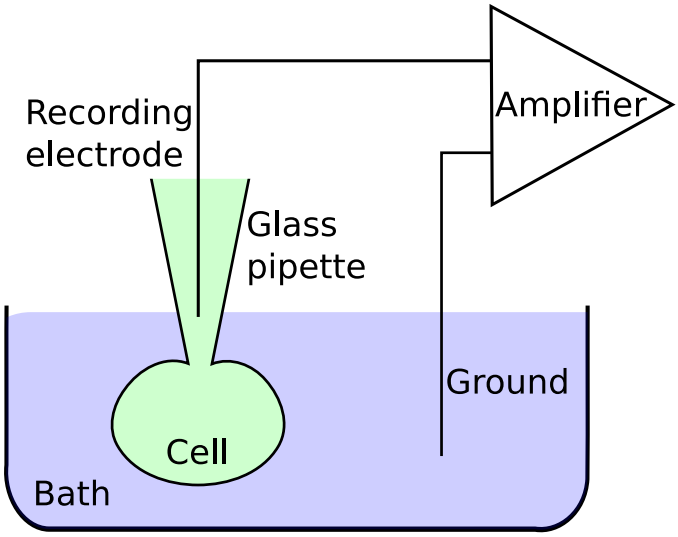

Figure 6.1: Experimental setup for whole-cell patch-clamp experiments. Note that this diagram is not to scale, in reality the volume of the cell is much smaller than that of the pipette.

6.1.2 Stem-cell derived cardiomyocytes

The stem cell-derived cardiomyocytes I worked with at Roche were provided by Cellular Dynamics International (CDI), from their "iCell" line. The iCells were plated and grown on fibronectin-coated glass cover slips according to the manufacturer's instructions with a density of 33,000 cells/cm^2. Medium change (iCell Maintenance Medium, CDI) was performed every 2 days.

The extracellular solution contains (in mM): NaCl 150; KCl 4; $CaCl_2$ 1.2; $MgCl_2$ 1; HEPES 10; pH 7.2–7.6 with NaOH, osmolarity 290–330 mOsm. The internal solution contains (in mM): KCl 125; NaCl 10; $MgCl_2$ 1; HEPES 10; EGTA 5; Mg-ATP 5; Na_3-GTP 0.1; pH = 7.0–7.4 with KOH, osmolarity 260–300 mOsm.

6.2 Experimental results

On each day of action potential recording, coverslips with the cardiomyocytes were placed at the bottom of the experimental chamber filled with the extracellular

solution. The action potential recording in single cells was performed using the patch-current-clamp technique in the whole-cell configuration at 35–37°C using the EPC-10 triple amplifier (HEKA Elektronik GmbH, Germany) and associated Patch MasterPro software (HEKA Elektronik GmbH, Germany). Stimulated action potentials were elicited by injecting 1 nA current pulses of 2 ms duration starting at 50 ms. Recorded signals were low-pass filtered using the internal Bessel filter of the EPC 10 and are digitized at 25 kHz.

Some of the cells were observed to be spontaneously beating without being externally stimulated. Once giga-seals were achieved and the cell membranes opened, most of the cells displayed resting potentials of -80 – -65 mV, with some being as high as -50 mV. The cells' resting potentials fluctuated in a 10–15 mV range of their mean resting potential, possibly due to the seal with the pipette deteriorating once the cell membrane was opened.

An example of such a scenario is given in Figure 6.2. Action potentials elicited from regular 1 Hz pacing are plotted, coloured according to time, where yellow traces are earlier and violet traces are later. There is clear shortening of action potential duration, possibly due to the cell settling into a new steady state. However, there is also an upward drift of baseline potential as well as lower peak voltages attained.

The reasons for the cells' instability are unclear. It is possible that the internal solution used was not close enough to the cells' internal composition, and therefore was not able to maintain normal cell behaviour. Alternatively, it might be that the cells themselves are inherently unstable. It is also possible that we might have obtained better results had I spent longer in Basel performing experiments.

Following these difficulties, we chose to prioritise 1 Hz pacing and primarily tried this frequency on cells that responded to stimulation. As in Chapter 5, we tried to find series of action potentials that appeared fairly similar to each other, suggesting the cell was in a nearly-steady state. The subset of action potential traces that we chose to examine further and model in this chapter is a series of 16 contiguous action potentials, starting from #100, recorded from a single cell. We

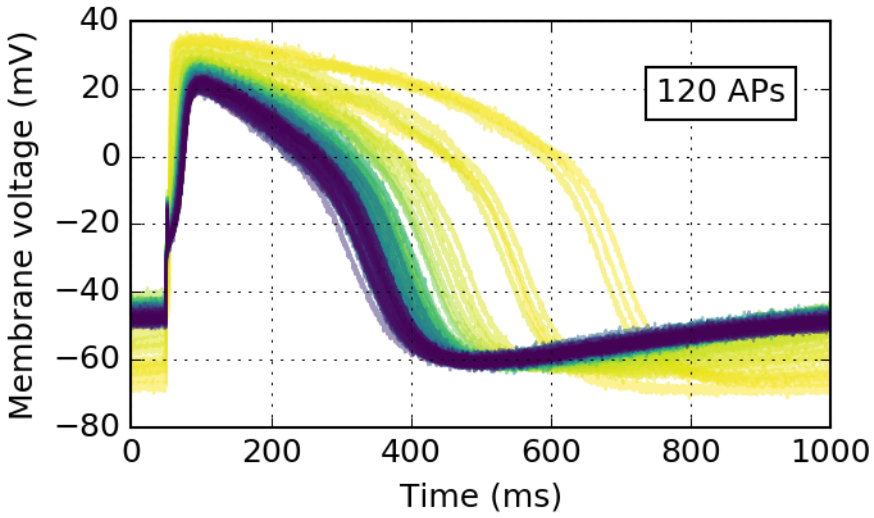

Figure 6.2: 120 action potentials recorded from a CDI iCell regularly paced at 1 Hz. The yellow traces are earlier and the violet ones are later. The cell appears unstable because of the rising resting potential after forming the seal, possibly due to degradation of the micropipette seal, or the condition of the cell itself.

therefore characterise the differences between action potentials again as beat-to-beat variability, instead of considering cell-to-cell variability.

An example of more-promising action potential recordings is given in Figure 6.3. As before, action potentials elicited from regular 1 Hz pacing are plotted, coloured according to time where yellow traces are earlier and violet traces are later. There is clear shortening of action potential duration, however the baseline potential is much more stable than in Figure 6.2.

We chose a subset of these action potential recordings to model using the techniques discussed in previous chapters. We used action potentials #100-115 because the earlier action potentials showed clear signs that the cell was adapting to the stimulation protocol, and we wanted to model action potentials closer to a steady state. These 16 actions potential recordings are shown in Figure 6.4, along with a zoomed-in portion. As before, the earliest trace is yellow and the latest trace is violet. The zoomed-in plot shows that the trend of shortening action potential duration has ended, and so we believe that the cell is approximately in a steady state.

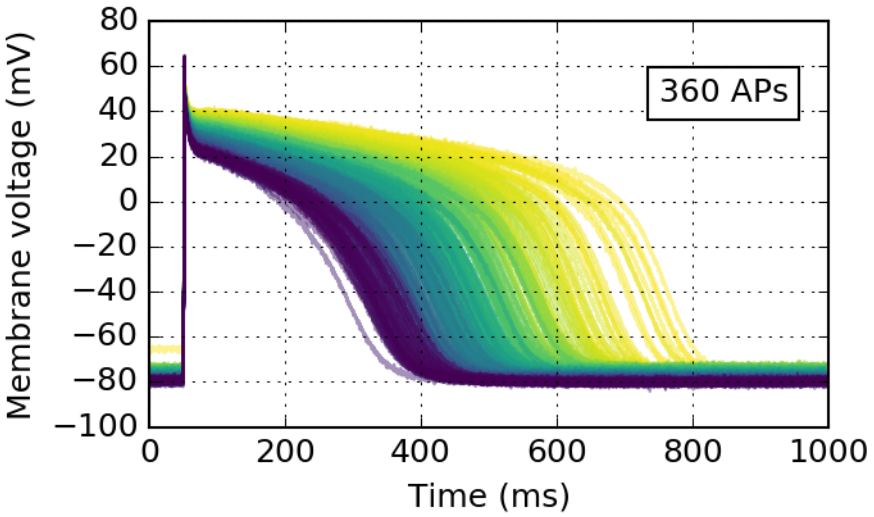

Figure 6.3: 360 action potentials recorded from a CDI iCell regularly paced at 1 Hz. The yellow traces are earlier and the violet ones are later. The cell appears more stable than the one in Figure 6.2 with respect to its resting potential.

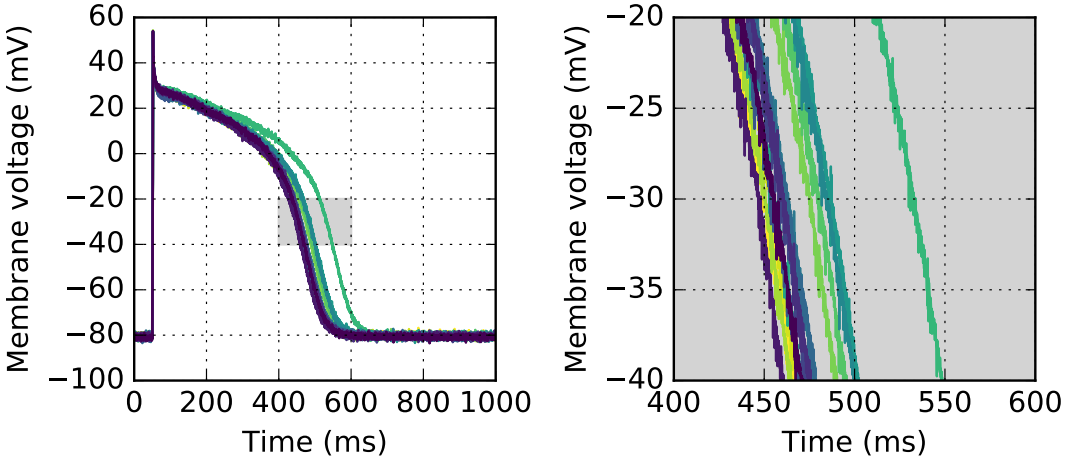

Figure 6.4: 16 action potential recordings starting from #100 recorded from 1 Hz pacing. The earliest recording is yellow and the latest is violet. The right-hand plot shows that the trend of decreasing action potential duration over time has is much less apparent.

6.2.1 Cell capacitances

I measured capacitances for 13 different iCells that produced action potentials when externally stimulated. These values are given in Table 6.1, and a histogram of these values is given in Figure 6.5.

Cell capacitance (pF)						
54.93	43.11	56.31	66.63	50.33	65.23	88.06
90.71	40.71	128.39	77.97	58.72	75.13	

Table 6.1: Capacitances of 13 different iCells that produced action potentials when externally stimulated.

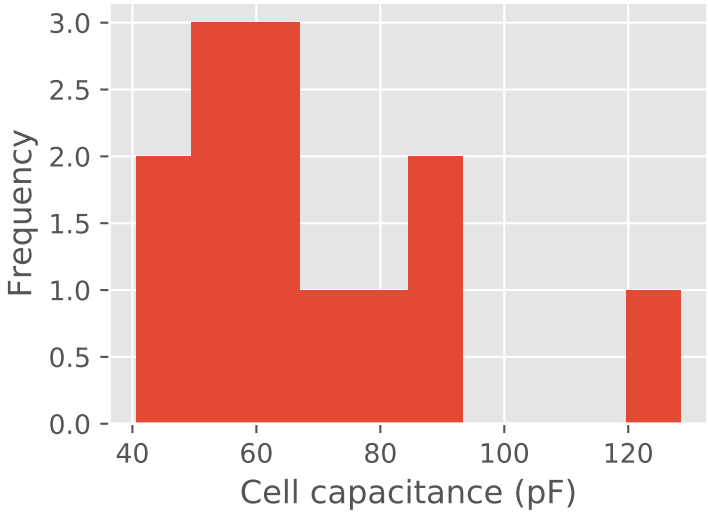

Figure 6.5: Histogram of the cell capacitance values given in Table 6.1.

