



This is a repository copy of *Fitting two human atrial cell models to experimental data using Bayesian history matching*.

White Rose Research Online URL for this paper:
<https://eprints.whiterose.ac.uk/134989/>

Version: Published Version

Article:

Coveney, S. and Clayton, R.H. orcid.org/0000-0002-8438-7518 (2018) Fitting two human atrial cell models to experimental data using Bayesian history matching. *Progress in Biophysics & Molecular Biology*, 139. pp. 43-58. ISSN 0079-6107

<https://doi.org/10.1016/j.pbiomolbio.2018.08.001>

Reuse

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:
<https://creativecommons.org/licenses/>

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

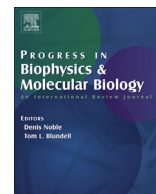


eprints@whiterose.ac.uk
<https://eprints.whiterose.ac.uk/>



Contents lists available at ScienceDirect

Progress in Biophysics and Molecular Biology

journal homepage: www.elsevier.com/locate/pbiomolbio

Fitting two human atrial cell models to experimental data using Bayesian history matching

Sam Coveney, Richard H. Clayton*

Insigneo Institute for in-silico Medicine and Department of Computer Science, University of Sheffield, UK

ARTICLE INFO

Article history:

Received 19 January 2018

Received in revised form

2 August 2018

Accepted 6 August 2018

Available online 24 August 2018

ABSTRACT

Cardiac cell models are potentially valuable tools for applications such as quantitative safety pharmacology, but have many parameters. Action potentials in real cardiac cells also vary from beat to beat, and from one cell to another. Calibrating cardiac cell models to experimental observations is difficult, because the parameter space is large and high-dimensional. In this study we have demonstrated the use of history matching to calibrate the maximum conductance of ion channels and exchangers in two detailed models of the human atrial action potential against measurements of action potential biomarkers. History matching is an approach developed in other modelling communities, based on constructing fast-running Gaussian process emulators of the model. Emulators were constructed from a small number of model runs (around 10^2), and then run many times ($> 10^6$) at low computational cost, each time with a different set of model parameters. Emulator outputs were compared with experimental biomarkers using an implausibility measure, which took into account experimental variance as well as emulator variance. By repeating this process, the region of non-implausible parameter space was iteratively reduced. Both cardiac cell models were successfully calibrated to experimental datasets, resulting in sets of parameters that could be sampled to produce variable action potentials. However, model parameters did not occupy a small range of values. Instead, the history matching process exposed inputs that can covary across a wide range and still be consistent with a particular biomarker. We also found correlations between some biomarkers, indicating a need for better descriptors of action potential shape.

© 2018 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Cardiac cell models have become valuable research tools, underpinning models of electrical excitation in cardiac tissue and increasingly applied to drug safety testing (Colatsky et al., 2016; Mirams et al., 2012). Cardiac cell models have become increasingly detailed, with components representing not only ion channels, pumps and exchangers in the cell membrane, but also Ca^{2+} storage and release (Fink et al., 2011); the number of parameters has increased correspondingly. Even the relatively simple Beeler-Reuter model of the ventricular action potential can be considered to have as many as 63 parameters (Dokos and Lovell, 2004).

For an ion channel model, these parameters include maximum conductances and parameters specifying the dynamics of transitions between states, such as open, closed, and inactive. During

model development these parameters are fitted to experimental data, which may have uncertain provenance (Niederer et al., 2009). Since action potential shape is determined by the overall balance of many different inward and outward currents, different combinations of currents can produce the same shapes. There are several published examples of identical action potentials produced by models with different parameter sets (Zaniboni et al., 2010; Sarkar and Sobie, 2010).

A further problem is that fitting of model parameters to experimental data on single ion channels does not usually take account of variability in experimental measurements (Pathmanathan et al., 2015), which is composed of natural variability and experimental errors. Cardiac action potentials are variable, with small variations from one beat to the next within the same cell and with larger variations between different cells (Zaniboni et al., 2000; Krogh-Madsen et al., 2015). This variability is important to take into account for safety-critical applications such as drug safety testing, and can be incorporated into parameter fitting by generating models of particular cells (Groenendaal et al., 2015).

* Corresponding author.

E-mail addresses: s.coveney@sheffield.ac.uk (S. Coveney), r.h.clayton@sheffield.ac.uk (R.H. Clayton).

Various approaches to global parameter inference for cardiac cell models have been described, and these include gradient descent (Dokos and Lovell, 2004), genetic algorithms (Groenendaal et al., 2015; Cairns et al., 2017), particle swarm (Loewe et al., 2015), multivariate regression (Sarkar and Sobie, 2010), and Markov chain Monte Carlo (Johnstone et al., 2015). Many of these approaches are computationally intensive, requiring large numbers of model evaluations. Another more general approach is to generate populations of models, where ranges of parameter values are established that produce action potential features consistent with a set of variable experimental observations (Britton et al., 2013; Muszkiewicz et al., 2015; Tixier et al., 2017; Lawson et al., 2018). A population of models recognises natural variability in model parameters such as maximal conductances. These parameters represent ion channel density in the cell membrane, which is regulated but variable (Balse and Boycott, 2017), and so is not a physical constant. However, the number of model evaluations used in these studies to generate an initial population of models is relatively small (typically of the order 10^4), and may not be sufficient to effectively explore the high dimensional parameter space.

An approach called history matching has been developed to address similar problems of parameter identification in models of galaxy formation (Vernon et al., 2010), disease transmission (Andrianakis et al., 2015), and plant physiology (Vernon et al., 2018). The computational model is supplemented by a fast-running emulator (surrogate model) such as a Gaussian Process (GP) (Rasmussen and Williams, 2006), which is built with simulation data obtained from the computational model. The emulator predicts the model output surface with an associated confidence, and can be rapidly evaluated (typically $> 10^6$ times) with different model inputs in order to thoroughly explore parameter space. History matching proceeds iteratively in a series of waves. For each wave, the computational model is run using sets of inputs sampled from the currently-known non-implausible parameter space. These model runs are used to train a set of emulators, which in turn are evaluated for a large number of samples of the current parameter space. The emulator outputs are compared with experimental observations, taking into account the variance (uncertainty) on the emulator predictions as well as experimental variability. The parameter space consistent with experimental observations is iteratively reduced at each wave. In this way, simulations runs are focused in the regions more consistent with experimental observations, making the process very efficient. The aim of the present study was to demonstrate the feasibility of history matching with GP emulators for calibrating two models of the human atrial action potential against previously published experimental data sets (Sánchez et al., 2014).

2. Background

2.1. Gaussian process emulators

A cardiac cell model can be thought of as a *simulator*, where *outputs* (either action potentials or a set of biomarkers e.g. action potential duration) depend on model parameters or *inputs* (e.g. maximum ion channel conductances). This relation can be described as:

$$\mathbf{y} = f_S(\mathbf{x}). \quad (1)$$

The simulator $f_S(\mathbf{x})$, for which we have no closed form expression, can be substituted with a fast running surrogate model $f(\mathbf{x})$ called an emulator. Given a set of model inputs and corresponding outputs, called *design data* $\{x_i, y_i\}$, the emulator is trained to reconstruct the output surface of the simulator, and so can make

rapid predictions for outputs \mathbf{y}^* from inputs \mathbf{x}^* that have not been tested with the simulator. A GP emulator, which we have previously used to emulate cardiac cell models (Chang et al., 2015; Johnstone et al., 2015), is a statistical model which also provides uncertainty on predictions. For a given input \mathbf{x}^* , the emulator output \mathbf{y}^* is composed of expected values for the outputs, as well as variances that encode the confidence of the emulator. It is also possible to designate the inputs \mathbf{x}^* to also be uncertain.

