8-17-2022

Bayesian inference and prediction in cardiac electrophysiology models with an application to representing variability

Alejandro Nieto Ramos
axn2780@rit.edu

SCHOOL OF MATHEMATICAL SCIENCES
COLLEGE OF SCIENCE
ROCHESTER INSTITUTE OF TECHNOLOGY
ROCHESTER, NEW YORK

CERTIFICATE OF APPROVAL

Ph.D. DEGREE DISSERTATION

The Ph.D. Degree Dissertation of *Alejandro Nieto Ramos* has been examined and approved by the dissertation committee as satisfactory for the dissertation requirement for the Ph.D. degree in Mathematical Modeling.

Dr. Ben Zwickl, Committee Chair

Dr. Flavio Fenton , Committee Member

Dr. Ernest Fokoué, Committee Member

Dr. Laura Muñoz, Committee Member

Dr. Elizabeth Cherry, Thesis Advisor

Date

Rochester Institute of Technology

Dissertation

Bayesian inference and prediction in cardiac electrophysiology models with an application to representing variability

Author:
Alejandro Nieto Ramos

Supervisor:
Elizabeth Cherry

A dissertation submitted in fulfillment of the requirements for the degree of Doctor of Philosophy

in the

School of Mathematical Sciences

August 17, 2022

Declaration of Authorship

I, Alejandro Nieto Ramos, declare that this thesis titled, "Bayesian inference and prediction in cardiac electrophysiology models with an application to representing variability" and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed:

_____

Date:

_____

"*Rien au monde ne vaut qu'on se détourne de ce qu'on aime.*"

Albert Camus.

ROCHESTER INSTITUTE OF TECHNOLOGY

Abstract

College of Science

School of Mathematical Sciences

Doctor of Philosophy

Bayesian inference and prediction in cardiac electrophysiology models with an application to representing variability

by Alejandro Nieto Ramos

Many different techniques have been used for parameter estimation in cardiac electrophysiology models, from optimization algorithms to heuristic and frequentist statistical methods. However, the fixed parameter values obtained from such approaches cannot provide a complete description of variability within an individual or across a population. To overcome this shortcoming, in this work we adopt a Bayesian approach by applying the Hamiltonian Monte Carlo (HMC) algorithm to cardiac electrophysiology models and data for the first time through three studies. (i) Using HMC, we fit synthetic and experimental cardiac voltage data from different pacing rates and find the probability distributions of the parameters of two relatively low-dimensional models, the Mitchell-Schaeffer (MS) and Fenton-Karma (FK) models. We successfully fit synthetic and experimental voltage traces and build populations of action potentials with the posterior probability distributions of the parameters. (ii) We compare the performance of HMC with that of the main Bayesian approach used previously for similar applications, the Approximate Bayesian Computation Sequential Monte Carlo (ABC SMC) algorithm. Both techniques are able to describe the dynamics of synthetic and experimental voltage data using the MS and FK models, with HMC more consistent and ABC SMC more versatile and easier to implement. (iii) We study the variability of cardiac action potentials in space within an individual. We use HMC and a novel approach employing a Gaussian process prior for one spatially varying MS model parameter along with a hierarchical model for the remaining parameters, considered spatially invariant. Using this approach, we do inference and prediction on synthetic cardiac voltage data, exploiting the spatial correlations in cardiac tissue that arise from cellular coupling to use voltage information from a small number of sites to predict parameter value distributions and families of voltage data in other locations. Together these three studies show the potential of Bayesian inference and prediction in providing a framework to represent variability within cardiac electrophysiology modeling.

Acknowledgements

The best portion of a good man's life is his little, nameless, unremembered acts of love and kindness.

William Wordsworth

Nobody does anything alone, and I would not be here if it were not for all the people who have helped me or taken care of me, and overall, loved me. I will start chronologically since I started the PhD and I apologize in advance if I miss someone since there are so many people I want to thank.

I want to start by thanking all the members of the committee for taking the time to read my thesis and provide guidance, first during the proposal and now during the defense: Elizabeth, Flavio, Laura, Professor Zwickl and Professor Fokoué. Even though Chris Krapu was not a member of the thesis committee, I want to thank him for having believed in me during a moment of my life when I really needed it. We worked so well that what we did became Chapter 6 of this thesis. Since I did my internship with Chris, I started believing that things that I had thought in the past were not possible were actually possible. I have had much good news since a year ago that I started the internship at Oak Ridge and I am very grateful for that opportunity.

I want to thank my official advisor Elizabeth Cherry and my unofficial advisor Flavio Fenton at Georgia Tech. I met Flavio in Mexico City, but Elizabeth became my advisor. Who would have guessed that in less than two years after I started the PhD, I was going to end up working with both? Not only that, especially here in Atlanta, Flavio and Elizabeth were the closest I had to a family all these three years, because not only academically, but also personally, they always made sure that I and the rest of their students were doing well. They have taken me shopping, celebrated with me, and they were with me in the hospital, first Elizabeth in Rochester, and then both in Atlanta.

Even if it would not seem possible for someone as impractical as me, I have learned from Elizabeth to be more pragmatic and not be so afraid of showing something even if it is not perfectly finished: papers and presentations need many iterations and that is the way

things are. I especially want to thank Elizabeth as she has seen me at my worst but she never judged me; she always gave me space to recover. She was always very sweet in those dark moments, and I have always known how she cares about me. I am very grateful for all she has done for me.

I want to thank Flavio for his generosity. He gave me a space in Howey even though we practically just continually worked for the summer when I arrived in Atlanta. He also encouraged me to apply to RIT, for an internship, for a job and start working even if I did not have the degree if the opportunity arrived; I was very happy in Flavio's lab since they day I arrived because there were many people I could talk to, and from the beginning, all of them seemed friendly to me.

Of the members of the lab, I especially want to thank Conner for always being able to talk about science, sharing data with me and teaching me anything of what he knows; I always had ideas after talking to him. I also want to thank Abouzar and Elly because they always cheered me up, saw the positive in me, invited me to go out or even to their house; to the other members of the lab who were always kind to me and whom I had nice talks with: Héctor, Noah, Kamal, Shaun, Andrea and Ilija. I also want to thank all the nice people in Howey and Coda that treated me like I was part of GaTech: JC, Martin, Alan, Reneé, Arlene and Nirvana.

Moving now to Rochester, I want to thank the best two friends I made there: Pancy and Professor Fokoué. Trevor and I might not be friends, but he helped me very much because he always saw the positive in me; same with Aaron and Danielle.

Pancy and I have a special bond, maybe because we understood what it means to study for a PhD in a foreign country, especially when we were members of the first cohort. After I moved to Atlanta, Pancy never stopped texting me or talking to me in such a way that when we met again at the end of last year, I felt like I had just seen her a couple of days ago.

What can I say about Professor Fokoué? I am so fortunate to have found him in my life. Professor Fokoué has been my spiritual, academic and life guide. To a great extent, I

have become freer because of all the things he has taught me, and I can say with certainty that with him I have learned what unconditional love is.

I also want to thank all the kind people I met at RIT, starting with Professor Cahill. He is a generous man and I want to thank him for all the nice things that he has done for me and the rest of the PhD students. I also want to thank Laura and Niels for the nice time that we spent in Journal Club and the social meetings that we used to have on Fridays; I really enjoyed them. I also want to thank Professors Ross and Hoffman for the encouragement. That helped me a lot when I did not feel very confident. Finally, I want to thank all the people who always helped me during the PhD with all the processes that needed to be done, from enrollment to stipend payments: to Kate, Tina, Corinne, Ginny, Susan and Shawna; thank you so much.

Moving now to Mexico, even though not everybody is still there, I want to thank to all my dear friends who never left me alone; they always listened to me, encouraged me to keep going, gave me advice or just made me laugh. Starting with Rólex and Amparo, who since I was studying for the master's degree encouraged me, and Rólex always called me to know how I was and encouraged me when I was down—which was so often—and the rest of the nerds: especially Hebert, who also called me often, Lalo, who never forgets my birthday and Memo and Raúl (not part of the nerds), who more than once gave me medical advice for me or my mother. To my friends from RIT, UAM, ITAM, UDLAP and all the academic and non-academic places I have visited: Sandra, Güeri, María Elena, Inútil, Becario, Carmencita, Do Re Mi Fa Jo, Flavie, Noris, Lore, Chucho, Citallila, Monty, Sofía, Azrilla, Karlilla, Charles, Eri, la Tía, Pablirijillo, Ana, Pollín, Ivanovski, Jason and Sim (the last two from the US). Thank you for loving me.

I have thought many times that I would probably have never finished the bachelor´s degree if it had not been for Beatriz Rumbos and Patricia Medina. Thank you very much for trusting me.

Finally, I want to thank the people who have been with me for the longest time, my family—my mother, my sister and my father—and my friends who have treated me very

much like I was part of their family: Alma, Salud and Françoise.

Contents

List of Abbreviations

Action potential	**AP**
Action potential duration	**APD**
Approximate Bayesian computation	**ABC**
Approximate Bayesian computation sequential Monte Carlo	**ABC SMC**
Bifurcation plot	**BP**
Cycle length	**CL**
Gaussian process	**GP**
Hamiltonian Monte Carlo	**HMC**
Markov Markov chain Monte Carlo	**MCMC**
Ordinary differential equation	**ODE**
Uncertainty quantification	**UQ**

A mi hermana.

A mi mamá, que me enseñó que no importa qué haga, siempre lo haga lo mejor que pueda.

Chapter 1

Introduction

1.1 Motivation

Heart disease is the leading cause of death in the United States, causing about one in four deaths [68]. Examples of heart disease include heart failure, sudden cardiac arrest and stroke, all of which can be caused by non-treated arrhythmias. Arrhythmias like atrial flutter, atrial fibrillation and ventricular fibrillation arise from disruptions to the heart's electrical signaling, which triggers contraction. Despite being a subject of study since at least 1847 [57], the origin of arrhythmias is not completely understood.

Mathematical models can be used to improve our understanding of real-world phenomena such as the electrophysiology and dynamics of the heart. As a simplification of a physical or engineered system, a model can provide insights into observed behavior, describe patterns, or make predictions. While the form of the model equations can express qualitative relationships among the system variables, well-chosen parameter values are needed for a quantitative understanding. The first mathematical model of cardiac electrophysiology was published in 1962 by Noble [73] as a modification of the influential neuronal model of the squid giant axon developed by Hodgkin and Huxley in 1952 [44] to study the electrical dynamics of cardiac Purkinje cells. The Hodgkin-Huxley model, and consequently, the Noble model, were based on the analogy of a circuit mathematically represented by an ordinary differential equation, where the excitation of cells starts with a sudden change in the electrical potential across the cell membrane as the result of a flux of charged ions

passing across the cell membrane, and where the initiation and propagation of electrical signals is caused by controlled opening and closing of ion channels in the membrane [41].

Among other properties, the Noble model was able to reproduce alternans, a behavior that can occur after applying a sufficiently strong periodic stimulus to a cardiac cell, where the duration of an action potential—the time that it takes for a cardiomyocyte to depolarize and repolarize for a given threshold—alternates between long and short. Because alternans are known to be precursors of arrhythmias ([77]), cardiac electrophysiology models can help in understanding alternans, and in turn, in the initiation and cessation of arrhythmias [56]. Although cardiac electrophysiology models typically are developed to describe the cellular level, they can be used to study the propagation of electrical waves in the heart by adding a diffusion term (in a cable or in a two dimensional space) to represent intercellular coupling.

Since the Noble model, many other cardiac models have been developed to study cells from different regions of the heart (e.g., Purkinje network [73, 26], ventricles [9], atria [62], sinoatrial node [86], and atrioventricular node [47]) in different species (e.g., guinea pig [55, 54, 18], human [97, 9, 75], dog [101], rabbit [43], rat [76], and mouse [4]). Some models have been developed from first principles [84] while others are phenomenological [50, 31, 66]; some models have only a few parameters [46, 66], while others include more than a hundred [48]. These different models demonstrate different dynamics and properties; for example, despite its importance, alternans is only exhibited by a minority of models [15].

Cardiac cell models can be classified by the types of components they include as well as by the breakthroughs [74] they have represented in describing cardiac cells. First-generation models are limited to a small number of ionic currents and in many cases a description of the intracellular calcium concentration. As an example, the Beeler-Reuter model (1977) [1] has four main currents, including a calcium current that had not been described when earlier models like the Noble model were developed along with a calcium concentration based on that current. Second-generation models include more components

such as pump and exchanger currents as well as additional intracellular ion concentrations and can enforce charge conservation, for example. The Luo-Rudy II model (1994) [54] is a widely used second-generation model developed as a modification of the earlier first-generation Luo-Rudy I model [55], which in turn was based on the Beeler-Reuter model. In addition, some second-generation models use Markov models of ion channels to describe transitions among a small set of discrete states [18]. In third-generation models, which are still being developed, a cardiac cell is considered a spatially extended entity instead of a point; this approach allows a spatiotemporal treatment of the intracellular calcium concentration, which has an effect on the excitation–contraction coupling dynamics of cardiac cells. The Greenstein-Winslow model [38], which is an updated version of the Winslow model from 1999 [101], is a third-generation model that was designed to incorporate more detailed descriptions of intracellular calcium handling [79].

All of the models described above have been published with a single set (or occasionally a few sets) of parameter values that represent a good fit to the data used during model development. However, different individuals have differences in properties such as ion channel expression that lead to differences that typically are captured in models through changed parameter values. Thus, researchers may want to determine individualized parameter values in order to use a model to customize a treatment or therapy. In addition, in the case of drug development, parameter values are desired not only to represent the population mean, but also to express variability across a population. Furthermore, since many models are non-identifiable [23], having just a set of fixed parameters oversimplifies the information that a probability distribution of parameters could give. As a result, having robust tools to identify personalized parameter values from real-world data is of great importance.

Nevertheless, parameter estimation is a notoriously challenging problem. For cardiac electrophysiology models, many different techniques have been used, from optimization algorithms to heuristic and statistical methods [100]. Traditional optimization approaches provide several advantages including simplicity and relatively quick execution; however, they are less suitable for uncertainty quantification and probabilistic modeling. For these

purposes, a statistical modeling framework may be desirable. The modern frequentist approach has been widely used for around a century, and its popularity is partially due to the fact that it has been able to address many kinds of problems but also because of its computational tractability. Despite these advantages, well-known limitations like the use of p values and the fact that experimental design most be established in advance restricts the applicability of the frequentist approach [49].

To avoid the limitations of traditional optimization and frequentist-based statistical-modeling approaches, this thesis is focused on the use of a Bayesian approach for describing parameter values based on available data. Bayesian methods can provide a more complete description of randomness in parameter values as an effect of all the sources of uncertainty involved in the modeling and fitting of cardiac electrical data (e.g., model variability and structural variability) [65]. One type of Bayesian approach, Markov Chain Monte Carlo (MCMC) methods, has been used successfully toward this end in a variety of fields, including pharmacometrics ([72]). However, such approaches have known limitations when used to do inference in nonlinear differential equations models; for instance, Metropolis-Hastings can become stuck in local minima [63]. As an alternative, Approximate Bayesian Computation methods, which use an approximation of the likelihood by employing a distance function between summary statistics [96], can be used. These methods can avoid some problems with full Bayesian methods like MCMC, including poor scaling with the number of parameters, and have been used in cardiac and neuronal electrophysiology models [21], but result in an approximation of the true posterior distribution.

In this thesis, we consider the application of a full Bayesian method designed to surpass the limitations of traditional MCMC algorithms, like Metropolis-Hastings and Gibbs sampling, to the problem of parameter estimation for cardiac models. The specific algorithm we use, which was developed in 1987 [28] as hybrid Monte Carlo and later called Hamiltonian Monte Carlo (HMC) [3], is known for exploring the parameter space in an efficient manner, since it uses the gradient of the likelihood and improved scalability and convergence. Here we apply HMC to fit synthetic and experimental cardiac voltage data

in three separate studies. (i) As an initial proof of concept, we use HMC to find the distributions of the parameters of two low-dimensional cardiac electrophysiology models, the Mitchell-Schaeffer [66] and Fenton-Karma [31] models. (ii) Using the same type of data and the two models from the previous study, we compare the performance of HMC and the approximate Bayesian computation sequential Monte Carlo algorithm. (iii) As a final demonstration of the usefulness of HMC in cardiac electrophysiology modeling, we study the variability of cardiac action potentials in space within an individual. Using HMC and a novel approach employing a Gaussian process prior in one of the parameters of the MS model and a hierarchical model in the non-spatial parameters, we do inference and prediction on synthetic cardiac action potentials. We exploit the correlation found in contiguous cells using training points in a few locations, where only voltage measurements known, to do inference on the locations and prediction in the rest of the locations in the domain.

1.2 Outline

The rest of this thesis is organized as follows. In Chapter 2 we present an overview of previous work on inference and variability in cardiac electrophysiology, some basic concepts of cardiac action potential modeling and terminology, the main ideas behind the Bayesian approach and some Gaussian process concepts. Chapter 3 introduces the cardiac electrophysiology models used in this work, the Mitchell-Schaeffer (MS) and Fenton-Karma (FK) models, as well as the two algorithms employed in this work, Hamiltonian Monte Carlo (HMC) and Approximate Bayesian Computation Sequential Monte Carlo (ABC SMC). In Chapter 4, we fit the MS and FK models to synthetic and experimental data using HMC and present our results. In Chapter 5, we compare the performance of HMC and ABC SMC, explore the robustness of these methods and suggest cases for their use. In Chapter 6, we study variability in space of cardiac electrophysiology using the MS model to fit synthetic data, employing a Gaussian prior on one of the parameters considered in space and a hierarchical model on the rest of the non-spatial parameters. Finally, in Chapter 7, we present our conclusions and some limitations and discuss future

directions this work might take.

Chapter 2

Background

In this chapter we present an overview of previous work on inference and variability in cardiac electrophysiology models. We later introduce some concepts related to cardiac electrophysiology and the modeling of cardiac action potentials. Finally, the main ideas behind the Bayesian approach and Bayesian estimation in dynamical systems are presented, and some concepts associated with Gaussian processes are discussed at the end.

2.1 Inference and variability

In this section we review studies where inference and variability have been explored in cardiac electrophysiology, highlighting the Bayesian approaches that researchers have taken.

2.1.1 Parameter estimation in cardiac electrophysiology models

A broad range of models of cardiac action potentials have been developed [30] to describe the heart's complex electrical dynamics across multiple species and regions of the heart. These models also vary in complexity and level of detail, from simple two-variable models with a few parameters up to as many as dozens of variables and hundreds of parameters. In most cases, these models have been published with a single set of parameter values. Nevertheless, there is often a desire to customize the models by obtaining parameter values that can be used to match particular experimental recordings [39], to represent individual patients [6], or to create virtual cohorts [70].

The challenges of finding parameter values to match specific input data or properties are well known. Within cardiac electrophysiology modeling, many approaches have been used, from optimization techniques to heuristic methods. Some examples include least squares variations [27, 103], sequential quadratic programming [9], genetic algorithms [94, 5, 11], automatic optimization [29, 51], and a hybrid method combining particle swarm optimization with a gradient-based local algorithms [53]. More recently, parameterizations have been obtained using Bayesian approaches, such as history matching [19], Bayesian active learning [102], and a combination of Metropolis-Hastings and Gibbs sampling [90].

Nevertheless, models that include a single set of parameter values, however well fitted, lack deeper information that can be included when a distribution of parameter values is obtained. For example, when multiple data sets are available, simplification by fitting to the mean may misrepresent properties of the data [78]. In addition, models with a single set of parameter values neglect the fact that many models are not fully identifiable from the necessarily limited input data used for fitting [27, 22, 11].

2.1.2 Bayesian algorithms for parameter estimation

While the optimization-centric approaches in the previous section yield several advantages including simplicity and relatively quick execution, they are less suitable for uncertainty quantification and probabilistic modeling. For these purposes, a statistical modeling framework may be desirable.

Adopting a Bayesian approach is useful for two primary reasons; first, by using Monte Carlo methods, it is possible to freely use any statistical model of choice for analyses without functional or distributional restrictions. Second, this approach allows rigorous uncertainty quantification (UQ) to be conducted via characterization of the posterior distribution over the model parameters given the data and also permits appropriate specification of prior distributions over model parameters. This characterization can be performed using Markov chain Monte Carlo (MCMC) to draw samples from the posterior distribution. The objective of MCMC methods within the context of Bayesian modeling is to design a Markov chain in such a way that the stationary distribution of the chain

coincides with that of the posterior distribution. In the case of parameter estimation for an ordinary differential equation (ODE), the interest lies in finding the distribution of parameters used as inputs for the ODE model. In [80], a review of methods to parametrize ODEs, some of which are Bayesian, is presented.

A diverse range of MCMC methods have been developed, some of which are widely applicable and generic, while others are problem-specific. The Metropolis-Hastings algorithm [64, 42] is perhaps the simplest method and requires only evaluating the model's non-normalized posterior density a single time for each Monte Carlo iteration. Gibbs sampling [34] is another popular MCMC method and is applied by iteratively selecting a single parameter and sampling conditionally on values of all other parameters.

These traditional full Bayesian MCMC methods have limitations, including sensitivity of the desired posterior distribution to the full Bayesian specification [99]. Also, they do not make use of gradient information from the posterior and may exhibit random walk behavior, which slows convergence to the posterior distribution, as well as poor scalability with the dimensionality of parameter space [69]. As a consequence, it may be advantageous to use gradients of the log-posterior for models with a continuous parameter space. Hamiltonian Monte Carlo (HMC) [28, 69] makes Monte Carlo proposals guided by these gradient terms and exhibits much more rapid convergence to the target distribution [2, 61]. An alternative paradigm is to identify a simplified surrogate distribution for the posterior distribution and employ optimization to minimize a variational loss function. Generic methods to do this for arbitrary probability models with continuous parameter spaces are now widely available [52] as a complement to MCMC methods. If the likelihood is unknown or intractable, approximate methods like Approximate Bayesian Computation [60] can be used.

2.1.3 Applications of Bayesian modeling in electrophysiology

Several methods have been developed to overcome the computational limitations of traditional Bayesian methods. One approach, as mentioned above, is the use of approximate Bayesian methods [91], where the likelihood, which can be expensive to compute, is not

used, resulting in greater efficiency at the cost of obtaining an approximation of the posterior distributions. An example is Approximate Bayesian Computation Sequential Monte Carlo (ABC SMC) [95, 25], which has been applied to obtain parameter distributions for the Hodgkin-Huxley neural model [22] and for the O'Hara et al. model [75] of cardiac cells [21]. HMC has been used successfully for model calibration in ecology [67] and pharmacometry [59], but its utility for cardiac action potential models [71] compared to approximate methods has not been studied.

2.1.4 Variability in cardiac electrophysiology models

Appropriate selection of parameter values can be challenging for systems that exhibit variability. In biology, variability occurs naturally across all levels, from the microscopic to the macroscopic. Cardiac cells in particular exhibit variability in the shape and duration of their action potentials, the electrical signals that trigger contraction, not only among individuals, but also in space and time within a single individual.

The presence of variability makes it difficult to predict outcomes, such as the effects of a drug or treatment on a population, because different individuals may exhibit different behavior. Some aspects of natural variation can be represented in a model through the use of different parameter sets that describe separate individuals. In the previous section, we mentioned the problems that can arise when models of cardiac action potentials do not consider any kind of variability and are published using a single set of parameter values.

Several approaches have been used previously to describe variability in cardiac electrophysiology [65]. Britton et al. [7] developed a population of models by selecting random sets of parameters for a model of interest and keeping only those with properties falling within defined ranges. The population then could be used to characterize variability in responses to specific interventions like treatment with antiarrhythmic drugs. Sánchez et al. [85] used a similar approach to characterize variability in the properties of cells from patients under normal sinus rhythm and chronic atrial fibrillation. The populations created were used to identify the ionic determinants of variability in action potential shape

and duration among the patients. In a different study, Sarkar et al. [87] used a linear regression approach based on parameter sensitivities to relate parameter space to output space as specified through measured biomarkers.

However, these approaches have some limitations. For example, the population-based approaches are inefficient, as many of the generated candidate parameter sets do not satisfy the inclusion criteria (nearly 98 percent for the study by Britton et al. [7]). The assumption by Sarkar et al. [87] of a linear relationship between the sets of parameters used as inputs and the responses obtained may be too restrictive. In addition, none of these studies addressed intra-subject variability, and variability was considered only at the cellular level, not at the tissue or organ level. Further, none of these approaches provides a method for obtaining different sets of parameter values consistent with given experiments.

Following a statistical approach, Pathmanathan et al. [78] developed a description for sodium channel inactivation that included terms to describe mean behavior as well as individual-level variation. Other relevant studies have looked at parameter sensitivity analysis via multivariate regression [92] and parameter estimation under multiple action potential models [58]. As we mentioned earlier, Bayesian approaches have been used previously with cardiac models for parameterization [19, 22, 90] and parameter sensitivity [13, 19]. However, Bayesian techniques have not been used to study spatiotemporal variability.

2.2 Cardiac electrophysiology

In this section we introduce some basic cardiac electrophysiology concepts needed to understand the ideas behind the mathematical modeling of cardiac action potential models, their dynamics and the estimation of their parameters.

2.2.1 Some basic concepts

A cardiomyocyte is a muscle cell in the heart. If a cardiomyocyte receives a sufficiently strong electrical stimulus, its membrane potential changes and, in response, ions move across the membrane through specialized proteins called ion channels, following electrochemical gradients, thereby inducing an electric current. Individual ion channels stochastically open and close, a phenomenon called gating, to allow or inhibit ion transport, with the open probability depending on properties such as the electrical potential across the membrane or ion concentrations. If the applied stimulus increases the cell's membrane potential above a threshold value, the cell depolarizes rapidly followed by a slow (hundreds of ms for large mammals) return to its resting potential, a response called an action potential (AP). After an action potential, the cell must recover to be able to respond to another stimulus with a similar action potential. If an electrical stimulus is applied too soon after a previous stimulus, the cell may respond with an abbreviated action potential or even may fail to respond.

The sinoatrial node spontaneously generates an action potential at a species-dependent rate, typically around once per second for humans. In many experimental preparations, the sinoatrial node is not present and thus cardiomyocytes are periodically stimulated externally to mimic the natural heart rhythm. The period of stimulation, which can be varied to elicit different dynamical behavior, is also known as the cycle length (CL). One way to quantify differences in action potentials, such as characteristic shortening for shorter CLs, is by measuring the action potential duration (APD), which is the time it takes for a cell to return to its rest state after initiating an action potential. After waiting some time for the cell to reach a steady state, the APDs of consecutive action potentials typically are constant for long CLs, but rapid pacing can produce a shortening of action potential duration and, at sufficiently fast rates, alternans (Fig. 2.1), a phenomenon in which action potentials elicited from a constant CL alternate between long and short durations. At extremely short CLs, every other stimulus may arrive well within the preceding action potential, resulting in block, which is the failure of an action potential to develop. The different dynamics can be represented in a bifurcation plot (BP) where

CL is the bifurcation parameter; for CLs with alternans, two branches are present in the bifurcation plot.

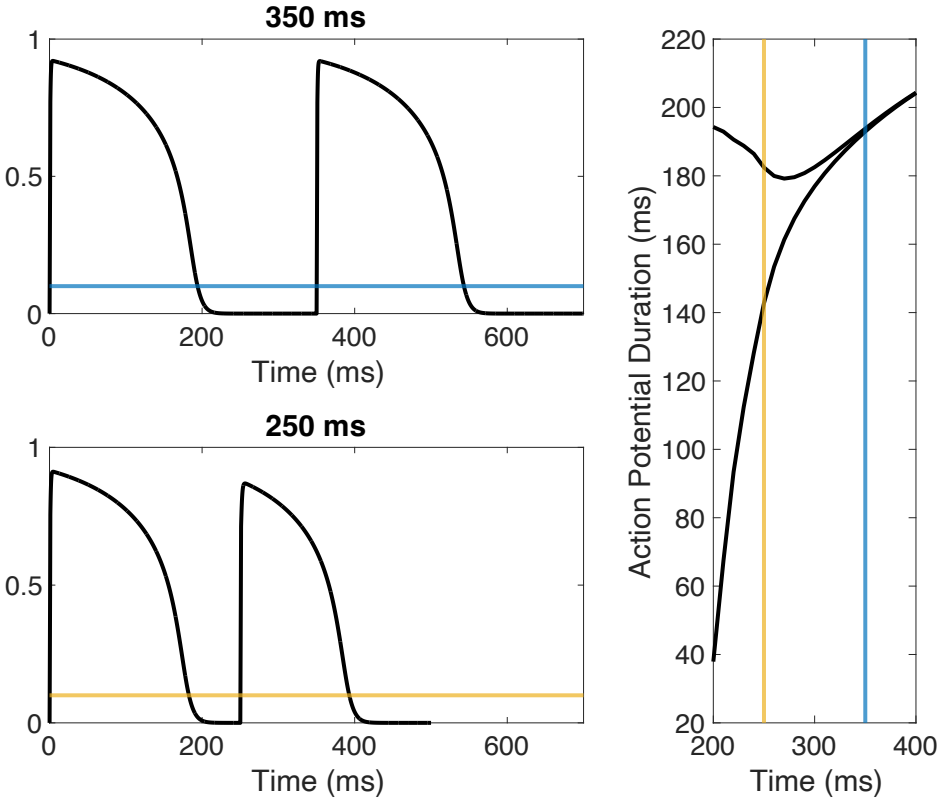


FIGURE 2.1: Action potential duration as a function of cycle length using the Mitchell-Schaeffer model introduced below. The left panel shows action potentials at CLs of 350 and 250 ms, with alternans present for the shorter CL. Horizontal lines correspond to APDs at 90%.

To observe the electrophysiological heterogeneity of the heart, it is important to characterize the spatiotemporal dynamics of cardiac action potentials. Optical mapping is an experimental technique that can be used towards this end. More specifically, optical mapping is a method used to indirectly record the electrical activity of the heart at high spatio-temporal resolution through the use of voltage (or calcium) fluorescent dyes. It allows for mapping the propagation of electrical signals across the myocardium, thereby allowing characterization of electrical dynamics through surface imaging (Figure [82]). After these signals are recorded, action potential durations may be calculated and mapped in order to visualize the variability of action potentials in space at specific CLs.

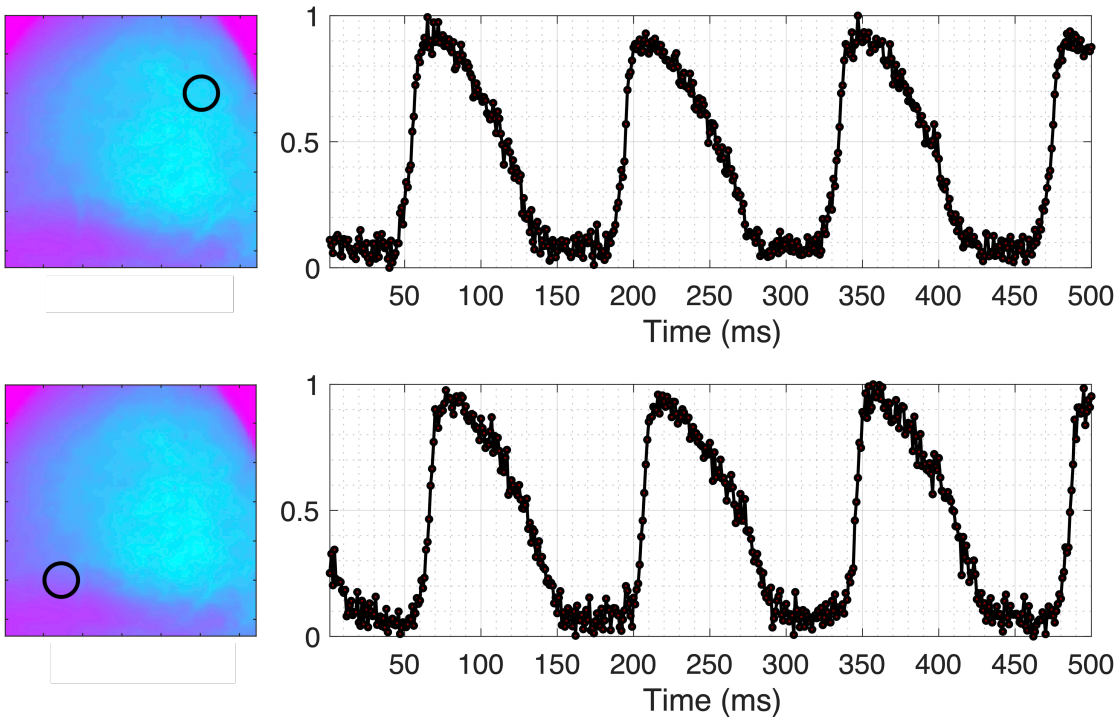


FIGURE 2.2: Voltage mapping obtained with optical mapping from a rabbit heart paced at 280 ms (left), where two series of action potentials are shown at two different locations represented by the center of the black circles.