6.3 Modelling results

6.3.1 Stimulus protocol

In Chapter 5, we used only the data-clamp method to elicit action potentials when solving the action potential model. This is because we did not perform the experiments ourselves and therefore did not originally know the stimulus protocol applied or the cell's membrane capacitance. However, I performed the experiments to obtain the action potential data discussed in this chapter, and so I was in control

of the stimulus current being applied to the cells, and I was able to measure the cells' membrane capacitances when patching the cells.

The importance of the cell's membrane capacitance when modelling action potentials can be seen from Equation (2.5). The conductance parameters in an action potential model can be scaled by the membrane capacitance and evaluate to the same value in Equation (2.5). Different published action potential models handle the membrane capacitance differently: some keep the membrane capacitance on the right-hand side in Equation (2.5), and others divide through by the membrane capacitance and re-define the conductance parameters so that only the stimulus current is divided by the membrane capacitance. Because of this, it is important to manually check the formulation and units of each action potential model that we choose to consider, and scale the stimulus amplitude by the cell's membrane capacitance when necessary.

We also note that there is an artefact in the membrane voltage action potential recordings introduced at the start of the application of the stimulus current. At the cell membrane's resting potential, we expect that there are no currents flowing across the membrane. Therefore, when the stimulus current is applied, the membrane voltage should increase linearly according to Equation (2.2). However, we noticed that when stimulating the cells, the membrane voltage initially increased quickly before settling in to the expected linear increase. This is illustrated in Figure 6.6, where we have plotted the same 16 action potentials from Figure 6.3. The stimulus current is applied during the time between the vertical red lines in the left-hand plot. We applied the correctly-scaled stimulus current in the action potential models for the duration of this time period. To account for the initial non-linear increase in membrane voltage, we apply the data-clamp, discussed in Chapter 5, for a short period of time described by the vertical green lines in the right-hand plot of Figure 6.6.

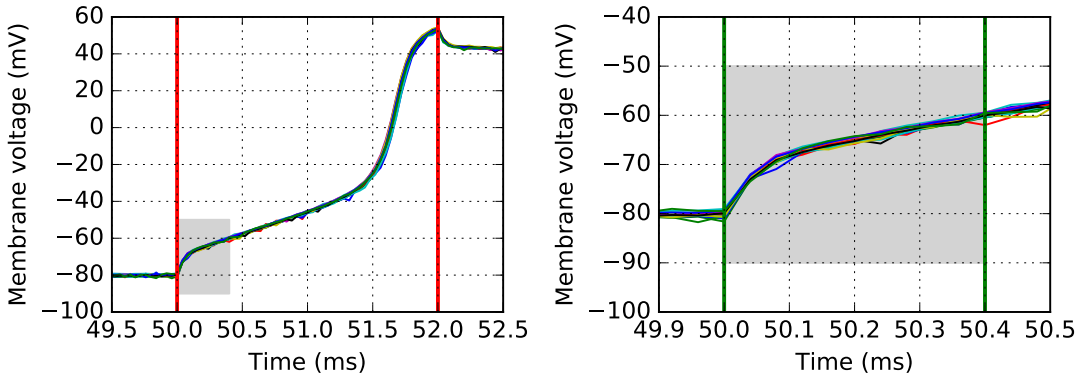

Figure 6.6: The same 16 action potentials from Figure 6.3. The membrane voltage increases quicker than expected before settling in to the expected linear increase. Left: the vertical red lines show the start and end times of the stimulus current. Right: the vertical green lines show where the data-clamp was applied.

6.3.2 Model selection

As discussed in previous chapters, it is important to select the best model for a particular application. We had originally planned to use the Paci *et al.* (2013) model, since it was fitted to data collected from CDI iCells (Ma *et al.*, 2011). In particular, the Paci *et al.* (2013) model has a model formulation for the "funny current", which acts as a pacemaker current for the cell, which regulates spontaneous beating. However, since we are fitting an action potential model only to individual stimulated action potential traces, this current may not be having a large enough effect to justify its inclusion in the model. We therefore chose to apply the same model selection process discussed in Section 5.2.3 of Chapter 5, and compute the Bayesian Information Criterion (BIC) for fits provided by two different action potential models to each of the 16 individual action potential traces shown in Figure 6.4.

We chose to compare the ten Tusscher *et al.* (2004) and Paci *et al.* (2013) action potential models. These action potential models have both have 12 maximal ion current density parameters. However, based on the results of Section 4.1.2 of Chapter 4 (Table 4.3), we do not expect to be able to infer information about some of the conductance parameters from individual action potential traces. We therefore limit the currents whose maximal conductance parameters we will allow to vary to $I_{Na}, I_{CaL}, I_{K1}, I_{Kr}, I_{Ks}$, and I_{to}. We will also include the funny current I_f in the

Paci *et al.* (2013) model, as this is the most notable difference from the ten Tusscher *et al.* (2004) model. We are therefore fitting 6 parameters of the ten Tusscher *et al.* (2004) model, and 7 parameters of the Paci *et al.* (2013) model, in addition to an observation noise standard deviation parameter. The remaining conductance parameter values were left as the original values given in the published models. While we might expect the greater number of parameters to allow a better fit to the experimental data, the BIC penalises models with more parameters. The fit provided by the Paci *et al.* (2013) will have to provide a sufficiently-improved fit than the ten Tusscher *et al.* (2004) to justify the inclusion of the additional parameter.

To address the issue of unknown initial conditions, we perform the log-likelihood CMA-ES optimisation by solving the action potential for two paces, and comparing only this second pace with the experimental data. That is, at each CMA-ES iteration, we set the state variables to those provided in the original model and then solve the action potential model from $t = 0\,\text{ms}$ to $t = 2000\,\text{ms}$ with a regular $1\,\text{Hz}$ stimulus (including the short data-clamp application discussed in Section 6.3). We then take the second half of the simulated membrane voltage ($1000 \leqslant t \leqslant 2000$) and use it as the 'test solution' to compare with the experimental trace. $2000\,\text{ms}$ is not long enough for the model to reach a steady state, but allows for more-realistic behaviour. For example, given a set of conductance parameter values, the model's resting potential might be higher than the initial value of the membrane voltage. Therefore, before the stimulus current is applied, the membrane voltage will drift up to this resting potential value, but this behaviour is never observed. Solving the action potential for two paces allows the cell to reach its resting potential before comparing with the experimental recording.

For all 16 action potential recordings, the Paci *et al.* (2013) model obtained the lowest BIC and was therefore selected over the ten Tusscher *et al.* (2004) model. For two of the experimental traces, the best-fit action potentials from the ten Tusscher *et al.* (2004) and Paci *et al.* (2013) models are shown in Figures 6.7 and 6.8, respectively. The fits from the ten Tusscher *et al.* (2004) do not attain the maximum membrane voltage in the experimental recordings and take longer to settle in to

their resting potentials after repolarisation. The fits from the Paci *et al.* (2013) model do attain closer to the maximum membrane voltage and settle in to their resting potentials quicker, but the model behaviour observed immediately after the upstroke is not present in the experimental recordings. The BICs computed for all 16 action potential traces with both action potential models are given in Table 6.2.

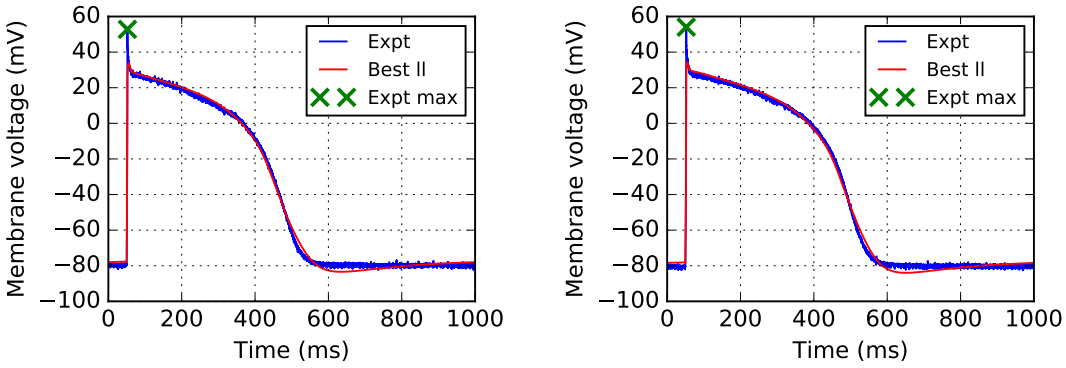

Figure 6.7: Best-fit action potentials obtained by fitting the ten Tusscher *et al.* (2004) model to two different experimental recordings.

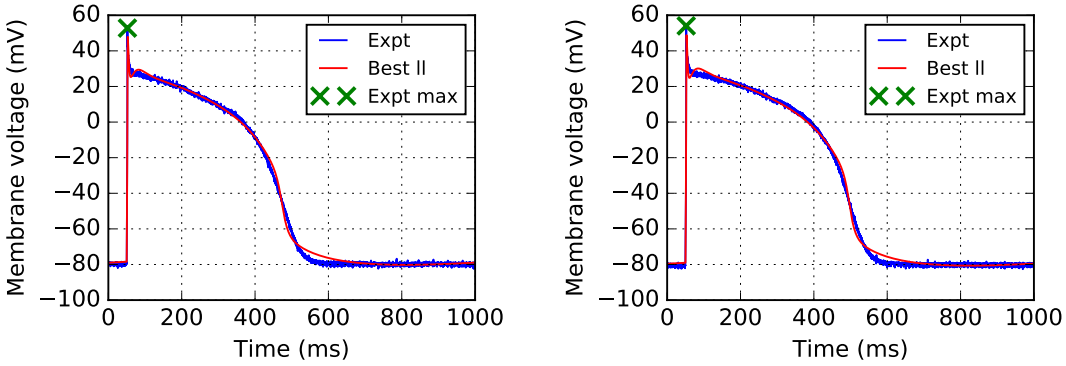

Figure 6.8: Best-fit action potentials obtained by fitting the Paci *et al.* (2013) model to the same two experimental recordings as in Figure 6.7.

6.3.3 Parameter inference

Following the results of Section 6.3.2, where the Paci *et al.* (2013) model was selected over the ten Tusscher *et al.* (2004) model using the BIC, we apply the same single-level MCMC methods discussed in Chapters 4–5 to infer a joint probability

Trace #	ten Tusscher *et al.* (2004)	Paci *et al.* (2013)
100	62668	55329
101	65937	54394
102	63999	50836
103	64286	55449
104	64467	55842
105	64396	58491
106	65524	52081
107	64463	55793
108	66014	54698
109	63127	56221
110	63495	55377
111	62660	57390
112	64180	54378
113	62983	55802
114	62431	52558
115	61887	55491

Table 6.2: BICs computed for all 16 action potential traces from 4 candidate models. Red: the highest (worst) BIC. Green: the lowest (best) BIC.

distribution for the subset of ion current conductance parameters discussed in Section 6.3.2 for the Paci *et al.* (2013) action potential model.