A common emulator configuration is a mean consisting of explicit basis functions combined with a zero-mean GP (Rasmussen and Williams, 2006):

$$f(\mathbf{x}) = m(\mathbf{x}) + g(\mathbf{x}); \quad (2)$$

where $m(\mathbf{x})$ is a mean function, taken here to be linear

$$m(\mathbf{x}) = h(\mathbf{x})^T \beta, \quad (3)$$

$$= \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p, \quad (4)$$

and $g(\mathbf{x})$ a zero-mean GP

$$g(\mathbf{x}) \sim \mathcal{GP}(0, k(\mathbf{x}, \mathbf{x}')), \quad (5)$$

with a radial basis function kernel k

$$k(\mathbf{x}, \mathbf{x}') = \sigma^2 \left(\exp \left[- \sum_{p=1}^P \left\{ \frac{(x_p - x'_p)^2}{\delta_p} \right\} \right] \right). \quad (6)$$

In these expressions the input vector $\mathbf{x} = (x_1, x_2, \dots, x_p)$ denotes P inputs, and the emulator is defined by three sets of hyperparameters; β and δ are each vectors of length P , and σ^2 is a scalar. Therefore the correlation between outputs y and y' depends on the distance between inputs x and x' . The hyperparameter values are optimised given a set of design data by maximizing the log-likelihood (Kennedy and O'Hagan, 2000). Design data are obtained by running the simulator for a parameter space filling set of inputs $\{\mathbf{x}\}$, designed to evenly sample the range of plausible values for each input.

Prediction of an emulator output given input vector \mathbf{x}^* takes the form of a posterior probability distribution, given by algebraic expressions (Oakley and O'Hagan, 2004) detailed in the supplementary information to Chang et al. (2015). This probability density is characterised by a posterior expectation $E^*[f(\mathbf{x}^*)]$ and a posterior variance $Var^*[f(\mathbf{x}^*)]$ on the output ($*$ represents quantities with respect to the emulator). For models with more than one output $n = 1..N$, a separate emulator can be constructed for each output y_n .

Since the emulator can be evaluated algebraically, it can be used to perform rapid uncertainty quantification and sensitivity analysis. Furthermore, in the case of normally distributed inputs, the mean and variance of the emulator output can be calculated explicitly (Oakley and O'Hagan, 2004). A variance-based sensitivity index, defined as the proportion of output variance attributable to the input variance, can be calculated to determine the influence that each input has on the output. In this study, we used a linear mean function $h(\mathbf{x})^T \beta = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p$ since this allowed direct calculation of sensitivity indices by our software (see 3.3).

2.2. History matching

A detailed background to the motivation and methodology of the history matching approach is given in Andrianakis et al. (2015) and Vernon et al. (2018). It is an iterative technique, which exploits fast running emulators to explore model input space efficiently,

with the aim of identifying regions of input space that produce outputs consistent with uncertain experimental observations. We designate the different model outputs by n , and build an emulator for each model output. At each iteration step, or *wave*, the emulators are evaluated at a large number of points in the input space and the emulator outputs compared with experimental observations using an implausibility measure:

$$I_n^2(\mathbf{x}) = \frac{(E^*[f_n(\mathbf{x})] - z_n)^2}{\text{Var}^*[f_n(\mathbf{x})] + \text{Var}(e_n) + \text{Var}(md)} \quad (7)$$

I_n^2 is the squared implausibility measure at the input space location \mathbf{x} for output n , which is the ratio of the square of the difference between the emulator mean $E^*[f_n(\mathbf{x})]$ and observations z_n , to the sum of the emulator variance $\text{Var}^*[f_n(\mathbf{x})]$, observation error $\text{Var}(e_n)$ and model discrepancy error $\text{Var}(md)$ (representing known discrepancy between the model and reality).

To distinguish implausible points in the input space, a threshold $I_{\text{threshold}}$ is introduced, so that if $\max[I_n(\mathbf{x})] > I_{\text{threshold}}$ (the maximum implausibility taken across all outputs for a given test input \mathbf{x}) then \mathbf{x} is considered implausible. This threshold or “cut-off” is commonly taken as 3.0 based on Pukelsheim's 3-sigma rule (Pukelsheim, 1994). A large value of $I_n(\mathbf{x})$ compared to the threshold denotes a point in input space that is implausible because the inputs \mathbf{x} produce an output ($E^*[f_n(\mathbf{x})]$) that deviates significantly from the experimental observation z_n , even accounting for the uncertainties expressed in the denominator. However, a small value of $I_n(\mathbf{x})$ compared to the threshold does not necessarily imply plausibility, since the emulator variance may be large. Rather, $\max[I_n(\mathbf{x})] \leq I_{\text{threshold}}$ means that the inputs \mathbf{x} are *non-implausible*.

History matching progressively reduces the region of non-implausible input space in waves. Implausibilities are calculated for a large number of input points \mathbf{x} within the space currently deemed non-implausible. We designate the input space remaining at the end of each wave as Not-Ruled-Out-Yet (NROY) input space (Andrianakis et al., 2015). In wave 1, the initial NROY input space is likely to be a hypercube extending beyond all plausible values of the inputs, determined by known constraints on the inputs or by expert knowledge. In later waves, NROY input space is determined by the non-implausible inputs found in the previous wave. For each wave, the simulator is evaluated for a space-filling design in NROY space, and these design data are used to construct the next wave of emulators. NROY input space can then be extensively evaluated with the emulators, and implausible input points discarded in order to reduce NROY space. As NROY space shrinks, the simulation data become denser in input space and so the emulators become more accurate and confident. Hence re-evaluation of NROY with the updated emulators will cause more points to be discarded. In this way, history matching can gradually refine regions of non-implausible input space using a limited number of simulation runs.

3. Methods

3.1. Cell models and model inputs

For the present study, we chose to investigate models of human atrial cells developed by Courtemanche et al. (1998) and Maleckar et al. (2009). Both models aim to simulate the human atrial action potential and represent a similar set of ion channel, pump and exchanger currents, but each model was developed using slightly different sets of experimental data and as a result the action potentials have different shapes (Wilhelms et al., 2012; Krogh-Madsen et al., 2015).

A simulator for each model was implemented using Matlab

(Mathworks Inc.) with code automatically generated from CellML (www.cellml.org). The model ordinary differential equations were solved using the Matlab ode15s adaptive solver for stiff systems, with relative tolerance set to 10^{-5} , absolute tolerance set to 10^{-3} and maximum time step set to 1 ms. In the Courtemanche model the intracellular concentrations of Na^+ and K^+ were held constant to avoid drift of intracellular ion concentrations arising from an imbalance of currents (Wilhelms et al., 2012). In the Courtemanche model, a new parameter fG_{Kur} was introduced. This input had a default value of 1.0, and was used to scale the I_{Kur} current.

A subset of model inputs were selected based on sensitivity analysis performed using emulators fitted to a larger set of design data, to identify inputs with a sensitivity index greater than 0.05. These eight inputs were then identified as most active. The default values for each input are listed in Table 1 and are discussed further below. All other model inputs were set to the default values as specified in the original model publications.

3.2. GP emulator design data

For each wave of history matching, a set of design data were generated by running the cell model simulators using a set of inputs sampled from NROY space within the ranges specified in Table 1. For the first wave, the inputs were selected with Latin Hypercube (LHC) sampling using the Matlab function `lhsdesign`, set to maximise the minimum distances between points over 1000 iterations. We selected LHC sampling to obtain good coverage of the input space, although other approaches such as orthogonal sampling may offer more efficient coverage (Bingham et al., 2009). Subsequent waves sampled inputs from the NROY region for design data inputs, and details of this procedure are given below. The number of design data required for a GP emulator is generally considered to be at least $10\times$ the number of inputs (Loeppky et al., 2009). We therefore chose to use 120 simulator runs for each wave of history matching, so that we could discard any runs that produced abnormal action potentials, described in more detail below.