2.2.2 Modeling cardiac action potentials

The excitation of a cardiac myocyte (or neuron) starts with an abrupt change in the electrical potential across the cell membrane as the result of a flux of charged ions passing though it. The initiation and propagation of electrical signals is caused by controlled opening and closing of ion channels. Hodgkin and Huxley [44] created the first mechanistic model for the neuronal action potential of a squid axon in 1952. This model is based on the analogy of a circuit [41]. The action potential evolution is described by the differential equation

$$C_m \frac{dV}{dt} + I_{ion} = I_{app}, \tag{2.1}$$

where C_m represents the membrane capacitance per unit area, V is the voltage, I_{ion} is the sum of ionic currents and I_{app} is the applied external current. Each individual ion current, that in this particular case were defined as sodium (Na), potassium (K) or leakage (L), is modeled according to Ohm's law

$$I_i = g_i(V - V_i),$$

where i =Na, K, L, $g_i = \hat{g}_i(t; V)$ is a conductance function that depends on the time t and voltage, and V_i is the Nernst potential for the ions species i. The Nernst potential is the potential at which the electrical and chemical gradients across the membrane are balanced for current i, producing no net movement of ions as a result.

Different models may incorporate different descriptions of intracellular ion concentrations or other processes, as we will see in the models presented in the next chapter. To incorporate spatial coupling, a diffusion term $D\frac{\partial^2 V}{\partial x^2}$ can be added to equation 2.1, where x represents space and D is a diffusion function, usually assumed constant.

Denis Noble developed the first cardiac action potential model in 1962 to simulate the action potential dynamics of Purkinje cells [73]. This model was itself a modification of the Hodgkin-Huxley model. Since then, several types of cardiac action potential models have been developed, representing different kinds of cells from diverse species. Over time, the models have become more complex, usually including more ionic currents, and as a consequence, more differential equations and parameters. They have also been modified to include cell communication and spatiotemporal calcium cycling dynamics. Depending on what is being studied, phenomenological models or more detailed models developed from first principles can be used [79].

2.3 The Bayesian approach

In this section we explain how to estimate the parameters of cardiac electrophysiology models to fit action potential recordings using a Bayesian approach. We introduce different Markov chain Monte Carlo algorithms that have traditionally been used for this end.

2.3.1 Bayesian estimation in dynamical systems

Given the system $\boldsymbol{x}'(t) = \boldsymbol{f}(\boldsymbol{x}, t, \boldsymbol{\theta})$ where $\boldsymbol{x} = \boldsymbol{x}(t) \in \mathbb{R}^n$, $t \in \mathbb{R}$, and $\boldsymbol{\theta} \in \mathbb{R}^m$ is a vector of parameters, we consider the statistical model

$$y_{ij} = x(t|\boldsymbol{\theta}) + \epsilon_{ij},$$

where y_{ij} is the j-th observation of the i-th individual being considered for the state $x(t|\boldsymbol{\theta}) \in \mathbb{R}$ registered at time t_{ij} and ϵ_{ij} is a random error corresponding to outcome y_{ij}.

Because the values of parameters used in cardiac models are usually difficult to measure experimentally, they often are inferred by fitting the model to data recorded by specific devices. In this work, we will infer the distributions of the parameters of a given model by applying different Bayesian approaches. Essentially, Bayesian inference is the process of fitting a probability model to a set of data where the result is summarized by a probability distribution on the parameters of the model and on unobserved quantities (e.g., predictions for new observations). We note that the parameters $\boldsymbol{\theta}$ are fixed but unknown. By choosing a Bayesian approach, the uncertainty in the parameter values is described by a probability model. Therefore, the parameters are considered random variables and posterior sampling can be applied.

By Bayes' theorem,

$$p(\boldsymbol{\theta}|\boldsymbol{y}) \propto p(\boldsymbol{y}|\boldsymbol{\theta})p(\boldsymbol{\theta}),$$

where $p(\boldsymbol{\theta}|\boldsymbol{y})$ is the final distribution (also known as the target or posterior distribution), $p(\boldsymbol{y}|\boldsymbol{\theta})$ is the likelihood, and $p(\boldsymbol{\theta})$ is the prior or initial distribution.

To sample from the final distribution, either a full Bayesian approach can be applied, meaning that we use the likelihood function on the right-hand side of Bayes' rule along with the prior, or an approximate Bayesian scheme can be used where instead of considering the likelihood, an approximate function is used. Examples of full Bayesian approaches are Markov Chain Monte Carlo (MCMC) methods [33] like Metropolis-Hastings and Hamiltonian Monte Carlo (HMC) algorithms. Approximate Bayesian Computation (ABC) methods are examples of the approximate approach [60].

2.3.2 Markov Chain Monte Carlo (MCMC)

To sample from the final distribution, Markov Chain Monte Carlo (MCMC) methods can be applied. A Markov chain is a sequence of points (for our application, in parameter space) generated by sequentially applying a random map. The objective of MCMC methods is to design a Markov chain in such a way that the stationary distribution of the chain coincides with that of the target distribution; in our case, we are interested in the final distribution of the parameters of a specified model. The goal is for the sample obtained with the Markov chain also to be a sample of the target distribution.

The Metropolis-Hastings algorithm [16] is the simplest of the MCMC methods. To obtain a sample from the final distribution, the chain is given an initial value; next, a proposal (also called transition) distribution from which it is easy to sample (q) is used to generate a candidate whose probability is conditional on the previous state. The value is accepted or rejected according to a rule and a correlated sample of the target distribution is obtained. To make an independent sample from the final distribution, only every k-th element of the correlated sample for some number k is taken. This procedure is known as lagging. Because the rule for this algorithm involves a ratio of probabilities, there is no need to have a normalized distribution from which to sample, which is an advantage.

Metropolis-Hastings Algorithm

Given $\boldsymbol{\theta}^{(t)}$ [a],

1. Draw $\boldsymbol{\theta}^* \sim q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)})$

2. Take $\boldsymbol{\theta}^{(t+1)} = \begin{cases} \boldsymbol{\theta}^* & \text{with probability } \alpha, \\ \boldsymbol{\theta}^{(t)} & \text{with probability } 1-\alpha, \end{cases}$

   where

$$\alpha = \min\left(1, \frac{p(\boldsymbol{\theta}^*)}{p(\boldsymbol{\theta}^{(t)})} \frac{q(\boldsymbol{\theta}^{(t)}|\boldsymbol{\theta}^*)}{q(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(t)})}\right).$$

[a] $t \in |mathbbN$ here, so it means θ at step t. The same applies for the next algorithm.

When a Markov chain is initialized, the chain will move in the state space around it. Therefore, if several chains are generated using different initial states, at first, the generated states of these chains will be close to one another. The period in which this phenomenon happens is known as the warm-up period or burn-in. A crucial property of Markov chains is that the initial state does not affect the state of the chain after a sufficiently long succession of transitions, assuming that certain conditions about the Markov chain are fulfilled[1]. When this state has been reached, it is said that the chain has reached a steady state and that the states reflect samples of the stationary distribution.

Another MCMC method is Gibbs sampling. This is a procedure in which all samples are accepted, which results in an improvement in computational efficiency over Metropolis-Hastings. An additional advantage of this method is that a transition distribution is not needed. However, the procedure can only be applied when all conditional distributions of each multivariate distribution parameter are known given the rest of the parameters.

Gibbs sampling

Given $\boldsymbol{\theta}^{(t)}$, obtain

1. $\theta_1^{(t+1)} \sim q(\theta_1|\theta_2^{(t)}, \ldots, \theta_N^{(t)})$,

   $\vdots$

[1] This property is a consequence of the *Ergodic Theorem*, which states that if a Markov chain is homogeneous, irreducible and aperiodic, it will reach an equilibrium distribution [33].

$i.\ \theta_i^{(t+1)} \sim q(\theta_i|\theta_1^{(t+1)}, \dots \theta_{i-1}^{(t+1)}, \theta_{i+1}^{(t)}, \dots, \theta_N^{(t)}),$

$\vdots$

$N.\ \theta_N^{(t+1)} \sim q(\theta_N|\theta_1^{(t+1)}, \dots, \theta_{N-1}^{(t+1)}).$

2.4 Gaussian processes

We present the multivariate normal distribution and its properties. This distribution is intrinsically related to Gaussian processes, which we also define in this section.

2.4.1 The multivariate normal and its properties

A multivariate normal distribution is an extension of the real random normal variable X with parameters μ and σ, whose probability density function is defined as

$$p_X(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right).$$

A random variable $\boldsymbol{X} \in \mathbb{R}^d$ is said to have a multivariate normal or Gaussian distribution with mean $\boldsymbol{\mu} \in \mathbb{R}^d$ and covariance matrix $\boldsymbol{\Sigma} \in S_{++}^d$ if

$$p_{\boldsymbol{X}}(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{\sqrt{(2\pi)^d |\boldsymbol{\Sigma}|}} \exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right),$$

where S_{++}^d represents the set of positive definite matrices. Let $\boldsymbol{X} \in \mathbb{R}^d$ with $X \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Let $\boldsymbol{x}$ be partitioned into $\boldsymbol{x}_A = (x_1, \dots, x_r)^T \in R^{d-r}$ and $\boldsymbol{x}_B = (x_{r+1}, \dots, x_d)^T \in R^{d-r}$ (and similarly for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$) such that

$$\boldsymbol{x} = \begin{pmatrix} \boldsymbol{x}_A \\ \boldsymbol{x}_B \end{pmatrix} \quad \mu = \begin{pmatrix} \boldsymbol{\mu}_A \\ \boldsymbol{\mu}_B \end{pmatrix} \quad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{AA} & \boldsymbol{\Sigma}_{AB} \\ \boldsymbol{\Sigma}_{BA} & \boldsymbol{\Sigma}_{BB} \end{pmatrix}.$$

Fig. 2.3 shows an example of a bivariate normal distribution sample where the two random variables involved are correlated

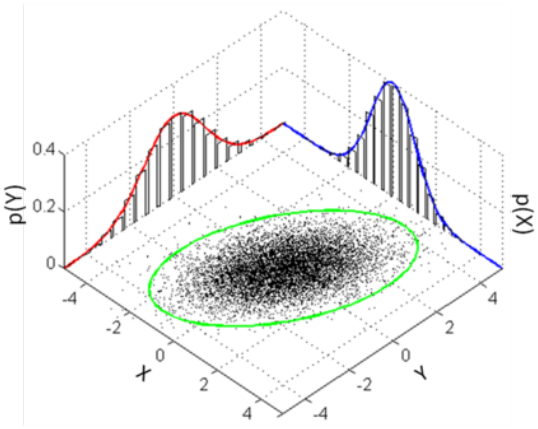


FIGURE 2.3: Sample of a normal distribution in $\mathbb{R}^2$. The marginal distributions (in blue and red) are also normally distributed.

The multivariate normal distribution has useful mathematical properties, because its marginal and conditional distributions are normally distributed. The sum of two univariate normally distributed random variables is also normally distributed.

- (marginals):

$$\boldsymbol{X}_A \sim \mathcal{N}(\boldsymbol{\mu}_A, \boldsymbol{\Sigma}_{AA}), \quad \boldsymbol{X}_B \sim \mathcal{N}(\boldsymbol{\mu}_B, \boldsymbol{\Sigma}_{BB}).$$

- (conditionals):

$$\boldsymbol{X}_A | \boldsymbol{X}_B \sim \mathcal{N}(\boldsymbol{\Sigma}_{AB}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_B - \boldsymbol{\mu}_B), \boldsymbol{\Sigma}_{AA} - \boldsymbol{\Sigma}_{AB}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{AB}^T)$$

$$\boldsymbol{X}_B | \boldsymbol{X}_A \sim \mathcal{N}(\boldsymbol{\Sigma}_{BA}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_A - \boldsymbol{\mu}_A), \boldsymbol{\Sigma}_{BB} - \boldsymbol{\Sigma}_{BA}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{BA}^T)$$

- (sum transformation):

  If $\boldsymbol{X}, \boldsymbol{Z} \in \mathbb{R}^d$ such that $\boldsymbol{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $\boldsymbol{Z} \sim \mathcal{N}(\boldsymbol{\mu}', \boldsymbol{\Sigma}')$ then

$$\boldsymbol{X} + \boldsymbol{Z} \sim \mathcal{N}(\boldsymbol{\mu} + \boldsymbol{\mu}', \boldsymbol{\Sigma} + \boldsymbol{\Sigma}').$$

2.4.2 Gaussian process

A collection of random variables $\{f(x) : x \in X\}$ is said to be drawn from a Gaussian process (GP) with mean function $m(\cdot)$ and covariance function $k_{ij} = k(x_i, x_j)$ if for

any finite set of elements $x_1, ..., x_n \in X$, the associated finite set of random variables $f(x_1), ..., f(x_n)$ have distribution

$$\boldsymbol{f} \sim \mathcal{N}(m(\cdot), \boldsymbol{K}) \text{ where } \boldsymbol{K} = (k_{ij})$$

or

$$f(\cdot) \sim \mathcal{GP}(m(\cdot), k(\cdot, \cdot)).$$

Any real-valued function $m(\cdot)$ is acceptable, but for the function $k(\cdot, \cdot)$ called kernel, it must be the case that the resulting matrix $\boldsymbol{K}$ is positive definite.

2.4.3 The squared exponential kernel

There exist many kernel distributions, but the most popular is the squared exponential kernel. Consider a zero-mean $\mathcal{GP}$, $f(\cdot) \sim \mathcal{GP}(0, k(\cdot, \cdot))$; the squared exponential kernel function is defined as

$$k_{SE}(x, x'|\alpha, \rho) = \alpha^2 \exp\left(\frac{1}{2\rho^2}\|x - x'\|_2^2\right).$$

The parameter ρ is called the length-scale and *alpha* is called the marginal standard deviation. The first corresponds to the frequency of the functions represented by the Gaussian process prior with respect to the domain; *alpha* adjusts the magnitude of the range of the function represented by the Gaussian process.

Since the covariance matrix needs to be invertible, the kernel $k(x, x') = \delta(x, x')\,\sigma^2$ needs to be added to the previous kernel. In Fig. 2.4 we see different graphs of the squared exponential function when the length-scale parameter is varied.

2.4.4 Gaussian processes in a Bayesian context

A Gaussian process generates data located throughout some domain such that any finite subset of the range follows a multivariate Gaussian distribution. Bayesian algorithms do not attempt to identify best-fit models of the data, nor do they make best-guess

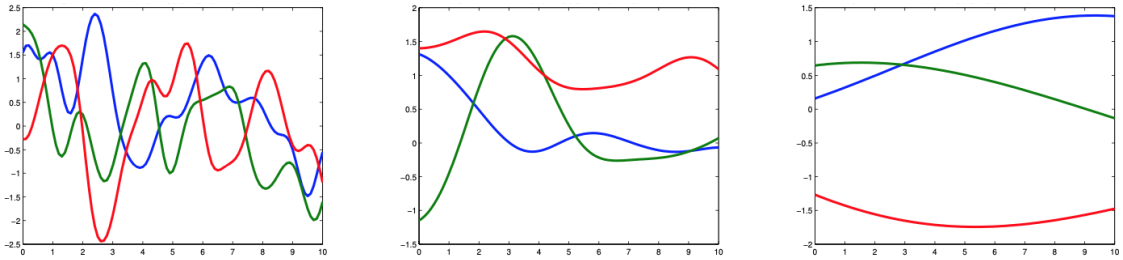


FIGURE 2.4: Different graphs of the square exponential kernel varying the length-scale parameter ρ .

predictions for new test inputs. They compute a posterior distribution over models or, equivalently, compute posterior predictive distributions for new test inputs. These distributions provide a useful way to quantify uncertainty in model estimates, and to exploit knowledge of this uncertainty in order to make more robust predictions for new test points.

2.4.5 Predictive inference with a Gaussian process

Bayesian predictive inference for Gaussian processes with Gaussian observations can speed up calculations by deriving the posterior analytically, then directly sampling from it. Suppose for a given sequence of inputs $\boldsymbol{x}$ that the corresponding outputs $\boldsymbol{y}$ are observed. Given a new sequence of inputs $\tilde{\boldsymbol{x}}$, the posterior predictive distribution of their labels is computed by sampling outputs $\tilde{\boldsymbol{y}}$ according to

$$\tilde{\boldsymbol{Y}}|\tilde{\boldsymbol{X}}, \boldsymbol{X}, \boldsymbol{Y} \sim \mathcal{N}(\boldsymbol{K}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{y}, \boldsymbol{\Omega} - \boldsymbol{K}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{K}) \tag{2.2}$$

where

$$\boldsymbol{\Sigma} = (k(x_i, x_j)) = \boldsymbol{K}(\boldsymbol{x}|\alpha, \rho, \sigma),$$

$$\boldsymbol{\Omega} = (k(\tilde{x_i}, \tilde{x_j})) = \boldsymbol{K}(\tilde{\boldsymbol{x}}|\alpha, \rho),$$

$\boldsymbol{K}$ is the covariance matrix between $\boldsymbol{x}$ and $\tilde{\boldsymbol{x}}$.

Chapter 3

Methods

Below we describe the cardiac action potential models used for fitting the data, the datasets to be fit, and the methods used for Bayesian inference in this thesis, including details of our implementations. For our study in variability, some of the details and the theory will be given in the related chapter.

3.1 Cardiac action potential models

In this work we seek parameter value distributions for two cardiac action potential models: the Mitchell-Schaeffer model and the Fenton-Karma model. Because the model parameters will be referred to frequently, we include the model equations in full.

The Mitchell-Schaeffer (MS) model [66] uses two variables, the voltage u and inactivation gating variable h, along with inward, outward, and stimulus currents (I_{in}, I_{out}, and I_{stim}, respectively) to describe the transmembrane currents that give rise to action potentials.

$$\frac{du(t)}{dt} = I_{in}(u,h) + I_{out}(u) + I_{stim}(t),$$
$$\frac{dh(t)}{dt} = \begin{cases} \frac{1-h}{\tau_{open}}, & u < u_{gate} \\ \frac{-h}{\tau_{close}}, & u > u_{gate}, \end{cases}$$

where

$$I_{in} = h\frac{u^2(1-u)}{\tau_{in}},$$
$$I_{out} = -\frac{u}{\tau_{out}}.$$

The five parameters include a threshold u_{gate} that determines the dynamics of the gating variable h; the remaining four parameters are time constants that effectively govern the durations of the depolarization (τ_{in}) and repolarization (τ_{out}) phases as well as the closing (τ_{close}) and opening (τ_{open}) of the gate. Initial values were set to $u = 0$ and $h = 1$ for each cycle length used. The stimulus current, whose units are $\frac{1}{ms}$, was applied periodically according to the specified cycle length (CL) for 1ms with a magnitude of 0.66. In Fig. 3.1 we can see the graphs of the two variables of the model after reaching steady state.

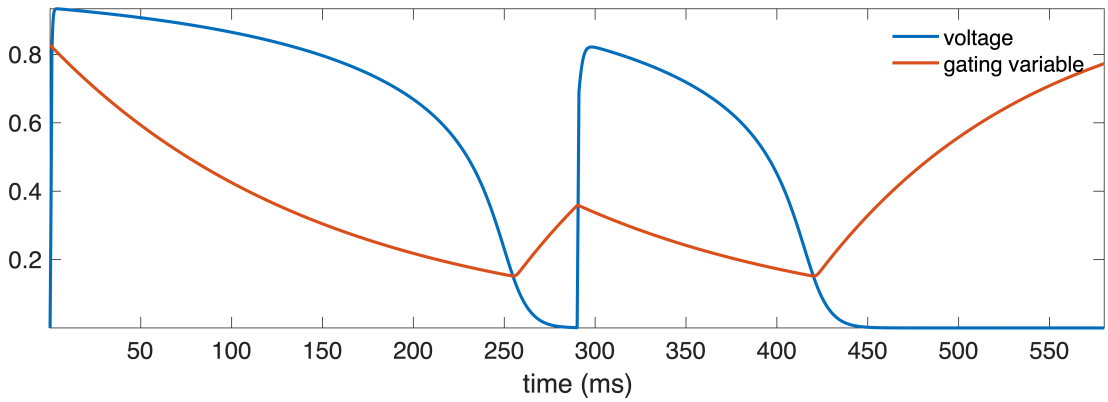


FIGURE 3.1: Mitchell-Schaeffer model solution after reaching steady state. The system was paced with a period of 276ms.

We also used the Fenton-Karma (FK) model [31], which is a phenomenological model that describes cardiac action potentials. It includes three state variables (voltage u and gating variables v and w) and 13 parameters.

$$\frac{du(t)}{dt} = -(I_{fi} + I_{so} + I_{si}) + I_{stim},$$

$$\frac{dv(t)}{dt} = \begin{cases} -\frac{v}{\tau_v^+}, & u \geq u_c \\ \frac{1-v}{\tau_{v1}^-}, & u_c > u \geq u_v, \\ \frac{1-v}{\tau_{v2}^-}, & u < u_v \end{cases}$$

$$\frac{dw(t)}{dt} = \begin{cases} -\frac{w}{\tau_w^+}, & u \geq u_c \\ \frac{1-w}{\tau_w^-}, & u < u_c \end{cases}$$

where

$$I_{fi} = \begin{cases} -\frac{(1-u)(u-u_c)v}{\tau_d}, & u \geq u_c \\ 0, & u < u_c \end{cases}$$

$$I_{so} = \begin{cases} \frac{1}{\tau_r}, & u \geq u_c \\ \frac{u}{\tau_0}, & u < u_c \end{cases}$$

$$I_{si} = -\frac{w}{2\tau_{si}}(1 + \tanh(k(u - u_c^{si}))).$$

The fast inward current I_{fi}, the slow outward current I_{so}, and the slow inward current I_{si} represent summary sodium, potassium, and calcium transmembrane currents, respectively. The magnitude of the 1ms-long stimulus current was 0.35. Initial values were set to $u = 0$, $v = 1$, and $w = 1$ for each cycle length considered. In Fig. 3.2 we can see the graphs of the three variables of the model after reaching steady state.

3.2 Numerical solution of the electrophysiology models

To avoid discontinuities, the Heaviside functions in the MS and FK models were replaced by smooth functions when using HMC or when comparing HMC vs. ABC SMC, to make a fair comparison. For instance, in the case of the MS model, the equation for the gating

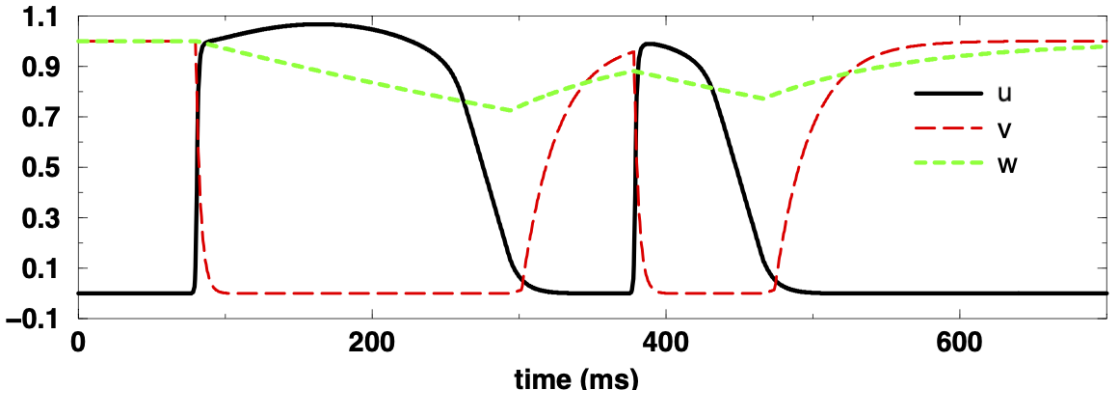


FIGURE 3.2: Fenton-Karma model solution after reaching steady state. The system was paced with a period of 375 ms.

variable was modified to

$$\frac{dh(t)}{dt} = \frac{1-h}{\tau_{open}}(1-p) - \frac{h}{\tau_{close}}p,$$

where $p = \frac{1}{2}(1 + \tanh(s(u - u_{gate})))$. The steepness parameter s was set to a large value of 50 for a steep transition resembling the Heaviside function.

3.3 Datasets

In this thesis, we use synthetic and experimental datasets to test the different statistical algorithms. The datasets include significant changes in action potential shape and durations as a result of rate changes, including a bifurcation to alternans (alternating long and short APDs despite a constant CL) at the shortest CLs. The specifications for the data set used for each study are described in the following subsections.

3.3.1 Datasets for the HMC study

Synthetic data were generated for the MS model using the parameter set given in the first column of Table 3.1, where the values for τ_{open} and τ_{close} were interchanged from the original parameters found in [66]. For the FK model, the parameter set can be seen in Table 3.2. In this section and the following, the sets of parameters used to generate the synthetic data for each model are subsequently referred to as the true values.

Pacing was applied for six beats at the chosen CLs and time series of points were taken from the last two of the six action potentials at a resolution of 0.1ms for the first 4ms, to ensure adequate temporal resolution of the upstroke, followed by a resolution of 15ms until the next stimulus was applied. The reason of our choice for the resolution was to accelerate the computational time while giving enough data information for the algorithm. Gaussian noise was then added with a mean of 0 and standard deviation of 0.03. The choice for the variance comes from the estimation of the noise shown in experimental data we used for fitting. For both models, the differential equations were solved using an adaptive forward Euler scheme, with a timestep size of 0.1 ms for the first 4 ms after the beginning of the stimulus followed by an increase to 0.5 ms in all cases.

TABLE 3.1: MS model parameter values for the HMC study.

	Dataset	Center values for HMC folded normal priors	Initial values for HMC
τ_{in}	0.3	0.3	0.3
τ_{out}	6	6	6
τ_{open}	150	150	150
τ_{close}	120	120	120
v_{gate}	0.13	0.31	0.31

TABLE 3.2: FK model parameter values for the HMC study.

	Dataset	Center values for HMC folded normal priors	Initial values for HMC
τ_d	0.48	0.48	0.3
τ_r	89	89	110
τ_0	26	26	20
τ_{si}	276	276	280
τ_v^+	44	44	27
τ_{v1}^-	82	82	80
τ_{v2}^-	589	589	350
τ_w^+	200	200	200
τ_w^-	215	215	200
u_c	0.17	0.17	0.2
u_v	0.01	0.01	0.01
u_c^{si}	0.37	0.37	0.45
k	4.5	4.5	5

The experimental dataset consisted of microelectrode recordings of voltage from zebrafish

hearts obtained previously [88]. Hearts were stimulated for a series of CLs until steady-state dynamics were achieved and included transitions to alternans. The last two APs from the same three CLs of 350 ms, 300 ms, and 276 ms were included in the final experimental dataset, with the smaller two producing alternans. Measurements were considered to include random Gaussian error with mean 0 and standard deviation σ (which was the standard deviation of added noise for the synthetic dataset and which represents uncertainty for the experimental dataset). The original resolution of the data was 0.1ms.

Data points were spaced every 0.1 ms (0.5 ms) for the first 4 ms of each AP and otherwise every 15 ms (10 ms) when using the MS model with the synthetic (experimental) dataset. With the FK model, data points were 0.5 ms apart for the first 4 ms of each AP and 1 ms apart for the next 3 ms, with a spacing of 15 ms otherwise, for both datasets.

3.3.2 Datasets for the comparative study

As for the previous study, both synthetic and experimental datasets were used to test the methods. Synthetic data were generated for the MS model using the default parameter set [66], with values given in the first column of Table 3.3. For the FK model, parameter set 4 from [32] was used; see the first column of Table 3.4. Pacing was applied as described in the previous section as well the Gaussian noise added to the solution.

For both models, the differential equations were solved using an adaptive forward Euler scheme, with a timestep size of 0.1 ms for the first 4 ms after the beginning of the stimulus followed by an increase to 0.5 ms in all cases except when fitting the MS model to synthetic data, in which case the time step was increased to 0.25 ms.

The experimental dataset is the same as the one used in the previous section; examples of action potentials from the dataset along with the bifurcation plot are shown in Fig. 3.3. To form the dataset in this case, a nonuniform resolution of 0.5ms was used for the first 4ms after applying the stimulus followed by an increase to 15ms until the next stimulus was applied. This approach allowed us to reduce the size of the dataset (and correspondingly

TABLE 3.3: MS model parameter values and intervals for the comparative study.

	Dataset	Intervals for initial SMC priors	Center values for HMC folded normal priors	Initial values for HMC
τ_{in}	0.3	(0.1,1)	0.3	0.3
τ_{out}	6	(3,15)	6	6
τ_{open}	120	(50,250)	120	120
τ_{close}	150	(100,200)	150	150
v_{gate}	0.13	(0.1,0.5)	0.13	0.13

TABLE 3.4: FK model parameter values and intervals for the comparative study.

	Dataset	Intervals for initial SMC priors	Center values for HMC folded normal priors	Initial values for HMC
τ_d	0.407	(0.03,1)	0.44	0.3
τ_r	34	(1,209)	84.29	100
τ_0	9	(1,50)	26.69	20
τ_{si}	26.5	(5,300)	283.38	200
τ_v^+	3.33	(1,100)	50.81	27
τ_{v1}^-	15.6	(1,300)	84.81	80
τ_{v2}^-	5	(1,2500)	591.19	350
τ_w^+	350	(1,800)	199.52	200
τ_w^-	80	(1,500)	218.63	200
u_c	0.15	(0.01,0.3)	0.15	0.2
u_v	0.04	(0.001,0.04)	0.01	0.01
u_c^{si}	0.45	(0.1,1.5)	0.36	0.45
k	15	(1,50)	4.25	5

the computational time) while retaining good accuracy during the upstroke. Voltage values were normalized.

For both synthetic and experimental datasets, the last two action potentials in a series of six were fitted to minimize transient behavior. After performing a series of initial experiments with different numbers and selections of CLs, we chose to use three CLs close to the bifurcation point, one at a long CL without alternans and two at shorter CLs within the alternans regime. Although the datasets to be fit utilized only three CLs, for comparisons of results at other CLs, voltage data were available at any CL of interest for synthetic data and for a broad range of CLs from the experiment; see Fig. 3.3. When generating bifurcation plots, CLs were decreased until block was reached, and action potential durations (APDs) were measured using a fixed threshold of $u = 0.1$.