As before, we will perform parameter inference using an adaptive Metropolis-Hastings algorithm varying log(conductance) parameters. The only difference here is the initial conditions we give to the action potential model before each solve. We take the best log-likelihood found by the CMA-ES optimiser (those shown in Table 6.2), and save the state variables at the end of the computed action potential. These state variables are then used as the initial conditions for each solve during the MCMC run, where we solve for one pace at each iteration.

As in Chapter 5, we compare the parameters' marginal histograms from each experiment. These histograms, as well as the approximate predictive distributions (discussed in Section 4.2.7 of Chapter 4) and Normal prior distributions are given in Figure 6.9.

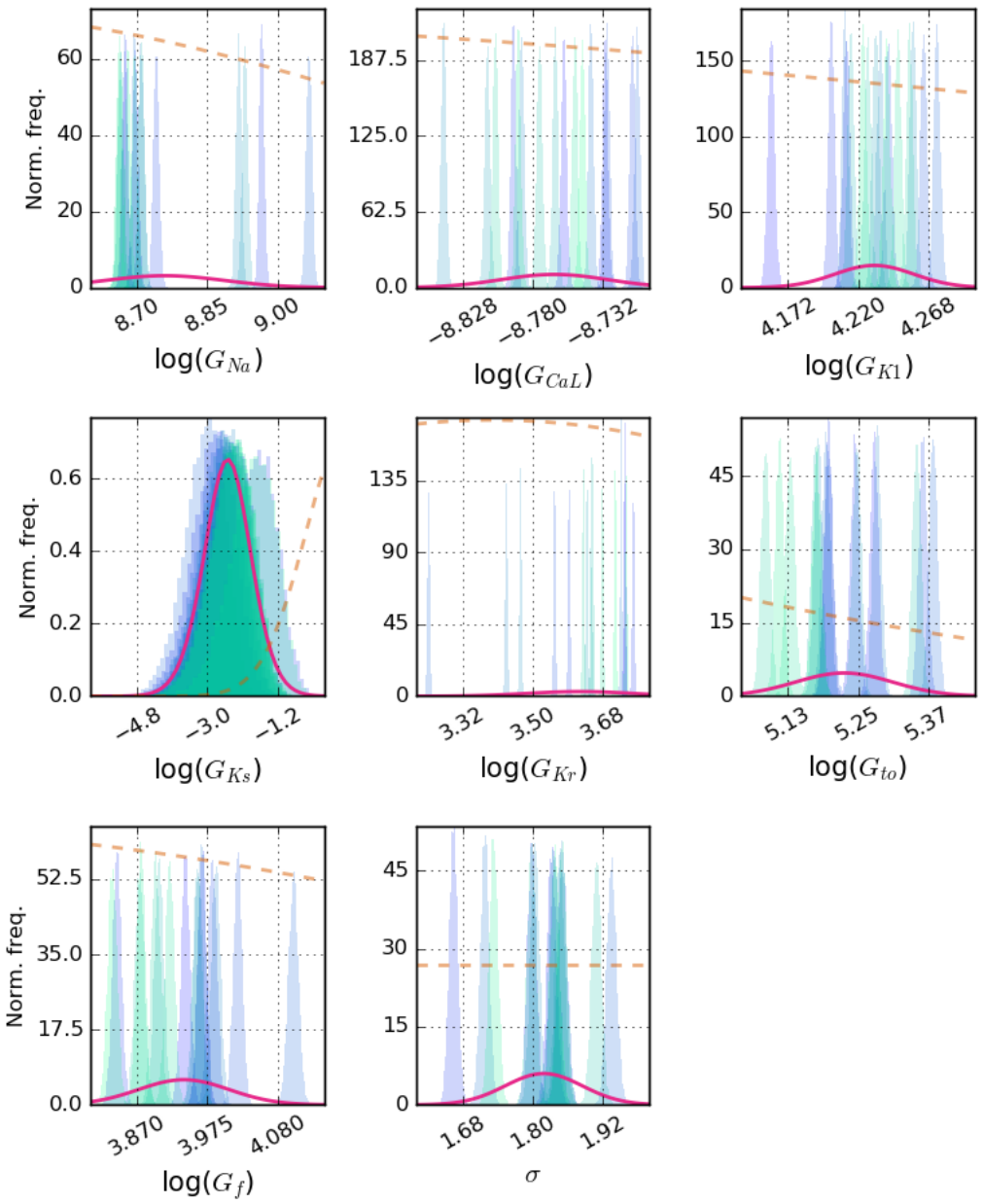

Figure 6.9: Normalised histograms for log(conductance) parameters from fitting the Paci *et al.* (2013) model to 16 separate action potential recordings. The fuchsia lines are the approximate predictive distributions (inferred from fitting the Paci *et al.* (2013) model to the action potential recordings), plotted on the same y-axis as the normalised histograms. The dashed orange lines are the Normal prior distributions (our subjective belief about the parameters before considering the experimental data), plotted on a different y-scale where the top of the axis is approximately the maximum value obtained by these priors.

6.4 Prediction of drug block on action potential behaviour

As part of the CiPA initiative, discussed in Chapter 2, computational action potential models are to be used in predicting cardiac drug safety along with ion channel screening data. We want to use the methods discussed in Chapters 4–6 to quantify and propagate some of the uncertainties associated with action potential modelling, along with the methods discussed in Chapter 3 to characterise some of the uncertainties associated with ion channel screening data and dose-response modelling. In this section, we will use the output of the single-level parameter inference performed in Chapter 3 as well as the output from Section 6.3.3 to predict possible ranges of action potential behaviour under different concentrations of Dofetilide. We will also consider model selection for the dose-response model defined in Equation (3.8), and using the transformation defined in Equation (3.10) so that the parameters in the dose-response model are pIC_{50} and $Hill$. The two versions of the dose-response model are

1. Allow pIC_{50} to vary and fix $Hill = 1$, and

2. Allow both pIC_{50} and $Hill$ to vary.

After constructing the approximate predictive distributions (inferred by fitting the Paci *et al.* (2013) model to the 16 action potential recordings), plotted as fuchsia lines in Figure 6.9, we sample from them to obtain log(conductance) parameter values. These values are exponentiated to obtain conductance parameter values. We then take random samples from the output of our MCMC algorithm applied to the ion channel screening data for Dofetilide for each of the 7 ion channels. This gives us pIC_{50} values when using Model 1, and gives us both pIC_{50} values and Hill coefficients when using Model 2. These parameter values, together with a chosen drug concentration, give % block of the 7 ion channels. We then scale the relevant ion current conductance parameters by these % blocks. These new conductance parameter values are passed into the action potential model to simulate an action

potential. This process is repeated many times and plotted on the same axes to build up a distribution of possible action potentials.

We wish to compare the difference in predictions (if any) when using outputs from the two dose-response models when simulating action potential behaviour under different concentrations of a drug. As discussed in Section 3.4 of Chapter 3, we compute the Bayes factors for comparing dose-response Models 1 and 2 when considering Dofetilide ion channel screening data (Crumb *et al.*, 2016). These Bayes factors as given in Table 6.3. In one case (hERG), Model 2 (vary pIC_{50} and $Hill$) is selected with "strong" evidence, according to Table 2.1 in Chapter 2. The best-fit dose-response curves of Dofetilide block on hERG for both models are shown in Figure 6.10. In one other case (Kv4.3), Model 1 (vary pIC_{50} and fix $Hill = 1$) is selected with "substantial" evidence, according to Table 2.1 in Chapter 2. The best-fit dose-response curves of Dofetilide block on Kv4.3 for both models are shown in Figure 6.11. In the remaining 5 cases, there is not evidence in favour of either model.

Channel	B_{12}	$B_{21} = 1/B_{12}$
hERG	0.06	17.38
Kv4.3	4.48	0.22
Cav1.2	0.98	1.02
Kir2.1	0.99	1.01
Nav1.5-late	0.98	1.02
KvLQT1/mink	0.98	1.02
Nav1.5-peak	1.05	0.95

Table 6.3: Bayes factors comparing dose-response Models 1 (vary pIC_{50} and fix $Hill = 1$) and 2 (vary pIC_{50} and $Hill$) for Dofetilide acting on 7 different ion channels. B_{12} is evidence in favour of Model 1 over Model 2.

We selected Dofetilide concentrations of $10^{-3}\,\mu$M and $1\,\mu$M as these concentrations give approximately 40–50% and 100% hERG (I_{Kr}) block, respectively. We sampled from both dose-response models and plotted the resulting action potentials. These plots are given in Figure 6.12. In each case, we used the outputs from each model on every channel. However, Table 6.3 suggests that we might use Model 1 to model I_{to} block, and use Model 2 to model the remaining ion current blocks.

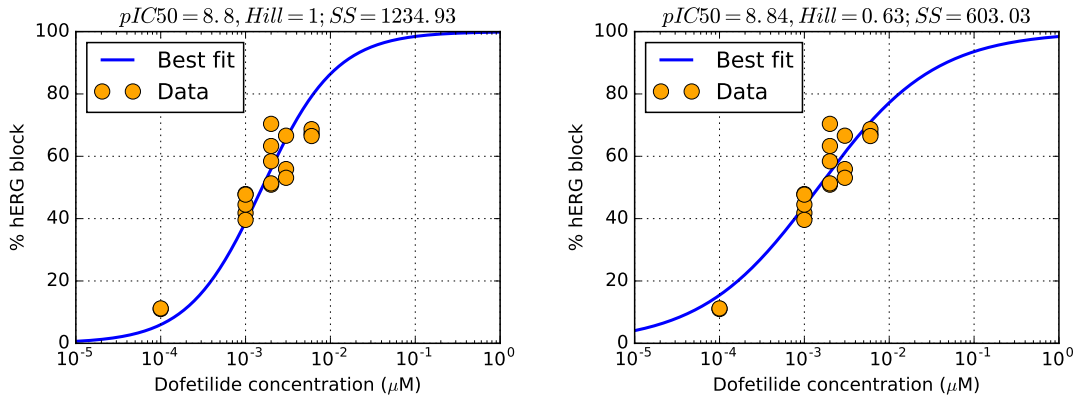

Figure 6.10: Best-fit dose-response curves fitted to ion channel screening data of Dofetilide block on hERG published by Crumb *et al.* (2016). Left: Model 1 (vary pIC_{50} and fix $Hill = 1$). Right: Model 2 (vary pIC_{50} and $Hill$). As shown in Table 6.3, Model 2 is selected by the Bayes factor.

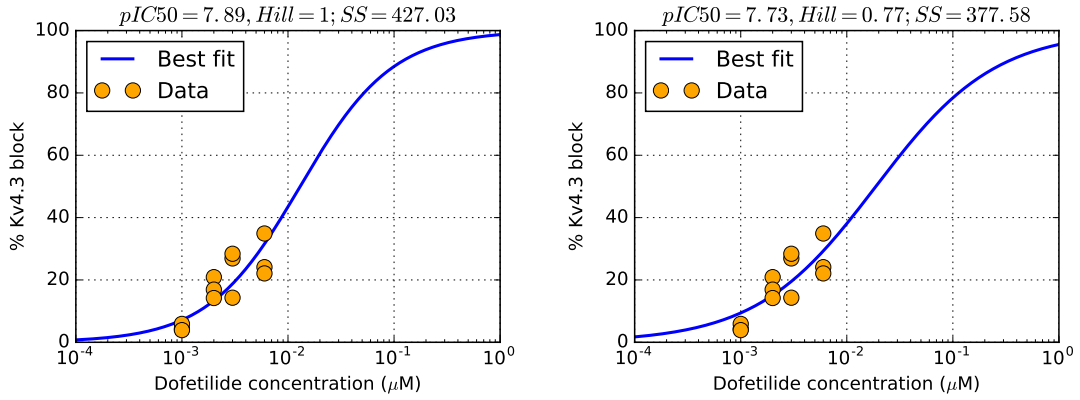

Figure 6.11: Best-fit dose-response curves fitted to ion channel screening data of Dofetilide block on Kv4.3 published by Crumb *et al.* (2016). Left: Model 1 (vary pIC_{50} and fix $Hill = 1$). Right: Model 2 (vary pIC_{50} and $Hill$). As shown in Table 6.3, Model 1 is selected by the Bayes factor.