The range for each input was initially set at 10% to 200% of the default value (i.e. $0.1\times$ to $2.0\times$ default), in order to bracket all likely variation of these parameters for the purposes of history matching. However the ranges for G_{Na} , G_{K1} , G_{Kur} , and $G_{\text{b,Ca}}$ were reduced in both models in order to reduce the number of simulator runs that resulted in pacemaking activity or a failure of the model to repolarise. Each simulator run comprised 40 beats at a cycle length of 1000 ms. For each simulator run, seven outputs were calculated from the final action potential, which correspond to the biomarkers used by Sánchez et al. (2014):

- dV/dt_{max} – Maximum slope of the action potential upstroke.
- V_{amp} – Amplitude of the action potential, measured as the difference between peak voltage and resting membrane potential.
- V_{20} – Membrane voltage measured at 20% of APD_{90} .
- APD_{20} , APD_{50} , and APD_{90} – Action potential duration (APD) at 20%, 50%, and 90% of repolarisation.
- RMP – Resting membrane potential, calculated as the average membrane voltage over a 10 ms period, 100 ms prior to the action potential upstroke.

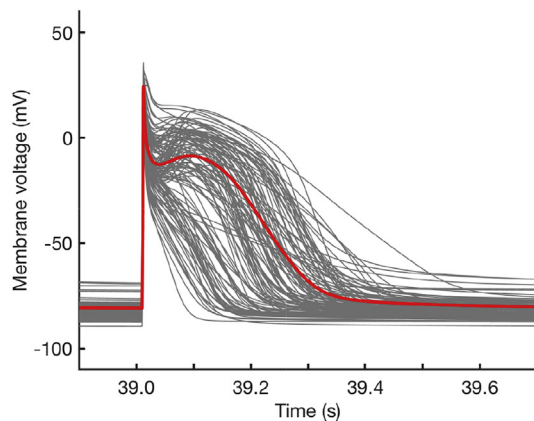
Fig. 1 shows the final action potential from the initial set of 120 simulator runs for each cell model. Superimposed on these traces is the final action potential from a simulator run with the inputs set to their default values. A wide range of action potential shapes are produced by simulator, and for the Maleckar model some simulator runs resulted in abnormal action potentials.

Simulator runs with spontaneous pacemaking activity or failure to repolarise indicated by $RMP > -50 \text{ mV}$ and/or $APD_{90} > 600 \text{ ms}$,

Table 1
Default values of each input, and the range used to sample inputs for emulator design data.

Courtemanche model				
Input	Default value	Range	Percentage of default	Units
G_{Na}	7.8	3.90–11.70	(50 – 150%)	nS/pF
G_{K1}	0.09	0.0675–0.1800	(75 – 200%)	nS/pF
G_{to}	0.1652	0.0165–0.3330	(10 – 200%)	nS/pF
fG_{Kur}	1.0	0.3–2.0	(30 – 200%)	none
G_{Kr}	0.0294	0.0029–0.0588	(10 – 200%)	nS/pF
$G_{Ca,L}$	0.1237	0.0124–0.2475	(10 – 200%)	nS/pF
$G_{b,Ca}$	0.0011	0.0001–0.0012	(10 – 110%)	nS/pF
$i_{NaK,Max}$	0.5993	0.0599–1.1987	(10 – 200%)	pA/pF
Maleckar model				
Input	Central value	Range		Units
P_{Na}	0.0018	0.0009–0.0027	(50 – 150%)	nL/s
G_{K1}	3.1	1.550–4.6500	(50 – 150%)	nS
G_t	8.250	0.825–16.500	(10 – 200%)	nS
G_{Kur}	2.250	0.675–4.500	(30 – 200%)	nS
G_{Kr}	0.5	0.05–1.00	(10 – 200%)	nS
$G_{Ca,L}$	6.750	0.675–13.500	(10 – 200%)	nS
$G_{b,Ca}$	0.0786	0.0079–0.0865	(10 – 110%)	nS
$i_{NaK,Max}$	68.55	6.855–137.100	(10 – 200%)	pA

(a) Courtemanche model



(b) Maleckar model

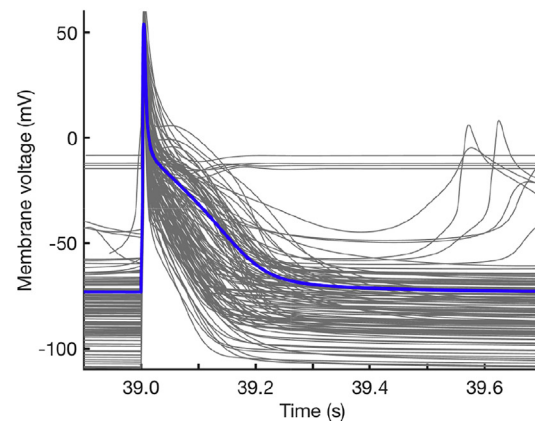


Fig. 1. Initial set of design data for (a) Courtemanche and (b) Maleckar models. Each panel shows the final action potential in a sequence of 40, with pacing at a cycle length of 1000 ms. Bold lines indicate the model behaviour for baseline values of the inputs, grey lines show design data.

or APD alternans defined as a difference in APD_{90} of successive beats of more than 5%, were classed as abnormal action potentials and were excluded from the design data used to train the emulators. The number of abnormal simulator runs was less than 12 in all waves of history matching, except for wave 2 with the Maleckar model where it was 25.

3.3. GP emulator fitting procedure

Fitting of the GP emulators, sensitivity analysis, and history matching were all done using “maGPy”, a Python package available with a GPL 3.0 license (<https://github.com/samcoveney/maGPy>).

A separate GP emulator was fitted to each of the eight model outputs identified above, with all of the model inputs considered to be ‘active’ (each model input corresponds to an index p in equation (6)). The emulator hyperparameters were optimised by maximising log-likelihood given the design data (Kennedy and O’Hagan, 2000), using the Python routine `scipy.optimize.minimize` with

method = L-BFGS-B implemented in the SciPy library. The optimisation was repeated 20 times with a different initial guess, and the solution with greatest log-likelihood was selected. Most repeats converged to the same log-likelihood, indicating that the most optimal solution had been found.

The quality of emulator fit was assessed by calculating variance-based sensitivity indices for all input-output combinations, and these indices are supplied as supplementary data. If the sum of sensitivity indices for a single output was close to 1, then this can indicate that output variability has been captured by the input variability assuming no interaction between inputs, which indicates goodness of fit. It is noteworthy that the emulator fits show improvement as history matching proceeds in each wave, because the training data becomes more dense in input space and because the output surface varies less across a smaller input space.

3.4. Experimental data

We used experimental datasets published by Sánchez et al. (2014), which comprise measurements made in cells obtained from patients with normal hearts in sinus rhythm (SR), and from patients with hearts that have undergone remodelling due to chronic atrial fibrillation (cAF). For each action potential biomarker in these datasets, the median was used as the observed output z_n for history matching, and the square of the median absolute deviation (MAD) as the experimental error variance $Var(e_n)$. We selected these measures because the datasets were not well described using the mean and standard deviations, due to outliers and long tails in the distributions which allowed apparent ‘matches’ which clearly did not represent the population data in any meaningful way. These median-based measures provide a much stricter and representative description of the population data. We set the model discrepancy term ($Var(md)$ in equation (7)) to zero, since we had no knowledge of how each cell model may systematically deviate from ‘real world’ outputs.

3.5. History matching procedure

Our history matching procedure was as follows:

1. If wave 1, select 120 simulator inputs with (maximin optimized) Latin Hypercube (LHC) sampling in the range described in Table 1. If wave > 1 , sample the previous wave NROY space to obtain 120 simulator inputs (see below for details).
2. Run simulator at each set of inputs and obtain corresponding outputs, discarding any abnormal action potentials. The remaining inputs and outputs were the design data for the current wave.
3. Fit emulators to the new design data, one emulator for each output. Design data for up to 4 previous waves were also included, in order to retain accurate predictions at the edges of NROY space.
4. If wave 1, run each emulator at each of 3,000,000 points in the input space, obtained from LHC in same ranges used for wave 1 design data. If wave > 1 , run emulator for all input points in the previous wave NROY space.
5. Calculate corresponding I_n for each output, and apply cut-off threshold to the highest maximum I_n to determine if input point lies in NROY space for current wave. If less than 100,000 points remain in this NROY space, then generate more non-implausible points (see below).
6. If wave < 11 , return to step 1

To generate additional points in NROY input space, we used an iterative ‘cloud’ method: for every point in NROY space, we generated new points by sampling from a multi-normal distribution centered on that point and scaled into the current range of the known NROY points, and then further scaled by a factor so that only about 20% of the new points were non-implausible (ensuring new points are sufficiently far from the generating points). We added these points to the NROY space and repeated the process until we had at least 100,000 points in NROY space.