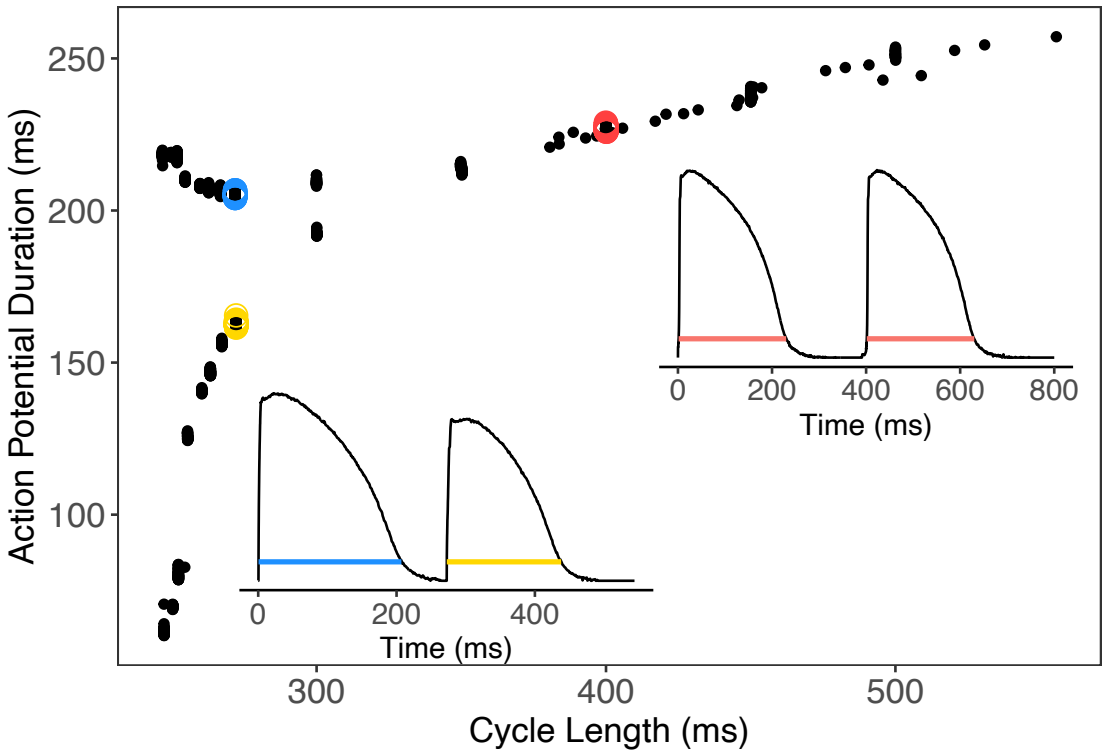


FIGURE 3.3: Action potential duration as a function of cycle length for the full zebrafish dataset. Insets show action potentials at CLs of 400 and 275 ms, with alternans present for the shorter CL. Horizontal lines correspond to APDs.

3.4 Bayesian Inference

3.4.1 Hamiltonian Monte Carlo

Hamiltonian Monte Carlo (HMC) is a Metropolis method that uses gradient information to explore the parameter space of a model more efficiently than conventional MCMC methods [3]. For a closed system, the Hamiltonian usually represents the total energy of a physical system: the sum of its kinetic and potential energies. The kinetic energy of an object is related to its motion, and the potential energy is related to its relative position. Hamilton's equations define the evolution in time of the physical system.

The physical interpretation of Hamiltonian dynamics can be used to provide insight into how the HMC algorithm works. In two dimensions, the dynamics can be thought of as similar to a ball with no friction sliding over a surface whose height varies. The state of

the system consists of the position of the ball given by a vector $\boldsymbol{x} \in \mathbb{R}^2$ and the momentum of the ball, defined as the product of its mass and its velocity, given by $\boldsymbol{p} \in \mathbb{R}^2$. The potential energy $U(\boldsymbol{x})$ of the ball is proportional to the height of the surface at its current position, and its kinetic energy is defined as $K(\boldsymbol{p}) = \boldsymbol{p}^T\boldsymbol{p}/(2m)$, where m is the mass of the ball. Depending on where the ball is (e.g., on a level part or a rising slope of the surface), its potential and kinetic energy will compensate to keep the total energy of the system constant.

In a non-physical MCMC application of Hamiltonian dynamics, the position corresponds to the variables of interest: the parameters given the data. The potential energy is the negative of the log of the probability density for the variables of interest, and for simplicity we assume that the mass is one. Momentum variables, one for each position variable, are introduced artificially to take advantage of the properties of a Hamiltonian system[1] and to work in the $(\boldsymbol{x}, \boldsymbol{p})$ space, which also makes it easy to recover the original variables $\boldsymbol{x}$.

At a high level, Hamiltonian Monte Carlo works as follows. If the probability density function P of a system can be written as

$$P(\boldsymbol{x}) = \exp(-E(\boldsymbol{x}))/Z,$$

where E is a function of $\boldsymbol{x}$ (like the potential energy $U(\boldsymbol{x})$ mentioned previously) and Z is an integration constant, $E(x)$ and $\nabla E(\boldsymbol{x})$ can be calculated to find the states with higher probability. In physics, P is called a canonical distribution[2] and E an energy function. The Hamiltonian (and the potential energy as well) is an energy function and thus a probability density function can be written using the Hamiltonian following the above form. Conversely, a distribution with density function $P(\boldsymbol{x})$ can be written as a canonical distribution by setting $E(\boldsymbol{x}) = -\log(P(\boldsymbol{x})) - \log(Z)$, where Z is a positive constant.

Using previous notation involving parameters $\boldsymbol{\theta}$ and data $\boldsymbol{y}$, the canonical distribution can

[1] Hamiltonian dynamics have many properties that are crucial in constructing MCMC updates such as reversibility, invariance and symplecticness [8]. The scope of this work does not allow us to explore these concepts.

[2] More generally, a canonical distribution considers the the exponential of the negative of an energy function divided by the temperature; here we assume that the temperature is 1.

be written as $P(\boldsymbol{\theta}) = p(\boldsymbol{\theta}|\boldsymbol{y})$ (even though the target distribution, $p(\boldsymbol{\theta}|\boldsymbol{y})$, is a conditional probability, it is a function of $\boldsymbol{\theta}$). In this way, taking $\boldsymbol{\theta}$ as the state space and augmenting it by a momentum variable $\boldsymbol{q}$, the Hamiltonian function can be written as

$$H(\boldsymbol{\theta}, \boldsymbol{q}) = E(\boldsymbol{\theta}) + K(\boldsymbol{q}),$$

where $K(\boldsymbol{q})$ is a kinetic energy like $K(\boldsymbol{q}) = \boldsymbol{q}^T\boldsymbol{q}/2$.

Because the Hamiltonian depends on two variables and HMC is a Metropolis algorithm (therefore it needs a proposal distribution to generate a candidate), there will be an alternation of two types of proposals, and, therefore, of two candidates. First, a sample is taken from the first proposal to obtain the momentum $\boldsymbol{q}$, leaving the parameter values $\boldsymbol{\theta}$ unchanged; then, we use the second proposal to obtain a candidate for both the momentum and parameter variables using the Hamiltonian dynamics as defined by $H(\boldsymbol{\theta}, \boldsymbol{q})$. Using these two candidates, we obtain a sample from the joint density

$$\begin{aligned} P_H(\boldsymbol{\theta}, \boldsymbol{q}) &= \exp(-H(\boldsymbol{\theta}, \boldsymbol{q}))/Z_H \\ &= \exp(-E(\boldsymbol{\theta}))/Z_E \exp(-K(\boldsymbol{q}))/Z_K, \text{ where } Z_H = Z_E \cdot Z_K, \\ &= P(\boldsymbol{\theta})P(\boldsymbol{q}); \end{aligned}$$

here Z_H, Z_E and Z_K are integration constants for the densities $p_H(\boldsymbol{\theta})$, $p(\boldsymbol{\theta})$ and $p(\boldsymbol{q})$, respectively. Because the canonical distribution $P_H(\boldsymbol{\theta}, \boldsymbol{q})$ can be written as the product of the canonical distributions $P(\boldsymbol{\theta})$ and $P(\boldsymbol{q})$, it is straightforward to recover the marginal density $P(\boldsymbol{\theta})$ and to obtain a sample for the parameters given the data.

The algorithm requires the following steps. A momentum is sampled from the normally distributed density $P(\boldsymbol{q}) = \exp(-K(\boldsymbol{q}))/Z_k$ with $\boldsymbol{\mu} = \mathbf{0}$ and $\boldsymbol{\Sigma}$ and the candidate is always accepted. The covariance matrix is set as $\boldsymbol{\Sigma} = \boldsymbol{I}$ because we consider the mass to be 1. For the second proposal, the momentum variables determine where the state $\boldsymbol{\theta}$ goes, and $\nabla E(\boldsymbol{\theta})$, which can be given explicitly or calculated using automatic differentiation,

determines how the momentum $\boldsymbol{q}$ changes according to the system

$$\dot{\theta}_i = q_i, \tag{3.1}$$

$$\dot{q}_i = -\frac{\partial}{\partial \theta_i} E(\boldsymbol{\theta}), \tag{3.2}$$

which represents the Hamiltonian dynamics. This system is usually solved using a finite-difference method known as the the leapfrog integrator, a numerical algorithm specifically designed to solve dynamical systems in classical mechanics. For each component of the position and momnetum variables i, it is defined as follows:

$$\begin{aligned} q_i\left(t + \frac{\epsilon}{2}\right) &= q_i(t) - \frac{\epsilon}{2}\frac{\partial}{\partial \theta_i} E(\boldsymbol{\theta}(t)), \\ \theta_i(t+\epsilon) &= \theta_i(t) + \epsilon\, q_i\left(t + \frac{\epsilon}{2}\right), \\ q_i(t+\epsilon) &= q_i\left(t + \frac{\epsilon}{2}\right) - \frac{\epsilon}{2}\frac{\partial}{\partial \theta_i} E(\boldsymbol{\theta}(t+\epsilon)). \end{aligned}$$

By applying l steps, a total of $l\epsilon$ time units are simulated for the leapfrog method, with the goal that this total time will be long enough to explore parameter space thoroughly but not so long that regions already explored are reconsidered. After solving the Hamiltonian dynamics system with the leapfrog method, we will have values for the candidate $(\boldsymbol{\theta^*}, \boldsymbol{q}^*)$, as is required for a Metropolis algorithm.

Because the leapfrog method, like any finite-difference approach to solving Eqs. 3.2, has numerical error, an additional step beyond generating a momentum vector is required. In particular, a Metropolis acceptance step is applied to take into account the numerical errors during integration, where the probability of accepting the candidate $(\boldsymbol{\theta}^*, \boldsymbol{q}^*)$ generated by transitioning from $(\boldsymbol{\theta}, \boldsymbol{q})$ is $\alpha = \min(1, \exp(H(\boldsymbol{\theta}, \boldsymbol{q}) - H(\boldsymbol{\theta}^*, \boldsymbol{q}^*)))$, which is the rejection rule from the Metropolis algorithm, and $\exp(H(\boldsymbol{\theta}, \boldsymbol{q}) - H(\boldsymbol{\theta}^*, \boldsymbol{q}^*))$ is 1 when there is no numerical error.

We note that a different kinetic energy can be used for HMC. The covariance matrix $\boldsymbol{\Sigma}$ of the normally distributed proposal for the momentum $\boldsymbol{q}$ acts as a Euclidean metric to rotate and scale the target distribution [3]; this matrix may be set to the identity matrix

or estimated from warmup draws [93] and optionally restricted to a diagonal matrix. The inverse $\mathbf{\Sigma}^{-1}$ is known as the mass matrix, and can be a diagonal matrix, a dense matrix if $\mathbf{\Sigma}$ is respectively a diagonal or a dense, or the identity matrix. Therefore, more generally, the kinetic energy can be defined as $K(\boldsymbol{q}) = \boldsymbol{q}^T \mathbf{\Sigma}^{-1} \boldsymbol{q}/2$.

The steps of the full HMC method are given below.

Hamiltonian Monte Carlo: Given $\boldsymbol{\theta} \in \mathbb{R}^n$ and the function E, calculate $e = E(\boldsymbol{\theta})$ and $\boldsymbol{g} = \nabla E(\boldsymbol{\theta})$.

1. Sample $\boldsymbol{q} \sim N(\mathbf{0}, I_n)$ and set $H = e + \boldsymbol{q}^T \boldsymbol{q}/2$.
2. Set $\boldsymbol{\theta}^* = \boldsymbol{\theta}$ and $\boldsymbol{g}^* = \boldsymbol{g}$.
3. Define ϵ and apply the leapfrog algorithm l times.
   (a) $\boldsymbol{q} = \boldsymbol{q} - \epsilon\, \boldsymbol{g}^*/2$
   (b) $\boldsymbol{\theta}^* = \boldsymbol{\theta}^* + \epsilon\, \boldsymbol{q}$
   (c) $\boldsymbol{g}^* = \nabla E(\boldsymbol{\theta}^*)$
   (d) $\boldsymbol{q} = \boldsymbol{q} - \epsilon\, \boldsymbol{g}^*/2$
4. Set $e^* = E(\boldsymbol{\theta}^*)$.
5. Set $H^* = e^* + \boldsymbol{q}^T \boldsymbol{q}/2$.
6. Set $dH = H^* - H$.
7. If $dH < 0$ or $u < \exp(-dH)$ where $u \sim U(0, 1)$,
   then set $\boldsymbol{g} = \boldsymbol{g}^*$; $\boldsymbol{\theta} = \boldsymbol{\theta}^*$; $e = e^*$.

3.4.2 Approximate Bayesian Computation (ABC)

When the likelihood $p(\boldsymbol{y}|\boldsymbol{\theta})$ is difficult to evaluate or it is not available, approximate Bayesian computation algorithms offer a useful solution. This method can also be used if a full Bayesian approach is feasible but computationally inefficient. The most basic approach is a simple rejection scheme that accepts only those candidates that produce

output sufficiently close to the data. The entire sample $\{\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}, \ldots \boldsymbol{\theta}^{(N)}\}$ is called a population.

ABC Rejection

Given a set of data $\boldsymbol{y}$, a system $\boldsymbol{x}' = f(\boldsymbol{x}, t|\boldsymbol{\theta})$, a distance function $\rho(\boldsymbol{x}, \boldsymbol{y})$, a tolerance $\epsilon > 0$, a desired population size N, and a chosen prior $p(\theta)$:

1. Set $i = 1$.
2. Sample $\boldsymbol{\theta}^*$ from $p(\boldsymbol{\theta})$ and solve $\boldsymbol{x}' = \boldsymbol{f}(\boldsymbol{x}, t|\boldsymbol{\theta}^*)$ to get $\boldsymbol{x}^*$.

   If $\rho(\boldsymbol{x}^*, \boldsymbol{y}) \geq \epsilon$, repeat this step.
3. $\boldsymbol{\theta}^{(i)} = \boldsymbol{\theta}^*$.

   if $i < N, i = i + 1$, go to step 2.

Approximate Bayesian Computation Sequential Monte Carlo

ABC rejection is computationally inefficient because its acceptance rate typically is very low. However, it is possible to build a sample iteratively by taking into account information from the previous population. One ABC method that does so is called ABC Sequential Monte Carlo (ABC SMC) [95]. Sample points are weight-sampled from the previous population and then perturbed using a random walk with a kernel K_t, which may be uniform or Gaussian.

ABC SMC

Given a set of data $\boldsymbol{y}$ and a series of tolerances $\epsilon_1 > \epsilon_2 > \ldots \epsilon_T \geq 0$, and a chosen prior $p(\boldsymbol{\theta})$:

1. Get a first population of size N with tolerance ϵ_1, $\{\boldsymbol{\theta}_1^{(i)}\}$ and assign weights $w_1^{(i)} = \frac{1}{N}$.
2. Weight-sample members from the previous population and then perturb each using a random walk with kernel K to obtain $\{\boldsymbol{\theta}_2^{(i)}\}$ with tolerance ϵ_2. Calculate weights as

$$w_t^{(i)} = \frac{p(\boldsymbol{\theta}_t^{(i)})}{\sum_j w_{t-1}^{(i)} K(\boldsymbol{\theta}_{t-1}^{(j)}, \boldsymbol{\theta}_t^{(i)})}.$$

3. Normalize weights and repeat the process until $\{\boldsymbol{\theta}_T^{(i)}\}$ is obtained with ϵ_T.

3.5 Implementation

As mentioned in section 3.4.1, HMC requires the two parameters l and ϵ, introduced by the use of the leapfrog integrator, to be tuned. Hoffman and Gelman [45] created the No-U-Turn sampler (NUTS) to avoid the need of tuning the forementioned parameters and it is implemented in Stan [12, 93], the statistical platform that we use for our study. In NUTS, l is determined adaptively at each iteration and there is also a procedure for adaptively setting both the mass matrix $\boldsymbol{\Sigma}^{-1}$ and the step size ϵ. HMC also requires calculating the gradient of the target distribution, which is done in Stan through automatic differentiation. Stan includes several useful tools to assess different characteristics of the obtained samples, like the convergence of the chains through the $\hat{R}$ statistic, or the tail and bulk effective sample size (ESS), to mention a few. In Stan, it is also possible to run more than one chain depending on the number of cores of the computer used to run the programs. As alternatives, PyMC, Pyro and JAGS have NUTS implementations. Stan also has its own solvers for differential equations, but we reduced the required computational time by writing our own solvers.

Chapter 4

Inference in cardiac action potential models using Hamiltonian Monte Carlo

4.1 Introduction

Fitting cardiac action potential (AP) models to reproduce the dynamics of experiments is a critical task for producing trustworthy model predictions. However, experimental data are accompanied by uncertainty and variability both within and across individuals. In general, models do not account for such uncertainty in parameter values and instead use various parameter-fitting methods to obtain a single value for each parameter being fit.

Modern statistical methods approach the problem differently by seeking to obtain a multivariate distribution of the parameter values consistent with the available data. Typically, Bayesian methods are used to iteratively evolve a population of parameterizations given by a prior to obtain the target posterior distribution conditioned on the data. Markov Chain Monte Carlo (MCMC) methods are a standard approach, but traditional MCMC approaches, like Metropolis-Hastings or Gibbs sampling, become computationally infeasible as the number of parameters increases beyond a small number. One alternative is the

use of approximate Bayesian computation (ABC) methods [21], where the computationally expensive step of calculating the likelihood is not needed and the target distribution obtained is an approximation of the true distribution being sought.

Although ABC approaches significantly reduce computation time compared to standard MCMC methods, it would be useful to obtain the true distribution while retaining computational efficiency. The Hamilton Monte Carlo (HMC) method [69] was developed for this purpose and it has been recently used to fit cardiac electrophysiology data [71]. Here we show that HMC can be used to produce a set of samples from the true distribution of parameter values for cardiac AP models in a manner that remains computationally efficient even with a moderate number of parameters.

4.2 Methods

As described in detail in section 3.3.1, we analyze synthetic and experimental data from zebrafish hearts at different CLs: 350, 300 and 276 ms, where the shortest two exhibit alternans. For the two sets of data, we chose a resolution that is sufficiently fine to capture the shape of the action potentials while keeping the computational time reasonable (computational time will be addressed in more detail in the next chapter). We used the MS and FK models to fit the data and verify how well the method along with the specific model was able to capture the dynamics of the system.

4.3 Results

Figure 4.1A-C show results for the synthetic dataset using the MS model. APs using the true values for the three CLs included in the dataset (350 ms, 300 ms, and 276 ms) are shown along with superimposed AP traces of 100 randomly selected parameter samples from the 500 samples generated using HMC. APs generated using those samples closely match the true APs well, with very little variation in AP shape and repolarization timing. A bifurcation plot using the true values for the MS model is shown in Figure 4.1D in black along with 100 bifurcation plots obtained using the same parameter samples as in the AP

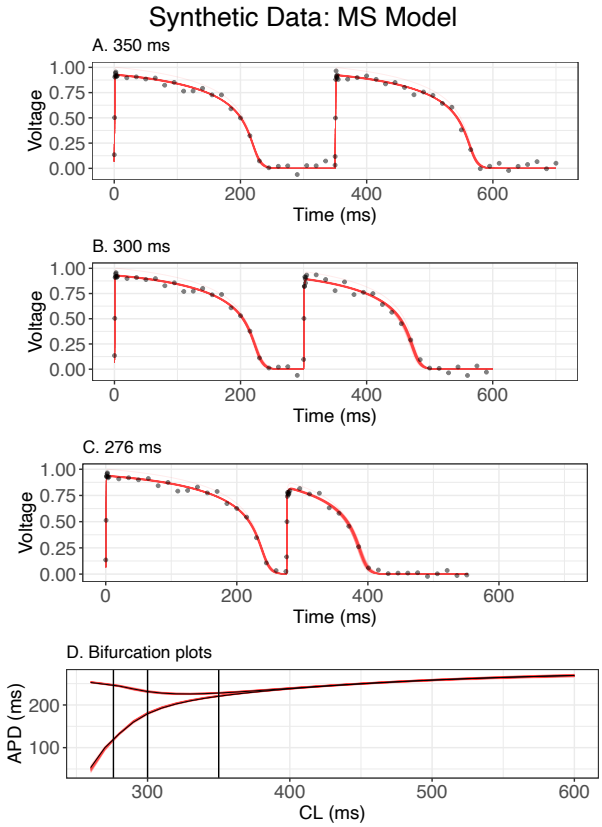


FIGURE 4.1: A-C: MS model-derived synthetic action potentials (black) and 100 action potentials obtained using MS model parameterization samples from the HMC population (red) for three CLs (vertical black lines). D: Bifurcation plots showing APDs as a function of cycle length for the synthetic data (black) and the same samples (red) as above.

plots (red); all cases considered a minimum CL of 150 ms. Little variability is observed for CLs that were not made available to HMC, including long CLs and short CLs past the bifurcation point.

The marginal distributions of the five MS parameters are shown in Figure 4.2. Each distribution is unimodal and tightly centered around values very close to the true parameter values; the use of the same model for generating the data and for fitting, together with the simple structure of the model, cause HMC to behave more like a parameter fitting-algorithm. In addition, the standard deviation of the Gaussian noise in the observations, as estimated through HMC, is quite small, with a mean of about 0.029. HMC is able to recover the true parameters with a high level of accuracy; the mean posterior predictive

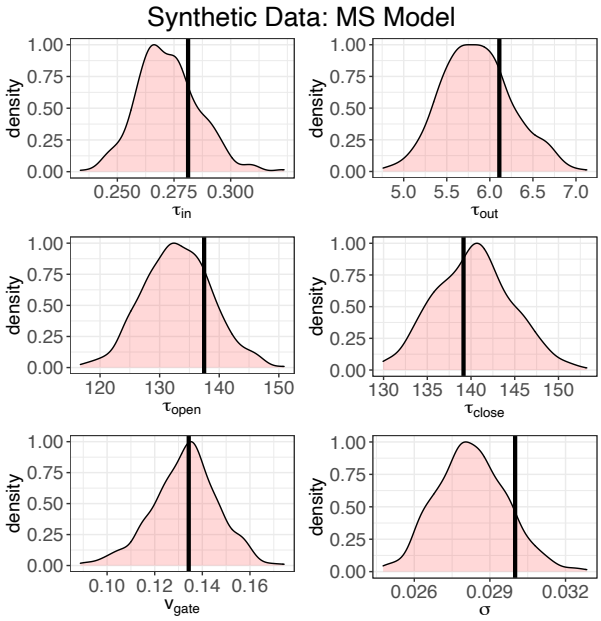


FIGURE 4.2: Distributions obtained for all five MS model parameters using HMC for the synthetic dataset along with the standard deviation σ of the noise. Vertical lines represent the true values used to generate the synthetic data.

values lie between the 2.5 and 97.5 percentiles. Thus, HMC is capable of recovering parameters for a model-derived dataset with a high degree of accuracy, even with multiple sources of noise included. The modes of the distributions displayed have a maximum error of 9%.

HMC's contributions beyond simple parameter fitting become more evident when considering experimental data, which may not be as well fit by the simple MS model. Figure 4.3A-C show zebrafish APs from three CLs (350 ms, 300 ms, and 276 ms) along with APs generated from the MS model using 100 of the 500 parameter samples generated using HMC. As in the synthetic data, alternans is present for CLs of 300 ms and 276 ms. Here good agreement is obtained but with greater variations; for example, repolarization times differ by as much as 30 ms. Figure 4.1D shows the bifurcation plot of the experimental data together with bifurcation plots obtained using the same parameterization samples (down to a minimum CL of 150 ms). Here again, greater variability can be seen than was observed for the model-derived dataset, especially at the longest and shortest CLs.

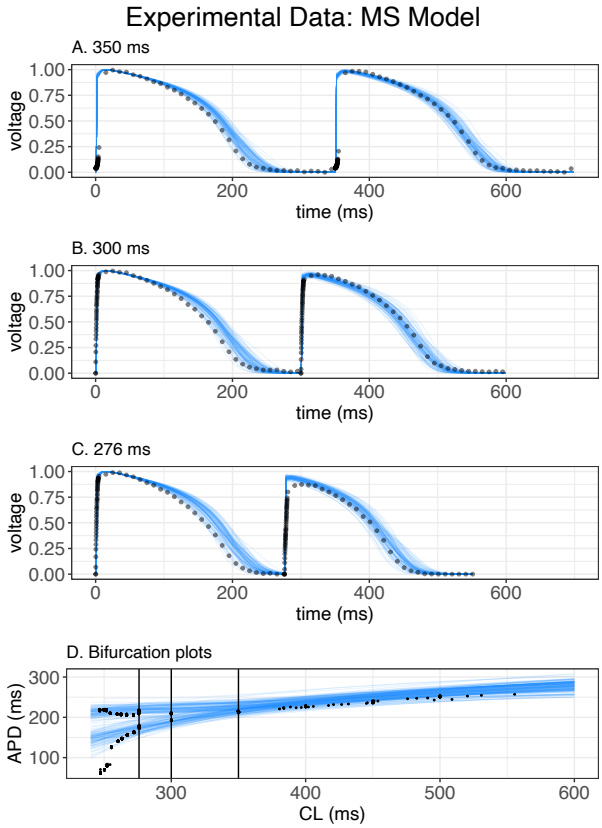


FIGURE 4.3: A-C: Normalized zebrafish action potentials (black) and 100 action potentials obtained using MS model parameterization samples from the HMC population (blue) for three CLs (vertical black lines). D: Bifurcation plots showing APDs as a function of CL for the synthetic data (black) and the same samples as above (blue).

Marginal distributions obtained using HMC for all five MS parameters are shown in Figure 4.4. As with the synthetic dataset, each distribution is unimodal, but now the values are spread across a broader range. In addition, the noise standard deviation σ is about an order of magnitude larger than it was estimated to be for the synthetic dataset.

We also considered use of the FK model within HMC for fitting an FK model-derived dataset with APs from three CLs, 350, 300 and 276 ms that included alternans (Figure 4.5), as well as the same zebrafish dataset (Figure 4.7). Results were similar, with errors in voltage traces obtained using the modes of the 13 parameter distributions less than 0.6 percent. Fitting the experimental dataset in HMC with the FK model also resulted in a low level of error of 3.4%. The distributions of the parameters can be seen in Figure 4.6 for the synthetic data and in Figure 4.8 for the experimental data.

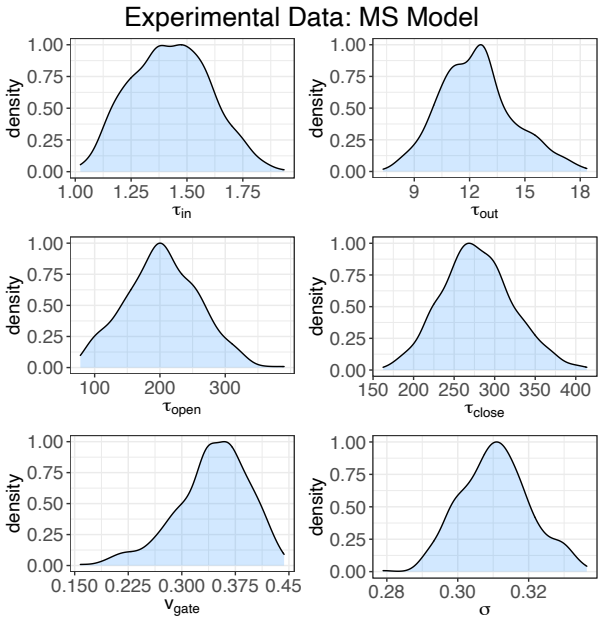


FIGURE 4.4: Distributions obtained for all five MS model parameters using the zebrafish data along with the standard deviation σ of the noise.

The main differences between using the MS and FK models can be seen when fitting the experimental data. The FK model was not able to fit the upper part of the upstroke compared to the MS model, especially for the CLs where alternans were present. However, the FK model was able to better capture the dynamics, as can be seen on the bifurcation plot in Figure 4.5.

4.4 Discussion

In this study, we have shown that HMC can produce probability distributions of model parameters that can reproduce the shapes and dynamics of cardiac APs, including during alternans, with good fidelity. When obtaining parameter distributions for a dataset derived from the same model used by HMC, narrow distributions that include the true values within the 2.5% and 97.5% quantiles were obtained, demonstrating the accuracy of the approach. For the experimental dataset, HMC also recovered unimodal distributions, but in this case they were much broader, reflecting a greater level of uncertainty. In addition, the dynamics obtained using the samples within the distribution showed greater variability, especially for CLs well beyond those included in the dataset when using the

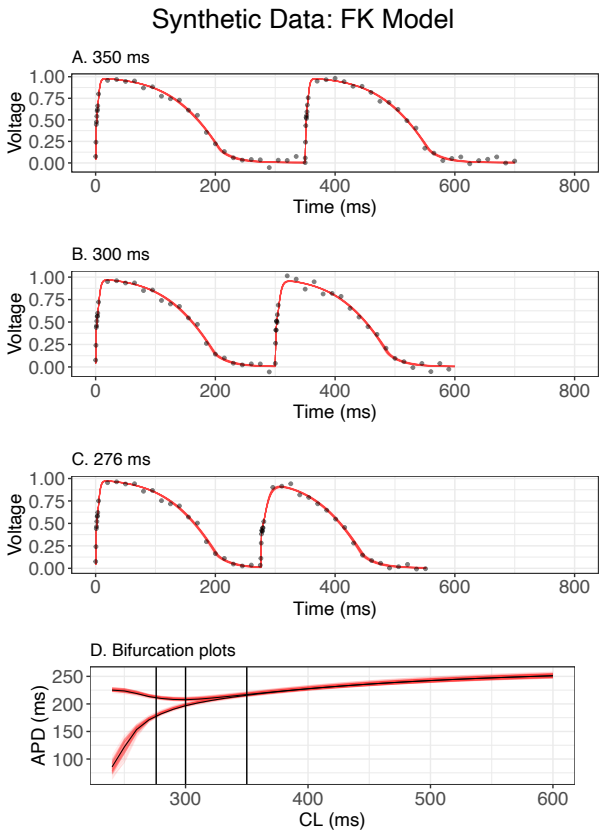


FIGURE 4.5: A-C: FK model-derived synthetic action potentials (black) and 100 action potentials obtained using FK model parameterization samples from the HMC population (red) for three CLs. D: Bifurcation plots showing APDs as a function of cycle length for the synthetic data (black) and the same samples (red) as above.

MS model. HMC also estimated the standard deviation of the noise in the datasets, which agreed with the known value for the synthetic data and was estimated about ten times larger for the experimental dataset when the MS model was used. We believe the larger standard deviation for the experimental dataset, despite its smoother appearance compared to the synthetic dataset, arises because the model is not a complete description of zebrafish AP dynamics.

Although HMC via the NUTS version implemented in the Stan platform was able to find populations of APs and BPs that in some cases were very close to the fitted profiles (so much so that for the synthetic data the profiles were completely superimposed), HMC is very sensitive to the priors used and the initial points specified for the algorithm. As will be discussed in the next chapter, we made use of an Approximate Bayesian Computation

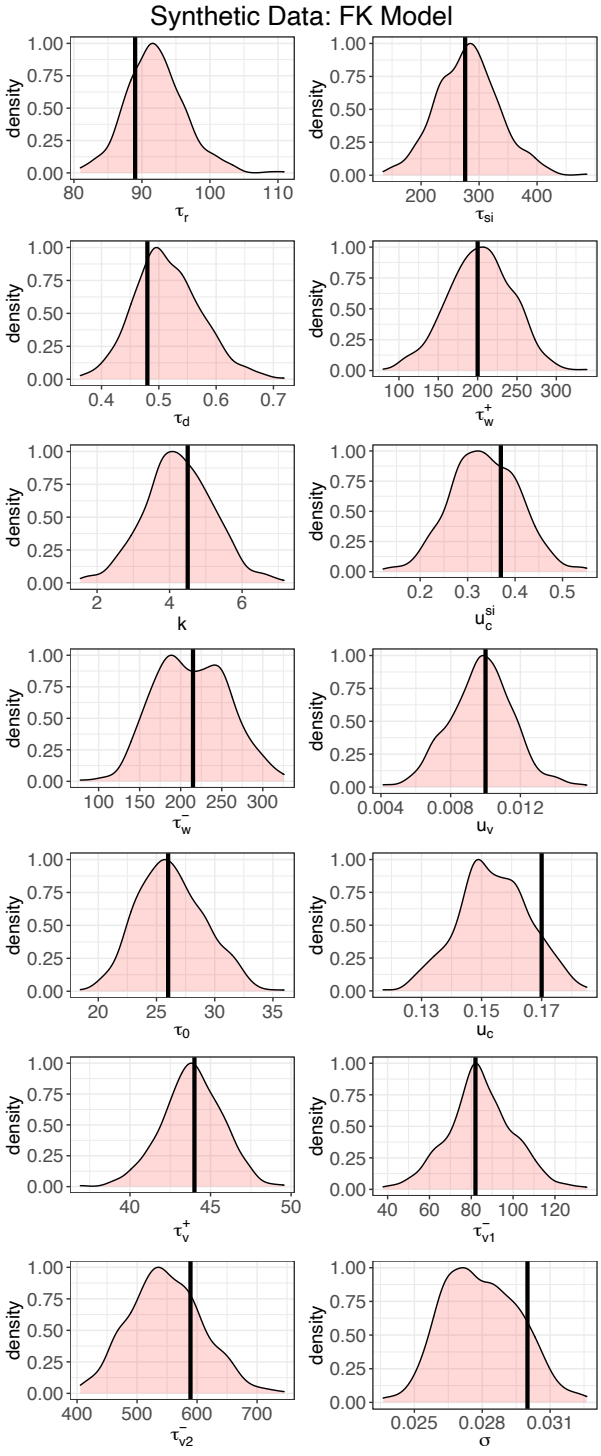


FIGURE 4.6: Distributions obtained for all 13 FK model parameters using HMC for the synthetic dataset along with the standard deviation σ of the noise. Vertical lines represent the true values used to generate the synthetic data.