The predictions shown in Figure 6.12 provide a concrete example of future work to test, where we would record multiple action potentials from an iCell with the relevant concentration of Dofetilide applied and observe how many of these action potentials fall within the distributions plotted in Figure 6.12. It is likely that a new member of our group will perform this work.

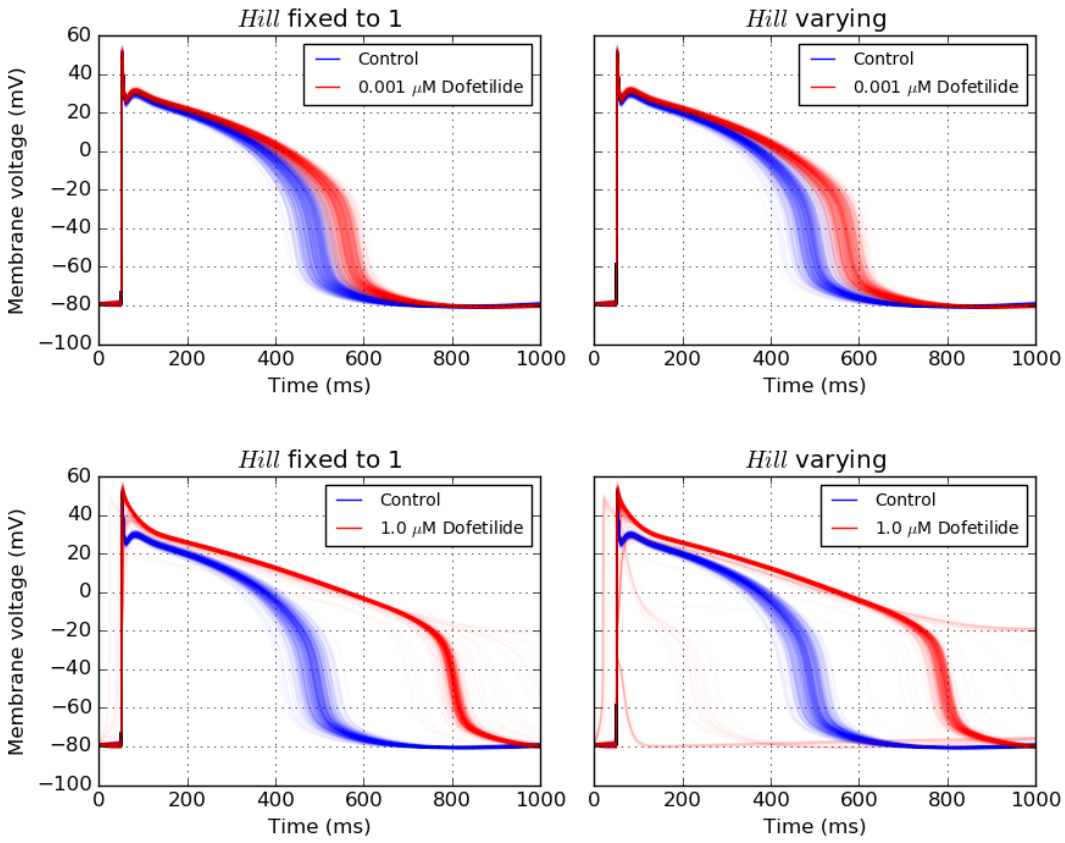

Figure 6.12: Predicted action potentials inferred from 16 experimental recordings using the Paci *et al.* (2013) model. Left: varying pIC_{50} and fixing $Hill = 1$; right: varying pIC_{50} and $Hill$. Blue: control; red: Dofetilide (top: $10^{-3}\,\mu$M; bottom: $1\,\mu$M).

6.5 Discussion

In this chapter, we have discussed the experimental work I undertook at Roche in Basel recording membrane voltage action potentials from stem cell-derived cardiomyocytes. There was a considerable amount of learning required, and as such I only obtained useful measurements during my second month there. In addition, to collect action potential recordings themselves, my time in the Roche lab gave me a greater appreciation for how unpredictable, and sometimes unreliable, biological experiments can be. It also gave me a greater understanding of what is possible in action potential-recording experiments, and what kinds of experimental protocols are generally infeasible.

We then applied the methods discussed in Chapters 4–5 to select an action

potential model (the choice of model was discussed in Section 6.3.2) for the obtained action potential recordings and to infer probability distributions for a subset of the models' log(conductance) parameters informed by the results of Section 4.1.2 of Chapter 4. Finally, we used the methods discussed in Chapter 3 to predict how two different doses of Dofetilide might affect action potential behaviour of these stem cell-derived cardiomyocytes. These predictions contain uncertainty that has been propagated through both the Dofetilide ion channel screening data and the action potential model parameterisation. This gives an initial attempt at a complete implementation of the process described in Figure 1.1 from Chapter 1.

We believe that, based on the available experimental data, Figure 6.12 give plausible estimates for Dofetilide response, taking into account beat-to-beat variability. The immediate next step would be to obtain stable action potential recordings from iCells that include the relevant concentrations of Dofetilide being applied. Comparing action potential model predictions with these new experimental recordings would be the validation step.

The experimental recordings discussed in this chapter came from a single cell. It would be interesting and useful to quantify uncertainty introduced through cell-to-cell variability, as well as considering cells from different batches from the manufacturer. It would also be useful to obtain recordings of different stimulus protocols, to see if we can extract more information about different ion currents. This would require a giga-seal to be maintained for several minutes and for the cell to not undergo obvious rundown. This is certainly plausible, but might require much longer in the lab for me to find a well-behaved cell.

Ultimately, the methodology discussed in this chapter might fit into a larger framework of inferring unexpected effects of pharmaceutical compounds on whole-cell electrophysiology. After re-parameterising an action potential model to describe the behaviour of a cell, we would introduce a drug at some concentration and record its new action potential behaviour. Based on these new observations, we would attempt to infer new distributions for the conductance parameters. Comparing these distributions with the pre-drug distributions allows us to infer which ion

currents the drug is affecting. Alternatively, instead of inferring new distributions for conductance parameters, we can introduce scaling parameters which would more directly represent levels of channel block, and then infer distributions for these scaling parameters. From these distributions of scaling parameters, we can conclude whether the observed behaviour is consistent with ion channel screening. If not, it is possible that the drug may be blocking additional ion channels, or that the screening should be repeated. There can then be an iterative process where action potential predictions inform ion channel screening experiments, which then inform new action potential predictions.

7 Discussion and future work

7.1 Summary and conclusions

In this thesis, we have discussed aspects of mathematical and computational modelling that are being incorporated into drug safety assessment for cardiac proarrhythmic risk, most notably as part of the CiPA initiative. These techniques will eventually allow for safer and more cost-effective drug development by identifying potentially proarrhythmic drugs reliably before human trials. In particular, we have focused on ideas of uncertainty characterisation and propagation which provide more information for model predictions and allow for more-informed decision-making about a drug's proarrhythmic risk. We have adopted a Bayesian framework, in which we can treat model parameters as random variables and make use of credible intervals, which are more intuitive than classical confidence intervals.

We have considered uncertainty in model inputs (ion channel screening data) and in model parameters (conductance parameters in action potential models) by inferring probability distributions describing a drug's effect on various ion channels, and probability distributions describing the action potential model's conductance parameters. By sampling from these distributions and making predictions about action potential behaviour using these samples, we can construct distributions of this behaviour. These distributions of action potential behaviour capture possible

ranges of behaviour along with associated probabilities. Using only the most-likely sets of inputs and model parameters to generate a single prediction leads us to ignore other probable outcomes — even potentially dangerous outcomes when a point prediction is safe. Instead, these probabilities allow us to make better-informed predictions about how a drug will affect a cardiac cell's electrical behaviour.

After providing an introduction to cardiac electrophysiology in Chapter 2, we introduced the key concepts of uncertainty in computational modelling of biological processes. We then described the Bayesian framework we chose to adopt, and outlined the key algorithms used to infer probability distributions for model parameters throughout the thesis.

In Chapter 3, we applied the statistical methods discussed in Chapter 2 to ion channel screening data that was published by the FDA. We showed how inferring probability distributions for model parameters in a Bayesian framework can account for uncertainty in the form of inter-experiment variability, and provide predictions for what would happen in a future experiment. We considered how a hierarchical statistical model can 'separate' some of the uncertainty caused by inter-experiment variability and by observation error, and how this might help us find the 'true' behaviour underlying the experimental observations. We also performed model selection in a Bayesian framework by computing Bayes factors for the ion channel screening data using two different dose-response models.

In Chapter 4, we applied the statistical methods discussed in Chapter 2 to quantify parameter uncertainty in action potential models. We generated synthetic data so that we were using the 'correct' action potential model, *i.e.* it was the best-case scenario. This provided a first estimate for how much information was present in datasets obtained under different experimental protocols. This approach can aid experimental design by suggesting which experiments will provide data that contains information about particular parameters of interest. We also considered a hierarchical statistical model to describe how the different datasets were related, but we believe that this method did not provide enough additional information relative to the increased complexity of implementation.

In Chapters 5–6, we applied the single-level statistical methods to experimentally-obtained action potential recordings. We demonstrated that Markov chain Monte Carlo techniques can be applied to infer distributions for conductance parameters in action potential models. In particular, in Chapter 6, we modelled experimental data that I obtained while at Roche in Basel, Switzerland. The choice of conductance parameters that we modelled was informed by the results from Chapter 4 where we were able to successfully infer distributions for a subset of the conductance parameters.

Finally, in Chapter 6, we used the results from Chapter 3 to propagate uncertainty from ion channel screening data and from action potential model parameters into predictions of how different concentrations of Dofetilide will affect action potential behaviour. We obtained a range of different action potentials, which constructed a probability distribution over predicted action potential behaviour. Such distributions contain more information than single point estimates.

7.2 Future work

The work presented in this thesis is a first attempt at using a Bayesian statistical framework to characterise and propagate uncertainty in ion channel screening and action potential modelling for use in the cardiac drug safety process. There are areas that can be improved upon, but this work demonstrates the plausibility and usefulness of such an approach in this field.

The first extension of this work would be to perform the analysis from Chapter 6 on action potentials recorded from multiple different cells. Initially, we would consider beat-to-beat variability within each cell separately. This could be extended to consider action potentials from different cells simultaneously, so that the final predictions would represent how another cell from the same batch of cells might behave. Due to the unreliable nature of biological experiments, I was not able to record stimulated action potentials from enough cells to perform this analysis. Spending longer periods of time performing patch-clamp experiments would increase the probability of obtaining useful recordings, or gathering previously-recorded action potential data from different sources would help. The latter option is less

desirable, however, since we would like to control as much of the experimental setup as possible, and because we would like to model data from as few sources as possible to increase reliability of predictions for particular situations.

Initially, we would focus on obtaining more recordings from regular 1 Hz pacing in both a control and drug environment. However, another experimental extension is to perform different experiments on the cells, such as different pacing protocols and taking longer recordings. We expect some of the parameters that we were unable to identify in Chapter 4 would become more accessible under different or longer experimental protocols.