The method of choosing a new space filling design in NROY space to use for more simulations for the next wave is a difficult and open problem. We used the python package diversipy (<https://pypi.python.org/pypi/diversipy>) for this, using the greedy_minmax routine for approximately 25% of selected points and psa_select for 75% of selected points, which balances filling the space against obtaining enough points around the edges.

The cut-off for implausibility used to designate regions of non-implausible space was gradually reduced with each wave

(Andrianakis et al., 2015; Vernon et al., 2018). The minimum threshold is usually considered to be 3.0, based on Pukelsheim's 3-sigma rule (Pukelsheim, 1994), and so we used this cutoff as our targeted ‘final’ NROY space. By starting with cutoff 4.0 and gradually reducing the cutoff with each wave (see Table 2), emulator inaccuracies in earlier waves were unlikely to result in regions of ‘true’ final NROY space (defined by cutoff 3.0) being discarded initially. Furthermore, this method allowed us to better focus simulation efforts in the region of interest, since large volumes input space can be deemed implausible even with cutoff 4.0 despite the relative sparsity of simulation data across input space in early waves.

A indication of when to stop history matching is that the emulator variance has fallen below the other variances, such that further waves do not significantly reduce NROY space. In our case, the experimental variance was the population variance (squared median absolute deviation) and was relatively large, so this criterion was not useful. We stopped history matching when NROY space had approximately converged.

4. Results

In the case of the Maleckar model, it was not possible to match V_{amp} to the median-based measures of either population, although we found it was possible to match to mean and variance based measures of V_{amp} . However, it was clear that matches to the mean and variance were only possible due to a very large variance caused by the spread of the population, and that these matches nonetheless failed to reproduce the population observations in any meaningful way. For this reason, V_{amp} was excluded from the history matching for the Maleckar model, and we conceded that it was not possible to match this output.

Table 2 shows the progressive reduction of the non-implausible region of NROY input space with each wave of history matching. The Maleckar model showed a rapid reduction of NROY space in the early waves. In both cell models for both data sets, there was a cumulative reduction of NROY by $\leq 0.1\%$ between waves 9 and 10, indicating that 10 waves were sufficient.

Fig. 2 shows plots of minimum implausibility projected onto the $G_{b,Ca} - I_{NaK,Max}$ input plane for the Courtemanche and Maleckar models during fitting to the SR dataset. The maximum implausibility across outputs was calculated for all input points as described above. The minimum of these values taken across all inputs falling within each 2D-pixel (hexagon) was then used to color the pixel (see colorbars). Grey regions indicate where no points have been tested in the current wave, since no points remained in NROY after the previous wave.

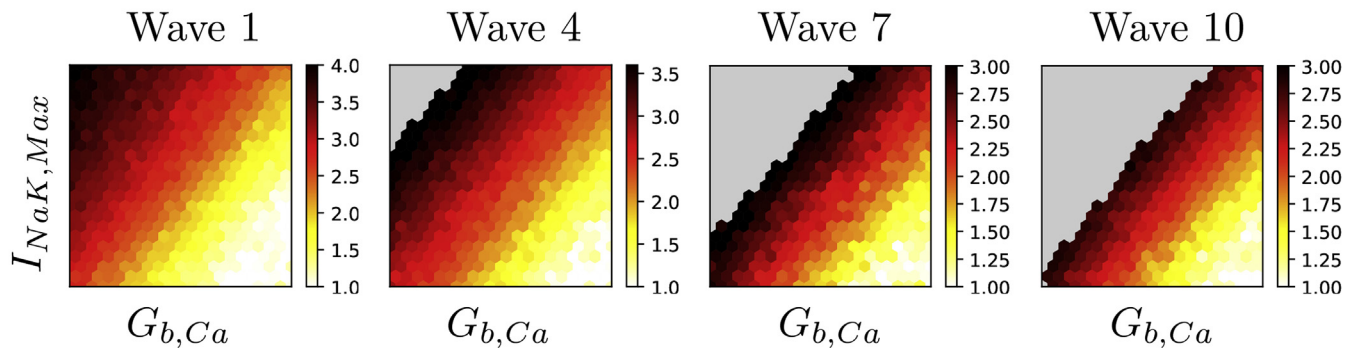
In the examples shown in Fig. 2, history matching indicates that the experimental observations were consistent with a correlation between $G_{b,Ca}$ and $I_{NaK,Max}$. For the Courtemanche model, all non-implausible $I_{NaK,Max}$ values fell below a threshold that was linear in $G_{b,Ca}$, such that for high $G_{b,Ca}$ values all $I_{NaK,Max}$ values were non-implausible, and low $G_{b,Ca}$ could only be paired with low values of $I_{NaK,Max}$. For the Maleckar model (Fig. 2a) the association is somewhat different, with low (high) values of $G_{b,Ca}$ consistent with low (high) values of $I_{NaK,Max}$, and higher values of $I_{NaK,Max}$ consistent with $G_{b,Ca}$ compared to the Courtemanche model. It is important to highlight that as history matching progresses, these plots did not show a convergence towards particular individual values of the outputs, but rather to broad regions of coupled $I_{NaK,Max} - G_{b,Ca}$ input space. We discuss this observation in more detail below.

As the NROY space was progressively reduced in size by both restricting the cutoff and reduction in emulator variance as simulator data became more densely located in input space, the set of simulator outputs obtained using inputs sampled from the NROY

Table 2
Implausibility cut-off threshold and size of NROY space at each wave of history matching. Numbers in brackets show cumulative reduction of NROY space.

Courtemanche model			
Wave	cutoff SR	NROY SR (%)	NROY cAF (%)
1	4.0	23.24 (23.24)	29.55 (29.55)
2	4.0	87.35 (20.30)	85.06 (25.14)
3	3.8	76.96 (15.62)	69.14 (17.38)
4	3.6	71.02 (11.09)	62.94 (10.94)
5	3.4	66.95 (7.43)	58.03 (6.35)
6	3.2	61.59 (4.58)	49.03 (3.11)
7	3.0	57.50 (2.63)	46.44 (1.45)
8	3.0	88.76 (2.34)	81.74 (1.18)
9	3.0	94.32 (2.20)	72.81 (0.86)
10	3.0	96.55 (2.13)	87.83 (0.76)
Maleckar model			
Wave	cutoff	NROY SR (%)	NROY cAF (%)
1	4.0	2.19 (2.19)	17.87 (17.87)
2	4.0	57.63 (1.26)	73.26 (13.09)
3	3.8	64.12 (0.80)	76.10 (9.96)
4	3.6	66.52 (0.53)	79.04 (7.87)
5	3.4	61.11 (0.32)	82.98 (6.53)
6	3.2	59.74 (0.19)	74.33 (4.86)
7	3.0	53.84 (0.10)	74.51 (3.62)
8	3.0	90.49 (0.09)	95.64 (3.46)
9	3.0	93.94 (0.08)	96.24 (3.33)
10	3.0	95.15 (0.08)	97.44 (3.25)