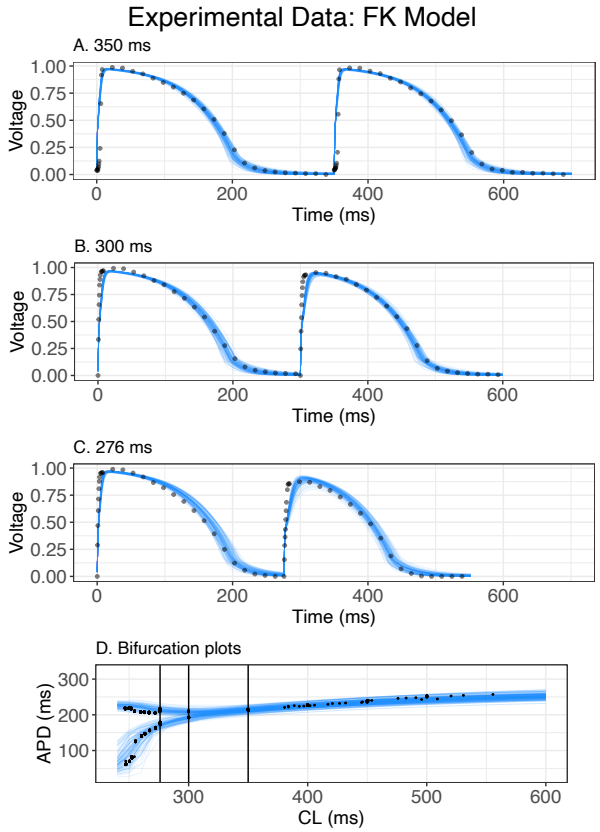


FIGURE 4.7: A-C: A-C: Zebrafish action potentials (black) and 100 action potentials obtained using FK model parameterization samples from the HMC population (blue) for three CLs. D: Bifurcation plots showing APDs as a function of CL for the synthetic data (black) and the same samples as above (blue).

algorithm to select the support of the priors. We also tried different priors (folded normal, gamma and uniform distributions), and tuning the algorithm for the chains to converge was time-consuming. For the different sets of data and models used, the fittings obtained using uniform priors were not satisfactory, because the chains did not converge or the fittings were far from the data to fit. We also found that in certain cases, the algorithm took longer to fit synthetic than experimental data. We will address this behavior in the next chapter. We aimed to apply HMC to describe variability in the shape and duration of synthetic and experimental APs among individuals and in space and time within a single individual, and we will present results of the latter using synthetic data in Chapter 6.

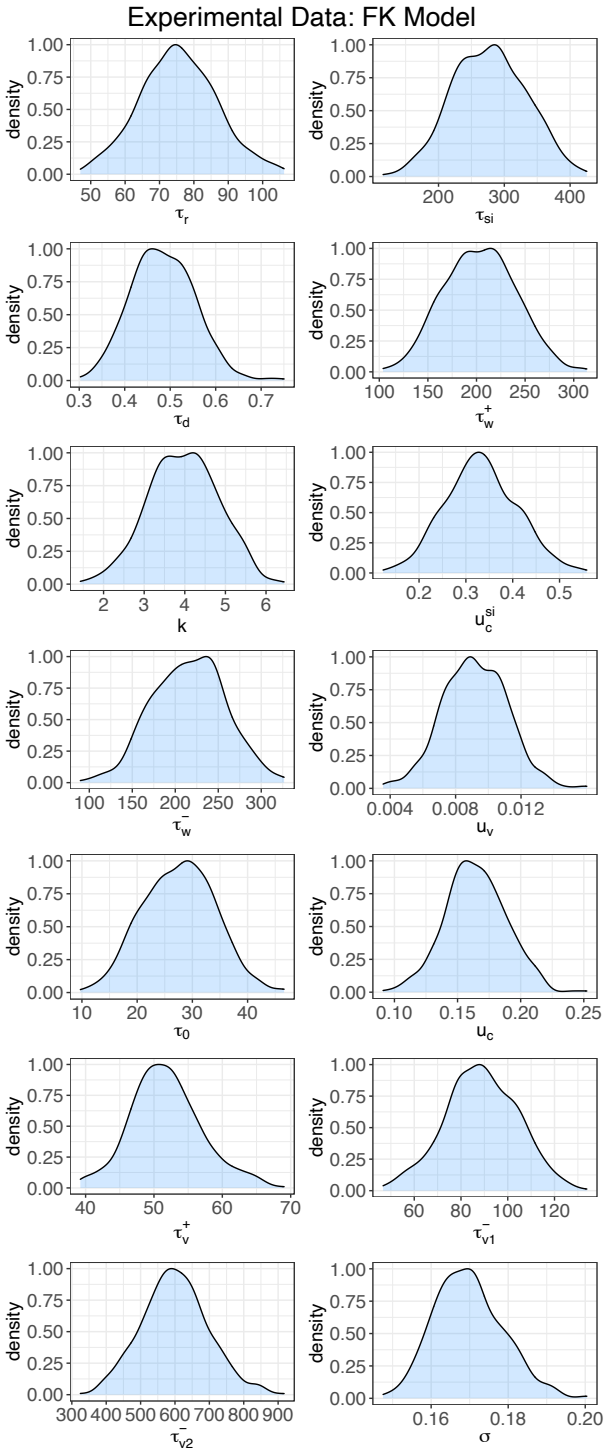


FIGURE 4.8: Distributions obtained for all five FK model parameters using the zebrafish data along with the standard deviation σ of the noise.

Chapter 5

Comparative study of Bayesian inference in cardiac action potential models

5.1 Introduction

In this chapter, we seek to obtain distributions of parameter values for cardiac electrophysiology models; these distributions can be used to represent variability and to diagnose parameter unidentifiability in models. Here, we build on the work of Chapter 4 by comparing the usefulness of the Hamiltonian Monte Carlo (HMC) algorithm and the Approximate Bayesian Computation Sequential Monte Carlo (ABC SMC) algorithm, which has been used to find distributions of a selection of parameter values for neural and cardiac models [22, 21, 23]. Specifically, we use both methods to find probability distributions for all parameters of two fairly low-dimensional cardiac action potential models: the Mitchell-Schaeffer model [66], which has two variables and five parameters, and the Fenton-Karma model [31], which has three variables and 13 parameters. We test these two Bayesian methods using synthetic data and experimental recordings from zebrafish hearts and compare their performance.

5.2 Implementation

This comparative study builds on the work described in the previous chapter of the first proof of concept for the application of HMC to cardiac model parameter estimation. However, SMC was used both for the present study as well as the previous study to obtain the appropriate support for the priors, as explained in the following sections.

5.2.1 HMC implementation

HMC was implemented in R through NUTS in Stan. A warm-up period of 1000 iterations was chosen. We point out that each time that HMC was applied, the $\hat{R}$ statistic in Stan was used to verify that the chains were converging to the target distribution [35, 98, 93]. A sample size of 500 was used for the posterior distributions; only 100 randomly selected members of the posterior sample were used to generate the population figures involving action potentials and APDs to improve clarity.

Because we found HMC to be quite sensitive to the prior distributions chosen, it was necessary to obtain somewhat informative priors in order to achieve reliable results. For synthetic data using the MS and FK models, the priors used were folded normal distributions extending 20% around the true values. Initial values needed for HMC were set as the modes of the priors. For experimental data, priors were obtained following a process that involved using ABC SMC which is more accepting of less informative priors, with wide uniform priors that contained but were not centered around the values used to generate the synthetic data, then using the modes of the resulting marginal posterior distributions as the values the HMC priors are centered around. The initial priors used by ABC SMC were defined on the intervals given in the second column of Table 3.3. The resulting priors used for HMC for the MS model with experimental data in the results to be shown were folded normals extending 20% around the values given in the third column of Table 3.3; these values also were used as the initial values for HMC, as indicated in the last column.

For the FK model, the process of obtaining the priors was the same as for the MS model,

with ABC SMC used with wide uniform priors to generate a more informative prior subsequently used by HMC. The initial uniform priors used by ABC SMC were defined on the intervals given in the second column of Table 3.4. The priors used for HMC with the FK model were folded normals extending 20% around the values given in the third column of Table 3.4, and the initial values were set to the values in the last column of the table.

5.2.2 ABC SMC implementation

The ABC SMC implementation we used was custom written in R and utilized the modifications made to the ABC SMC algorithm introduced by [95] and also found in [21], including adaptive tolerances and the use of the effective sample size for each population. In addition, to improve convergence and the exploration of the parameter space, we used a decreasing sequence of values for the scale factor for perturbing the populations. As suggested in [21], we used a probability density function as the distance function to generate the first population. Later populations were constrained to produce output closer to the target data using the sum of squared error by choosing a series of smaller tolerances. The tolerance reduction could be modified adaptively if found to be too restrictive and the algorithm stopped when the tolerance reduction was smaller than a specified value; see Ref. [21] for more detailed information.

The probability density function used as a distance to obtain the first population was $\rho_{\epsilon_1}(\theta) = \exp\left(\frac{-SE(\theta)/2(\epsilon_1\sigma)^2}{c_{\epsilon_1}}\right)$, where $SE(\theta)$ represents the sum of squared error between the data to fit and the model solution at each time point. Here $\epsilon_1 = 1$ is the first tolerance value, σ is the standard deviation of the error or noise assumed in the data measurements, and c_{ϵ_1} is a normalizing constant set to 1000 for all cases. The use of a probability density function can help in avoiding an overestimation of the variance in the first population that can occur otherwise. For all subsequent populations, the distance function was simply $SE(\theta)$. Once the first population was calculated, the value of the first tolerance was updated to the sum of squared error between the data and the model solution obtained using the modes of the distributions for each parameter, and the subsequent tolerance was

set to be $\frac{3}{4}$ of the updated first tolerance. The value by which the tolerance was reduced for subsequent populations was decreased by a factor of 0.5 but could be modified if too strict, following Ref. [21], and the minimum tolerance reduction serving as a stopping criterion was set to 1.5625×10^{-3}.

The population size chosen for ABC SMC was 500, ensuring a posterior distribution the same size as for HMC. As for HMC, a random selection of only 100 population members was used to generate the action potential and APD figures to improve clarity. For synthetic data using the MS and FK models, the priors used were folded normals extending 20% around the true values. As discussed in the previous section, it was necessary to obtain more informative priors for use with HMC in conjunction with the experimental data, and ABC SMC was used with an initial uniform prior to generate these priors. However, for the results shown below, ABC SMC uses the same priors as those generated for use by HMC to allow for a fair comparison.

5.2.3 Accuracy measurements

For the synthetic data, to show that the parameters could be recovered, we verified that the mean values for the posterior parameter distributions were between quantiles 10 and 90. Although with HMC the noise parameter can be estimated, ABC SMC makes no assumptions about the error; therefore, we also used the coefficient of determination $R^2 = 1 - \frac{SS_{res}}{SS_{tot}}$, where SS_{res} is the residual sum of squares (the same as $SE(\hat{\boldsymbol{\theta}})$, where the argument is the mean of the posterior distribution) and SS_{tot} is the total sum of squares comparing the data points with their average. This measure shows how well observed outcomes are replicated by the model, based on the proportion of total variation of outcomes explained by the model [17] (note that R^2 here does not have the same meaning it has for a linear regression problem).

5.3 Results

5.3.1 Synthetic datasets

First we consider parameter distributions for the MS model. Fig. 5.1 shows the marginal distributions of the five MS parameters and of R^2 across the final population for HMC (blue) and ABC SMC (red); the true values used to generate the original APs are shown as black vertical lines for comparison. The distributions for all parameters are fairly well centered around the true values for both methods, but the distributions obtained using HMC are narrower than those from ABC SMC. Note also that the shapes of the distributions can be quite different for the two methods; in particular, the distributions from ABC SMC are less symmetric than those from HMC. The R^2 plots, which reflect the distributions across the populations for each method, show consistently high values, above 0.983 for ABC SMC and above 0.99 for HMC. In addition, the final population for HMC has less variability in R^2 than that for ABC SMC. The insets correspond to q-q plots of the distributions obtained using ABC SMC vs. HMC, where the ranges for each axis are given by the corresponding x-axis limits. For most parameters, the q-q plots suggest that the posteriors come from the same distribution, even though the ABC SMC distributions in general are wider, indicating that ABC SMC likely explores more of the parameter space.

The population members in each case produce action potentials that agree well with those of the dataset for all three CLs used, as shown for a random selection of 100 of the 500 population members in the posterior samples for each method in Fig. 5.2A-C. Fig. 5.2D shows that the populations obtained from both methods agree well with the data at CLs not used during the fitting, including the location of the bifurcation to alternans. Some parameterizations from the ABC SMC population can be paced slightly faster than the true parameters before block occurs.

For the FK model, the results from the two methods show more variation. Although the marginal distributions for the 13 model parameters largely are centered around the true values, as shown in Fig. 5.3, the shapes of the distributions for each parameter

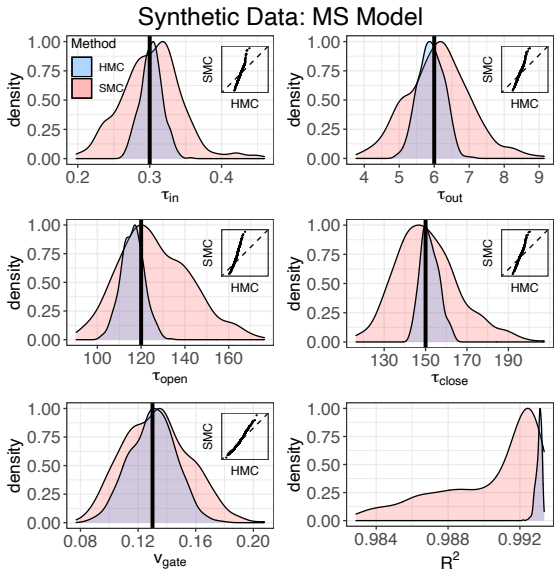


FIGURE 5.1: Marginal distributions for the MS model parameters along with the R^2 distribution when fitting to synthetic data (HMC, blue; ABC SMC, red). Vertical lines represent the true values used to generate the synthetic data. Insets show q-q plots for ABC SMC (vertical) vs. HMC (horizontal); dashed lines have slope one.

frequently are visibly different. Most of the marginal distributions obtained using ABC SMC are fairly broad, again suggesting greater exploration of parameter space. For some parameters, including τ_d, u_c^{si}, and τ_w^-, the distributions from HMC are narrow compared to ABC SMC, whereas for others the distributions are similarly broad. R^2 values also vary much more across the population for HMC than for ABC SMC, although even for ABC SMC the values are generally above 0.975. The q-q inset plots are still close to linear, even though they do not always have a slope of one, especially for τ_d, u_c^{si}, and τ_w^-, which feature wider ranges of values from ABC SMC compared to HMC. In the case of a few parameters like τ_w^+ and τ_0, the two methods produce distributions centered around different values, with the true value located between the peaks.

The fitted action potentials reflect the apparent increase in variability when using the FK model compared to the MS model. Fig. 5.4 shows a subsampling of the final populations obtained using each method and reflect the greater variation with the samples from ABC SMC. Nevertheless, good agreement is obtained across a wide range of CLs, even for CLs not close to those used for fitting. In particular, the bifurcation to alternans is well

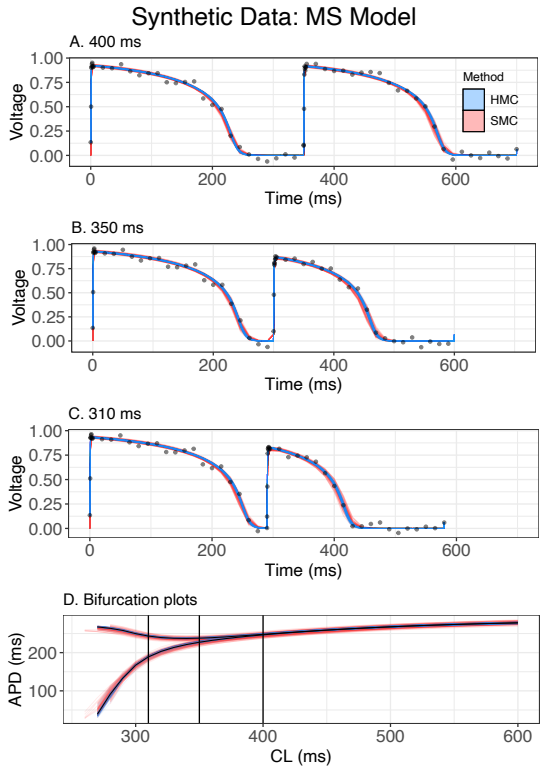


FIGURE 5.2: A-C. 100 action potentials from populations of size 500 using synthetic data for HMC (blue) and ABC SMC (red) compared with the data points for the MS model. CLs of (A) 400, (B) 350 and (D) 310 ms are included. D. 100 representative plots of APD as a function of CL for synthetic data with the MS model. CLs from 600 ms down to where block occurs are included, with HMC results in blue, ABC SMC results in red, and the true data points in black.

characterized for both methods.

5.3.2 Experimental dataset

For the experimental data, the results for the MS model show some differences compared to the synthetic data. As demonstrated in Fig. 5.5, the marginal distributions from HMC and ABC SMC are nearly identical for τ_{open}; for τ_{in} and τ_{out}, the distributions appear slightly offset relative to each other. For the remaining parameters, τ_{close} and u_{gate}, the distributions look similar to the results for the synthetic data, with a broader range of values included for ABC SMC. The agreement in marginal distribution shapes compared to the synthetic data case leads to linear q-q plots with slope values of nearly one in all cases. The R^2 values are lower when the data are not derived from the model being fit;

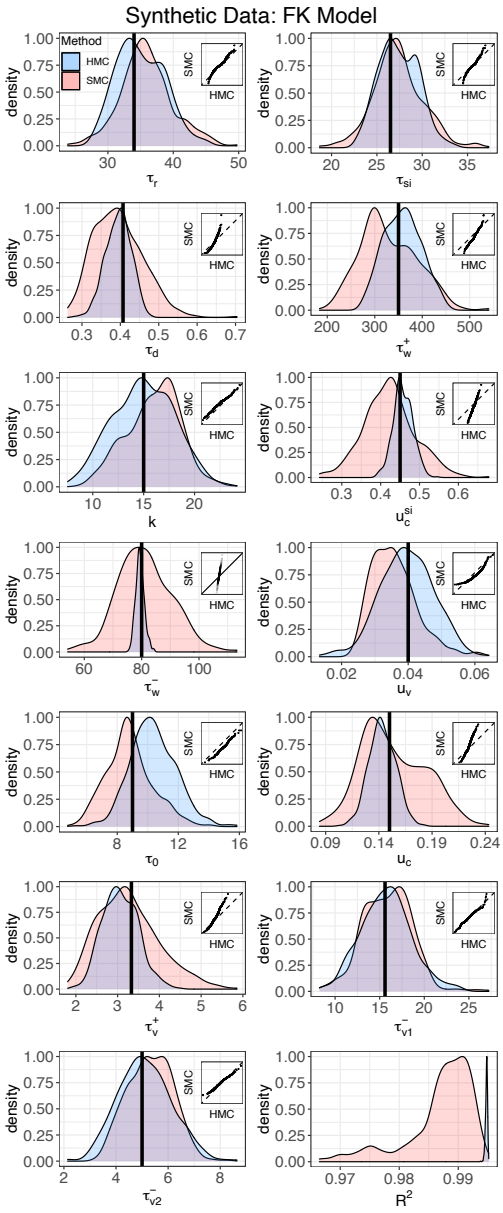


FIGURE 5.3: Marginal distributions for the FK model parameters along with the R^2 distribution when fitting to synthetic data (HMC, blue; ABC SMC, red). Vertical lines represent the true values used to generate the synthetic data. Insets show $q-q$ plots for ABC SMC (vertical) vs. HMC (horizontal); dashed lines have slope one.

however, R^2 remains above 0.65 for ABC SMC and above 0.7 for HMC.

Fig. 5.6A-B shows the fitting to experimental data for 100 of the 500 population members for two of the three CLs, 300 ms and 276 ms. Again, good agreement is seen overall, but with more variability in AP shapes and durations across the population with ABC

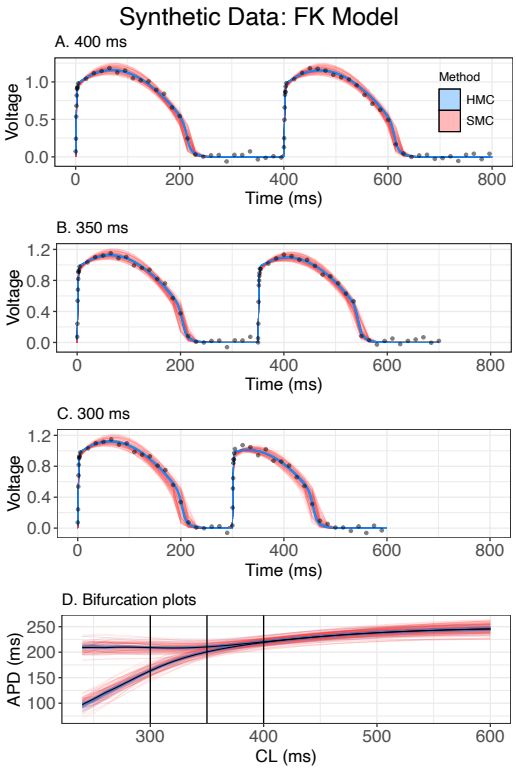


FIGURE 5.4: A-C. 100 action potentials from populations of size 500 using synthetic data for HMC (blue) and ABC SMC (red) compared with the data points for the FK model. CLs of (A) 400, (B) 350, and (C) 300 ms are included. D. 100 representative plots of APD as a function of CL for synthetic data with the FK model. CLs from 600ms down to where block occurs are included, with HMC results in blue, ABC SMC results in red, and the true data points in black.

SMC. Of particular note, the variability seen in the ABC SMC posterior sample includes longer and shorter APs, but the HMC sample is biased toward longer APs for the MS model. This trend can be observed across a broad range of CLs; HMC generally achieves longer APDs than occurred in the experiment. In addition, neither method accurately captures the dynamics associated with alternans, including the bifurcation point and the magnitude of the alternans at the shortest CLs.

For the FK model, the marginal distributions of the parameters show even more variability, as shown in Fig. 5.7, with wider ranges of values in many cases obtained using HMC. Although the distributions have similar peak values for some parameters, such as τ_r and τ_w^-, they appear shifted for others, such as τ_d, τ_w^+, and u_v. The q-q plots for these shifted cases look less than linear. Nevertheless, the R^2 values are somewhat higher than

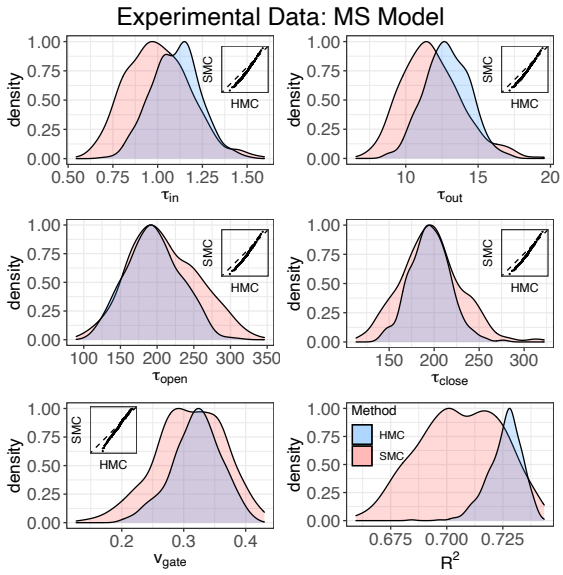


FIGURE 5.5: Marginal distributions for the MS model parameters along with the R^2 distribution when fitting to experimental data (HMC, blue; ABC SMC, red). Insets show q-q plots for ABC SMC (vertical) vs. HMC (experimental); dashed lines have slope one.

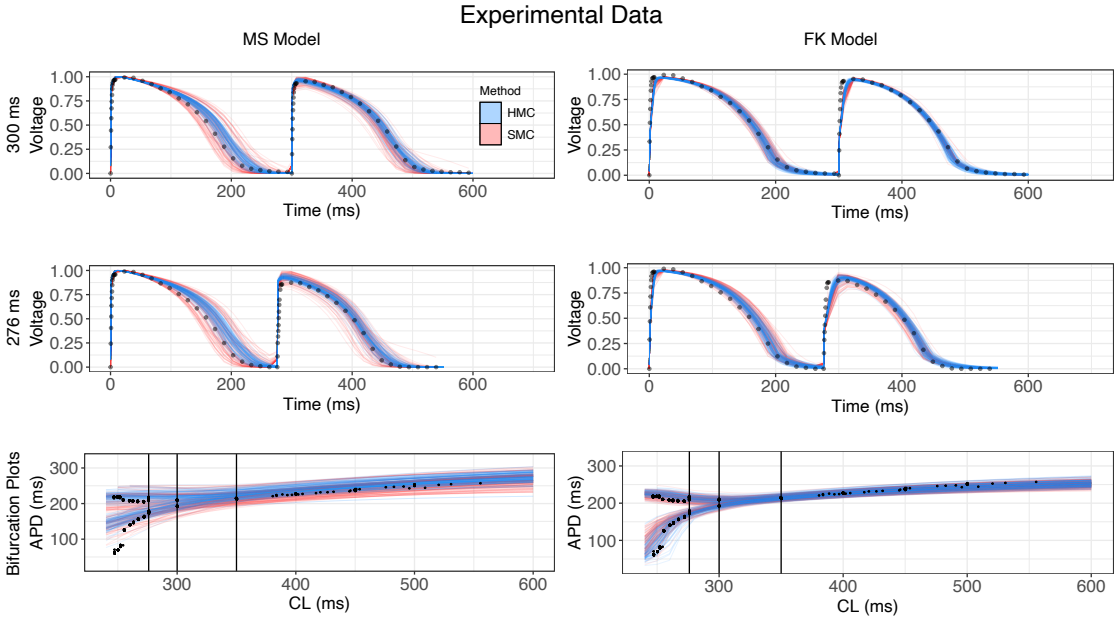


FIGURE 5.6: Fit action potentials and bifurcation plots for the MS model (left) and the FK model (right) using HMC (blue) and ABC SMC (red) for experimental data (shown in black). Top two rows show action potentials for CLs of 300 and 290 ms from a sample of 100 randomly selected population members. Bottom row shows APD as a function of CL for the same sample.

they were for the MS model for the same data. Despite this variability in the values of parameters, the fitted action potentials for the FK model closely fit the data, as shown in Fig. 5.6 (right column). Both methods show less variability across the sample for the FK model than for the MS model. However, there is a tendency for the upstroke to be fitted poorly for higher voltage values for CLs exhibiting alternans. Despite this fact, the dynamics overall, especially during alternans, are much better captured by the FK model than by the MS model.

5.3.3 Consistency and robustness

Both HMC and ABC SMC are dependent on randomization within their algorithms; as a result, it is possible that results could be different every time either algorithm is executed—for example, the algorithm could converge to a different posterior distribution. To assess whether the results are robust to different randomizations, we ran each algorithm ten times for each dataset and for each model. Fig. 5.8 summarizes the resulting distributions for the MS model for both the synthetic (left) and experimental (right) data. In both cases, the distributions obtained are fairly consistent for each method. For the synthetic data, the marginal distributions are unimodal and centered near the correct values. However, the marginal distributions for all time constants are wider for ABC SMC than for HMC; for v_{gate}, the distribution widths are similar for both methods. For the experimental data, there is even greater consistency across the ten runs for each method along with more similarity of the posterior distributions for the two methods. The distributions generally are wider with ABC SMC, but the difference in widths is less pronounced than with the synthetic data. Some differences in the locations of the peaks for τ_{in} and τ_{out} are evident, as discussed earlier when using experimental data for the MS model in Fig. 5.5.

For the FK model, more variation occurs for both datasets. As shown in Fig. 5.9, across the ten runs, the posteriors from HMC when fit to synthetic data are fairly consistent, but ABC SMC can yield distributions with more variation in shape and location. For example, the the peak location for the distributions of τ_r, τ_{si}, and τ_w^+ obtained using

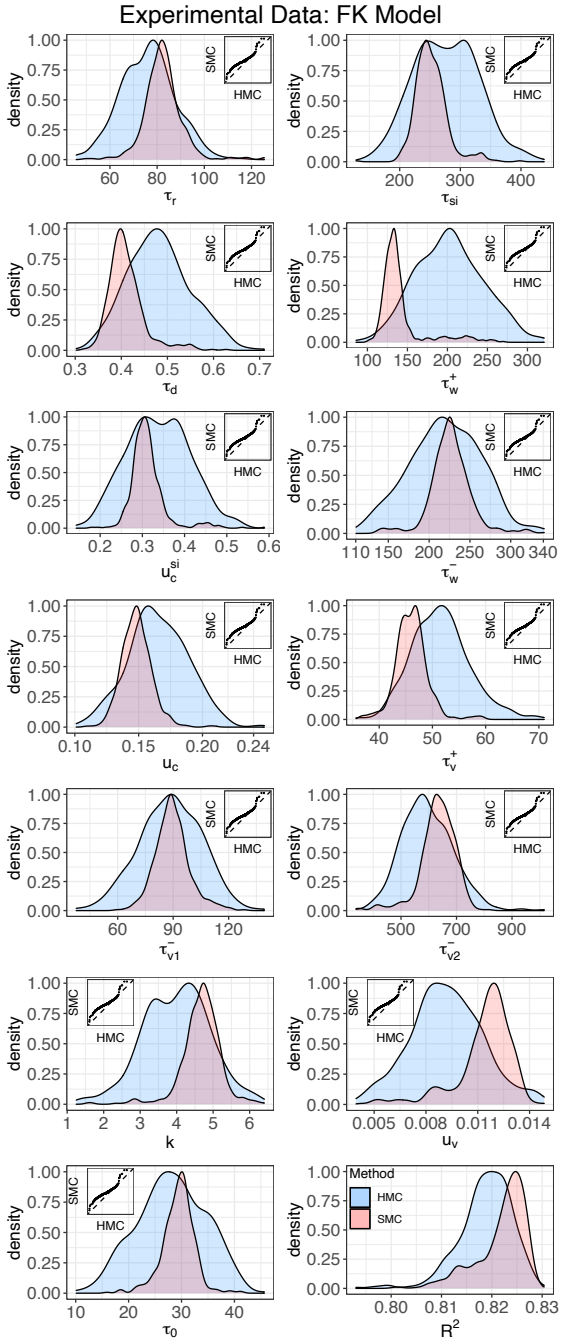


FIGURE 5.7: Marginal distributions for the FK model parameters along with the R^2 distribution when fitting to experimental data (HMC, blue; ABC SMC, red). Insets show q-q plots for ABC SMC (vertical) vs. HMC (horizontal); dashed lines have slope one.