A further experimental extension of this work would be to perform *dynamic clamp* experiments (Dorval *et al.*, 2001). The dynamic clamp allows the experimenter to inject artificial currents across the cell membrane, simulating new ion currents. The simulated currents are formulated as a function of the membrane voltage (and possibly other quantities) and during the experiment these currents are computed and injected back across the cell membrane in real time. This would allow us to test the model formulations of particular ion currents. For example, we could record action potentials from a cell in a control environment. We then apply a hERG-blocking compound to eliminate I_{Kr}. The dynamic clamp could use the formulation for I_{Kr} from the Paci *et al.* (2013) model to re-add the current into the cell. If the newly-recorded action potentials match the control recordings, then we can be confident in the model formulation for I_{Kr}. Similar experiments can be performed for different ion currents. A more ambitious use of the dynamic clamp would be to inject simulated currents scaled with randomly-sampled conductance parameters, leading to MCMC being run and informed by the cell in real time. A more conservative approach would be to use the dynamic clamp along with an optimiser with Hessian information to find a best-fit set of conductance parameter values and the approximate spread of these parameters. This optimisation would require fewer iterations than running an MCMC algorithm, and so would not require a stable cell for as long.

In addition to experimental considerations, there are several avenues for investigation on the modelling side. When fitting action potential models to experimental data, the initial conditions of the model can have an important effect on the model output. Typically, the initial values of the state variables supplied with the models are steady-state values for regular 1 Hz pacing, or similar. However, these steady-state values also depend on model parameter values, such as the conductance parameters. As such, when solving an action potential model with different conductance parameter values, the initial state variables no longer correspond to steady-state variables. While performing MCMC, we are changing the conductance parameter values at each iteration, and therefore need to take the initial conditions into consideration. Further investigation into different methods of choosing the initial conditions at each iteration will be important for more accurately capturing the cell's behaviour during experiments. In particular, in Chapter 6 we treated 16 action potential recordings as independent datasets. In reality, however, these were subsequent paces from a single experiment. Therefore, the state variables from the end of one pace should be equal to the state variables at the start of the following pace. When fitting to one particular pace, we might consider clamping the membrane voltage to one or more of the preceding paces while solving the model, then using these state variables as the initial conditions for the pace of interest.

MCMC algorithms typically require lots of computation time; running the algorithm for more iterations will allow the chain to better approximate the target distribution. There are alternatives to the MCMC methods discussed in this thesis. For example, instead of solving the action potential model for a full membrane voltage trace at each iteration, we might use a Gaussian process emulator (or "surrogate model"). A Gaussian process estimates model outputs based on the model inputs, *e.g.* action potential duration will be estimated from conductance parameters by solving after training the emulator on a set of design data. Inputs and outputs are represented by Normal distributions, and so uncertainty is propagated quickly. This method was used by our collaborators in Sheffield in Johnstone *et al.* (2016b). An interesting approach would be to apply Gaussian process emulators

to the same situation addressed in Chapter 6 with propagating uncertainty from ion channel screening and action potential modelling into drug predictions. A comparison of the two methods would show how much time can be saved, or how much information is lost, by Gaussian process emulators. The faster inference might be necessary when using the data clamp where computations must occur in real time and fed back into the electrical circuit.

As shown in Figure 4.2 in Chapter 4, ODE solver tolerances can dramatically affect the computed membrane voltage. This affects the 'correctness' of the solution, but also creates a very uneven surface in parameter space for the Markov chain to explore. This can lead to the chain getting stuck in artefactual wells and so more probability mass is placed in these regions of parameter space where it should be more evenly distributed. To address this, we simply lowered the solver tolerances until we obtained a smooth target surface. However, in doing so we have not taken solver uncertainty into account. There have been efforts to address this source of uncertainty and propagate it into model predictions (Chkrebtii *et al.*, 2016). Applying similar methods to our approach would reduce reliance upon solver tolerances by incorporating uncertainty into model solutions, introducing additional variance to deal with the 'spikiness' observed in Figure 4.2. We would ultimately be able to perform faster ODE solves, at the cost of wider posteriors in model predictions. These posteriors may still be tight enough to prove useful in certain applications.

7.3 Concluding remarks

In this thesis, we have shown how mathematical, statistical and computational techniques for uncertainty quantification (UQ) can be adapted for use in cardiac electrophysiology modelling. These powerful techniques can be used to quantify uncertainty when fitting (conductance) parameters to data as well as the uncertainty present in model inputs (ion channel screening data), and then to examine the consequences of this uncertainty on model predictions. UQ allows us to place probabilities on model predictions, and hence is a first step in allowing us to say how much trust we should place in model predictions. A better understanding of

uncertainty in cardiac action potential models will yield new insights into the sources of variability, its mechanism(s) and consequences, in real cardiac cells and tissue. We believe that electrophysiology models should have uncertainty quantification performed for each use, to allow us to place confidence on model predictions, and allow us to quantify how the predictions depend on the assumptions in the models, and therefore how much we should rely on them.

It is valuable to compare different models when assessing different sources of uncertainty. For example, using a hierarchical model to describe ion channel screening data can help to unpick inter-experiment variability where there is intra-experiment correlation. Similarly, it is helpful to consider beat-to-beat variability as a source of uncertainty in conductance parameters to predict how a future beat may behave.

Formal UQ will improve our quantitative understanding of the sources and characteristics of natural variability in physiological systems, together with quantitative understanding of the consequences of such variability. Overall, UQ applied to cardiac action potential models will be tremendously important for developing trustworthy models and will provide a means for developing deeper physiological understanding of variability. Rigorous uncertainty assessment will be crucial for safety in clinical applications. It will provide more information about effects of a drug on cardiac electrophysiology, allowing for better-informed decision-making at the clinical level.

Appendices

A

Approximating Bayes factors

Here we introduce one method for approximating Bayes factors as used in Chapter 3. The likelihood of the data given model $\mathcal{M}_i$ is obtained by integrating over the parameter space of $\mathcal{M}_i$:

$$p(\boldsymbol{y}|\mathcal{M}_i) = \int_{\boldsymbol{\theta_i}} p(\boldsymbol{y}|\boldsymbol{\theta_i}, \mathcal{M}_i) p(\boldsymbol{\theta_i}|\mathcal{M}_i) d\boldsymbol{\theta_i} \tag{A.1}$$

$$= \mathbb{E}_{p(\boldsymbol{\theta_i}|\mathcal{M}_i)}[p(\boldsymbol{y}|\boldsymbol{\theta_i}, \mathcal{M}_i)] \tag{A.2}$$

where $p(\boldsymbol{\theta_i}|\mathcal{M}_i)$ is the prior density of the parameters. $\mathbb{E}_\pi(X)$ is the expectation, or mean, of X under the probability distribution π.

We wish to compute the marginal likelihood $p(\boldsymbol{y}|\mathcal{M}_i)$ as defined in Section 2.5.5, where we drop the $\mathcal{M}_i$-dependence for notational simplicity. In general, $p(\boldsymbol{y})$ is an integral over a high-dimensional space, and so is difficult to compute (van Haasteren, 2014). However, we are able to use *thermodynamic integration* (Gelman & Meng, 1998) to approximate $\log[p(\boldsymbol{y})]$, from which we can compute an approximation for $p(\boldsymbol{y})$. There are other methods to approximate $p(\boldsymbol{y})$, but the advantage of thermodynamic integration is that it makes use of our pre-existing MCMC algorithms, with minor alterations. We follow the method presented by Friel & Pettitt (2008).

To begin, we define the *power posterior* as

$$p_t(\boldsymbol{\theta}|\boldsymbol{y}) = \frac{p(\boldsymbol{y}|\boldsymbol{\theta})^t p(\boldsymbol{\theta})}{z(\boldsymbol{y}|t)}, \tag{A.3}$$

where

$$z(\boldsymbol{y}|t) = \int_{\boldsymbol{\theta}} p(\boldsymbol{y}|\boldsymbol{\theta})^t p(\boldsymbol{\theta}) \mathrm{d}\boldsymbol{\theta}. \tag{A.4}$$

Note that

$$\frac{z(\boldsymbol{y}|t=1)}{z(\boldsymbol{y}|t=0)} = \frac{\int_{\boldsymbol{\theta}} p(\boldsymbol{y}|\boldsymbol{\theta}) p(\boldsymbol{\theta}) \mathrm{d}\boldsymbol{\theta}}{\int_{\boldsymbol{\theta}} p(\boldsymbol{\theta}) \mathrm{d}\boldsymbol{\theta}} = \frac{p(\boldsymbol{y})}{1} = p(\boldsymbol{y}). \tag{A.5}$$

We derive an expression for $\log[p(\boldsymbol{y})]$:

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t} \log[z(\boldsymbol{y}|t)] &= \frac{1}{z(\boldsymbol{y}|t)} \frac{\mathrm{d}}{\mathrm{d}t} z(\boldsymbol{y}|t) && \text{(A.6)} \\
&= \frac{1}{z(\boldsymbol{y}|t)} \frac{\mathrm{d}}{\mathrm{d}t} \int_{\boldsymbol{\theta}} p(\boldsymbol{y}|\boldsymbol{\theta})^t p(\boldsymbol{\theta}) \mathrm{d}\boldsymbol{\theta} && \text{(A.7)} \\
&= \frac{1}{z(\boldsymbol{y}|t)} \int_{\boldsymbol{\theta}} \log[p(\boldsymbol{y}|\boldsymbol{\theta})] p(\boldsymbol{y}|\boldsymbol{\theta})^t p(\boldsymbol{\theta}) \mathrm{d}\boldsymbol{\theta} && \text{(A.8)} \\
&= \int_{\boldsymbol{\theta}} \log[p(\boldsymbol{y}|\boldsymbol{\theta})] \left(\frac{p(\boldsymbol{y}|\boldsymbol{\theta})^t p(\boldsymbol{\theta})}{z(\boldsymbol{y}|t)} \right) \mathrm{d}\boldsymbol{\theta} && \text{(A.9)} \\
&= \mathbb{E}_{\boldsymbol{\theta}|\boldsymbol{y},t}[\log[p(\boldsymbol{y}|\boldsymbol{\theta})]], && \text{(A.10)}
\end{aligned}$$

which is the expectation of $\log[p(\boldsymbol{y}|\boldsymbol{\theta})]$ with respect to the probability distribution $p_t(\boldsymbol{\theta}|y)$ (Equation (A.3)). Integrate over t from 0 to 1:

$$\log[z(\boldsymbol{y}|t=1)] - \log[z(\boldsymbol{y}|t=0)] = \int_{t=0}^{1} \mathbb{E}_{\boldsymbol{\theta}|\boldsymbol{y},t}[\log[p(\boldsymbol{y}|\boldsymbol{\theta})]] \mathrm{d}t, \tag{A.11}$$

and take the log of Equation (A.5) to obtain

$$\log[p(\boldsymbol{y})] = \int_{t=0}^{1} \mathbb{E}_{\boldsymbol{\theta}|\boldsymbol{y},t}[\log[p(\boldsymbol{y}|\boldsymbol{\theta})]] \mathrm{d}t. \tag{A.12}$$

We now discretise the t-interval [0,1] and run our MCMC algorithm for each discrete value of t to approximate this integral. Suppose we choose $\{t_0 = 0 < t_1 < t_2 < ... < t_n = 1\}$ to be our discretisation. We then approximate Equation (A.12) by

$$\log[p(\boldsymbol{y})] \approx \sum_{i=0}^{n-1} \frac{1}{2} (t_{i+1} - t_i) (\mathbb{E}_{\boldsymbol{\theta}|\boldsymbol{y},t_{i+1}}[\log[p(\boldsymbol{y}|\boldsymbol{\theta})]] + \mathbb{E}_{\boldsymbol{\theta}|\boldsymbol{y},t_i}[\log[p(\boldsymbol{y}|\boldsymbol{\theta})]]), \tag{A.13}$$

using the trapezium rule. Each individual expectation in Equation (A.13) is approximated by running an MCMC algorithm and computing

$$\mathbb{E}_{\boldsymbol{\theta}|\boldsymbol{y},t_i}[\log[p(\boldsymbol{y}|\boldsymbol{\theta})]] \approx \frac{1}{T-B} \sum_{j=B+1}^{T} \log[p(\boldsymbol{y}|\boldsymbol{\theta}_i^{(j)})], \tag{A.14}$$

where $\boldsymbol{\theta}_i^{(j)}$ is the j^{th} iteration from the MCMC-approximated power posterior distribution $p_{t_i}(\boldsymbol{\theta}|y)$, where we have run our MCMC algorithm for T iterations and are discarding the first B as a burn-in.