(a) Courtemanche model



(b) Maleckar model

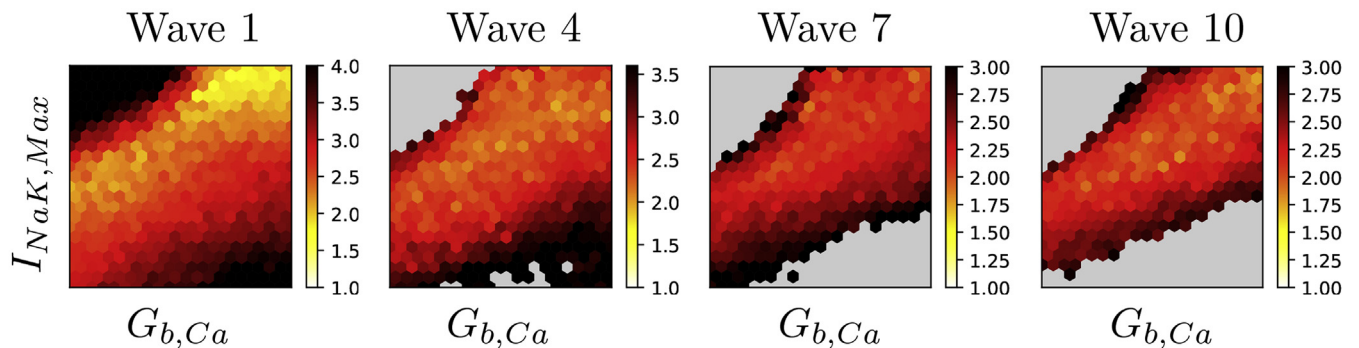


Fig. 2. : Implausibility measures for each cell model at waves 1, 4, 7, and 10, for SR dataset. Each plot pixel shows the minimum, across all tested points for the combination of $I_{NaK,Max}$ and $G_{b,Ca}$ in that pixel, of the maximum implausibility across all output. The maximum of the colourscale in each panel is the implausibility cut-off shown in table 2. Regions in grey contained no points to test for that wave of history matching, since all points in that region were discarded in previous waves. Each axis spans the initial range of each input.

space was gradually constrained. Fig. 3 shows the sets of 120 model outputs obtained from design data inputs for each wave in the Courtemanche model during history matching to the SR dataset, as

well as the experimental data from Sánchez et al. (2014). A similar plot for the Maleckar model is given in (Fig. 4). The data for wave 1 indicate the initial set of design data, and the data indicated by

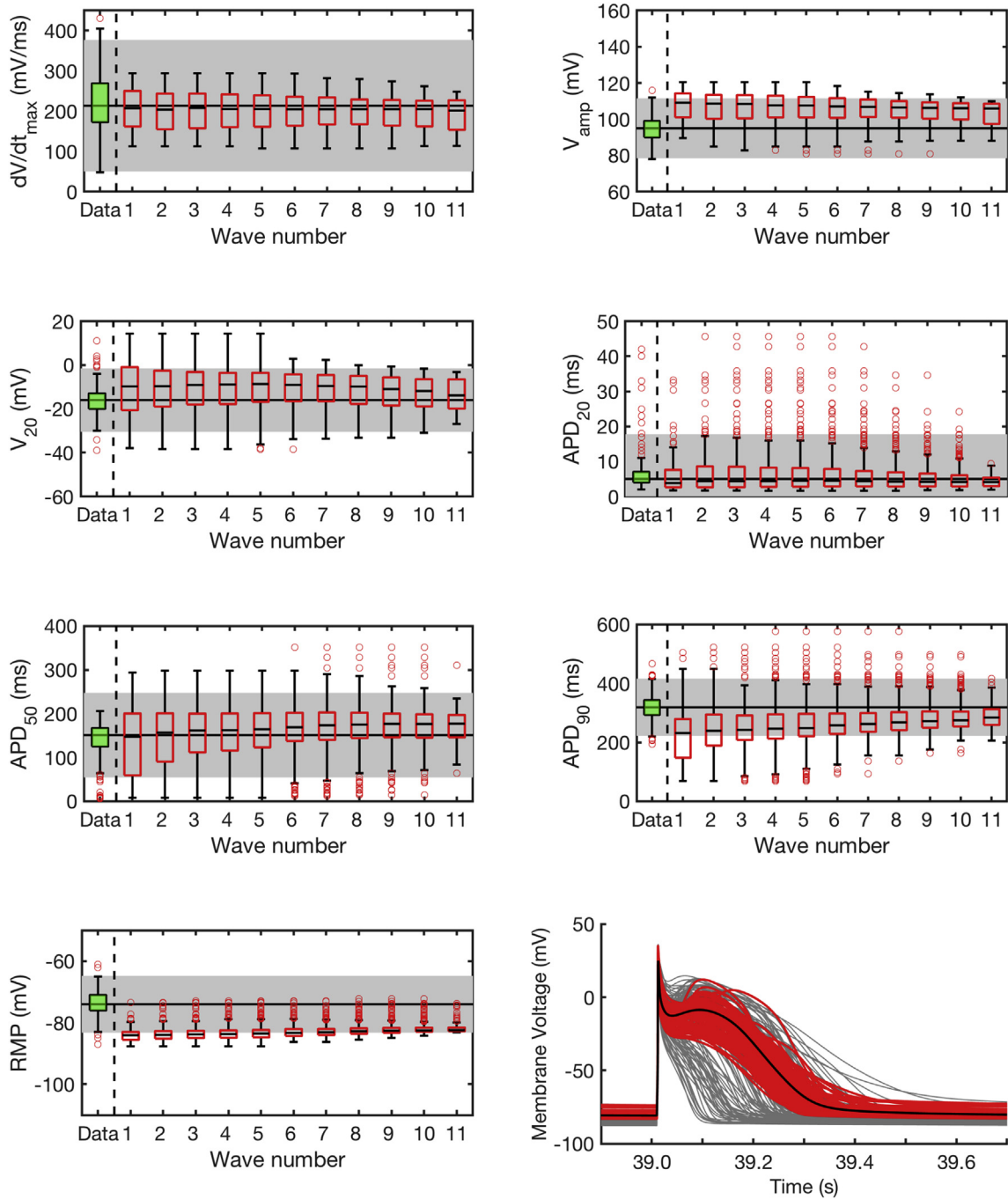


Fig. 3. Changes in design data outputs for each wave of history matching in the Courtemanche model. Experimental data are shown as a green filled box, model outputs are shown as red open boxes. Solid horizontal lines indicate the medians, and boxes denote first and third quartiles. Whiskers indicate either $1.5 \times$ the interquartile range, or the data point with the greatest deviation from the median, whichever is the smallest. The plot in the bottom right shows action potentials in the original design data (grey lines), and for the final wave design data (red lines). Superimposed in black is the action potential obtained using the default mode inputs.

wave 11 show the final set of design data obtained from sampling the NROY space produced from wave 10 of the history matching procedure. As expected there is a progressive reduction in the range of each output, and the action potentials produced by the simulator using as inputs the design data obtained from sampling the NROY space produced from wave 10 (bottom right in each figure) show convergence towards an action potential shape with spike and dome morphology and an APD_{90} of around 300 ms. However, the spread of action potential shapes and durations is greater for the Courtemanche compared to the Maleckar model, and a small

number of the Maleckar model solutions in the final wave show some depolarisation just prior to the action potential upstroke.

The history matching procedure resulted in compression of simulation data from each wave into the ranges defined by the median and MAD of the population. *A priori* a trend towards the median of each observation is not expected, because the outputs from the models co-varied. Nonetheless, the Courtemanche model outputs mostly seemed to converge towards the median of the experimental observations, while for the Maleckar model dV/dt_{max} and V_{20} deviated quite significantly away from the experimental