ABC SMC vary substantially. For other parameters, such as τ_w^- and u_c^{si}, the results for HMC and ABC SMC have similar peak locations, but much broader distributions for ABC SMC. With experimental data, even greater variability across the ten distributions

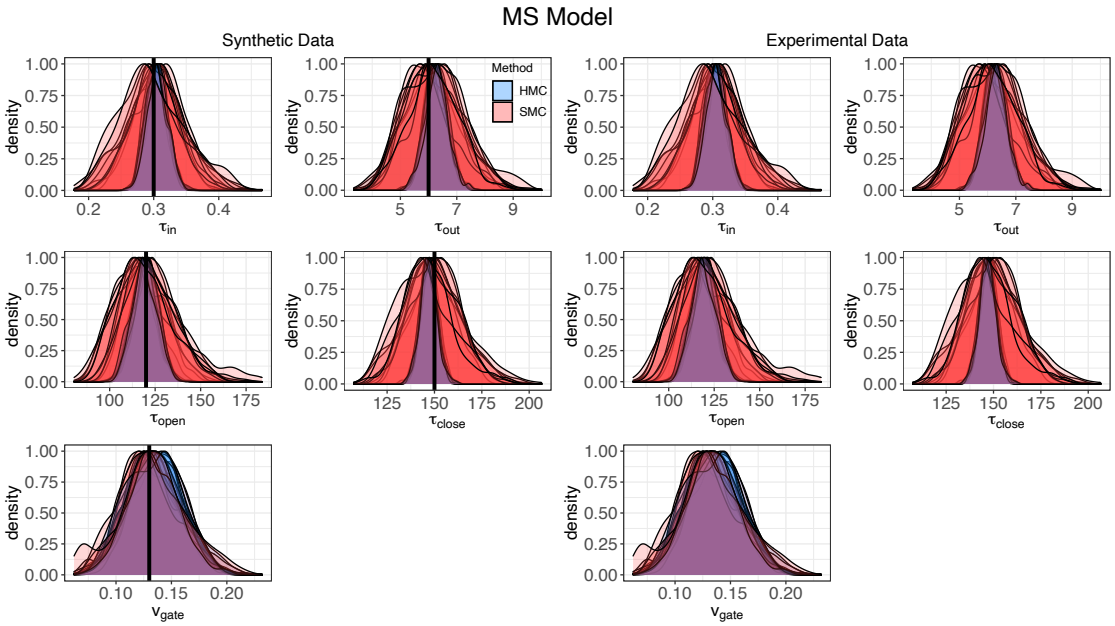


FIGURE 5.8: Superimposed marginal distributions for the MS model parameters when fitting to (left) synthetic and (right) experimental data (HMC, blue; ABC SMC, red) from 10 runs of each algorithm (HMC, blue; ABC SMC, red). Vertical lines represent the true values used to generate the synthetic data.

obtained using ABC SMC is evident, with noticeable peak shifts for nearly all parameters. HMC produces more consistent results, indicating that this method is less sensitive to randomization effects.

As a final measure of robustness, Fig. 5.10 shows the distribution of R^2 across the posterior sample for each of the ten runs using each method for the two models and the two datasets. In all scenarios, there is more variability in the R^2 value across the posterior sample for ABC SMC than for HMC, which achieves consistent narrow distributions. In the case of synthetic data and the ABC SMC algorithm, the distributions are wider for the FK model fit than for the MS model; the FK model also has lower and more variable R^2 values. HMC achieves high R^2 values for both models. In comparison with the results for synthetic data, the R^2 values for the experimental data are lower and the distributions are wider for all cases, but with ABC SMC distributions wider than those of HMC for both models. In contrast to the synthetic data case, however, the values of R^2 are higher for the FK model than for the MS model, a result that strongly suggests, consistent with

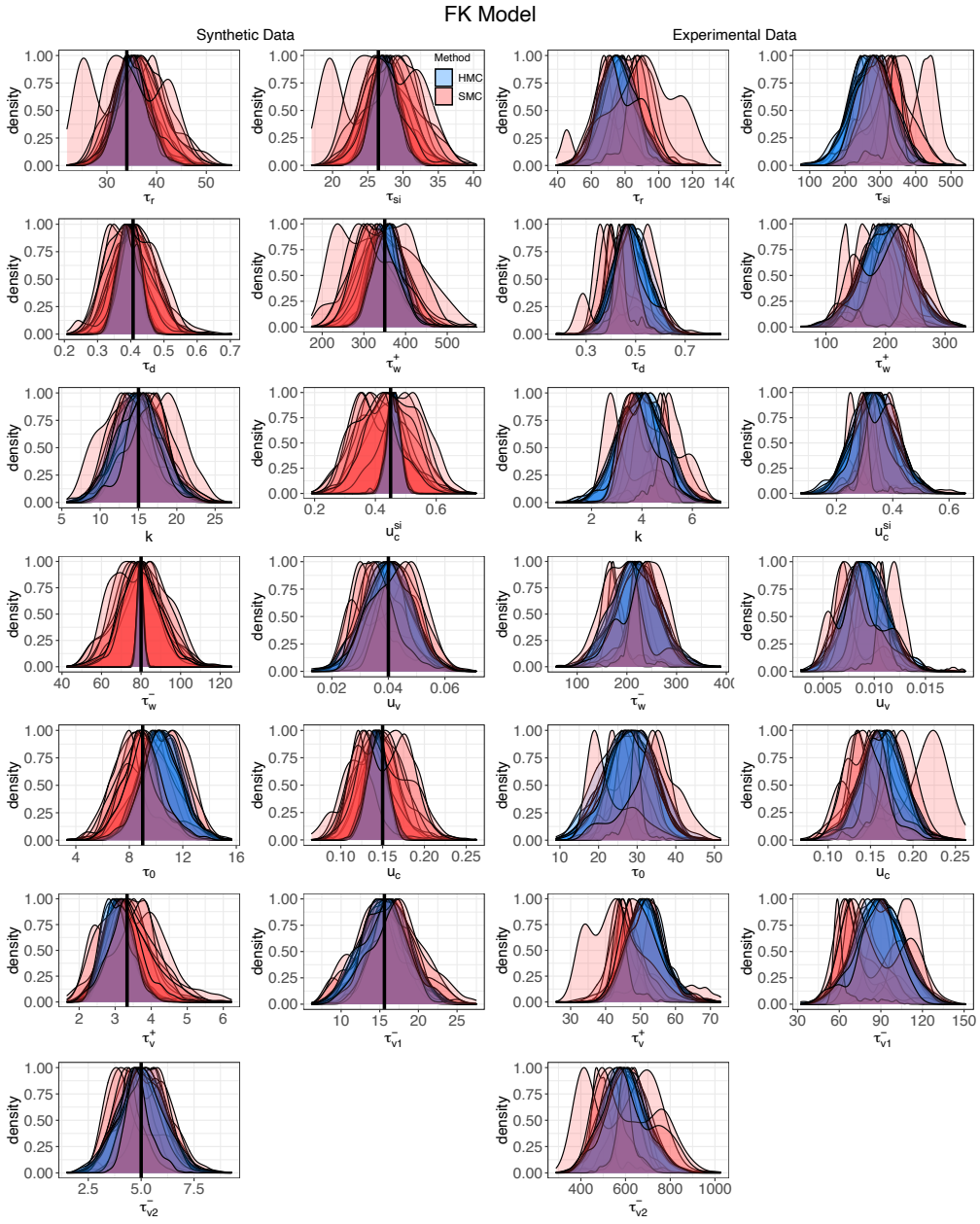


FIGURE 5.9: Superimposed marginal distributions for the FK model parameters when fitting to (A) synthetic and (B) experimental data (HMC, blue; ABC SMC, red) from 10 runs of each algorithm (HMC, blue; ABC SMC, red). For synthetic data, the vertical lines represent the true values.

the results shown in Fig. 5.6, that the FK model is better able to fit the experimental data than the MS model.

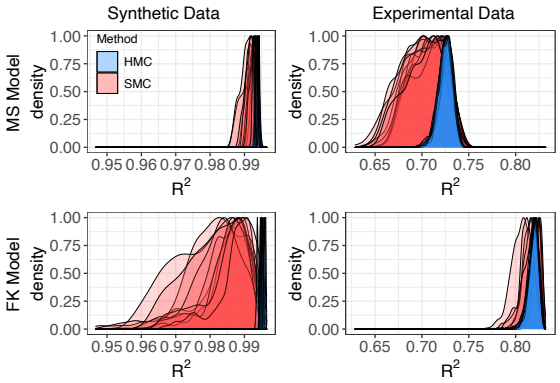


FIGURE 5.10: Superimposed distributions of R^2 from 10 runs of each algorithm (HMC, blue; ABC SMC, red) for the MS (upper) and FK (lower) models and for synthetic (left) and experimental (right) data.

5.4 Discussion

In this section, we compare results obtained for the different models, datasets, and Bayesian methods. We also describe some limitations of this study.

5.4.1 Influence of dataset and model choices

In all cases, the distributions found using HMC tended to be consistent, with more variability in the resulting distributions occurring with ABC SMC. For synthetic data, HMC often, but not always, gave rise to narrower distributions than ABC SMC, whereas for experimental data, HMC's distributions sometimes were wider in such a way as to cover more of the variability of the ABC SMC distributions.

With regard to models, we found clear differences in results for the two models considered. With synthetic data, for both methods, the peaks of the parameter distributions for the MS model usually coincided, with minimal shifting between the methods (despite variations in distribution widths). In contrast, for the FK model, greater variability arose, especially for the experimental data. The reduced clustering of ABC SMC distributions across runs for the FK model suggests the potential for identifiability issues, with the datasets not providing enough information to reliably constrain all parameter values for this model. Such behavior has been observed using approximate Bayesian methods previously [22, 24].

5.4.2 Bayesian method considerations

Overall, our results across the two methods were fairly consistent. In particular, the q-q plots of the ABC SMC vs. HMC samples do not support the idea that the posteriors obtained with the two algorithms come from different distributions other than the actual posterior. Nevertheless, each of the two methods considered has advantages and disadvantages. HMC performs exact inference and explores parameter space more efficiently than traditional methods like Metropolis-Hastings. Furthermore, HMC was consistent when running the programs multiple times, giving posteriors centered around the same value, for synthetic or experimental data. However, it can be difficult to use HMC because of the need to choose initial points and a prior distribution. Inappropriate selections for either initial points or the prior can affect convergence, but finding good choices can be time-consuming.

In contrast, although we show results only for the same priors used for HMC, ABC SMC could produce good results with a variety of different priors, including uniform, gamma and folded normal distributions. Even with relatively non-informative priors such as wide uniforms, ABC SMC was able to find a useful approximation to the posterior in all cases we tried, for synthetic or experimental data.

However, the lower bound for the R^2 distribution when fitting experimental data was smaller using wide uniform priors than folded normal distributions (MS: not lower than 0.48 vs. 0.65; FK: not lower than 0.67 vs. 0.79). While we expect that this result occurs because the uniforms are less informative than the folded normal distributions, more study would be needed to make a fair comparison. In addition to imposing less stringent requirements for the priors, ABC SMC could find an approximate population fit even if the population size was small (e.g., 100).

In terms of computational efficiency for the two methods, using experimental data with ABC SMC generally took around 5 minutes and more than twice as long for synthetic data when fitting the MS model. For the FK model, ABC SMC took only slightly longer to fit experimental data, around 6 minutes, and about 1.5 times longer than that to fit

synthetic data. For HMC, the experimental data could be fit to the MS model much more quickly, in about 3.5 minutes, with the fitting to synthetic data requiring about eight times longer. When fitting synthetic data, HMC took around 12 minutes to fit the MS model but around 4 hours to fit the FK model. The time it took for ABC SMC to fit synthetic data with the MS and FK models was respectively similar to the time it took fitting experimental data. A possible explanation for the long times required for HMC to fit synthetic data may be that the likelihood is very flat for some regions of parameter space, limiting choices for acceptable candidate parameterizations. Another possible reason that HMC is much faster for fitting experimental data may be the nature and extent of noise and variability in the dataset. Indeed, HMC is unable to obtain appropriate fittings for datasets with no or very low noise, which, along with the magnitude of noise in the experimental data, influenced our selection of the noise level in creating the synthetic datasets considered.

Overall, we found that ABC SMC was able to obtain good results under a broader range of conditions, whereas HMC imposed more constraints for reasonable performance. However, even ABC SMC saw benefit from the use of an informative prior. Because ABC SMC does not need initial points and accepts wide priors from several different types of distributions, it can be used to find feasible priors for HMC, and the initial points for HMC can be selected as the modes of the distributions obtained with the first ABC SMC pass. This hybrid approach, which was used to obtain the results shown here, can be useful to sidestep the difficulties of working with HMC while taking advantage of its ability to generate a population that closely fits the data with limited variability and of its performance of exact inference as a full Bayesian method.

5.4.3 Limitations

In this study we considered only a limited number of datasets. In particular, we chose data from three CLs, with one at a longer CL and two at shorter CLs within the alternans regime. We made a preliminary study to select a proper number of CLs that can be seen in Appendix A. It may be possible to optimize the selection of CLs beyond what was

chosen. In addition, it is possible that performance may change for noisier data or for data with different dynamics that may not be well captured by the model being fit.

Similarly, we only considered two models, and it is possible that performance could differ for a different selection of models. We also note that we have used a single approximate Bayesian method and that a different choice may result in different findings.

Within the models, we used a simple square-pulse stimulus. Use of a biphasic stimulus [11] could help prevent selection of large values for the excitability parameters (e.g., τ_{in} for the MS model and τ_d for the FK model) that would not allow propagation in tissue and thus may be unphysiological.

The tolerance reduction approach used for ABC SMC, while adaptive, nevertheless was fixed in advance, following Ref. [21]. It would be interesting to try a more sophisticated way to select the tolerances to improve efficiency and to facilitate working with different datasets.

We also note that to make the comparison between ABC SMC and HMC fair, we chose to use one chain for HMC in Stan, but we found the results were consistent when using the default number of chains, which was four.

5.4.4 Conclusion

In this chapter, we have used two Bayesian methods, HMC and ABC SMC, to find populations of cardiac action potential model parameters consistent with data used for fitting. We have shown that the methods can work effectively with both synthetic data derived from the models used as well as for an experimental dataset taken from a zebrafish heart. In nearly all cases, both methods find well-shaped marginal distributions with clear peaks for each model parameter for both the MS model, which has five parameters, and the FK model, which has 13. We also have shown through the use of q-q plots that the posterior distribution samples obtained by the two methods do not give any strong indication of being from different distributions; in other words, both methods appear to converge to the same distribution. In the case of synthetic data, where the true parameters

used to generate the dataset are known, those true values in general are well contained within the posterior distributions found, and across multiple runs of the algorithms the true values coincide well with the distribution peaks.

Given that both methods achieve similar results with no clear computational advantage, considerations other than accuracy may motivate the choice of method. ABC SMC is generally easier to work with, as it accepts different kinds of prior distributions, and those distributions may be broad. In addition, ABC SMC appears to explore the parameter space more fully than HMC. While HMC requires more effort to find an acceptable prior (and indeed we suggest that ABC SMC may be useful in this task), it tends to find narrower distributions, which may be advantageous in some cases.

Chapter 6

Gaussian process priors for studying spatial variability

6.1 Introduction

Cardiac cells exhibit variability in the shape and duration of the electrical signals that trigger muscle contraction in both space and time within a single individual. Some aspects of natural variation can be represented in a model through the use of nonstationary parameter sets that describe separate individuals. However, current models of cardiac action potentials avoid parameter estimation for a spatially varying set of parameters and are often employed using a single value for the entire spatial domain. It would be useful to be able to describe variability more systematically, and additional assumptions regarding the heterogeneity of model parameters in space or time may help provide a more faithful and less restrictive representation of the system of interest. Toward this end, in this chapter we aim to characterize admissible parameter sets that can represent variability variability of cardiac action potentials in space for one individual.

6.2 Methods

Before introducing the probability hierarchical model mathematically, we present an overview of the main concepts.

- To generate the synthetic data used as input for fitting, we create a two-dimensional grid of spatially correlated values for the MS model parameter τ_{in}. For a subset of grid locations chosen as training points, we generate a series of action potentials using the corresponding values of τ_{in}. The remaining parameters are based on spatially invariant fixed values but with spatially varying Gaussian noise, such that there are small variations in the values of the remaining parameters used to generate the training data but, unlike for τ_{in}, without spatial correlation. A small amount of Gaussian noise is added to the voltage data to represent measurement error. At the same time, to assess the performance of the technique, we also select a set of prediction sites where the values of τ_{in} are known from the initial setup and the remaining parameter values are set to the baseline spatially invariant fixed values used at the training locations, in this case without added noise. Action potential series plus Gaussian noise are also generated at the prediction sites for later comparison, and bifurcation plots of APD as a function of CL are also generated for later evaluation.

- After generating the action potential time series for the chosen training points, we fit the noisy voltage data and find the probability distributions for all of the MS model parameters at the training sites. To do so, we treat τ_{in} as a spatially correlated parameter; for this proof-of-concept study, the remaining parameters are treated as spatially varying but uncorrelated. The distributions are used to generate populations of action potentials that can characterize variability and uncertainty at the training sites.

- As a final step, we use the information obtained at the training sites to make predictions at the prediction sites. We obtain values for τ_{in} at each prediction site using the (posterior) distributions of τ_{in} at the training sites based on the assumption that the parameter values are correlated in space along with the training and predicted locations. For the remaining parameters, we use the spatially varying values at the testing sites to generate a single tissue-level distribution for each parameter that is then used for each prediction site. Because our statistical model allows us to use

spatially varying distributions of a parameter to obtain a single tissue-level distribution for that parameter, the model can be considered to include different "levels" (e.g., cell level vs. tissue level) arranged hierarchically—hence the description as a hierarchical model. Using the spatially varying distributions for τ_{in} for each site and the tissue-level distributions for the remaining parameters, populations of action potentials are generated at each of the prediction sites. Accuracy is assessed at each prediction site by comparing the predictions of τ_{in}, the population of action potentials, and the population of bifurcation plots with the corresponding reference data.

In the sections below, we describe how the above concepts are realized.

6.2.1 Hierarchical probability model

Given the system defined by the MS model, which can be written as

$$\boldsymbol{x}' = f(\boldsymbol{x}, t; \tau_{in}, \tau_{out}, \tau_{open}, \tau_{close}, u_{gate}),$$

where $\boldsymbol{x}(t) \in R^2$ and $t \in \mathbb{R}^+$, we consider

$$y_{ij} = u(t_{ij}; \tau_{in}^{(i)}, \tau_{out}, \tau_{open}, \tau_{close}, u_{gate}) + \epsilon_{ij}$$

where $i = 1, 2, \ldots, N_1$ indexes the time series of voltage data to be fitted at each of the N_1 positions $\boldsymbol{s}_i$, $u = x_1$ is the solution to the MS model representing the voltage, $j = 1, \ldots, T$ indexes the T time points available for each position, and the measurement at each location at each time has measurement error represented by $\epsilon_{ij} \sim \mathcal{N}(0, \sigma_i)$.

We define $\boldsymbol{\lambda}^{(i)} = (\tau_{out}^{(i)}, \tau_{open}^{(i)}, \tau_{close}^{(i)}, u_{gate}^{(i)})$ and τ_i as the parameter τ_{in} at each position $\boldsymbol{s}_i$ in a spatial domain $\mathcal{D} \subset \mathbb{R}^2$. As a small simplification, we assume that the measurement error depends only on position and is independent of time ($\epsilon_i \sim \mathcal{N}(0, \sigma_i)$). Let $\boldsymbol{U}$ be a

matrix of size $N_1 \times T$ where row i is of the form

$$(u(t_{i1}; \tau_i, \boldsymbol{\lambda}^{(i)}, \sigma_i), \ldots, u(t_{iT}; \tau_i, \boldsymbol{\lambda}^{(i)}, \sigma_i))$$

We formulate the inference problem as a hierarchical Bayesian model for the spatiotemporal voltage measurement y_{ij}. Our strategy is to instantiate one separate instance of the MS model at each location and use it to relate spatial variations in parameter values to resulting outputs in spatiotemporal voltage signals. For the parameter τ_{in} of the MS model, we use a Gaussian process prior to encode spatial autocorrelation among the values of $\boldsymbol{\tau} = (\tau_1, ..., \tau_{N_1})$ used to parameterize the individual AP models at locations $\boldsymbol{s}_1, ..., \boldsymbol{s}_{N_1}$. The specifications for the model parameters are summarized as

$$\begin{aligned}
\epsilon_i &\sim \mathcal{N}^+(0, 0.5) \\
\mu_k &\sim \mathcal{N}^+(\mathrm{tv}(\lambda_k^{(i)}), 0.2 * \mathrm{tv}(\lambda_k^{(i)})),\ k = 1, 2, 3, 4 \\
\lambda_k^{(i)} &\sim \mathcal{N}^+(\mu_k, 0.2 * \mathrm{tv}(\lambda_k^{(i)})),\ k = 1, 2, 3, 4 \\
\boldsymbol{\tau} &\sim \boldsymbol{\mathcal{N}}(\boldsymbol{\mu}^{(1)}, \boldsymbol{K}_{\alpha,\rho} + \sigma_{GP} \boldsymbol{I}_{N_1}) \\
\alpha &\sim \mathcal{N}^+(0, 1) \\
\rho &\sim \mathrm{Uniform}(1, b), \\
\sigma_{GP} &\sim \mathcal{N}^+(0, 1),
\end{aligned}$$

where $\mathcal{N}^+$ represents a folded normal distribution, $\mathrm{tv}(\cdot)$ is the value used to generate the synthetic data (*true value*), $\mathbb{I}_{N_1}$ is the $N_1 \times N_1$ identity matrix, and b is the maximum distance between any two positions on the grid (those in opposite corners of the grid). A schematic of the hierarchical model can be seen in Fig. 6.1 in plate notation [10]. The plate notation is used to represent the hierarchy of the parameters, meaning for instance, that the higher level represents the parameters for the population of positions as a whole (outside the bigger square) and the next level represents the parameters for each member

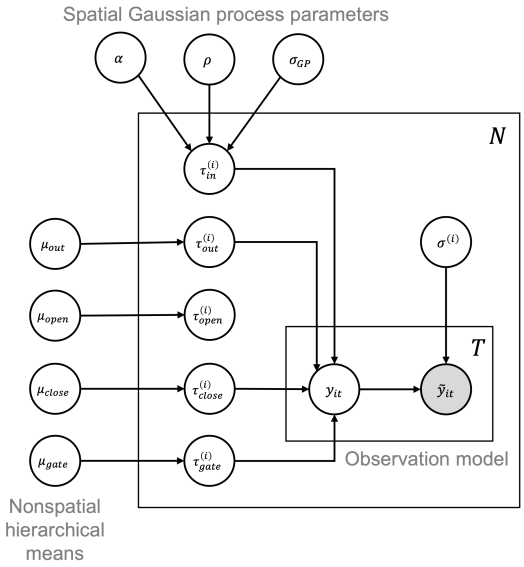


FIGURE 6.1: Probability model in plate notation. Circles indicate random variables while boxes (plates) indicate repetitions. We consider N spatial points, each of which has T observations associated with it. Hollow circles are unobserved or latent, while filled-in circles correspond to observational data.

of the population (inside the square). If we were to consider more than one individual, we could include a new level which would the the highest.

More specifically, the Gaussian process prior is

$$\tau(\boldsymbol{s}) \sim \mathcal{GP}(m(\boldsymbol{s}), C(\boldsymbol{s}, \boldsymbol{s}'; \rho, \alpha)), \tag{6.1}$$

where $m(\boldsymbol{s}) : \mathbb{R}^2 \to \mathbb{R}$ denotes a mean function and $C : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$ denotes a suitable covariance kernel function parameterized by a spatial correlation distance parameter ρ and a scale parameter α. We specify the mean function as a constant with a value of $m(\boldsymbol{s}_i) = 0$ for all $i = 1, 2, \ldots, N_1$. Then, the covariance function gives rise to the covariance matrix K via the construction

$$\boldsymbol{K}_{\alpha,\rho} = \begin{bmatrix} C(\boldsymbol{s}_1, \boldsymbol{s}_1) & C(\boldsymbol{s}_1, \boldsymbol{s}_2) & \cdots \\ \vdots & \ddots & \\ C(\boldsymbol{s}_1, \boldsymbol{s}_{N_1}) & & C(\boldsymbol{s}_{N_1}, \boldsymbol{s}_{N_1}) \end{bmatrix}. \tag{6.2}$$

In this work, we use the squared exponential covariance function, defined as

$$C(\boldsymbol{s}, \boldsymbol{s}'; \rho, \alpha) = \alpha^2 \exp\left(\frac{||\boldsymbol{s} - \boldsymbol{s}'||_2^2}{2\rho^2}\right), \tag{6.3}$$

where $||\boldsymbol{s} - \boldsymbol{s}'||_2$ denotes the Euclidean distance between two locations $\boldsymbol{s}$ and $\boldsymbol{s}'$. The role of the correlation distance parameter ρ is to decide roughly how far spatial correlations should extend, while the parameter α determines the relative scale of that variation. Alternative choices of covariance kernels include the Mátern or rational quadratic kernels. If additional spatial structure such as oscillations or trends in space are present, sinusoidal and linear kernels could also be employed. We note that this is a standard presentation of Gaussian process priors; we refer the reader to [81] for more detail. It is important to note that calculating the prior density associated with $\log \boldsymbol{\tau}$ requires inversion of the pairwise covariance matrix $\boldsymbol{K}_{\alpha,\rho}$, which can be numerically unstable. To circumvent this issue, we make use of the Cholesky decomposition $\boldsymbol{K}_{\alpha,\rho} = \boldsymbol{L}\boldsymbol{L}^T$ within `Stan`.

In total, the vector of free parameters $\boldsymbol{\theta}$ for the probability model for all N_1 spatial locations is

$$\boldsymbol{\theta} = (\boldsymbol{\tau}, \boldsymbol{\lambda}, \boldsymbol{\sigma}, \alpha, \rho, \sigma_{GP}),$$

where $\boldsymbol{\sigma} = (\sigma_1, ..., \sigma_{N_1}^2)$ denotes a vector of measurement noise variance parameters concatenated together and $\boldsymbol{\lambda} = (\boldsymbol{\lambda}^{(1)}, ..., \boldsymbol{\lambda}^{(N_1)})$. Then, with the abbreviation $[\cdot] = p(\cdot)$, the prior distribution as described previously factorizes as

$$p(\boldsymbol{\theta}) = \underbrace{[\boldsymbol{\tau}|\alpha, \rho, \sigma_{GP}]\,[\alpha]\,[\rho]\,[\boldsymbol{\sigma}]}_{\text{Spatial parameters}} \underbrace{\prod_{k=1}^{4}\prod_{i=1}^{N_1}\left[\boldsymbol{\lambda}^{(i)}\middle|\mu_k\right]\prod_{k=1}^{4}[\mu_k]}_{\text{Nonspatial hierarchical prior}},$$

where the index $k \in 1, ..., 4$ runs over the non-spatially varying parameters. Finally, by assuming independent Gaussian measurement error variates that are identically distributed

within spatial locations, the likelihood and its log-transformation are

$$p(\boldsymbol{Y}|\boldsymbol{\theta}) = \prod_{i=1}^{N_1}\prod_{j=1}^{T} \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left[\frac{-(y_{ij}-U_{ij})^2}{2\sigma_i^2}\right], \tag{6.4}$$

$$\log p(\boldsymbol{Y}|\boldsymbol{\theta}) = \frac{1}{2}\sum_{i=1}^{N} \frac{||y_{ij}-\boldsymbol{U}_i||_2^2}{\sigma_i^2} - \log\left(2\pi\sigma_i^2\right), \tag{6.5}$$

where $\boldsymbol{Y}$ represents the voltage data to fit and the L^2 norm as $||\cdot||_2$.

TABLE 6.1: Notation

$\boldsymbol{s}_1, ..., \boldsymbol{s}_N$	Spatial coordinates of observational data
t_{ij}	Temporal coordinates for observational data
N	Number of observation sites indexed by spatial coordinates
N_1	Number of training locations
N_2	Number of predicted locations
T	Number of time points for each voltage signal
U	$N \times T$ matrix of voltage values observed over space and time
$\boldsymbol{\tau}$	Vector of site-specific parameters of MS
$\boldsymbol{\lambda}^{(i)}$	Vector of non-site parameters of MS at position i
μ_k	Hierarchical mean of Mitchell-Schaeffer parameters pooling across locations

6.2.2 Parameter estimation

To estimate the parameters associated with the probability model of the previous section, we make use of Markov chain Monte Carlo methods for drawing samples from the posterior distribution with density $p(\boldsymbol{\theta}|\boldsymbol{Y}) \propto p(\boldsymbol{Y}|\boldsymbol{\theta})\, p(\boldsymbol{\theta})$. As is standard in the calculation of Monte Carlo estimators for Bayesian inference, we construct a Markov chain $\boldsymbol{\theta}^{(1)}, ..., \boldsymbol{\theta}^{(j)}$ of J samples drawn from the posterior distribution, with the posterior mean $\sum_{j=1}^{L} \boldsymbol{\theta}^{(j)}$ serving as our estimator of choice. In particular, we use a variant of Hamiltonian Monte Carlo as described in section 3.4.1, a method that constructs a fictitious Hamiltonian system for which the canonical coordinates are given by the probability model parameters, and the energy is given by the negative log posterior.

6.2.3 Synthetic data generation

We generate synthetic data using MS under the assumption that the parameter $\log \boldsymbol{\tau}$ is highly spatially correlated. In experiments, for an image obtained with optical mapping at a fixed time, each location is represented by a pixel. In reality, the locations are correlated because the electrical signals in the heart propagate by diffusion from one myocyte to the next. We fit the MS model to the last two APs in a series of 6 at N_1 locations that are represented on the grid by the vector $\boldsymbol{\tau}$. The rest of the parameters are not fixed but do not vary in space.

We generate spatial grids of $N = 25, 100$ and 900 locations with coordinates residing in the corresponding sets $\{1, ..., 5\}^2, \{1, ..., 10\}^2, \{1, ..., 30\}^2$. Each coordinate is assigned a raw independent standard normal noise variate; we then apply a Gaussian spatial filter to smooth the locations in space and thereby induce spatial correlations in the values of $\boldsymbol{\tau}$. We apply an affine transformation to the grid values in order to restrict them to be in the interval (0,1). Schematics for these setups are shown in Figures 6.2, 6.4 and 6.6. We then select $N_1 = 4, 6$, and 14 training locations, respectively, and drew values for the other model parameters by applying Gaussian noise, with σ set to 10% of the values of $\tau_{out}, \tau_{open}, \tau_{close}, u_{gate} = (6, 150, 120, 0.13)$. These values, including those used to generate the grid for τ_{in}, are based on those of the original MS model paper [66] (where τ_{in}=0.3).

Because of the alternans behavior, discussed in section 2.2.1, more than one CL is needed when fitting cardiac models to AP data. Therefore, we simulate voltage time series for CLs paced at $350, 300$ and 276ms at the N_1 training locations, then perturb them with Gaussian noise with mean 0 and standard deviation 0.03. Our objective with this synthetic data exercise is to determine whether or not the spatial parameter surface for $\boldsymbol{\tau}$ along with the system dynamics could be recovered given a limited number of voltage time series. The last two APs are taken in a series of 6—after reaching steady state—and the fitting is made using only the two shortest CLs from the alternans regime and the shortest CL before reaching the bifurcation. The resolution of the points selected from the APs to be fitted is 0.5 ms for the first 4 ms and 10 ms thereafter for each CL.

To summarize, we use a latent Gaussian process as a prior for the spatially varying parameter $\boldsymbol{\tau}$ and simulate data drawn from this prior. We then perform inference for a subset of $N_1 = 4, 6$ or 14 sampled locations—used as training points—and attempt to respectively predict the values of τ_{in} at the remaining $N_2 = 21, 94$ or 886 non-sampled locations employing Hamiltonian Monte Carlo to infer the model parameter distributions.