When discretising the integral in Equation (A.12), there is generally more variation in $\mathbb{E}_{\boldsymbol{\theta}|\boldsymbol{y},t}[\log[p(\boldsymbol{y}|\boldsymbol{\theta})]]$ closer to $t = 0$, and so it is helpful to take more values of t closer to 0 than to 1. Friel & Pettitt suggest taking temperatures of the form

$$t_i = \left(\frac{i}{n}\right)^c, \tag{A.15}$$

for $i = 0, \ldots, n$, and $c > 1$.

After running our MCMC algorithm to compute the expectations in Equation (A.14), we can them compute $\log[p(\boldsymbol{y})]$ as in Equation (A.13). We then exponentiate this to obtain $p(\boldsymbol{y})$ as defined in Equation (A.2).

B

Singularities in action potential models

B.1 Removable singularities in GHK equations

Across many different action potential models, a common formulation for several different components as a function of membrane voltage, V, is the GHK Equation (2.14), introduced in Section 2.2.3 of Chapter 2:

$$c(V) := \frac{f(V)(V - v_0)}{e^{B(V-v_0)} - 1}, \tag{B.1}$$

where B and v_0 are constants, and $f(V)$ is some function of V that may be a constant. For example, the background calcium current component $I_{Ca,b}$ in the Davies *et al.* (2011) model of a canine cardiomyocyte. To simplify notation, we define $f(V) := A(V) \times B$, and we partially nondimensionalise Equation (B.1) using the substitution $U = B(V - v_0)$. Then

$$c = A(V) \times \frac{U}{e^U - 1}. \tag{B.2}$$

We define

$$g(U) := \frac{U}{e^U - 1}. \tag{B.3}$$

The numerator and denominator of $g(U)$ are both equal to zero at $U = 0$, and so we must evaluate the limit as $U \to 0$ to check if this is a removable

singularity, or if $g(U)$ is not well-behaved. Using L'Hôpital's rule, we can evaluate the limit of $g(U)$ as $U \to 0$:

$$\lim_{U\to 0} g(U) = \lim_{U\to 0} \frac{1}{e^U} = 1. \tag{B.4}$$

Therefore $g(U)$ is well-defined for all U, and the singularity at $U = 0$ is removable. This is not a problem when we implement the function on a computer and take large enough steps along the U-axis to never get very close to $U = 0$, as shown in Figure B.1.

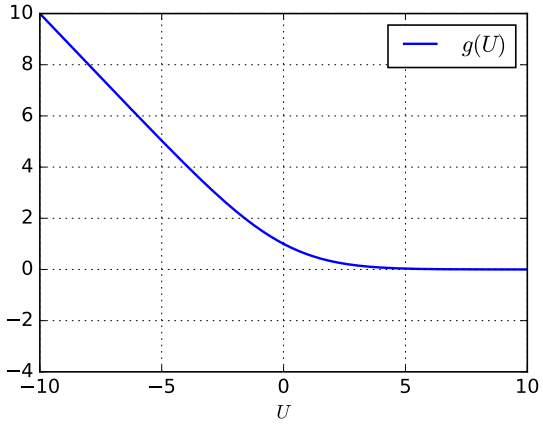

Figure B.1: Plotting $g(U)$ equally spaced from $U = -10$ to $U = 10$ with a step size of approximately 4×10^{-2}.

However, when we use much smaller steps close to $U = 0$, we cannot guarantee that the computer will be able to evaluate the limit to the correct value. This becomes a problem when using adaptive time-stepping methods (such as CVODE) that refine steps when steep gradients are encountered, as such methods can 'home in' on singularities and then fail. An example of such singular behaviour is given in Figure B.2.

To avoid these potential numerical problems, we can linearise $g(U)$ close to $U = 0$ and include a condition in the action potential model to use the linear approximation when U is sufficiently close to 0, and use the full expression when U is far away from 0.

We have that $g(0) = 1$ and compute that $\frac{\mathrm{d}g}{\mathrm{d}U}(0) = -1/2$. We therefore have

$$g(U) \approx 1 - \frac{1}{2}U =: h(U), \tag{B.5}$$

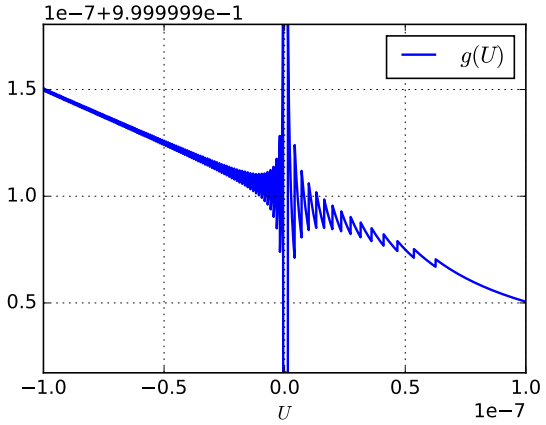

Figure B.2: Plotting $g(U)$ equally spaced from $U = -10^{-7}$ to $U = 10^{-7}$ with a step size of approximately 4×10^{-10}.

when U is sufficiently close to 0. This approximation is plotted in Figure B.3, on the same scale as in Figure B.2.

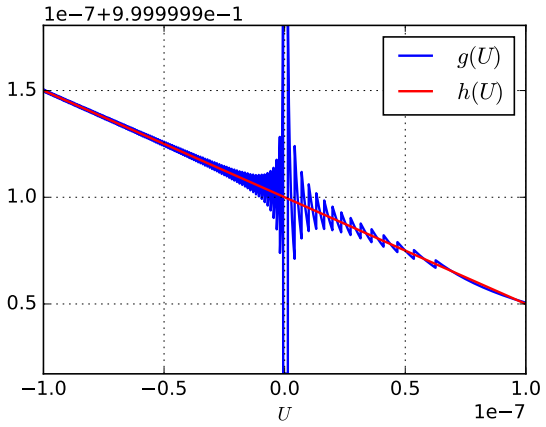

Figure B.3: Plotting $g(U)$ and its linear approximation, $h(U)$, equally spaced from $U = -10^{-7}$ to $U = 10^{-7}$ with a step size of approximately 4×10^{-11}.

However, we have to be careful not to use the linear approximation when far away from $U = 0$. Both $g(U)$ and $h(U)$ are plotted in Figure B.4, on the same scale as in Figure B.1.

B.2 CellML implementation

To implement this approximation in the action potential model, we introduce new variables A, B, v_0, and U as described above. When introducing new variables in

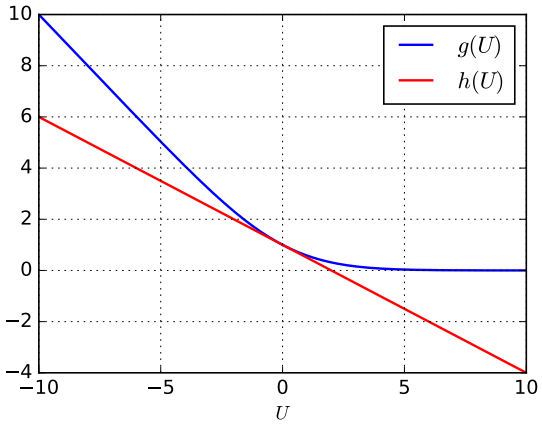

Figure B.4: Plotting $g(U)$ and its linear approximation, $h(U)$, equally spaced from $U = -10$ to $U = 10$ with a step size of approximately 4×10^{-2}.

CellML, their units must be specified. Clearly v_0 has the same units as V, which will be millivolts or volts. All exponents must be dimensionless, therefore U must be dimensionless. This means that B must have the same units as $1/V$, so will have units of $(\text{mV})^{-1}$ (or V^{-1}). From Equation (B.2), we can see that A must have the same units as the component c, *i.e.* $[A] = [c]$.

Once we have defined the new variables' units, we assign to them based on the component's format as in Equation (B.1). In particular, since we are given $f(V)$, B, and v_0 directly by $c(V)$, we assign $A = f(V)/B$ and $U = B(V - v_0)$.

After defining these new variables, we can redefine $c(V)$, so that it is given by

$$c = \begin{cases} A\left(1 - \frac{1}{2}U\right) & \text{if } -10^{-k} \leq U \leq 10^{-k}, \\ \dfrac{AU}{e^U - 1} & \text{otherwise,} \end{cases} \tag{B.6}$$

where k is chosen to provide sufficient accuracy. This nondimensionalisation means we can use the same value of k for all components without worrying about how other scaling factors affect the size of $c(V)$ as $V \to v_0$. We choose to use $k = 7$.

B.3 CellML action potential model example

We consider an example from the Davies *et al.* (2011) model. The original formulation of the component $I_{Ca,b}$ is commented out by "//" at the bottom.

```
def comp ICab as
    var V: mV {pub: in};
    var F: C_per_mole {pub: in};
    var FonRT: per_mV {pub: in};
    var{membrane_background_calcium_current_conductance
       } GCab: L_per_F_ms {init: 0.0000007980336};
    var Ca_i: mM {pub: in};
    var Ca_o: mM {pub: in};
    var{membrane_background_calcium_current} ICab:
       uA_per_uF {pub: out};
    var A: uA_per_uF;
    var B: per_mV;
    var v0: mV;
    var U: dimensionless;

    B = 2{dimensionless}*FonRT;
    v0 = 0{mV};
    A = GCab*F*FonRT*(Ca_i*exp(2{dimensionless}*V*FonRT
       )-0.341{dimensionless}*Ca_o)/B;
    U = B*(V-v0);

    ICab = sel
        case (-1e-7{dimensionless} <= U) and (U <= 1e
           -7{dimensionless}):
            A*(1{dimensionless}-0.5{dimensionless}*U);
        otherwise:
            A*U/(exp(U)-1{dimensionless});
    endsel;

    //ICab = GCab*V*F*FonRT*(Ca_i*exp(2{dimensionless}*
       V*FonRT)-0.341{dimensionless}*Ca_o)/(exp(2{
       dimensionless}*V*FonRT)-1{dimensionless});
enddef;
```

Bibliography

Ari J Alexandrou, Rona S Duncan, Anneli Sullivan, Jules C Hancox, Derek J Leishman, Harry J Witchel and Joanne L Leaney. April 2006. Mechanism of hERG K+ channel blockade by the fluoroquinolone antibiotic moxifloxacin. *British journal of pharmacology*, **147**, 905–16. ISSN 0007-1188. (doi:10.1038/sj.bjp.0706678)

Jacob Beal. 2017. Biochemical complexity drives log-normal variation in genetic expression. *Engineering Biology*, **1**, 55–60.