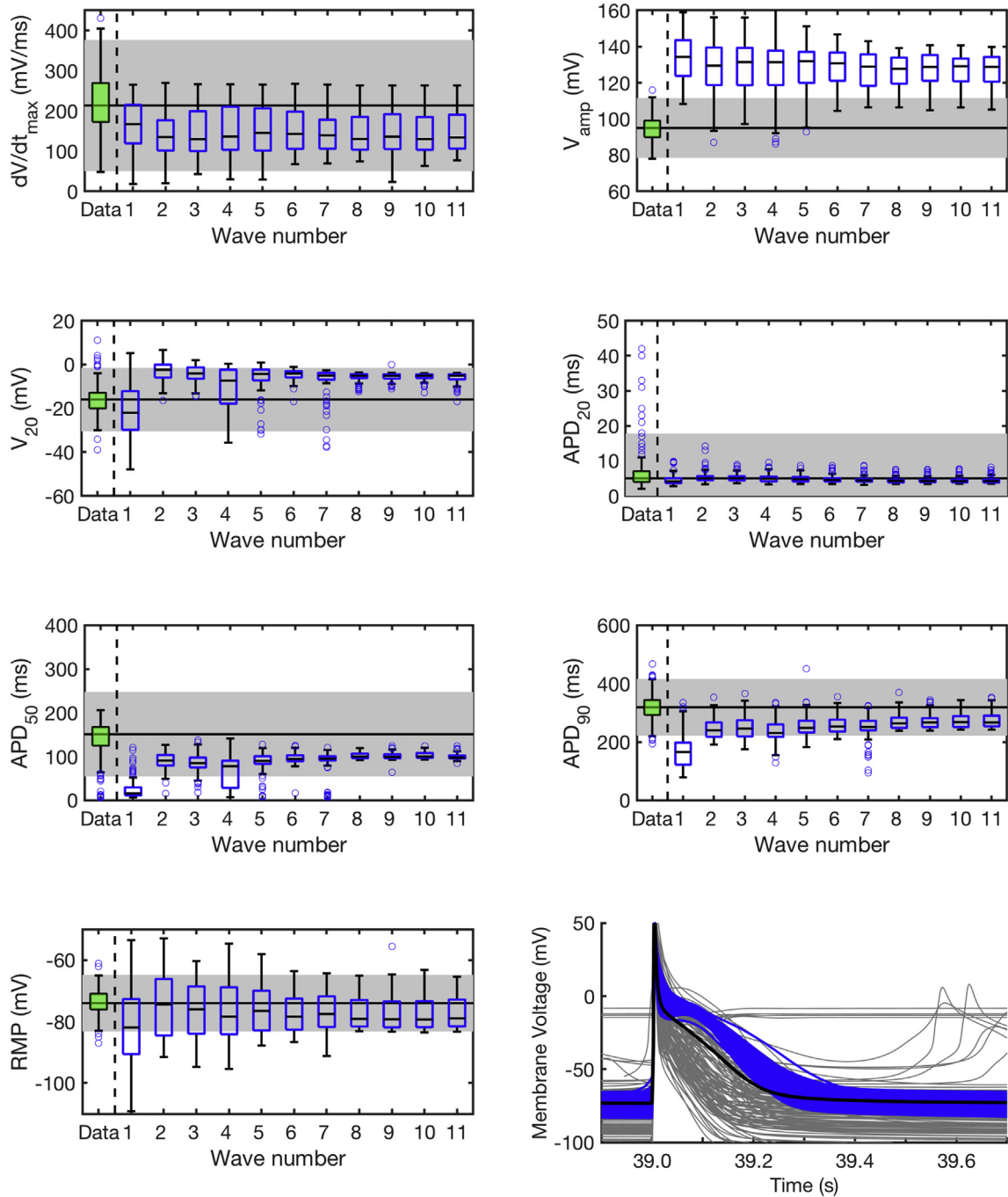


Fig. 4. Changes in design data outputs for each wave of history matching in the Maleckar model. Experimental data are shown as green filled box, model outputs are shown as blue open boxes. Solid horizontal lines indicate the medians, and boxes denote first and third quartiles. Whiskers indicate either $1.5 \times$ the interquartile range, or the data point with the greatest deviation from the median, whichever is the smallest. The plot in the bottom right shows action potentials in the original design data (grey lines), and for the final wave design data (blue lines). Superimposed in black is the action potential obtained using the default mode inputs.

median. For the Maleckar model, V_{amp} remained much higher than the experimental median throughout history matching, which was why it was excluded. Therefore, with our chosen set of inputs, the Maleckar model was not able to reproduce all of the biomarkers in the SR dataset.

The trend in design data inputs is shown in Figs. 5 and 6 for history matching to the SR dataset. Some of the inputs converged towards a reduced range. In the Courtemanche model (Fig. 5) median G_{Na} , G_{K1} , and $G_{b,Ca}$ trended towards their default values, while median $I_{NaK,Max}$ trended below its default value. In the Maleckar

model (Fig. 6), median P_{Na} , G_{K1} and G_{Kur} trended below their default values, while median G_{to} and G_{CaL} trended above their default values. However, although the other inputs continued to be spread over a wide range, the shape of NROY space confines the way that these inputs covary (Fig. 7).

Some of the trends identified in Figs. 5 and 6 can be interpreted mechanically in terms of the model. For example, in the Maleckar model there was a trend towards higher values of G_t and G_{CaL} compared to the default settings for these inputs when history matched to the SR dataset. Increased G_{to} results in a greater

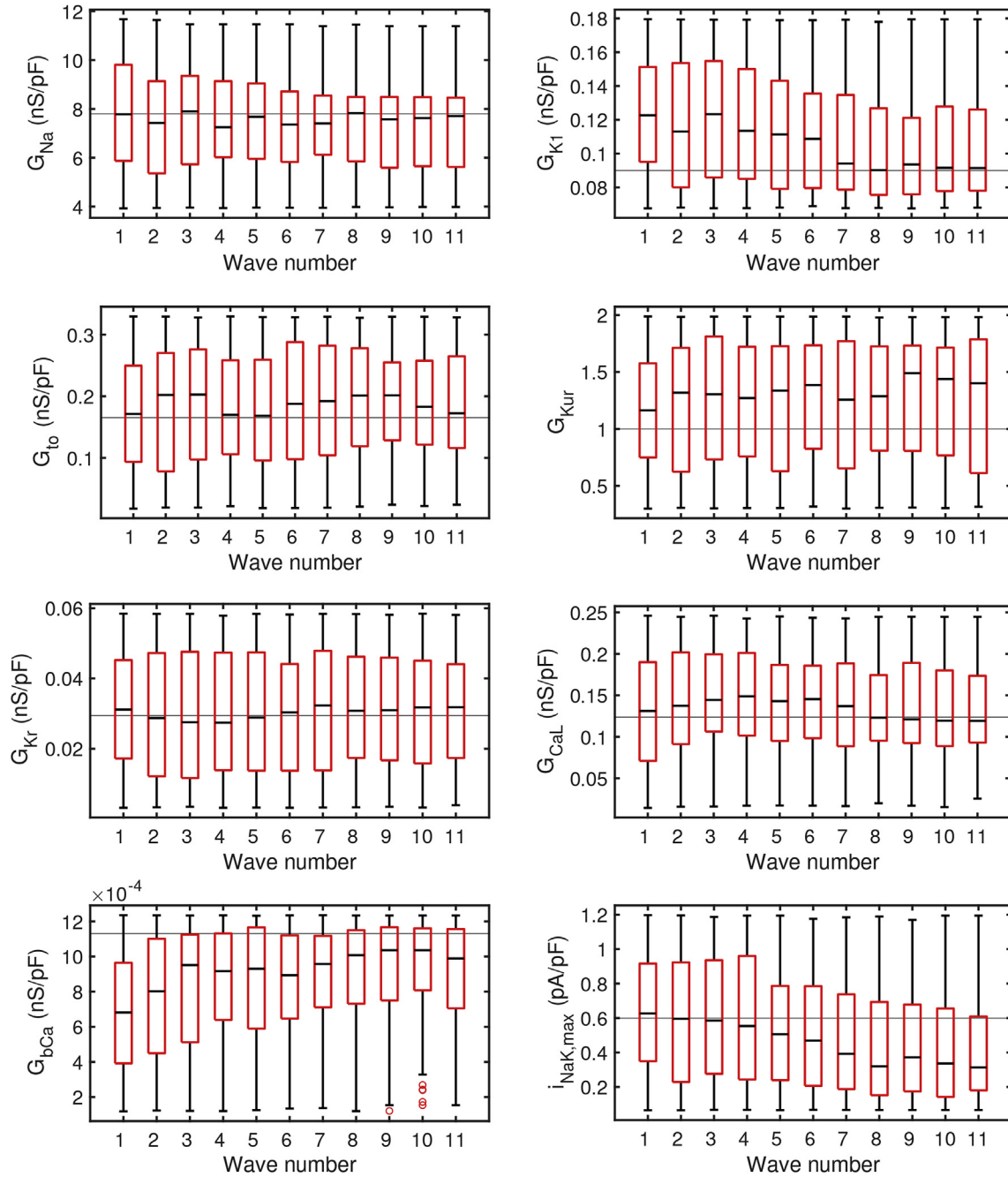


Fig. 5. Changes in each input during history matching of the Courtemanche model to data from normal atrial cells (SR). Solid horizontal lines extending across the plot indicate the default value of each input. Solid horizontal lines in each box indicate the medians, and boxes denote first and third quartiles. Whiskers indicate either $1.5 \times$ the interquartile range, or the data point with the greatest deviation from the median, whichever is the smallest.

transient outward current immediately following the action potential upstroke, and produces a more pronounced notch in the history matched action potentials (Fig. 4) compared to the default model (Fig. 1b). An increased G_{CaL} results in more inward current during the plateau phase of the action potential, leading to a longer plateau phase and increased APD_{90} . These trends cause the Maltackar model action potential shape to align more with the population observations.