6.2.4 Spatial parameter recovery

To obtain the grid of predicted values for τ_{in}, we take the mean of the N_2 posterior distributions for this parameter. Then, we calculate the Pearson correlation coefficient between the true and predicted values at the N_2 locations to assess whether there is a linear relationship between them. The Pearson correlation coefficient is defined as

$$r_{xy} = \frac{\Sigma_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\Sigma_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\Sigma_{i=1}^{n}(y_i - \bar{y})^2}}.$$

To assess the validity and usefulness of our method, for each CL used for fitting, we generate a population of action potentials at a desired location from the N_2 total locations and compared them with the true APs at that specific location. To generate the population, we use the voltage model solution at the sample values from the posterior distribution of τ_{in} at the desired location and the four posterior distributions for the rest of the parameters. To generate the true APs, we select τ_{in} from the grid of true values at the desired location, and for the four other model parameters, we assign the values that were used to generate the synthetic data. Since the desired locations are used as training points, there is no need to do inference at every location of the grid, which is computationally expensive. Inspired by heterogeneity of the heart such as what can be observed in optical mapping experiments, we take advantage of the correlated spatial structure on one of the parameters: the APD varies at each location and that measurement is what is represented in the mapping.

6.2.5 Implementation

We implemented our model in Stan. We solved the MS model numerically using an adaptive version of forward Euler with two step sizes for each CL. For the first 4 ms, corresponding to the upstroke (the stimulus was applied for the first 1 ms), we used a step size of 0.1 ms, and a step size of 0.25 ms was used otherwise. We implemented our programs using one chain, the sample size for all cases was equal to 500 and the warm-up period was equal to 1000. We used RStan in R and ran the programs in an Apple MacBook Pro with M1 processor and 16 GB of unified RAM.

6.3 Results

Fig. 6.2 show the 5×5 grid of the true and predicted values the values of τ_{in} at each location. In this case, the four training points were selected at coordinates (2,1), (4,2), (1,4) and (3,5) for equal spacing and are represented by black circles. For illustrative purposes, we selected two of the N_2 test (i.e., non-training) positions, (3,3) and (5,4), as indicated by black squares in Fig. 6.2 to generate 100 posterior predictive distributions of action potential and bifurcation plots for the cycle lengths of 350, 300, and 276 ms used for fitting. We then compared these with the true APs, represented by the black dots, which were not used during inference. Fig. 6.3 shows that the posterior predictive distribution of APs exhibits substantial variation representing the posterior uncertainty in the parameters of the MS model at the new locations. Unsurprisingly, this dispersion is dampened for spatial locations that are closer to the coordinates used for training data. For the data corresponding to position (3,3), the true bifurcation plot does not show a pronounced bifurcation, while some samples from the posterior predictive distribution do show a bifurcation. For position (5,4), the distribution displays more variation about true APs, and specifically for the 276 ms CL, this predictive distribution of APs appears to systematically underestimate the true AP.

Table 6.2 gives the mean posterior values for all the MS parameters (for each location in the case of τ_{in}) along with the true values. Some metrics to assess the convergence

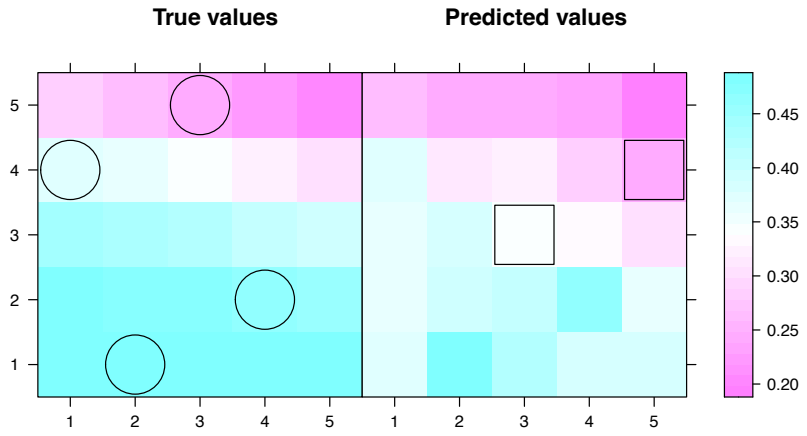


FIGURE 6.2: Grid of true (left) and predicted (right) values for $\tau_1 = \tau_{in}$ when the grid size was 5×5 and there were 4 training points (black circles). The squares show the two test points for which we predicted the APs and BPs in Fig. 6.3.

Parameter	Position	True	Predictive mean	$\hat{R}$	Bulk ESS	Tail ESS
τ_{in}	(3,3)	0.421	0.338	1.001	476.726	505.271
	(5,4)	0.304	0.246	1.004	491.023	467.781
τ_{out}	N/A	6	6.082	0.998	478.265	483.113
τ_{open}	N/A	150	153.891	1.002	467.087	496.537
τ_{close}	N/A	120	115.319	1.005	450.483	459.604
u_{gate}	N/A	0.13	0.124	1	518.237	411.188

TABLE 6.2: True and posterior mean estimates values and MCMC diagnostic metrics for the 5×5 grid.

of the chains and to observe the bulk and tail effective sample size for this example are also given. The predictive means were always between the 2.5% and 97.5% quantiles, meaning that the prediction lay in a 95% approximated credible set, the $\hat{R}$ value are very close to 1, indicating convergence of the chains, and the bulk and tail effective sample size approximations are respectively 450 and 411 in the worst case scenario, sizes not very far from the actual sample size.

Fig. 6.4 shows the 10×10 grid of the true and predicted values for each of the τ_{in} values at each location. The training points in this case ($N_1 = 6$) were (1,1), (8,2), (5,4), (2,6), (9,7) and (6,9) and are represented by circles. We selected 3 out the $N_2 = 94$ positions, (3,8), (6,6) and (9,4), shown as squares, to generate the AP populations. As in the

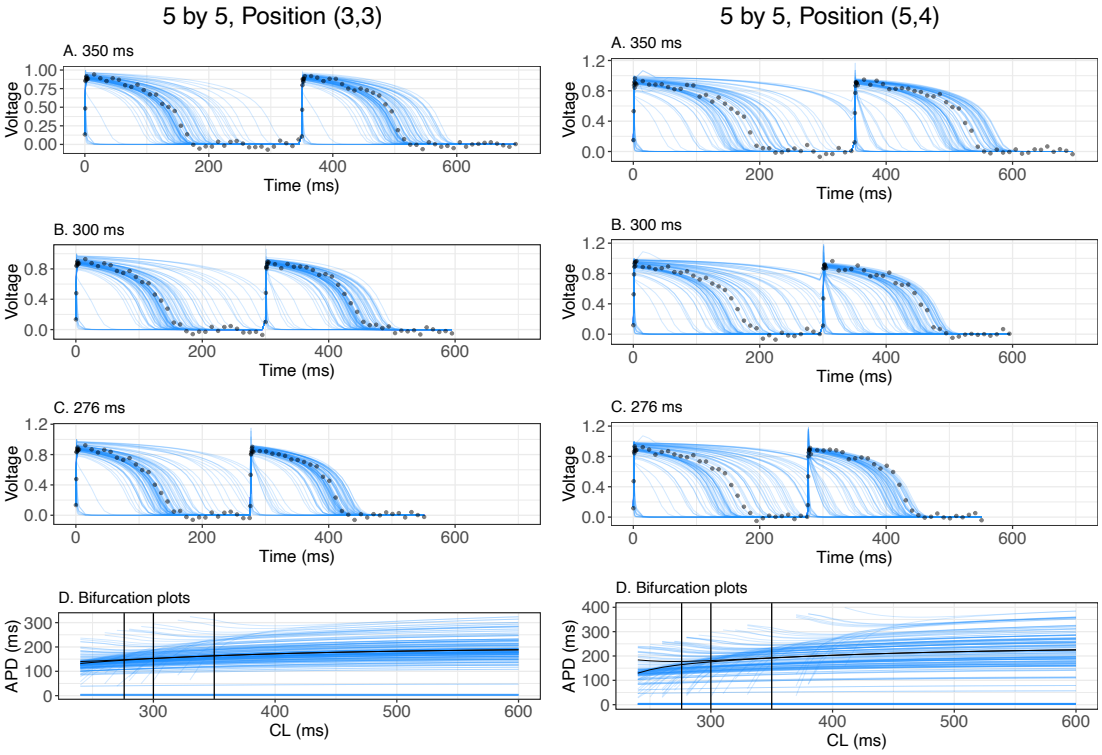


FIGURE 6.3: Comparison of true and predictive samples of action potential and bifurcation plots. Blue lines indicate samples from the noise-free posterior predictive distribution of the AP / BP values at new points in space. The black markers indicate the true values observed without noise and the true bifurcation plots also appear in black. The size of the grid in this case is 5×5 and the predicted locations are $(3, 3)$ and $(5, 4)$.

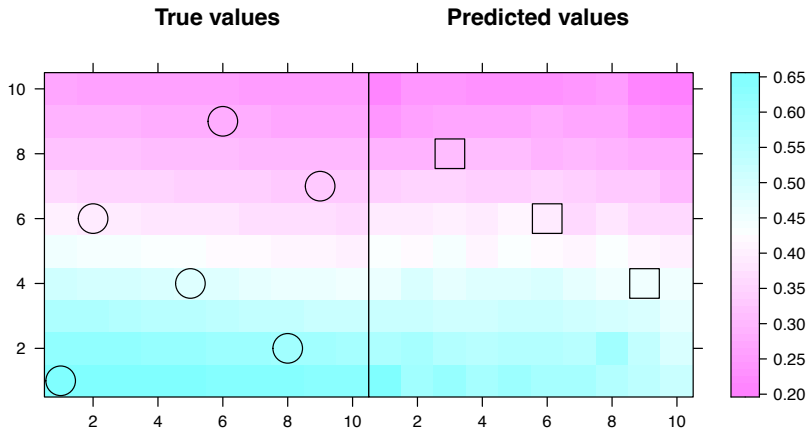


FIGURE 6.4: Grid of true (left) and predicted values (right) for $\tau_1 = \tau_{in}$ when the grid size was 10×10 and five training points (black circles) were used. The squares show the two test points for which the predicted APs and BPs are shown in Figure 6.5.

previous example, we also plotted a population of bifurcation plots to compare it with the true BP. The results are shown in Figure 6.5.

As in the previous case, the populations are able to capture the shapes and dynamics of the three selected positions. The populations of both the APs and the BPs demonstrate more variability for position (3,8), but the populations are again denser for the members that are closer to the true plots. Position (9,4) shows the least population variability of the three predicted positions. Table 6.3 lists the true and mean posterior predictive values of the parameters along with selected Stan metrics. As before, the predictive means were always in a 95% approximated credible set. The $\hat{R}$ values, which are very close to 1, indicate that the chains converged, and the bulk and tail effective sample size approximations are around the same size as the posterior sample size.

For the larger 30×30 grid, Fig. 6.6 shows the true and predicted values of τ_{in} at each location. The training points ($N_1 = 14$) were chosen in a regular pattern; they are represented by circles. For the pattern, we start by selecting location $(i_1, j_1) = (1, 1)$ and recursively defining the coordinates of each next training point location $(i_{m+1}, j_{m+1}$ based on the previous location (i_m, j_m) for $m = 1, ..., N_1 - 1$ as follows: $i_{m+1} = mod(i_m + 6, 30)$

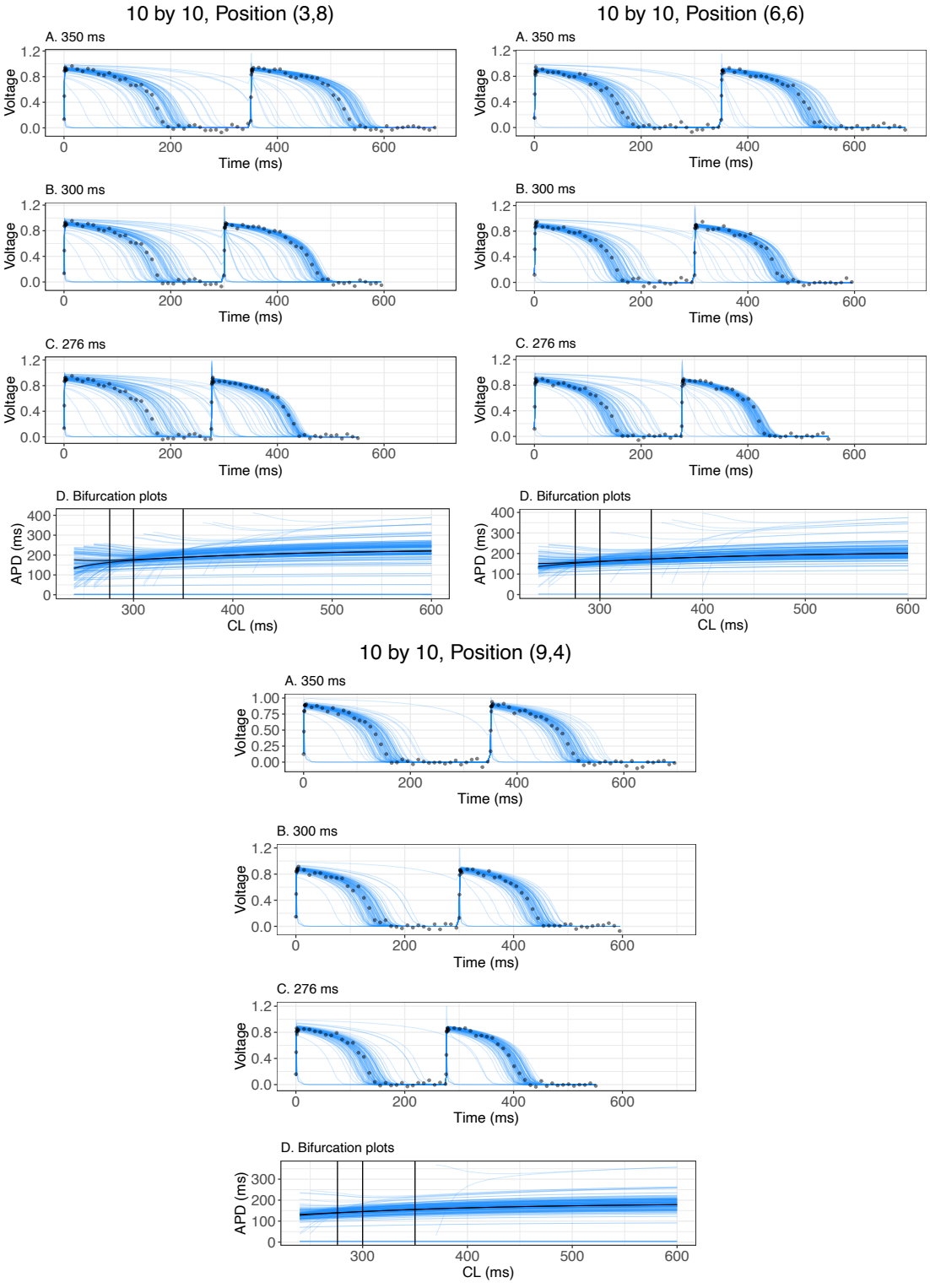


FIGURE 6.5: Predicted populations of APs and BPs (blue) along with the true APs to be predicted for the 3 CLs used for fitting. The size of the grid is 10×10 and the predicted locations are (3,8), (6,6) and (9,4).

Parameter	Position	True	Predictive mean	$\hat{R}$	Bulk ESS	Tail ESS
τ_{in}	(3,8)	0.52	0.541	0.999	451.991	434.566
·	(6,6)	0.493	0.507	0.998	400.749	514.387
·	(9,4)	0.463	0.466	0.998	400.749	514.387
τ_{out}	N/A	6.0	5.992	0.999	500.687	438.954
τ_{open}	N/A	150.0	147.418	0.998	586.925	483.624
τ_{close}	N/A	120.0	124.524	1.003	455.105	439.378
u_{gate}		0.13	0.127	1.000	422.59	385.778

TABLE 6.3: Real and mean posterior parameter values and selected MCMC diagnostics for the 10×10 grid.

and $j_{m+1} = j+1+\lceil((i+6)/30)\rceil$. We selected four of the $N_2 = 986$ non-training positions, (8,20), (10,8), (25,5) and (26,26), shown as squares, to generate the AP populations and compared them with the true APs for each CL. Figure 6.7 shows populations of bifurcation plots for the four test locations.

The predicted true parameter values, voltage values, and bifurcation plots in this case show much greater variation and lower accuracy overall than the previous examples. However, it is important to consider that in this case we used only 1.5% of all locations as training points (compared to the first and second examples, where the training points represented 16% and 6% of the total positions, respectively). Position (10,8) has the best fitting for both APs and BPs, and its populations also show the least variability. For position (25,5), the populations of APs are always below the true plots and consequently, the populations of BPs are always below the true plot. Position (26,26) shows the opposite behavior: the populations of APs are close but above the true APs, except for some cases for 276 ms. The true BP seems to be more centered, but the bifurcation is less pronounced compared to the population of predicted BPs. This behavior is consistent with the errors in the predicted values of τ_{in}. The locations where τ_{in} is predicted to be lower than the true value have longer APDs than the true values, and vice versa. This would be expected because τ_{in} governs the strength of the inward current; smaller τ_{in} means larger inward current and thus is expected to prolong APD.

Table 6.4 provides the mean posterior values for the MS model parameters, at each location in the case of τ_{in}, as well as the true values and selected Stan metrics. As before,

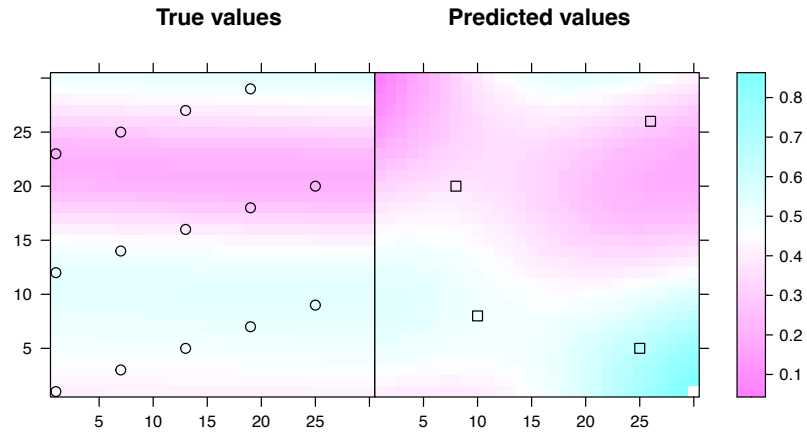


FIGURE 6.6: Grid of true (left) and predicted values (right) for $\tau_1 = \tau_{in}$ for the 30×30 grid with 14 training points (black circles). The squares indicate the four testing point for which the predicted APs and BPs are shown in Figure 6.7.

Parameter	Position	True	Predictive mean	$\hat{R}$	Bulk ESS	Tail ESS
τ_{in}	(8,20)	0.22	0.369	0.999	451.991	434.566
	(10,8)	0.523	0.499	0.998	400.749	514.387
	(25,5)	0.488	0.667	0.998	400.749	514.387
	(26,26)	0.365	0.301	0.998	400.749	514.387
τ_{out}		6	5.992	0.999	500.687	438.954
τ_{open}		150	147.418	0.998	586.925	483.624
τ_{close}		120	124.524	1.003	455.105	439.378
u_{gate}		0.13	0.127	1	422.59	385.778

TABLE 6.4: Real and mean posterior parameter values and selected Stan metrics for the 30×30 grid.

the predictive means were always between the 2.5% and 97.5% quantiles, the $\hat{R}$ indicate convergence of the chains and the bulk and tail effective sample size approximations are close to the size of the posterior sample.

The Pearson correlation coefficients (R) for the three grid sizes are given in Table 6.5, along with the computational runtime for the corresponding examples. Runtime did not increase linearly with the number of training points, since the time per location was 0.15, 0.45 and 2.9 hours for the 5x5, 10x10 and 30x30 cases. The Pearson correlation coefficient was close to one for the two smaller grid sizes, but not as close for the 30x30

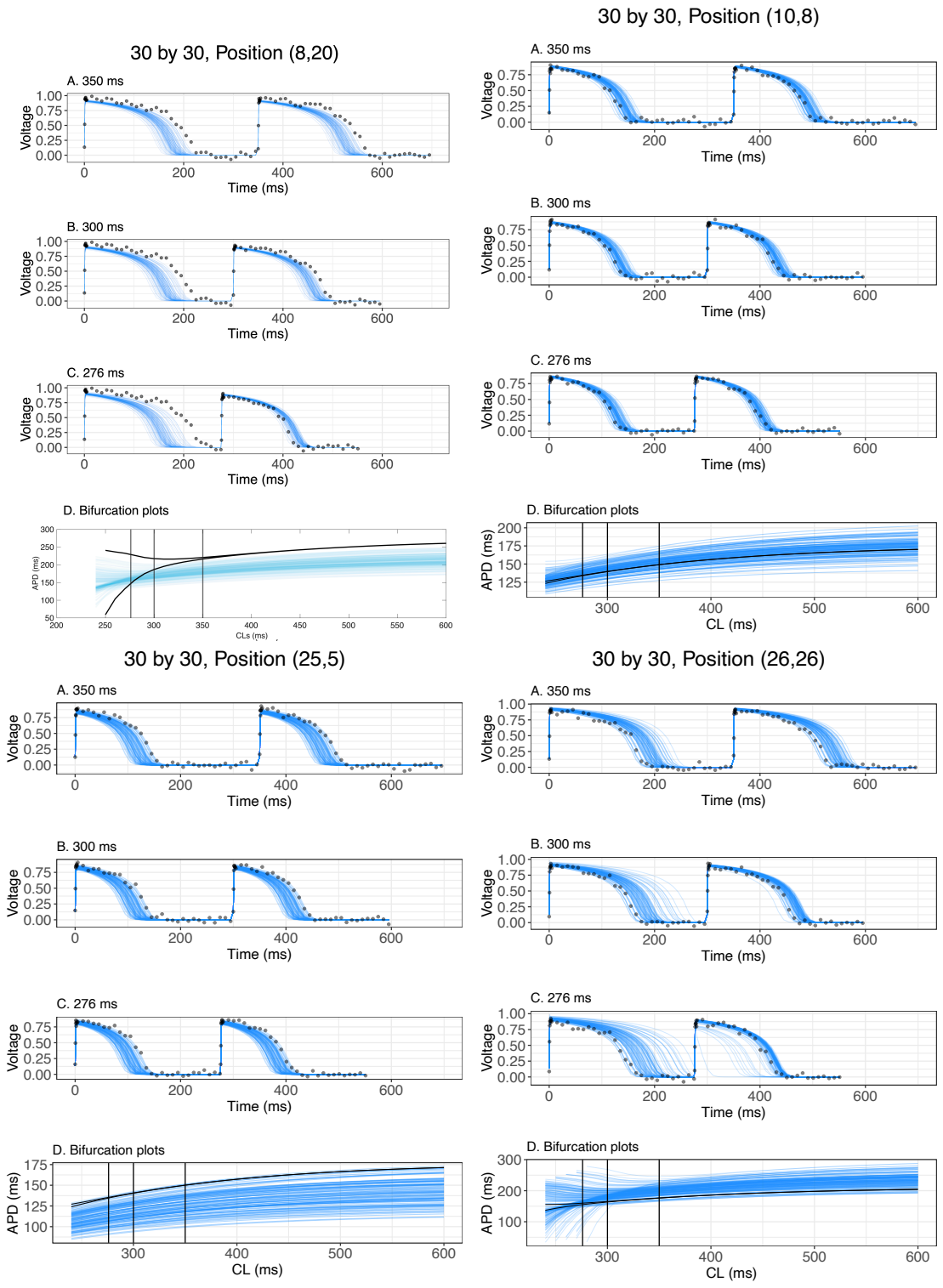


FIGURE 6.7: Predicted populations of APs and BPs (blue) along with the true AP for the 3 CLs used for fitting for the 30×30 grid. The predicted locations are (8,20), (10,8), (25,5) and (26,26).

grid size	5x5	10x10	30x30
N	25	100	900
N_1	4	6	14
r_{xy}	0.96	0.98	0.63
time(h)	0.6	1.8	11.5
divergences	396	0	0

TABLE 6.5: Computational time and Pearson correlation coefficients for the three grid sizes.

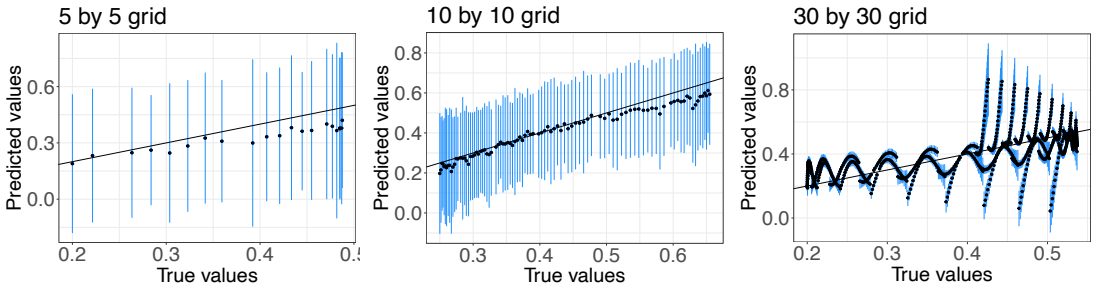


FIGURE 6.8: Correlation between predicted and true values for τ_{in}. The blue bars represent the standard deviation of the mean posterior predictive values and reference lines with slope 1 are shown in black.

case. The smallest grid size had many divergences, whereas the larger grids had none. The divergences point out numerical inaccuracies when solving the leapfrog algorithm, which can make the results to be non reliable. Fig. 6.8 shows the predicted values of τ_{in} plotted against the true values as well as the standard deviations of the means of the posterior distributions at each predicted location. For for the 5×5 and 10×10 grids, the lines through the plotted points have slopes very close to 1, which indicates a strong linear relationship between the true and the predicted values. For the 30×30 case, the results are not as good, and the plot shows some artifacts from the function used to assign training point locations. In particular, many of the true values above 0.4 are not well predicted.

6.4 Discussion

We successfully performed inference and made predictions on grids of different sizes when selecting a small number of training points (no more than 16% of the total positions).

The correlation between the predicted and true values was closer to one when the grid size was smaller, but note that the percentage of training points with respect to the grid size also decreased with increasing grid size. Even for the 30×30 case, the populations of action potentials were acceptable because at some of the positions shown, they were always below the voltage profiles to be fit, even though there were close.

The results improved when the non-training points were given as input to calculate the covariance matrix for the square exponential kernel when doing inference (results not shown). We do not include the non-training points when building the covariance matrix show those as examples in this work for two reasons. On the one hand, the number of training points we used was close to being the worst-case scenario as the grid size was gradually being incremented; on the other hand, we wanted to reduce the computational time and still obtain acceptable results. For instance, the computation time for the grid of size 100×100 decreased from 5.7 to 1.8 hr when not giving the non-training points as input for the covariance matrix, where the number of training points represented 6% of the total number of positions on the grid.

For the priors, we selected not only folded normals for $\tau_{out}, \tau_{open}, \tau_{close}$ and u_{gate}, but also gamma distributions and obtained comparable results (not shown). Because we used a hierarchical model, we selected folded normals so that the distributions of the parameters at the higher level represented the mean (and variance) for all the positions, as the two parameters of a gamma distribution represent shape and scale (or rate) but not directly the mean and variance of the random variable in question. As for the distance parameter ρ of the kernel, we used uniform and inverse gamma priors, selecting the parameters of the priors in such a way that the support included the minimum and maximum distance of any two points on the grid; the results presented here used uniform priors for the 5x5 and 30x30 case and an inverse gamma prior for the 10x10 case; we actually used these two kinds of priors for each grid size. In the 10x10 case, the inverse gamma gave a better approximation of the predicted grid. The other two cases gave equivalent results. Finally, even though we present results of the Gaussian process prior where the mean was taken as constant, we also used folded normals as priors for the mean and obtained results similar

to those shown in the examples. However, the time increased when using folded normal priors.

Our application of a hierarchical model to study variability in space shows that it should be possible to study variability of action potentials across different individuals or from the same individual over time. In this way, the parameters of the model at the higher level would represent the whole population; otherwise, they would represent each individual.

Chapter 7

Conclusion

In this thesis we demonstrated the usefulness of the Hamiltonian Monte Carlo (HMC) method as a tool for parameter inference in cardiac electrophysiology modeling through three studies. We applied HMC to fit synthetic and experimental data from cardiac action potentials to find the distributions of the parameters of the Mitchell-Schaeffer (MS) and Fenton-Karma (FK) cardiac action potential models. Using similar datasets and the same two models, we also compared the performance of HMC versus ABC SMC. Finally, we described the intra-variability of cardiac action potentials in space using HMC in a hierarchical model context to perform inference and prediction on voltage profiles by using information from a small number of locations to predict a spatially varying parameter value for the MS model and spatially invariant values for the remaining parameters at other locations in the domain. We begin this final chapter with a summary of our main findings for each study.

In chapter 4, we used HMC for the first time in the context of cardiac electrophysiology to fit the MS and FK model to synthetic and experimental data. For these two phenomenological models, we found the parameter probability distributions, which can be used to completely describe the different sources of uncertainty in a rigorous way. We also used these distributions to build populations of action potentials and measured how accurate they were with respect to the data to be fitted. We verified that the technique was able to also describe the dynamics of the two systems in the sense that the alternans behavior was captured for more than just the CLs used for the fitting. HMC also estimated the

standard deviation of the noise in the datasets, which agreed with the known value for the synthetic data and was estimated about ten times larger than that for the experimental dataset. Because selecting the priors for HMC and their support is difficult, we used the more flexible ABC SMC to generate distributions we then used to specify the support of the HMC priors.

In chapter 5, we used HMC and ABC SMC to find populations of cardiac action potential model parameters consistent with data used for fitting and compared the two methods. We showed that the methods were effective for both synthetic data derived from the models used as well as for an experimental dataset from a zebrafish hearts. In nearly all cases, both methods found quasi-symmetric marginal distributions with clear modes for the give parameters of the MS model and for the 13 parameters of the FK model. We showed through the use of *q*-*q* plots that the posterior distribution samples obtained by the two methods did not give any strong indication of being from different distribution families. In the case of synthetic data, where the true parameters used to generate the dataset were known, the true values in general were well contained within the posterior distributions found, and across multiple runs of the algorithms the true values coincided well with the distribution peaks. Given that both methods achieved similar results with no clear computational advantage, considerations other than accuracy may motivate the choice of method. ABC SMC generally was easier to work with, as it accepts different kinds of prior distributions, and those distributions may be broad. ABC SMC appeared to explore the parameter space more broadly than HMC, which may indicate that HMC was able to find the area of parameter space corresponding to the true distribution more efficiently. Although HMC required more effort to find an acceptable prior (and indeed we suggest that SMC may be useful in this task), it tended to find narrower distributions and is a full rather than approximate Bayesian method, which may be advantageous in some cases.

In chapter 6, we used HMC to describe the variability of action potentials in space within a single individual. By pairing HMC with a latent Gaussian process prior for one of the MS model parameters that varied spatially over a two-dimensional grid, along with a

hierarchical model for the rest of the parameters not considered correlated in space, we successfully applied HMC to do inference and prediction. Toward this end, we selected a very small percentage of training points from the whole two-dimensional space. Using a series of voltage traces at different pacing periods from those training points for different grid sizes, we found the distributions of the MS model parameters, and then we predicted the distributions of all the parameters at desired locations not considered for training. We also verified that the methodology was able capture the dynamics associated with alternans. Despite the small number of training points selected, the population of bifurcation plots predicted a bifurcation when it existed. This novel technique utilized spatial correlations in voltage that could arise through cell coupling, although the data were not generated from a partial differential equation. Using only ODEs reduced computational time, because less than 16 percent of the total number of grid points were used to do inference and to predict the distributions of the model parameters. In this way, populations of action potentials could be produced for any desired location not used for training.

7.1 Limitations

The results of our studies are accompanied by a number of limitations. In this thesis we considered only a limited number of datasets. In particular, we chose data from three CLs, with one at a longer CL and two at shorter CLs within the alternans regime. It may be possible to optimize the selection of CLs beyond what was chosen. In addition, it is possible that performance may change for noisier data or for data with different dynamics that may not be well captured by the model being fit.