Kylie A Beattie, Chris Luscombe, Geoff Williams, Jordi Munoz-Muriedas, David J Gavaghan, Yi Cui and Gary R Mirams. 2013. Evaluation of an in silico cardiac safety assay: using ion channel screening data to predict qt interval changes in the rabbit ventricular wedge. *Journal of pharmacological and toxicological methods*, **68**, 88–96.

G W Beeler and H Reuter. 1977. Reconstruction of the action potential of ventricular myocardial fibres. *J. Physiol.*, **268**, 177.

Luiz Belardinelli, Charles Antzelevitch and Marc A Vos. 2003. Assessing predictors of drug-induced torsade de pointes. *Trends in pharmacological sciences*, **24**, 619–625.

Saïd Bendahhou, Theodore R Cummins, Rabi Tawil, Stephen G Waxman and Louis J Ptácek. 1999. Activation and inactivation of the voltage-gated sodium channel: role of segment s5 revealed by a novel hyperkalaemic periodic paralysis mutation. *The Journal of neuroscience*, **19**, 4762–4771.

T Brennan, M Fink and B Rodriguez. 2009. Multiscale modelling of drug-induced effects on cardiac electrophysiological activity. *European Journal of Pharmaceutical Sciences*, **36**, 62–77.

Ben Calderhead and Mark Girolami. 2009. Estimating bayes factors via thermodynamic integration and population mcmc. *Computational Statistics & Data Analysis*, **53**, 4028–4045.

Icilio Cavero and Henry Holzgrefe. 2014. Comprehensive in vitro proarrhythmia assay, a novel in vitro/in silico paradigm to detect ventricular proarrhythmic liability: a visionary 21st century initiative. *Expert opinion on drug safety*, **13**, 745–758.

Oksana A Chkrebtii, David A Campbell, Ben Calderhead, Mark A Girolami *et al.* 2016. Bayesian solution uncertainty quantification for differential equations. *Bayesian Analysis*, **11**, 1239–1267.

Colleen E Clancy and Yoram Rudy. 2001. Cellular consequences of herg mutations in the long qt syndrome: precursors to sudden cardiac death. *Cardiovascular research*, **50**, 301–313.

Jonathan Cooper, Raymond J Spiteri and Gary R Mirams. January 2014. Cellular cardiac electrophysiology modeling with Chaste and CellML. *Frontiers in Physiology*, **5**, 511. ISSN 1664-042X. (doi:10.3389/fphys.2014.00511)

William J Crumb, Jose Vicente, Lars Johannesen and David G. Strauss. sep 2016. An evaluation of 30 clinical drugs against the comprehensive in vitro proarrhythmia assay (CiPA) proposed ion channel panel. *Journal of Pharmacological and Toxicological Methods*, **81**, 251–262. ISSN 10568719. (doi:10.1016/j.vascn.2016.03.009)

Autumn A Cuellar, Catherine M Lloyd, Poul F Nielsen, David P Bullivant, David P Nickerson and Peter J Hunter. 2003. An overview of cellml 1.1, a biological model description language. *Simulation*, **79**, 740–747.

Aidan C Daly, David J Gavaghan, Chris Holmes and Jonathan Cooper. 2015. Hodgkin–huxley revisited: reparametrization and identifiability analysis of the classic action potential model with approximate bayesian methods. *Royal Society open science*, **2**, 150499.

John Davey. 2004. G-protein-coupled receptors: new approaches to maximise the impact of gpcrs in drug discovery. *Expert opinion on therapeutic targets*, **8**, 165–170.

Mark Richard Davies, Hitesh Balvanta Mistry, Leyla Hussein, Chris E Pollard, J-P Valentin, Jonathan Swinton and Najah Abi-Gerges. 2011. An in silico canine cardiac midmyocardial action potential duration model as a tool for early drug safety assessment. *American Journal of Physiology-Heart and Circulatory Physiology*, **302**, H1466–H1480.

Keith F Decker, Jordi Heijman, Jonathan R Silva, Thomas J Hund and Yoram Rudy. 2009. Properties and ionic mechanisms of action potential adaptation, restitution, and accommodation in canine epicardium. *American Journal of Physiology-Heart and Circulatory Physiology*, **296**, H1017–H1026.

Socrates Dokos and Nigel H Lovell. 2004. Parameter estimation in cardiac ionic models. *Progress in biophysics and molecular biology*, **85**, 407–31. ISSN 0079-6107. (doi:10.1016/j.pbiomolbio.2004.02.002)

Alan D Dorval, David J Christini and John A White. 2001. Real-time linux dynamic clamp: a fast and flexible way to construct virtual ion channels in living cells. *Annals of biomedical engineering*, **29**, 897–907.

R. C. Elkins, M. R. Davies, S. J. Brough, D. J. Gavaghan, Y. Cui, N. Abi-Gerges and G. R. Mirams. 2013a. Variability in high-throughput ion-channel screening data and consequences for cardiac safety assessment. *Journal of Pharmacological and Toxicological Methods*, **68**, 112–122. (doi:10.1016/j.vascn.2013.04.007)

Ryan C Elkins, Mark R Davies, Stephen J Brough, David J Gavaghan, Yi Cui, Najah Abi-Gerges and Gary R Mirams. 2013b. Variability in high-throughput ion-channel screening data and consequences for cardiac safety assessment. *Journal of Pharmacological and Toxicological Methods*, **68**, 112–122.

M Fink, S A Niederer, E M Cherry, F H Fenton, J T Koivumaki, G Seemann, R Thul, H Zhang, F B Sachse, E J Crampin and N P Smith. 2011. Cardiac cell modelling: Observations from the heart of the Cardiac Physiome project. *Progress in Biophysics and Molecular Biology*, **104**, 2–21.

R F Fox. 1997. Stochastic versions of the Hodgkin-Huxley equations. *Biophysical journal*, **72**, 2068–74. ISSN 0006-3495. (doi:10.1016/S0006-3495(97)78850-7)

Nial Friel and Anthony N Pettitt. 2008. Marginal likelihood estimation via power posteriors. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **70**, 589–607.

Andrew Gelman and Xiao-Li Meng. 1998. Simulating normalizing constants: From importance sampling to bridge sampling to path sampling. *Statistical science*, pages 163–185.

Stuart Geman and Donald Geman. 1984. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on pattern analysis and machine intelligence*, pages 721–741.

S Geneser, R Kirby, D Xiu and Sachse F. 2007. Stochastic Markovian modeling of electrophysiology of ion channels: Reconstruction of standard deviations in macroscopic currents. *Journal of theoretical biology*, pages 627–637.

Wally R Gilks, NG Best and KKC Tan. 1995. Adaptive rejection metropolis sampling within gibbs sampling. *Applied Statistics*, pages 455–472.

Leon Glass. June 2009. Introduction to controversial topics in nonlinear science: is the normal heart rate chaotic? *Chaos (Woodbury, N.Y.)*, **19**, 028501. ISSN 1089-7682. (doi:10.1063/1.3156832)

Heikki Haario, Eero Saksman and Johanna Tamminen. 2001. An adaptive metropolis algorithm. *Bernoulli*, pages 223–242.

Jules C Hancox, Mark J McPate, Aziza El Harchi and Yi hong Zhang. 2008. The herg potassium channel and herg screening for drug-induced torsades de pointes. *Pharmacology & therapeutics*, **119**, 118–132.

Nikolaus Hansen. 2006. The cma evolution strategy: a comparing review. *Towards a new evolutionary computation*, pages 75–102.

Nikolaus Hansen. 2016. The cma evolution strategy: A tutorial. *arXiv preprint arXiv:1604.00772*.

A R Harmer, J-P Valentin and C E Pollard. September 2011. On the relationship between block of the cardiac Na^+ channel and drug-induced prolongation of the QRS complex. *British journal of pharmacology*, **164**, 260–73. ISSN 1476-5381. (doi:10.1111/j.1476-5381.2011.01415.x)

W Keith Hastings. 1970. Monte carlo sampling methods using markov chains and their applications. *Biometrika*, **57**, 97–109.

Jordi Heijman, Antonio Zaza, Daniel M Johnson, Yoram Rudy, Ralf LM Peeters, Paul GA Volders and Ronald L Westra. 2013. Determinants of beat-to-beat variability of repolarization duration in the canine ventricular myocyte: a computational analysis. *PLoS computational biology*, **9**, e1003202.

A V Hill. 1910. Proceedings supplement: The possible effects of the aggregation of the molecules of haemoglobin on its dissociation curves. *The Journal of Physiology*, **40**, iv–vii.

Bertil Hille *et al. Ion channels of excitable membranes*, volume 507. Sinauer Sunderland, MA, 2001.

A L Hodgkin and A F Huxley. 1952a. A quantitative description of membrane current and its application to conduction and excitation in nerve. *Journal of Physiology*, **117**, 500–544.

Allan L Hodgkin and Andrew F Huxley. 1952b. Currents carried by sodium and potassium ions through the membrane of the giant axon of loligo. *The Journal of physiology*, **116**, 449–472.

Thomas J Hund and Yoram Rudy. 2004. Rate dependence and regulation of action potential and calcium transient in a canine cardiac ventricular cell model. *Circulation*, **110**, 3168–74. ISSN 1524-4539. (doi:10.1161/01.CIR.0000147231.69595.D3)

International Committee on Harmonization. 2005. Guidance for industry: E14 clinical evaluation of qt/qtc interval prolongation and proarrhythmic potential for non-antiarrhythmic drugs. *Rockville, MD: Center for drug evaluation and research*.

Vivek Iyer, Reza Mazhari and Raimond L Winslow. 2004. A computational model of the human left-ventricular epicardial myocyte. *Biophysical journal*, **87**, 1507–1525.

M Saleet Jafri, J Jeremy Rice and Raimond L Winslow. 1998. Cardiac ca 2+ dynamics: the roles of ryanodine receptor adaptation and sarcoplasmic reticulum load. *Biophysical journal*, **74**, 1149–1168.

Ross H Johnstone, Rémi Bardenet, David J Gavaghan and Gary R Mirams. 2016a. Hierarchical bayesian inference for ion channel screening dose-response data. *Wellcome open research*, **1**.

Ross H Johnstone, Eugene TY Chang, Rémi Bardenet, Teun P De Boer, David J Gavaghan, Pras Pathmanathan, Richard H Clayton and Gary R Mirams. 2016b. Uncertainty and variability in models of the cardiac action potential: Can we build trustworthy models? *Journal of molecular and cellular cardiology*, **96**, 49–62.

Robert E Kass and Adrian E Raftery. 1995. Bayes factors. *Journal of the american statistical association*, **90**, 773–795.

Arnold M Katz. *Physiology of the Heart*. Lippincott Williams & Wilkins, 2010.

L. A. Lee, K. J. Pringle, C. L. Reddington, G. W. Mann, P. Stier, D. V. Spracklen, J. R. Pierce and K. S. Carslaw. September 2013. The magnitude and causes of uncertainty in global model simulations of cloud condensation nuclei. *Atmospheric Chemistry and Physics*, **13**, 8879–8914. ISSN 1680-7324. (doi:10.5194/acp-13-8879-2013)

Chon Lok Lei, Ken Wang, Michael Clerx, Ross H Johnstone, Maria P Hortigon-Vinagre, Victor Zamora, Andrew Allan, Godfrey Smith, David J Gavaghan, Gary R Mirams *et al.* 2017. Tailoring mathematical models to stem-cell derived cardiomyocyte lines can improve predictions of drug-induced changes to their electrophysiology. *Frontiers in Physiology*, **8**, 986.