The example implausibility plots shown in Fig. 2 indicate that history matching might expose pairs of inputs that co-vary to match the desired outputs. The design data inputs shown in Figs. 5 and 6 are therefore plotted again in Fig. 7, where the projection of the 8-

dimensional input space onto each pair of inputs are shown for history matching at wave 1 (open circles) and wave 11 (filled circles). These plots are not the same as the history matching plots in Fig. 2, because they show only the set of 120 design data (from simulation) sampled from NROY space rather than the extensive and dense sampling of the input space used for history matching. They serve to emphasise the point that history matching reduced the size of non-implausible input space, and did not necessarily identify particular values for the model inputs. Rather, complex configurations in high dimensional input space are revealed. It is therefore important to sample the high dimensional NROY space shown in Fig. 7 rather than the marginal spread of inputs shown in Figs. 5 and 6.

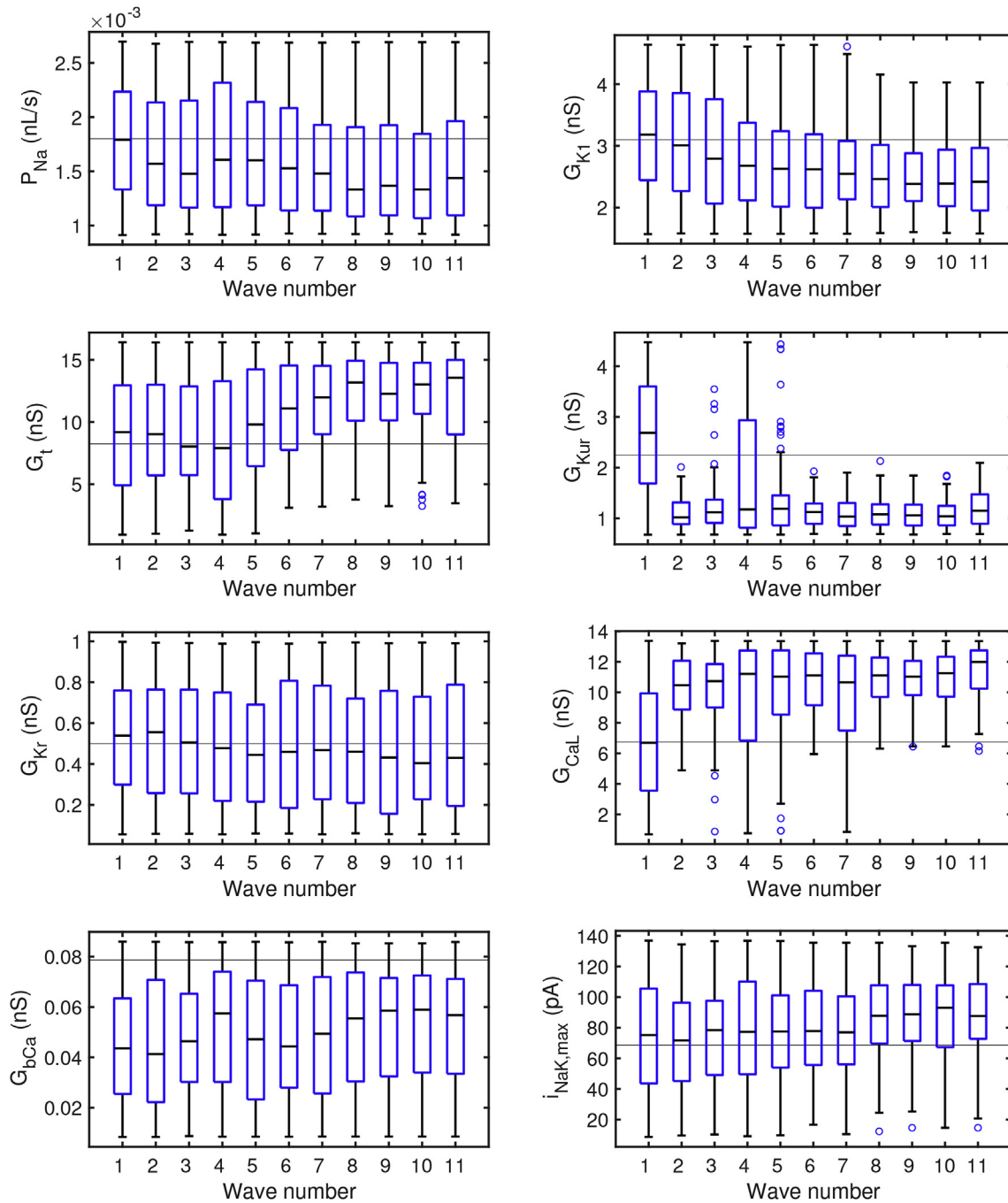


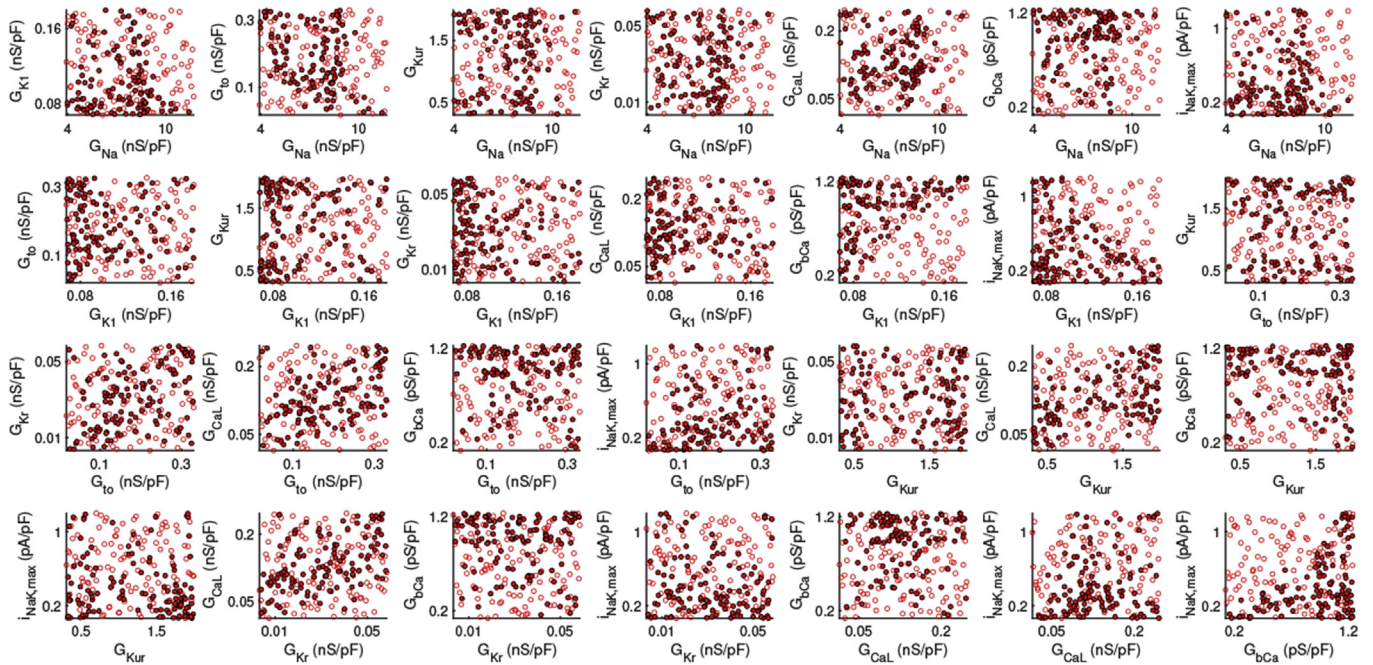
Fig. 6. Changes in each input during history matching of the Maleckar model to data from normal atrial cells (SR). Solid horizontal lines extending across the plot indicate the default value of each input. Solid horizontal lines in each box indicate the medians, and boxes denote first and third quartiles. Whiskers indicate either $1.5 \times$ the interquartile range, or the data point with the greatest deviation from the median, whichever is the smallest.