Similarly, we only considered two models, and it is possible that performance could differ for a different selection of models. We also note that we have used a single approximate Bayesian method and that a different choice may result in different findings. Within the models, we used a simple square-pulse stimulus. Use of a biphasic stimulus [11] could help prevent selection of large values for the excitability parameters (e.g., τ_{in} for the MS model and τ_d for the FK model) that would not allow propagation in tissue and thus may be unphysiological.

In the use of HMC, first as a proof of concept and then for comparison with an approximate method, we found that HMC was sensitive not only to the distribution family of the prior chosen, but also to its support. The NUTS implementation in Stan also needs good initial points for the chains to converge properly. One potential way to overcome this difficulty is to use the modes or the means of the given priors as initial points.

In the case of ABC SMC, we only used one kind of function for the parameter statistic: the distance function between the candidate parameter model and the synthetic data, which is related to the error between the two sets of data. Many other statistics can be used and choosing what kind of statistic might depend on the problem. We also measured the fitting error using the coefficient of determination; in the case of ABC SMC, this was an indirect way of verifying that the chains converged, as parameters with small error are more likely to be chosen as part of the posterior sample. Also, the tolerance reduction approach used for ABC SMC, while adaptive, nevertheless was fixed in advance, following Ref. [21].

The probability model presented in chapter 6 includes major structural and distributional assumptions encoded within the model. The assumption of independent noise variates may be quite poor if there is substantial unmodeled signal variance not captured by the MS model. In this case, the stochastic error is likely to be highly correlated, and the independent Gaussian noise model would be a poor fit. A potential remedy in such a situation would be to include autocorrelated noise via an error model, as commonly studied in time series analysis such as the AR(p) family of probability models. It is also possible that the model errors may be correlated across space and time; if so, usage of a Gaussian process observational model would be appropriate. We note that this approach differs from the proposed GP prior usage in this work by employing a Gaussian process for part of the likelihood, rather than as a prior over model parameters.

Another major assumption in this work is that a spatial prior is used for only one parameter out of the five MS model parameters; this choice was made primarily for computational tractability but also to highlight the differences in posterior inferences resulting

from a spatially correlated versus uncorrelated prior distributional assumption. A more comprehensive analysis would be to use a five-dimensional multivariate Gaussian process to model correlations between different parameters of the MS model as well as between instantiations of the same parameter at different locations in space.

For Bayesian predictive inference for Gaussian processes performed using the hierarchical probability model, we assumed Gaussian observations because we created the synthetic data, but that assumption may not be true for a different problem or application. In such a case, inference could still be done but the calculation time would increase because the posterior could not be derived analytically. Another limitation comes from the fact that to speed up the calculations, the training locations were not used to calculate the covariance matrix of the Gaussian process; however, this can be done.

7.2 Future work

The work presented in this thesis offers many opportunities for extensions in terms of the methodological framework, computational implementation, additional related studies, and other applications to cardiac electrophysiology.

7.2.1 Methodological improvements

It may be useful to optimize the data used for fitting to better constrain certain parameter values, just as we did using a different resolution for the upstroke and the rest of the action potential, per CL. For example, in the FK model, the parameter τ_{v1}^{-} helps to set the minimum diastolic interval; datasets that do not represent that information may have difficulty adequately constraining that parameter and related parameters like u_v.

Approximate Bayesian computation methods, which approximate the likelihood such that the posterior distributions obtained are an approximation of the true distributions, are not the only methods that are not fully Bayesian. Variational methods [36, 37] can also be used to perform Bayesian variational inference to approximate the posteriors. A comparison between the performance of these two kinds of approximate methods would

be useful, and a variational method could also be compared to a full Bayesian method such as HMC [83] to assess the trade-offs in terms of computational effort and accuracy.

7.2.2 Computational improvements

ABC SMC might benefit from a more sophisticated way of choosing tolerances, although other limitations of this method arise because the method is approximate. More sophisticated statistical tools could be used to ensure that the chains are converging. Also, ABC SMC is one of many ABC methods available; some comparisons between ABC sequential methods have already been performed. For instance, in Ref. [21], an adaptive ABC SMC method was compared to the existing ABC SMC. Therefore, it would be useful to know whether using a different ABC method would give more consistent results. Given ABC SMC's flexibility and ease of implementation compared with HMC,the variability study could be performed using ABC SMC instead of HMC to compare the performance of these two methods in this context.

In terms of the computational time and resources needed for these kinds of problems, especially when the number of operations can increase exponentially, taking advantage of the tools that Stan provides for parallel computing would be useful, particularly in the case of spatial variability. The largest grid size we used for inference and prediction in this work was 30x30, but the size of optical mapping recordings is 128x128, motivating the need for approaches with good scalability. In addition, even for smaller grid sizes, depending on the problem at questions, more training points may be required. Finally, using more chains may help to improve the robustness of results.

7.2.3 Additional related studies

Because finding appropriate priors in Stan through NUTS can be time-consuming, it would be useful to see if ABC SMC consistently helps in finding a good candidate for the HMC prior support. If so, then a hybrid algorithm that uses ABC SMC to select good priors could be used when doing inference with HMC. More research is needed to understand why certain families of priors did not work for HMC.

With respect to the variability study, possible future implementations of the proposed approach could be used with cardiac voltage data obtained from optical-mapping recordings. In such a scenario, it may be advantageous to extend the spatial GP formulation to include space-time correlations; the intricacies of this approach are reviewed in [40]. Another potential avenue is to employ a more complex electrophysiology model such as the Fenton-Karma model [31].

Some phenomenological models like the ones presented in this work are useful to capture the presence of alternans. However, other characteristics of cardiac electrophysiology, such as calcium transients, may be of interest. In those cases, other cardiac electrophysiology models, such as more biophysically detailed models or models with more parameters, could be used to fit datasets like those shown here.

7.2.4 Other applications

We expect that Bayesian methods like HMC will be useful for ongoing efforts, including efficient creation of model populations [7] and virtual patient cohorts [70] as well as addressing nonidentifiability of model parameters [20, 22, 89, 23] and uncertainty quantification [78, 14]. For instance, verifying that the mean of the parameters appears between specific quantiles backs up the idea that the technique can be used successfully with experimental data, as in the examples shown. The Bayesian approach also allows for quantifying an approximation of a credible set for the parameters given the particular priors selected.

Appendix A

Previous studies using ABC SMC

A.1 Preliminary Results

Below we briefly describe some preliminary results related to the topics of this thesis. First, we present a selection of results from a more comprehensive study of the use of ABC SMC to obtain parameters for the Fenton-Karma model using both synthetic and experimental data. Second, we show proof-of-concept results using HMC to obtain parameter values for the FHN model.

A.1.1 ABC SMC Using the Fenton-Karma Model

We conducted a series of studies to obtain values for model parameters using an approximate Bayesian method. Specifically, we used ABC SMC to parameterize the Fenton-Karma model for two scenarios: (1) using synthetic data, to assess the accuracy and overall performance of the algorithm, and (2) using microelectrode recordings of zebrafish action potentials, toward the goal of developing a mathematical model of this system. For the synthetic data scenario, we studied the accuracy, robustness and computational efficiency of the ABC SMC algorithm. In this case, the true values of the parameters were known, thereby allowing accuracy to be quantified. For the numerical solution of the Fenton-Karma model, we used an explicit Euler method with an adaptive timestep.

Algorithm and Model Settings

Data points used for fitting consisted of N voltage data points, either generated from simulations without added noise or obtained from experiments, for a specific set of M CLs:

$$N = \Sigma_i^M(n_i),$$

where n_i is the number of data points to be fitted per CL. Only the last two beats of ten at each CL were used for fitting to allow the model to reach the steady state specific to each CL. The ABC SMC algorithm also requires a distance function, a tolerance schedule and a kernel function to be specified in advance. The distance function used was

$$\rho(y_{ij}, y'_{ij}) = \frac{1}{MN}\Sigma_j^M\Sigma_i^{n_j}\left(|y_{ij} - y'_{ij}|\right),$$

where y_{ij} are the target synthetic voltage data and y'_{ij} the solution of the voltage variable using the candidate parameters. The initial values for the model variables were $u_0 = 1$, $v_0 = 0$ and $w_0 = 0$. We used uniform distributions for all the parameters; for the synthetic data case, we used $U(0.8*\theta_{truth}, 1.2*\theta_{truth})$, where θ_{truth} were the values used to generate the synthetic data. The specific settings for the zebrafish data are given in section A.1.1. For the kernel function, we used a uniform distribution $U(-1, 1)$ scaled by a parameter σ set to 0.1.

The algorithm proved to be sensitive to the initial tolerance for acceptance. Therefore, we developed a data-driven approach to set this parameter as the minimum average relative error from five random populations. Specifically, five random samples of size N were generated without rejection, and the average distance between the action potential profiles using each sample point and the profiles to be fitted was calculated. The minimum of those five averages was taken as the initial tolerance. The tolerance reductions followed

$$init\_tol * (0.9, 0.8, 0.7, 0.6, 0.5, 0.4, 0.3, 0.2, 0.1, 0.05, 0.025, 0.125),$$

and the algorithm was terminated when the acceptance rate (the percentage of accepted sample points) for a population was below 0.03 or after 12 iterations.

The different algorithm settings studied were the parameters to fit, the cycle lengths used in the fitting, the population size, and the tolerance schedule. Unless specified otherwise, all other settings used the default values as follows. The parameters to be fit in the synthetic-data scenario were $\tau_r, \tau_{si}, \tau_d$, and τ_w^+, with their true values set to 33.33, 29, 0.25 and 870 ms, respectively. The CLs used in the data to be fit were 330, 340 and 360 ms, and the default population size was 100.

We wrote our own custom programs to solve the system of equations and to implement the Bayesian algorithm in R. The system used to obtain the timing results given here was a 2.6 GHz Dual-Core Intel Core i5 MacBook Pro with 8 GB of memory.

Performance Assessment Metrics

To quantify the accuracy, robustness, and computational efficiency of the ABC SMC method, we used several metrics. For accuracy, we used the following four types of error measurements.

1. **Error in fit CLs:** Average (in time) relative error in the voltage averaged across all fit CLs with each parameter assigned its mode in the last population.
2. **Error in unfit CLs:** Average (in time) relative error in the voltage for CLs of 1000 ms and 300 ms with each parameter assigned its mode in the last population.
3. **APD error:** Average of absolute error in the last two APDs for all CLs.
4. **Relative error of the mode:** Error in the mode for each parameter compared to its true value, normalized to the true value.

In the present study, we emphasize accuracy; however, we include some preliminary discussion regarding robustness and efficiency. To assess robustness, we calculated the standard deviation of the relative error of the average mode for each parameter across ten runs and the standard deviation of the fit and unfit CLs errors and APD errors across

ten runs; analysis and interpretation of these results is ongoing. To quantify computational efficiency, we used the runtime, but because these results were not obtained in a controlled setting, further investigation of algorithm performance is needed.

Example of Parameter Estimation

We now show an example of the estimation of parameters where we used the default settings: four parameters, three CLs and a population size of 100. Fig. A.1 shows the action potential profiles obtained using all 100 candidate parameter sets superimposed for several generations. As can be seen, the dispersion of the profiles is reduced in later populations. Fig. A.2 shows how the parameter distributions change as the algorithm progresses. Over time, the distributions in general become narrower.

The action potential profiles for fit CLs using the parameter modes along with the data used for fitting and the mean of the posterior predictive distribution, which can be thought of as the mean action potential profile of the last generation, can be seen in Fig. A.3. The profiles obtained using the final population's parameter modes, given in Table A.1, show good agreement with the experimental data, and the parameter values themselves also agree well with the true values (relative error between 1 and 6%). Fig. A.4 shows the distributions of the final population for each parameter. The posteriors for τ_r , τ_{si}, and τ_w^+ are narrow compared to the priors, indicating a reduction in the variance. In the case of τ_d, although the posterior is not narrow, the mode of the distribution is close to the true value. The distributions for τ_{si} and τ_d contain the true values.

The action potential profiles from the entire last population for each parameter are shown superimposed in Fig. A.5. The voltage profiles show little variability, which indicates that the algorithm is able to constrain the action potential profiles. Table A.2 shows the number of populations generated, the acceptance criterion (acceptance rate less than 0.03), the fit CLs error, and the runtime. Action potential profiles for the whole last population are superimposed for unfit CLs in Fig. A.6, which also shows the fitted bifurcation plot. Table A.3 shows the errors in fit CLs, unfit CLS, and APDs. There is good accuracy in the voltage and bifurcation plots even though the parameter values still have error. For

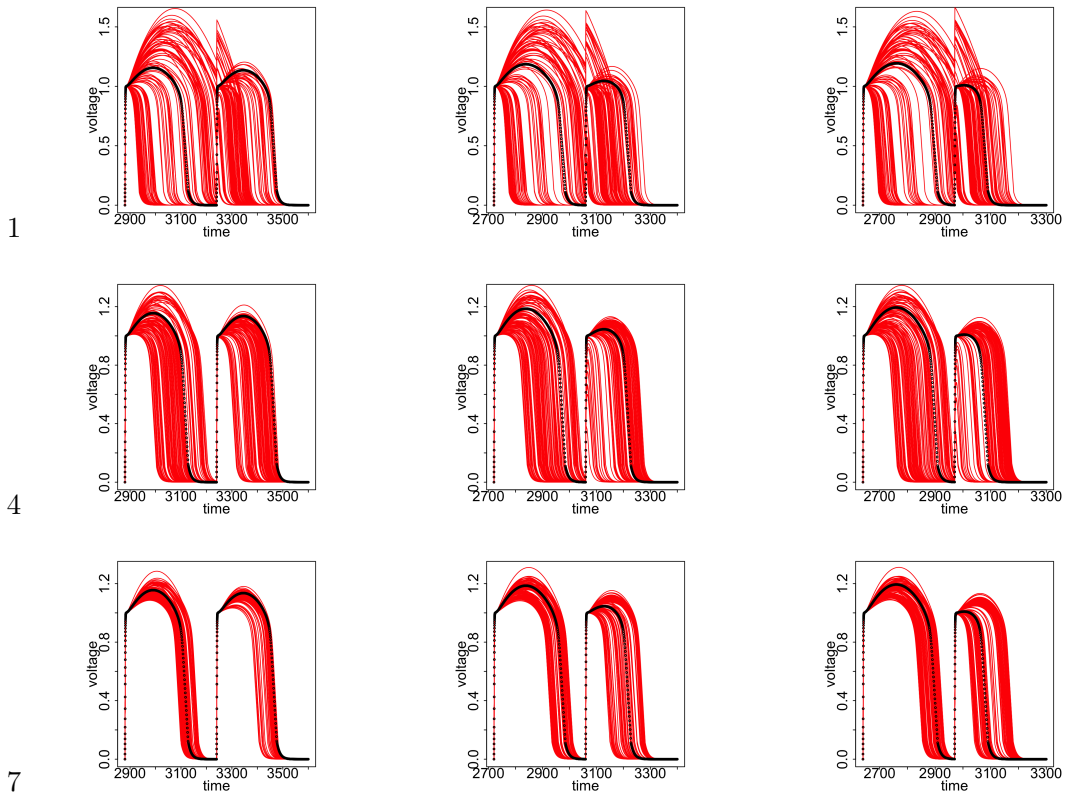


FIGURE A.1: Voltage profiles for populations 1, 4 and 7 (red) corresponding to the three CLs used for fitting along with the synthetic data (black).

all plots (voltage profiles and bifurcation plot), the errors are low enough to be visually consistent with the data.

Algorithm Performance

In this section, we summarize results obtained by varying fitting settings, including the number of CLs used in the fitting, the number of parameters fit, and the population size in the SMC ABC algorithm. To study the effect of changing the number of CLs, we used the CL sets $\{360, 340\}$, $\{360, 340, 330\}$, and $\{360, 350, 340, 330\}$ ms; the bifurcation point was at 350 ms. To assess the effect of the randomness of the algorithm on the outcome, we ran the program ten times and examined the extent to which the fitting results varied.

The effect of changing the number of CLs used for fitting produced an average relative error in each parameter mode less than 9 percent when compared to the true values, with

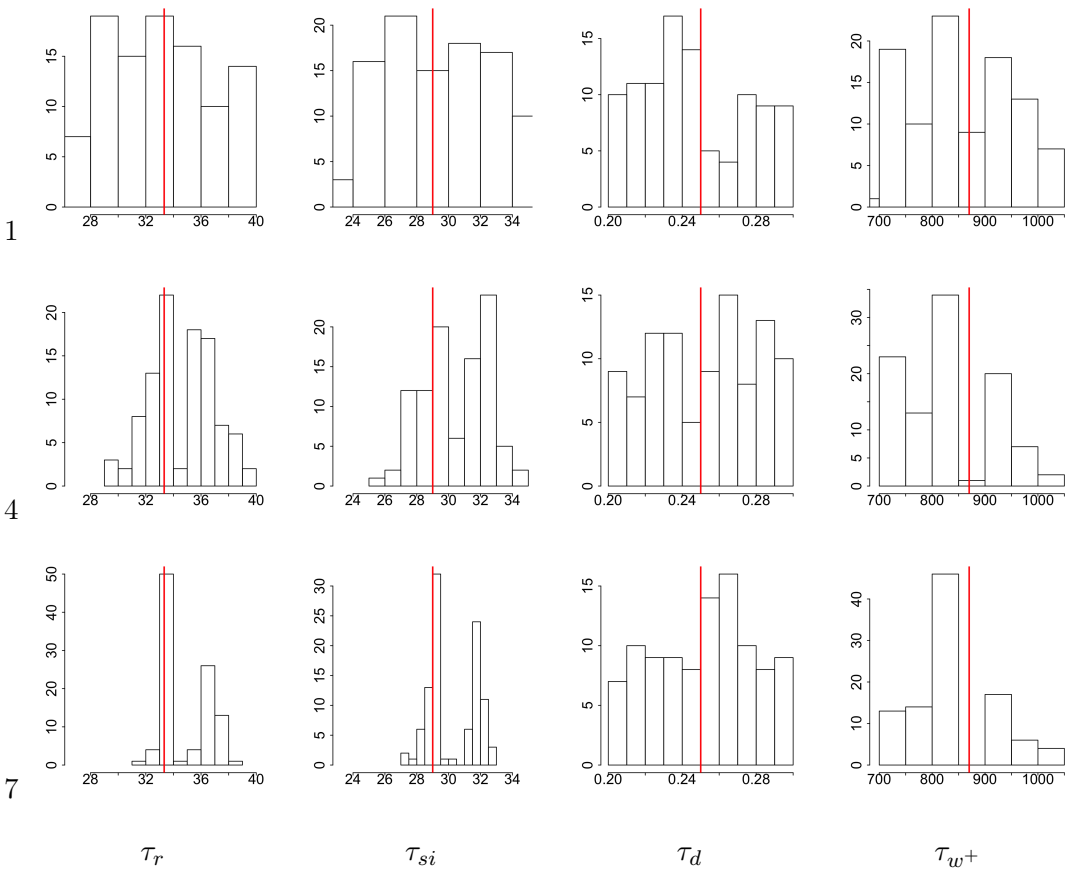


FIGURE A.2: Parameter distributions for populations 1, 4 and 7 for the different cycle lengths along with the true parameter values (red).

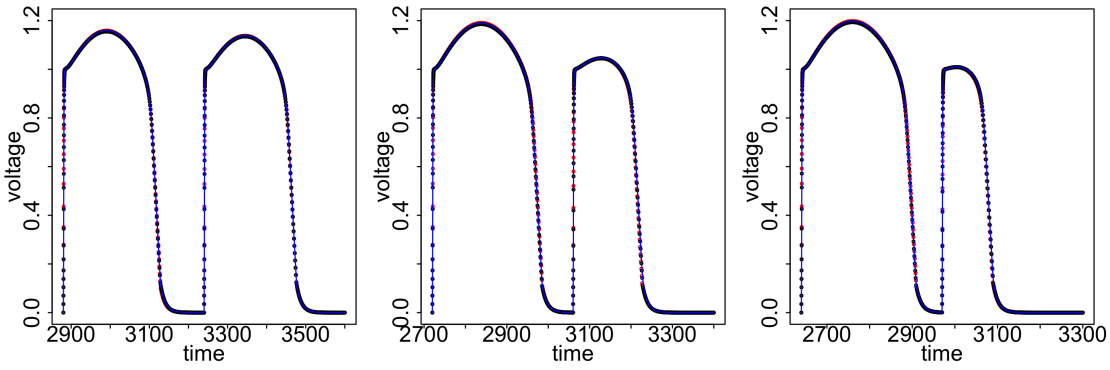


FIGURE A.3: Action potential profiles for fit CLs using the parameter modes (red), synthetic data (black) and posterior predictive (blue).

θ	truth	mode(θ)	relative error
τ_r	33.33	33.585	0.008
τ_{si}	29	29.125	0.004
τ_d	0.25	0.234	0.064
t_{w^+}	870	849.364	0.024

TABLE A.1: Parameter modes from the final distribution along with their relative errors.

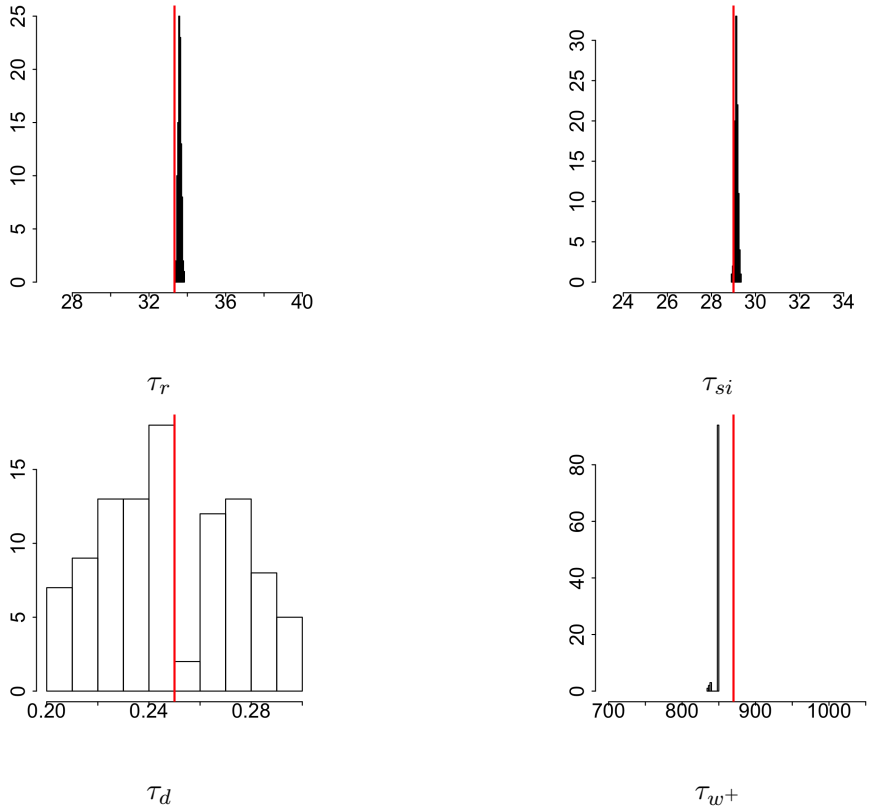
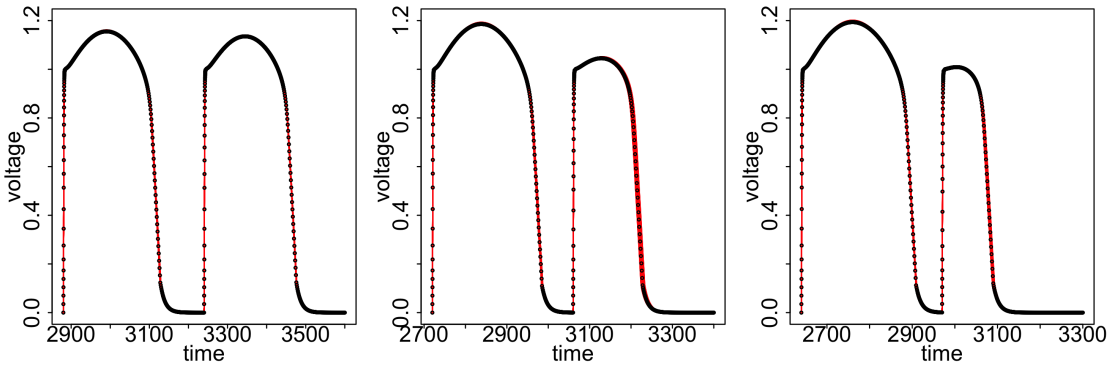


FIGURE A.4: Distribution for each parameter for the last population. Red vertical lines indicate the true values.

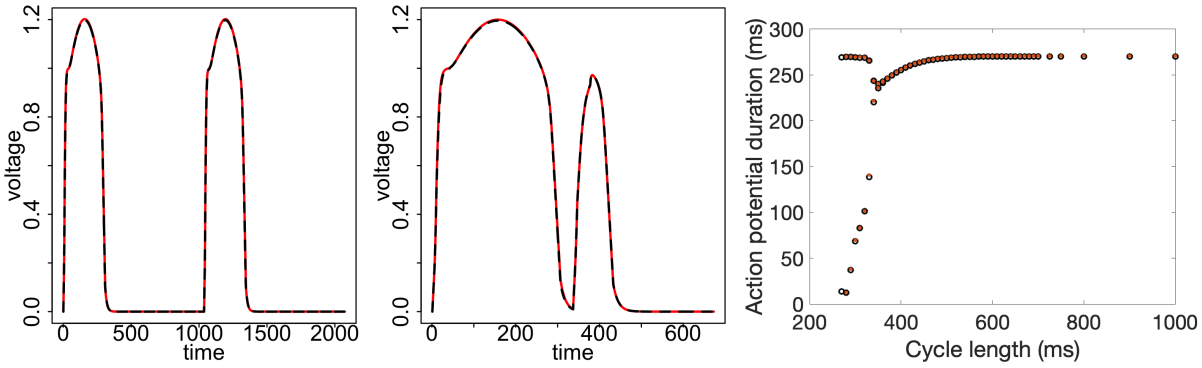
metric	value
fit CLs error	0.003
acceptance rate	0.029
population	12
time (s)	316.512

TABLE A.2: Error for the fit CLs, acceptance rate of the last population and runtime for the example.



360 ms 340 ms 330 ms

FIGURE A.5: Action potential profiles for fit CLs using the last population superimposed (red) and the synthetic data (black).



1000 ms 300 ms bifurcation plot

FIGURE A.6: Action potential profiles for unfit CLs and bifurcation plot obtained using the parameter values of the last population superimposed (red) compared with the corresponding profiles and bifurcation plot using the true parameter values (black).

metric	value
fit CLs error	0.003
unfit CLs error (300 ms)	0.001
unfit CLs error (1000 ms)	0.0
APD error (ms)	6.008

TABLE A.3: Fit and unfit CLs error along with APD error.

the largest error observed when two CLs were used for fitting. The standard deviation of the parameter modes, which was between 0.008 and 37.216, was the smallest for the two CLs case, as well as the standard deviation of the APD error, the fit CLs error and the unfit 300 ms CL error, which were between 0.001 and 6.6693 ms. To evaluate the robustness of our results more thoroughly, we would need to confirm that the populations obtained in the 10 runs have a high likelihood of having been drawn from the same distribution.

Increasing the number of parameters fit from four to six led to a decrease in accuracy, but did not change runtime; in this case, we just ran the program once. The relative error in the parameter modes when considering 4 parameters was between 0.4 and 6.4%; for 6 parameters, it was between 0.8 and 5.9%. Using the same stopping criterion as before (the acceptance rate), the algorithm terminated earlier (population 10 vs. 12). Greater accuracy could be obtained by fixing the number of populations, but the runtime would increase because of the additional populations, particularly because the last population typically is the slowest to compute.

Finally, by changing the size of the population from 100 to 200, the results improved overall. Although the relative error of the mode improved for only one parameter, the fit and unfit CL errors along with the APD error decreased when the size of the population increased. In particular, the accuracy in the unfit CLs cases was reduced by a factor of 8 and the APD error decreased by a factor of 80.

Zebrafish Action Potentials

In the second scenario considered, we fit action potential profiles obtained from zebrafish hearts (data courtesy of Conner Herndon and Flavio Fenton). We configured ABC SMC to use CLs of 400, 350, 300, 276, 272, and 267 ms with a population of size 100 and the same tolerance schedule given above. The priors for all the parameters were uniform distributions with the maximum and minimum values given in Table A.4.

Superimposed action potential profiles obtained using the parameters from the whole last

τ_r	τ_{si}	τ_d	t_w^+	k	u_c^{si}	τ_w^-	u_v	τ_0	u_c	τ_v^+	τ_{v1}^-	τ_{v2}^-
11	28	0.03	20	0.5	0.045	20	0.001	2	0.02	2.7	8	35
209	532	0.57	380	9.5	0.855	380	0.019	38	0.38	51.3	152	665

TABLE A.4: Values used for the uniform priors for the 13 parameters of the Fenton-Karma model.

population are shown in Fig. A.7 along with the experimental data. As can be seen, the voltage profiles for the whole population do not show much variability (although there is more variability than for the synthetic data case), indicating that the algorithm is able to constrain the action potential profiles.

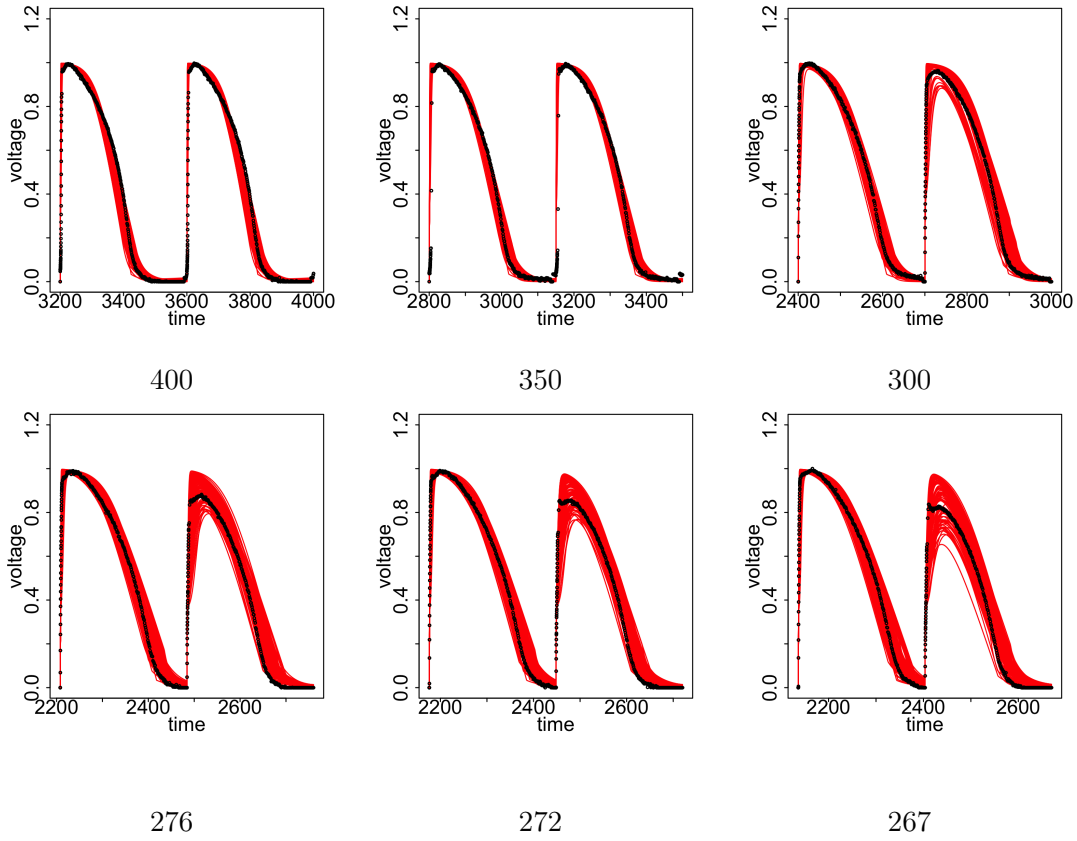


FIGURE A.7: Action potential profiles from the last population for the six CLs used for fitting.