Mathieu Lemay, Enno de Lange and Jan P Kucera. 2011. Effects of stochastic channel gating and distribution on the cardiac action potential. *Journal of theoretical biology*, **281**, 84–96.

C.M. Lloyd, J.R. Lawson, P.J. Hunter and P.F. Nielsen. 2008. The CellML model repository. *Bioinformatics*, **24**, 2122–2123. (doi:10.1093/bioinformatics/btn390)

C. H. Luo and Y. Rudy. 1991. A model of the ventricular cardiac action potential. Depolarization, repolarization, and their interaction. *Circulation Research*, **68**, 1501. ISSN 0009-7330. (doi:10.1161/01.RES.68.6.1501)

Ching-hsing Luo and Yoram Rudy. 1994. A dynamic model of the cardiac ventricular action potential. i. simulations of ionic currents and concentration changes. *Circulation research*, **74**, 1071–1096.

Junyi Ma, Liang Guo, Steve J Fiene, Blake D Anson, James A Thomson, Timothy J Kamp, Kyle L Kolaja, Bradley J Swanson and Craig T January. 2011. High purity human-induced pluripotent stem cell-derived cardiomyocytes: electrophysiological properties of action potentials and ionic currents. *American Journal of Physiology-Heart and Circulatory Physiology*, **301**, H2006–H2017.

Nicholas Metropolis, Arianna W Rosenbluth, Marshall N Rosenbluth, Augusta H Teller and Edward Teller. 1953. Equation of state calculations by fast computing machines. *The journal of chemical physics*, **21**, 1087–1092.

Gary R Mirams, Christopher J Arthurs, Miguel O Bernabeu, Rafel Bordas, Jonathan Cooper, Alberto Corrias, Yohan Davit, Sara-Jane Dunn, Alexander G Fletcher, Daniel G Harvey *et al.* 2013. Chaste: an open source c++ library for computational physiology and biology. *PLoS Comput Biol*, **9**, e1002970.

Gary R Mirams, Yi Cui, Anna Sher, Martin Fink, Jonathan Cooper, Bronagh M Heath, Nick C McMahon, David J Gavaghan and Denis Noble. 2011. Simulation of multiple ion channel block provides improved early prediction of compounds' clinical torsadogenic risk. *Cardiovascular research*, **91**, 53–61.

Gary R Mirams, Mark R Davies, Stephen J Brough, Matthew H Bridgland-Taylor, Yi Cui, David J Gavaghan and Najah Abi-Gerges. 2014. Prediction of Thorough QT study results using action potential simulations based on ion channel screens. *Journal of pharmacological and toxicological methods*, **70**, 246–254.

Gary R Mirams, Mark R Davies, Yi Cui, Peter Kohl and Denis Noble. 2012. Application of cardiac electrophysiology simulations to pro-arrhythmic safety testing. *British Journal of Pharmacology*, **167**, 932–945. ISSN 1476-5381. (doi:10.1111/j.1476-5381.2012.02020.x)

Kevin P Murphy. 2007. Conjugate bayesian analysis of the gaussian distribution. *def*, **1**, 16.

L Nalos, R Varkevisser, M K B Jonsson, M J C Houtman, J D Beekman, R van der Nagel, M B Thomsen, G Duker, P Sartipy, T P de Boer, M Peschar, M B Rook, T A B van Veen, M A G van der Heyden and M A Vos. January 2012. Comparison of the IKr blockers moxifloxacin, dofetilide and E-4031 in five screening models of pro-arrhythmia reveals lack of specificity of isolated cardiomyocytes. *British journal of pharmacology*, **165**, 467–78. ISSN 1476-5381. (doi:10.1111/j.1476-5381.2011.01558.x)

Carlo Napolitano, Silvia G Priori and Peter J Schwartz. 1994. Torsade de pointes. *Drugs*, **47**, 51–65.

National Research Council. *Assessing the Reliability of Complex Models: Mathematical and Statistical Foundations of Verification, Validation, and Uncertainty Quantification*. The National Academies Press, Washington, DC, 2012. ISBN 978-0-309-25634-6.

Andrew A Neath and Joseph E Cavanaugh. 2012. The bayesian information criterion: background, derivation, and applications. *Wiley Interdisciplinary Reviews: Computational Statistics*, **4**, 199–203.

S A Niederer, M Fink, D Noble and N P Smith. 2009. A meta-analysis of cardiac electrophysiology computational models. *Experimental Physiology*, **94**, 486–95. ISSN 1469-445X. (doi:10.1113/expphysiol.2008.044610)

D. Noble. 1962. A modification of the Hodgkin—Huxley equations applicable to Purkinje fibre action and pacemaker potentials. *Journal of Physiology (London)*, **160**, 317–352.

Denis Noble. *The initiation of the heartbeat*. Oxford University Press, USA, 1979.

W. Oberkampf and C. Roy. *Verification and validation in scientific computing*. Cambridge University Press, 2010.

William L Oberkampf, Timothy G Trucano and Charles Hirsch. 2004. Verification, validation, and predictive capability in computational engineering and physics. *Applied Mechanics Reviews*, **57**, 345–384.

Thomas O'Hara, László Virág, András Varró and Yoram Rudy. 2011. Simulation of the undiseased human cardiac ventricular action potential: model formulation and experimental validation. *PLoS computational biology*, **7**, e1002061.

Michelangelo Paci, Jari Hyttinen, Katriina Aalto-Setälä and Stefano Severi. 2013. Computational models of ventricular-and atrial-like human induced pluripotent stem cell derived cardiomyocytes. *Annals of biomedical engineering*, **41**, 2334–2348.

Pras Pathmanathan and Richard A. Gray. 2013. Ensuring reliability of safety-critical clinical applications of computational cardiac models. *Frontiers in Physiology*, **4**, 1–9. (doi:10.3389/fphys.2013.00358)

Pras Pathmanathan and Richard A Gray. 2014. Verification of computational models of cardiac electro-physiology. *International Journal for Numerical Methods in Biomedical Engineering*, **30**, 525–544. (doi:10.1002/cnm.2615)

Pras Pathmanathan, Matthew S Shotwell, David J Gavaghan, Jonathan M Cordeiro and Richard A Gray. 2015a. Uncertainty quantification of fast sodium current steady-state inactivation for multi-scale models of cardiac electrophysiology. *Progress in Biophysics and Molecular Biology*, **117**, 4–18.

Pras Pathmanathan, Matthew S Shotwell, David J Gavaghan, Jonathan M Cordeiro and Richard A Gray. 2015b. Uncertainty quantification of fast sodium current steady-state inactivation for multi-scale models of cardiac electrophysiology. *Progress in Biophysics and Molecular Biology*, **117**, 4–18.

Andreas Raue, Clemens Kreutz, Thomas Maiwald, Julie Bachmann, Marcel Schilling, Ursula Klingmüller and Jens Timmer. 2009. Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics*, **25**, 1923–1929.

Christian Robert. *The Bayesian choice: from decision-theoretic foundations to computational implementation*. Springer, 2007. ISBN 978-0-387-71598-8.

Christian Robert and George Casella. *Monte Carlo statistical methods*. Springer, 2013. ISBN 978-1-4419-1939-7.

Byron N Roberts, Pei-Chi Yang, Steven B Behrens, Jonathan D Moreno and Colleen E Clancy. October 2012. Computational approaches to understand cardiac electrophysiology and arrhythmias. *American Journal of Physiology (Heart and Circulatory Physiology)*, **303**, H766–83. ISSN 1522-1539. (doi:10.1152/ajpheart.01081.2011)

Gareth O Roberts, Andrew Gelman, Walter R Gilks *et al.* 1997. Weak convergence and optimal scaling of random walk metropolis algorithms. *The annals of applied probability*, **7**, 110–120.

Philip T Sager, Gary Gintant, J Rick Turner, Syril Pettit and Norman Stockbridge. 2014. Rechanneling the cardiac proarrhythmia safety paradigm: a meeting report from the cardiac safety research consortium. *American Heart Journal*, **167**, 292–300.

Michael C Sanguinetti and Martin Tristani-Firouzi. 2006. herg potassium channels and cardiac arrhythmia. *Nature*, **440**, 463–469.

Robert Schlaifer and Howard Raiffa. *Applied statistical decision theory.* 1961.

Gideon Schwarz *et al.* 1978. Estimating the dimension of a model. *The annals of statistics*, **6**, 461–464.

Rashmi R Shah. 2005. Drugs, qt interval prolongation and ich e14. *Drug Safety*, **28**, 115–125.

Ivo Siekmann, James Sneyd and Edmund J Crampin. 2012. Mcmc can detect nonidentifiable models. *Biophysical journal*, **103**, 2275–2286.

Ivo Siekmann, Larry E Wagner, David Yule, Colin Fox, David Bryant, Edmund J Crampin and James Sneyd. 2011. MCMC estimation of Markov models for ion channels. *Biophysical journal*, **100**, 1919–29. ISSN 1542-0086. (doi:10.1016/j.bpj.2011.02.059)

K. ten Tusscher, D. Noble, P. J. Noble and A. V. Panfilov. 2004. A model for human ventricular tissue. *Am. J. Physiol. Heart Circ. Physiol.*, **286**, 1573–1589.

KHWJ Ten Tusscher, D Noble, PJ Noble and Alexander V Panfilov. 2004. A model for human ventricular tissue. *American Journal of Physiology-Heart and Circulatory Physiology*, **286**, H1573–H1589.

N A Trayanova, Thomas O'Hara, Jason D Bayer, Patrick M Boyle, Kathleen S McDowell, Jason Constantino, Hermenegild J Arevalo, Yuxuan Hu and Fijoy Vadakkumpadan. November 2012. Computational cardiology: how computer simulations could be used to develop new therapies and advance existing ones. *Europace*, **14 Suppl 5**, v82–v89. ISSN 1532-2092. URL `http://www.ncbi.nlm.nih.gov/pubmed/23104919`. (doi:10.1093/europace/eus277)

MV Tyrode. 1910. The mode of action of some purgative salts.

Rutger van Haasteren. Marginal likelihood calculation with mcmc methods. In *Gravitational Wave Detection and Data Analysis for Pulsar Timing Arrays*, pages 99–120. Springer, 2014.

Ian Vernon, Michael Goldstein and Richard G. Bower. 2010. Galaxy formation: a Bayesian uncertainty analysis. *Bayesian Analysis*, **5**, 619–669. ISSN 1936-0975. (doi:10.1214/10-BA524)

Olivia J Walch and Marisa C Eisenberg. 2016. Parameter identifiability and identifiable combinations in generalized hodgkin–huxley models. *Neurocomputing*, **199**, 137–143.

James N Weiss. 1997. The hill equation revisited: uses and misuses. *The FASEB Journal*, **11**, 835–841.

Geoff Williams and Gary R Mirams. 2015. A web portal for in-silico action potential predictions. *Journal of Pharmacological and Toxicological Methods*, **75**, 10–16.

Raimond L Winslow, Jeremy Rice, Saleet Jafri, Eduardo Marban and Brian O'Rourke. 1999. Mechanisms of altered excitation-contraction coupling in canine tachycardia-induced heart failure, ii. *Circulation research*, **84**, 571–586.

Junying Yu, Maxim A Vodyanik, Kim Smuga-Otto, Jessica Antosiewicz-Bourget, Jennifer L Frane, Shulan Tian, Jeff Nie, Gudrun A Jonsdottir, Victor Ruotti, Ron Stewart *et al.* 2007. Induced pluripotent stem cell lines derived from human somatic cells. *Science*, **318**, 1917–1920.