ABC Summary

Overall, the ABC SMC algorithm consistently obtains parameter values that generate action potentials that match the target profiles, as quantified through error metrics and

low variability across the final population. For additional improvement, we could consider using informative priors for faster convergence to the target distributions.

Bibliography

[1] G. W. Beeler and H. Reuter. "Reconstruction of the action potential of ventricular myocardial fibres". In: *The Journal of Physiology* 268.1 (June 1977), pp. 177–210. ISSN: 0022-3751. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1283659/ (visited on 06/24/2022).

[2] Alexandros Beskos et al. "Optimal Tuning of the Hybrid Monte Carlo Algorithm". In: *Bernoulli* 19.5A (Nov. 2013), pp. 1501–1534. ISSN: 1350-7265. DOI: 10.3150/12-BEJ414.

[3] Michael Betancourt. "A conceptual introduction to Hamiltonian Monte Carlo". In: *arXiv:1701.02434 [stat]* (July 2018). URL: http://arxiv.org/abs/1701.02434.

[4] Vladimir E. Bondarenko et al. "Computer model of action potential of mouse ventricular myocytes". eng. In: *American Journal of Physiology. Heart and Circulatory Physiology* 287.3 (Sept. 2004), H1378–1403. ISSN: 0363-6135. DOI: 10.1152/ajpheart.00185.2003.

[5] Corina T. Bot et al. "Rapid genetic algorithm optimization of a mouse computational model: benefits for anthropomorphization of neonatal mouse cardiomyocytes". eng. In: *Frontiers in Physiology* 3 (2012), p. 421. ISSN: 1664-042X. DOI: 10.3389/fphys.2012.00421.

[6] P. M. Boyle et al. "Computationally guided personalized targeted ablation of persistent atrial fibrillation". In: *Nature Biomedical Engineering* 3.11 (2019), pp. 870–879.

[7] Oliver J. Britton et al. "Experimentally calibrated population of models predicts and explains intersubject variability in cardiac cellular electrophysiology". eng. In: *Proceedings of the National Academy of Sciences of the United States of*

America 110.23 (June 2013), E2098–2105. ISSN: 1091-6490. DOI: 10.1073/pnas.1304382110.

[8] Steve Brooks et al. *Handbook of Markov chain Monte Carlo*. CRC press, 2011.

[9] Alfonso Bueno-Orovio, Elizabeth M. Cherry, and Flavio H. Fenton. "Minimal model for human ventricular action potentials in tissue". en. In: *Journal of Theoretical Biology* 253.3 (Aug. 2008), pp. 544–560. ISSN: 0022-5193. DOI: 10.1016/j.jtbi.2008.03.029. URL: https://www.sciencedirect.com/science/article/pii/S0022519308001690.

[10] W. L. Buntine. *Operations for Learning with Graphical Models*. Number: arXiv:cs/9412102 arXiv:cs/9412102. Nov. 1994. DOI: 10.48550/arXiv.cs/9412102. URL: http://arxiv.org/abs/cs/9412102 (visited on 06/15/2022).

[11] Darby I. Cairns, Flavio H. Fenton, and E. M. Cherry. "Efficient parameterization of cardiac action potential models using a genetic algorithm". In: *Chaos* 27.9 (Sept. 2017), p. 093922. ISSN: 1054-1500. DOI: 10.1063/1.5000354. URL: https://aip.scitation.org/doi/10.1063/1.5000354.

[12] Bob Carpenter et al. "Stan: A probabilistic programming language". en. In: *Journal of Statistical Software* 76 (Jan. 2017), pp. 1–32. ISSN: 1548-7660. DOI: 10.18637/jss.v076.i01. URL: https://doi.org/10.18637/jss.v076.i01.

[13] Eugene T. Y. Chang, Mark Strong, and Richard H. Clayton. "Bayesian Sensitivity Analysis of a Cardiac Cell Model Using a Gaussian Process Emulator". en. In: *PLOS ONE* 10.6 (June 2015), e0130252. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0130252. URL: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0130252.

[14] Kelly C. Chang et al. "Uncertainty quantification reveals the importance of data variability and experimental design considerations for in silico proarrhythmia risk assessment". In: *Frontiers in Physiology* 8 (2017), p. 917. ISSN: 1664-042X. DOI: 10.3389/fphys.2017.00917. URL: https://www.frontiersin.org/article/10.3389/fphys.2017.00917.

[15] E. Cherry and F. Fenton. "A tale of two dogs: analyzing two models of canine ventricular electrophysiology." In: *American journal of physiology. Heart and circulatory physiology* (2007). DOI: 10.1152/AJPHEART.00955.2006.

[16] Siddhartha Chib and Edward Greenberg. "Understanding the Metropolis-Hastings Algorithm". In: *The American Statistician* 49.4 (1995), pp. 327–335. ISSN: 0003-1305. DOI: 10.2307/2684568. URL: https://www.jstor.org/stable/2684568.

[17] Davide Chicco, Matthijs J. Warrens, and Giuseppe Jurman. "The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation". In: *PeerJ Computer Science* 7 (July 2021), e623. ISSN: 2376-5992. DOI: 10.7717/peerj-cs.623. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8279135/.

[18] C. E. Clancy and Y. Rudy. "Linking a genetic defect to its cellular phenotype in a cardiac arrhythmia". eng. In: *Nature* 400.6744 (Aug. 1999), pp. 566–569. ISSN: 0028-0836. DOI: 10.1038/23034.

[19] Sam Coveney and Richard H. Clayton. "Fitting two human atrial cell models to experimental data using Bayesian history matching". eng. In: *Progress in Biophysics and Molecular Biology* 139 (Nov. 2018), pp. 43–58. ISSN: 1873-1732. DOI: 10.1016/j.pbiomolbio.2018.08.001.

[20] Dávid Csercsik, Katalin M. Hangos, and Gábor Szederkényi. "Identifiability analysis and parameter estimation of a single Hodgkin–Huxley type voltage dependent ion channel under voltage step measurement conditions". en. In: *Neurocomputing* 77.1 (Feb. 2012), pp. 178–188. ISSN: 0925-2312. DOI: 10.1016/j.neucom.2011.09.006. URL: http://www.sciencedirect.com/science/article/pii/S0925231211005145.

[21] Aidan C. Daly et al. "Comparing two sequential Monte Carlo samplers for exact and approximate Bayesian inference on biological models". In: *Journal of The Royal Society Interface* 14.134 (Sept. 2017), p. 20170340. DOI: 10.1098/rsif.2017.0340. URL: https://royalsocietypublishing.org/doi/10.1098/rsif.2017.0340.

[22] Aidan C. Daly et al. "Hodgkin-Huxley revisited: reparametrization and identifiability analysis of the classic action potential model with approximate Bayesian methods". eng. In: *Royal Society Open Science* 2.12 (Dec. 2015), p. 150499. ISSN: 2054-5703. DOI: 10.1098/rsos.150499.

[23] Aidan C. Daly et al. "Inference-based assessment of parameter identifiability in nonlinear biological models". In: *Journal of The Royal Society Interface* 15.144 (July 2018). Publisher: Royal Society, p. 20180318. DOI: 10.1098/rsif.2018.0318. URL: https://royalsocietypublishing.org/doi/10.1098/rsif.2018.0318 (visited on 06/24/2022).

[24] Aidan C. Daly et al. "Inference-based assessment of parameter identifiability in nonlinear biological models". In: *Journal of The Royal Society Interface* 15.144 (July 2018), p. 20180318. DOI: 10.1098/rsif.2018.0318. URL: https://royalsocietypublishing.org/doi/full/10.1098/rsif.2018.0318.

[25] Pierre Del Moral, Arnaud Doucet, and Ajay Jasra. "An adaptive sequential Monte Carlo method for approximate Bayesian computation". en. In: *Statistics and Computing* 22.5 (Sept. 2012), pp. 1009–1020. ISSN: 1573-1375. DOI: 10.1007/s11222-011-9271-y. URL: https://doi.org/10.1007/s11222-011-9271-y.

[26] D. DiFrancesco and D. Noble. "A model of cardiac electrical activity incorporating ionic pumps and concentration changes". eng. In: *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 307.1133 (Jan. 1985), pp. 353–398. ISSN: 0962-8436. DOI: 10.1098/rstb.1985.0001.

[27] Socrates Dokos and Nigel H. Lovell. "Parameter estimation in cardiac ionic models". en. In: *Progress in Biophysics and Molecular Biology*. Modelling Cellular and Tissue Function 85.2 (June 2004), pp. 407–431. ISSN: 0079-6107. DOI: 10.1016/j.pbiomolbio.2004.02.002. URL: https://www.sciencedirect.com/science/article/pii/S0079610704000306.

[28] Simon Duane et al. "Hybrid Monte Carlo". en. In: *Physics Letters B* 195.2 (Sept. 1987), pp. 216–222. ISSN: 0370-2693. DOI: 10.1016/0370-2693(87)91197-X. URL: https://www.sciencedirect.com/science/article/pii/037026938791197X.

[29] Alan Fabbri et al. "Computational analysis of the human sinus node action potential: model development and effects of mutations". In: *The Journal of Physiology* 595.7 (2017), pp. 2365–2396. DOI: https://doi.org/10.1113/JP273259. eprint: https://physoc.onlinelibrary.wiley.com/doi/pdf/10.1113/JP273259. URL: https://physoc.onlinelibrary.wiley.com/doi/abs/10.1113/JP273259.

[30] F H Fenton and E M Cherry. "Models of cardiac cell". In: *Scholarpedia* 3.8 (2008), p. 1868.

[31] Flavio Fenton and Alain Karma. "Vortex dynamics in three-dimensional continuous myocardium with fiber rotation: Filament instability and fibrillation". In: *Chaos* 8.1 (Mar. 1998), pp. 20–47. ISSN: 1054-1500. DOI: 10.1063/1.166311. URL: https://aip.scitation.org/doi/10.1063/1.166311.

[32] Flavio H. Fenton et al. "Multiple mechanisms of spiral wave breakup in a model of cardiac electrical activity". eng. In: *Chaos* 12.3 (Sept. 2002), pp. 852–892. ISSN: 1089-7682. DOI: 10.1063/1.1504242.

[33] Dani Gamerman and Hedibert F. Lopes. *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference, Second Edition*. en. May 2006. ISBN: 978-1-4822-9642-6.

[34] Alan E. Gelfand and Adrian F. M. Smith. "Sampling-Based Approaches to Calculating Marginal Densities". In: *Journal of the American Statistical Association* 85.410 (June 1990). Publisher: Taylor & Francis _eprint: https://www.tandfonline.com/doi/pdf/10.1080/
pp. 398–409. ISSN: 0162-1459. DOI: 10.1080/01621459.1990.10476213. URL: https://www.tandfonline.com/doi/abs/10.1080/01621459.1990.10476213 (visited on 04/25/2022).

[35] Andrew Gelman et al. *Bayesian Data Analysis*. English. 3rd edition. Boca Raton: Chapman and Hall/CRC, Nov. 2013. ISBN: 978-1-4398-4095-5.

[36] Ryan Giordano, Tamara Broderick, and Michael I. Jordan. *Covariances, Robustness, and Variational Bayes*. arXiv:1709.02536 [stat]. Oct. 2018. DOI: 10.48550/arXiv.1709.02536. URL: http://arxiv.org/abs/1709.02536 (visited on 06/30/2022).

[37] Nico S. Gorbach, Stefan Bauer, and Joachim M. Buhmann. "Scalable Variational Inference for Dynamical Systems". In: *Proceedings of the 31st International Conference on Neural Information Processing Systems*. NIPS'17. Long Beach, California, USA: Curran Associates Inc., 2017, pp. 4809–4818. ISBN: 9781510860964.

[38] Joseph L Greenstein and Raimond L Winslow. "An integrative model of the cardiac ventricular myocyte incorporating local control of Ca2+ release." In: *Biophysical Journal* 83.6 (Dec. 2002), pp. 2918–2945. ISSN: 0006-3495. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1201479/ (visited on 06/24/2022).

[39] Willemijn Groenendaal et al. "Cell-specific cardiac electrophysiology models". eng. In: *PLoS Computational Biology* 11.4 (Apr. 2015), e1004242. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1004242.

[40] Peter Guttorp and Alexandra M. Schmidt. "Covariance Structure of Spatial and Spatiotemporal Processes". In: *WIREs Computational Statistics* 5.4 (2013), pp. 279–287. ISSN: 1939-0068. DOI: 10.1002/wics.1259.

[41] S. Göktepe and E. Kuhl. "Computational modeling of cardiac electrophysiology: A novel finite element approach". en. In: *International Journal for Numerical Methods in Engineering* 79.2 (2009). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/nme.2571, pp. 156–178. ISSN: 1097-0207. DOI: 10.1002/nme.2571. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/nme.2571 (visited on 06/14/2022).

[42] W. K. Hastings. "Monte Carlo sampling methods using Markov chains and their applications". In: *Biometrika* 57.1 (Apr. 1970), pp. 97–109. ISSN: 0006-3444. DOI: 10.1093/biomet/57.1.97. URL: https://doi.org/10.1093/biomet/57.1.97 (visited on 04/25/2022).

[43] D. W. Hilgemann and Denis Noble. "Excitation-contraction coupling and extracellular calcium transients in rabbit atrium: reconstruction of basic cellular mechanisms". In: *Proceedings of the Royal Society of London. Series B. Biological Sciences* 230.1259 (Mar. 1987). Publisher: Royal Society, pp. 163–205. DOI: 10.1098/rspb.1987.0015. URL: https://royalsocietypublishing.org/doi/10.1098/rspb.1987.0015 (visited on 06/24/2022).

[44] A. L. Hodgkin and A. F. Huxley. "A quantitative description of membrane current and its application to conduction and excitation in nerve". In: *The Journal of Physiology* 117.4 (1952), pp. 500–544. DOI: https://doi.org/10.1113/jphysiol.1952.sp004764. eprint: https://physoc.onlinelibrary.wiley.com/doi/pdf/10.1113/jphysiol.1952.sp004764. URL: https://physoc.onlinelibrary.wiley.com/doi/abs/10.1113/jphysiol.1952.sp004764.

[45] Matthew D. Hoffman and Andrew Gelman. "The No-U-turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo". In: *The Journal of Machine Learning Research* 15.1 (Jan. 2014), pp. 1593–1623. ISSN: 1532-4435.

[46] P. J. Hunter, P. A. McNaughton, and D. Noble. "Analytical models of propagation in excitable cells". en. In: *Progress in Biophysics and Molecular Biology* 30 (Jan. 1976), pp. 99–144. ISSN: 0079-6107. DOI: 10.1016/0079-6107(76)90007-9. URL: https://www.sciencedirect.com/science/article/pii/0079610776900079 (visited on 06/24/2022).

[47] S. Inada et al. "One-dimensional mathematical model of the atrioventricular node including atrio-nodal, nodal, and nodal-his cells". eng. In: *Biophysical Journal* 97.8 (Oct. 2009), pp. 2117–2127. ISSN: 1542-0086. DOI: 10.1016/j.bpj.2009.06.056.

[48] Vivek Iyer, Reza Mazhari, and Raimond L. Winslow. "A Computational Model of the Human Left-Ventricular Epicardial Myocyte". In: *Biophysical Journal* 87.3 (Sept. 2004), pp. 1507–1525. ISSN: 0006-3495. DOI: 10.1529/biophysj.104.043299. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1304558/ (visited on 06/24/2022).

[49] Branimir K. Hackenberger. "Bayes or not Bayes, is this the question?" In: *Croatian Medical Journal* 60.1 (Feb. 2019), pp. 50–52. ISSN: 0353-9504. DOI: 10.3325/cmj.2019.60.50. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6406060/ (visited on 06/27/2022).

[50] Alain Karma. "Spiral breakup in model equations of action potential propagation in cardiac tissue". In: *Physical Review Letters* 71.7 (Aug. 1993). Publisher: American Physical Society, pp. 1103–1106. DOI: 10.1103/PhysRevLett.71.1103.

URL: https://link.aps.org/doi/10.1103/PhysRevLett.71.1103 (visited on 06/24/2022).

[51] Trine Krogh-Madsen et al. "Global Optimization of Ventricular Myocyte Model to Multi-Variable Objective Improves Predictions of Drug-Induced Torsades de Pointes". In: *Frontiers in Physiology* 8 (2017), p. 1059. ISSN: 1664-042X. DOI: 10.3389/fphys.2017.01059. URL: https://www.frontiersin.org/article/10.3389/fphys.2017.01059.

[52] Alp Kucukelbir et al. "Automatic Differentiation Variational Inference". In: *arXiv:1603.00788 [cs, stat]* (Mar. 2016). arXiv: 1603.00788 [cs, stat].

[53] Axel Loewe et al. "Parameter estimation of ion current formulations requires hybrid optimization approach to be both accurate and reliable". eng. In: *Frontiers in Bioengineering and Biotechnology* 3 (2015), p. 209. ISSN: 2296-4185. DOI: 10.3389/fbioe.2015.00209.

[54] C. H. Luo and Y. Rudy. "A dynamic model of the cardiac ventricular action potential. II. Afterdepolarizations, triggered activity, and potentiation". eng. In: *Circulation Research* 74.6 (June 1994), pp. 1097–1113. ISSN: 0009-7330. DOI: 10.1161/01.res.74.6.1097.

[55] C H Luo and Y Rudy. "A model of the ventricular cardiac action potential. Depolarization, repolarization, and their interaction." In: *Circulation Research* 68.6 (June 1991). Publisher: American Heart Association, pp. 1501–1526. DOI: 10.1161/01.RES.68.6.1501. URL: https://www.ahajournals.org/doi/10.1161/01.res.68.6.1501 (visited on 06/24/2022).

[56] Stefan Luther et al. "Low-energy control of electrical turbulence in the heart". eng. In: *Nature* 475.7355 (July 2011), pp. 235–239. ISSN: 1476-4687. DOI: 10.1038/nature10216.

[57] Berndt Lüderitz. "Historical perspectives of cardiac electrophysiology". eng. In: *Hellenic journal of cardiology: HJC = Hellenike kardiologike epitheorese* 50.1 (Feb. 2009), pp. 3–16. ISSN: 1109-9666.

[58] Stefan A. Mann et al. "Convergence of models of human ventricular myocyte electrophysiology after global optimization to recapitulate clinical long QT phenotypes". In: *Journal of Molecular and Cellular Cardiology* 100 (2016), pp. 25–34. ISSN: 0022-2828. DOI: https://doi.org/10.1016/j.yjmcc.2016.09.011. URL: https://www.sciencedirect.com/science/article/pii/S0022282816303637.

[59] Charles C. Margossian, Yi Zhang, and William R. Gillespie. "Flexible and efficient Bayesian pharmacometrics modeling using Stan and Torsten, Part I". en. In: *arXiv:2109.10184 [stat]* (Nov. 2021). URL: http://arxiv.org/abs/2109.10184.

[60] Jean-Michel Marin et al. "Approximate Bayesian computational methods". en. In: *Statistics and Computing* 22.6 (Nov. 2012), pp. 1167–1180. ISSN: 1573-1375. DOI: 10.1007/s11222-011-9288-2. URL: https://doi.org/10.1007/s11222-011-9288-2.

[61] Jonathan C. Mattingly, Natesh S. Pillai, and Andrew M. Stuart. "Diffusion Limits of the Random Walk Metropolis Algorithm in High Dimensions". In: *The Annals of Applied Probability* 22.3 (June 2012), pp. 881–930. ISSN: 1050-5164. DOI: 10.1214/10-AAP754. arXiv: 1003.4306.

[62] R. E. McAllister, D. Noble, and R. W. Tsien. "Reconstruction of the electrical activity of cardiac Purkinje fibres". eng. In: *The Journal of Physiology* 251.1 (Sept. 1975), pp. 1–59. ISSN: 0022-3751. DOI: 10.1113/jphysiol.1975.sp011080.

[63] Cody McRae. "Bayesian inference in nonlinear differential equation models". In: *Australian Mathematical Sciences Institute* (2014). URL: http://www.vrs.amsi.org.au/wp-content/uploads/sites/6/2014/09/MCRAE_Cody_Research_Paper.pdf.

[64] Nicholas Metropolis et al. "Equation of State Calculations by Fast Computing Machines". In: *The Journal of Chemical Physics* 21.6 (June 1953). Publisher: American Institute of Physics, pp. 1087–1092. ISSN: 0021-9606. DOI: 10.1063/1.1699114. URL: https://aip.scitation.org/doi/abs/10.1063/1.1699114 (visited on 04/25/2022).

[65] Gary R. Mirams et al. "Uncertainty and variability in computational and mathematical models of cardiac physiology: Uncertainty and variability in cardiac models". en. In: *The Journal of Physiology* 594.23 (Dec. 2016), pp. 6833–6847. ISSN: 00223751. DOI: 10.1113/JP271671. URL: http://doi.wiley.com/10.1113/JP271671 (visited on 05/02/2022).

[66] Colleen C. Mitchell and David G. Schaeffer. "A two-current model for the dynamics of cardiac membrane". eng. In: *Bulletin of Mathematical Biology* 65.5 (Sept. 2003), pp. 767–793. ISSN: 0092-8240. DOI: 10.1016/S0092-8240(03)00041-7.

[67] Cole C. Monnahan, James T. Thorson, and Trevor A. Branch. "Faster estimation of Bayesian models in ecology using Hamiltonian Monte Carlo". en. In: *Methods in Ecology and Evolution* 8.3 (2017), pp. 339–348. ISSN: 2041-210X. DOI: 10.1111/2041-210X.12681. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/2041-210X.12681.

[68] Sherry L Murphy et al. "Mortality in the united states, 2017". In: (2018).

[69] Radford M. Neal. "MCMC using Hamiltonian dynamics". In: *Handbook of Markov chain Monte Carlo*. Chapman & Hall/CRC Handb. Mod. Stat. Methods. 2011, pp. 113–162.

[70] S. A. Niederer et al. "Creation and application of virtual patient cohorts of heart models". In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 378.2173 (June 2020), p. 20190558. DOI: 10.1098/rsta.2019.0558. URL: https://royalsocietypublishing.org/doi/full/10.1098/rsta.2019.0558.

[71] A. Nieto Ramos et al. "Quantifying distributions of parameters for cardiac action potential models using the Hamiltonian Monte Carlo method". In: *Computing in Cardiology* 48 (2021), pp. 9662836–1 –9662836–4.

[72] Alejandro Nieto-Ramos. "Estimación Bayesiana en Modelos Farmacocinéticos". MA thesis. Mexico: Universidad Autónoma Metropolitana, 2017.

[73] D. Noble. "A modification of the Hodgkin—Huxley equations applicable to Purkinje fibre action and pacemaker potentials". In: *The Journal of Physiology* 160.2

(1962), pp. 317–352. DOI: https://doi.org/10.1113/jphysiol.1962.sp006849. eprint: https://physoc.onlinelibrary.wiley.com/doi/pdf/10.1113/jphysiol.1962.sp006849. URL: https://physoc.onlinelibrary.wiley.com/doi/abs/10.1113/jphysiol.1962.sp006849.

[74] Denis Noble, Alan Garny, and Penelope J. Noble. "How the Hodgkin-Huxley equations inspired the Cardiac Physiome Project". eng. In: *The Journal of Physiology* 590.11 (June 2012), pp. 2613–2628. ISSN: 1469-7793. DOI: 10.1113/jphysiol.2011.224238.

[75] Thomas O'Hara et al. "Simulation of the undiseased human cardiac ventricular action potential: Model formulation and experimental validation". In: *PLoS Comput Biol* 7.5 (May 2011), e1002061.

[76] S. V. Pandit et al. "A mathematical model of action potential heterogeneity in adult rat left ventricular myocytes". eng. In: *Biophysical Journal* 81.6 (Dec. 2001), pp. 3029–3051. ISSN: 0006-3495. DOI: 10.1016/S0006-3495(01)75943-7.

[77] J. M. Pastore et al. "Mechanism linking T-wave alternans to the genesis of cardiac fibrillation". eng. In: *Circulation* 99.10 (Mar. 1999), pp. 1385–1394. ISSN: 0009-7322. DOI: 10.1161/01.cir.99.10.1385.

[78] Pras Pathmanathan et al. "Uncertainty quantification of fast sodium current steady-state inactivation for multi-scale models of cardiac electrophysiology". ENG. In: *Progress in Biophysics and Molecular Biology* 117.1 (Jan. 2015), pp. 4–18. ISSN: 1873-1732. DOI: 10.1016/j.pbiomolbio.2015.01.008.

[79] Zhilin Qu et al. "Nonlinear and stochastic dynamics in the heart". In: *Physics Reports* 543.2 (2014). Nonlinear and Stochastic Dynamics in the Heart, pp. 61–162. ISSN: 0370-1573. DOI: https://doi.org/10.1016/j.physrep.2014.05.002. URL: https://www.sciencedirect.com/science/article/pii/S037015731400204X.

[80] J. O. Ramsay et al. "Parameter estimation for differential equations: a generalized smoothing approach". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 69.5 (2007), pp. 741–796. DOI: https://doi.org/10.1111/j.1467-9868.2007.00610.x. eprint: https://rss.onlinelibrary.wiley.

com/doi/pdf/10.1111/j.1467-9868.2007.00610.x. URL: https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9868.2007.00610.x.

[81] Carl Edward Rasmussen and Christopher K. I. Williams. *Gaussian Processes for Machine Learning*. en. Google-Books-ID: Tr34DwAAQBAJ. MIT Press, Nov. 2005. ISBN: 978-0-262-18253-9.

[82] David Rosenbaum and Jose Jalife, eds. *Optical Mapping of Cardiac Excitation and Arrhythmias*. English. 1st edition. Armonk, NY: Wiley-Blackwell, Dec. 2001. ISBN: 978-0-87993-481-1.

[83] Tim Salimans, Diederik P. Kingma, and Max Welling. *Markov Chain Monte Carlo and Variational Inference: Bridging the Gap*. arXiv:1410.6460 [stat]. May 2015. DOI: 10.48550/arXiv.1410.6460. URL: http://arxiv.org/abs/1410.6460 (visited on 06/30/2022).

[84] K. J. Sampson et al. "A computational model of Purkinje fibre single cell electrophysiology: implications for the long QT syndrome". eng. In: *The Journal of Physiology* 588.Pt 14 (July 2010), pp. 2643–2655. ISSN: 1469-7793. DOI: 10.1113/jphysiol.2010.187328.

[85] Carlos Sánchez et al. "Inter-Subject Variability in Human Atrial Action Potential in Sinus Rhythm versus Chronic Atrial Fibrillation". en. In: *PLOS ONE* 9.8 (Aug. 2014). Publisher: Public Library of Science, e105897. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0105897. URL: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0105897 (visited on 04/25/2022).

[86] Nobuaki Sarai et al. "Role of individual ionic current systems in the SA node hypothesized by a model study". eng. In: *The Japanese Journal of Physiology* 53.2 (Apr. 2003), pp. 125–134. ISSN: 0021-521X. DOI: 10.2170/jjphysiol.53.125.

[87] Amrita X. Sarkar, David J. Christini, and Eric A. Sobie. "Exploiting mathematical models to illuminate electrophysiological variability between individuals". eng. In: *The Journal of Physiology* 590.11 (June 2012), pp. 2555–2567. ISSN: 1469-7793. DOI: 10.1113/jphysiol.2011.223313.

[88] Shahrokh Shahi et al. "Long-Time Prediction of Arrhythmic Cardiac Action Potentials Using Recurrent Neural Networks and Reservoir Computing". In: *Frontiers in Physiology* 12 (2021). ISSN: 1664-042X. URL: https://www.frontiersin.org/article/10.3389/fphys.2021.734178 (visited on 06/07/2022).

[89] Matthew S. Shotwell and Richard A. Gray. "Estimability analysis and optimal design in dynamic multi-scale models of cardiac electrophysiology". en. In: *Journal of Agricultural, Biological, and Environmental Statistics* 21.2 (June 2016), pp. 261–276. ISSN: 1537-2693. DOI: 10.1007/s13253-016-0244-7. URL: https://doi.org/10.1007/s13253-016-0244-7.

[90] Ivo Siekmann et al. "MCMC Estimation of Markov models for ion channels". In: *Biophysical Journal* 100.8 (Apr. 2011), pp. 1919–1929. ISSN: 0006-3495. DOI: 10.1016/j.bpj.2011.02.059. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3077709/.

[91] Scott A Sisson, Yanan Fan, and Mark Beaumont. *Handbook of approximate Bayesian computation*. CRC Press, 2018.

[92] Eric A. Sobie. "Parameter Sensitivity Analysis in Electrophysiological Models Using Multivariable Regression". In: *Biophysical Journal* 96.4 (2009), pp. 1264–1274. ISSN: 0006-3495. DOI: https://doi.org/10.1016/j.bpj.2008.10.056. URL: https://www.sciencedirect.com/science/article/pii/S0006349508032256.

[93] Stan Development Team. *Stan Modeling Language Users Guide and Reference Manual, version 2.29*. en. 2022. URL: //mc-stan.org/.

[94] Z. Syed et al. "Atrial cell action potential parameter fitting using genetic algorithms". eng. In: *Medical & Biological Engineering & Computing* 43.5 (Sept. 2005), pp. 561–571. ISSN: 0140-0118.

[95] Tina Toni et al. "Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems". In: *Journal of The Royal Society Interface* 6.31 (Feb. 2009), pp. 187–202. DOI: 10.1098/rsif.2008.0172. URL: https://royalsocietypublishing.org/doi/10.1098/rsif.2008.0172.

[96] Brandon M. Turner and Trisha Van Zandt. "A tutorial on approximate Bayesian computation". en. In: *Journal of Mathematical Psychology* 56.2 (Apr. 2012), pp. 69–85. ISSN: 0022-2496. DOI: 10.1016/j.jmp.2012.02.005. URL: https://www.sciencedirect.com/science/article/pii/S0022249612000272 (visited on 06/27/2022).

[97] K. H. W. J. ten Tusscher et al. "A model for human ventricular tissue". eng. In: *American Journal of Physiology. Heart and Circulatory Physiology* 286.4 (Apr. 2004), H1573–1589. ISSN: 0363-6135. DOI: 10.1152/ajpheart.00794.2003.

[98] Aki Vehtari et al. "Rank-normalization, folding, and localization: An improved $\widehat{R}$ for assessing convergence of MCMC". In: *Bayesian Analysis* 16.2 (June 2021). ISSN: 1936-0975. DOI: 10.1214/20-BA1221. URL: http://arxiv.org/abs/1903.08008.

[99] Ian Vernon et al. "Bayesian uncertainty analysis for complex systems biology models: emulation, global parameter searches and evaluation of gene functions". In: *BMC Systems Biology* 12.1 (Jan. 2018), p. 1. ISSN: 1752-0509. DOI: 10.1186/s12918-017-0484-3. URL: https://doi.org/10.1186/s12918-017-0484-3.

[100] Dominic G. Whittaker et al. "Calibration of ionic and cellular cardiac electrophysiology models". eng. In: *Wiley Interdisciplinary Reviews. Systems Biology and Medicine* 12.4 (July 2020), e1482. ISSN: 1939-005X. DOI: 10.1002/wsbm.1482.

[101] Raimond L. Winslow et al. "Mechanisms of Altered Excitation-Contraction Coupling in Canine Tachycardia-Induced Heart Failure, II". In: *Circulation Research* 84.5 (Mar. 1999). Publisher: American Heart Association, pp. 571–586. DOI: 10.1161/01.RES.84.5.571. URL: https://www.ahajournals.org/doi/full/10.1161/01.RES.84.5.571 (visited on 06/24/2022).

[102] Md Shakil Zaman et al. "Fast Posterior Estimation of Cardiac Electrophysiological Model Parameters via Bayesian Active Learning". In: *Frontiers in Physiology* 12 (Oct. 2021), p. 740306. ISSN: 1664-042X. DOI: 10.3389/fphys.2021.740306. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8573318/ (visited on 04/15/2022).

[103] Massimiliano Zaniboni et al. "How different two almost identical action potentials can be: A model study on cardiac repolarization". In: *Mathematical Biosciences* 228.1 (2010), pp. 56–70. ISSN: 0025-5564. DOI: https://doi.org/10.1016/j.mbs.2010.08.007. URL: https://www.sciencedirect.com/science/article/pii/S0025556410001392.