In Fig. 8 the most implausible output during each wave of history matching to the SR dataset is shown for each model. APD_{20} was rarely the most implausible output, and so contributed little to the history matching process in each model. In the Courtemanche model, RMP was an important output at each wave. For the Maleckar model, V_{amp} was excluded throughout, and V_{20} and APD_{50} appeared to be the most important outputs for constraining NROY space, except in wave 1 where V_{20} was fairly negligible compared to the other outputs.

The most implausible outputs for history matching to the cAF dataset are shown in Fig. 9a for the Courtemanche model and

Fig. 9c for the Maleckar model. For the cAF dataset, dV/dt_{max} played a bigger role in fitting the Maleckar model. Action potentials produced by the simulator using the final set of design data obtained following history matching to the cAF dataset as inputs are shown in Fig. 9b for the Courtemanche model and Fig. 9d for the Maleckar model. Both sets of action potentials have a shorter APD_{90} compared to those resulting from history matching to SR data and shown in Figs. 3 and 4. However, the greater V_{amp} in the Maleckar model results in a much larger spike compared to the Courtemanche model, even though the V_{20} is fitted well to the experimental data.

(a) Courtemanche model



(b) Maleckar model

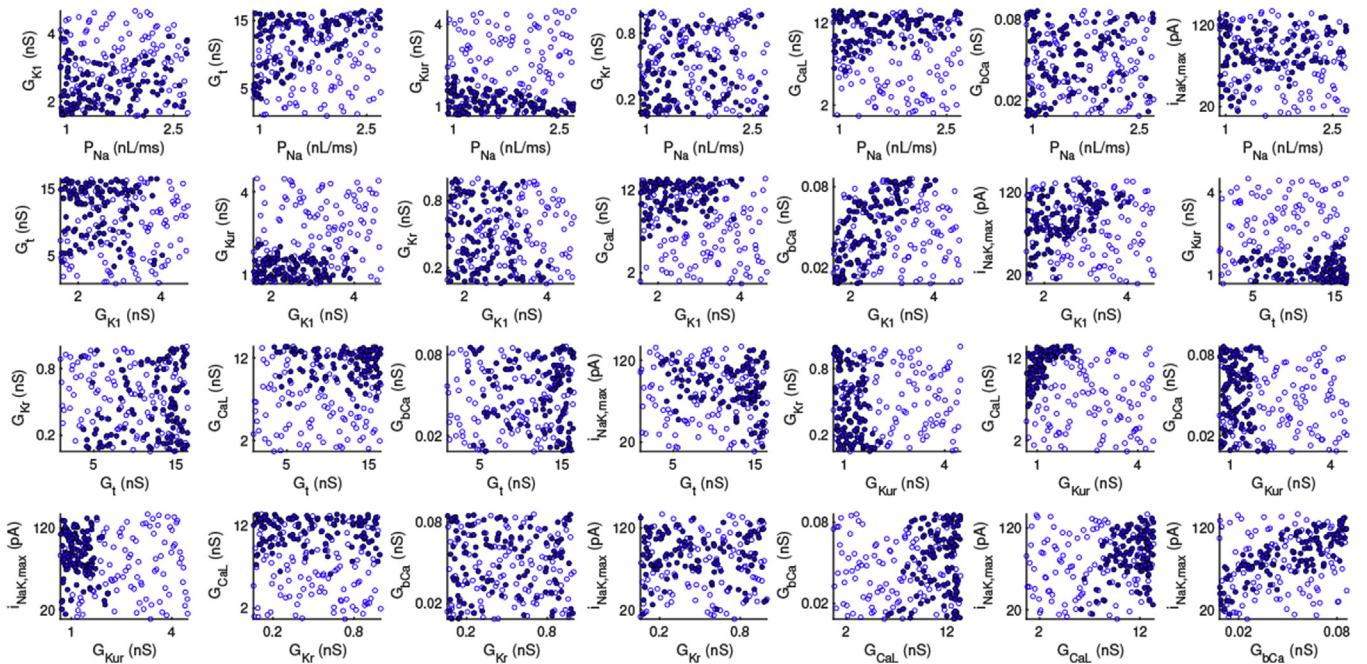
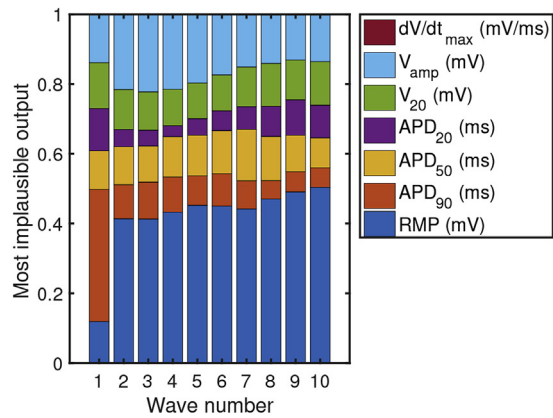


Fig. 7. Projections of the NROY space onto each pair of inputs in (a) the Courtemanche model (red), and (b) the Maleckar model (blue) arising from simulator runs in wave 1 and wave 11 of history matching to the SR dataset. Each point shows one pair of inputs in the design data; points from wave 1 are shown as open circles and points from wave 11 are shown as filled circles.

(a) Courtemanche model



(b) Maleckar model

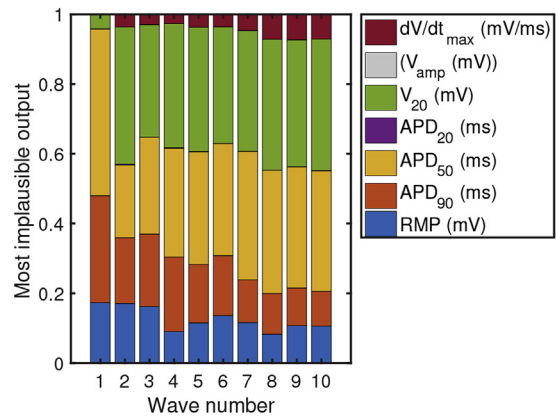
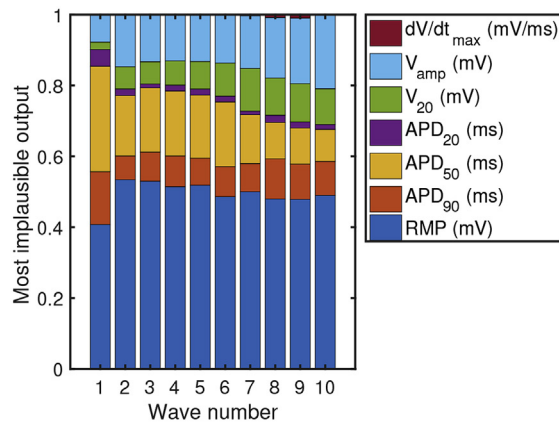
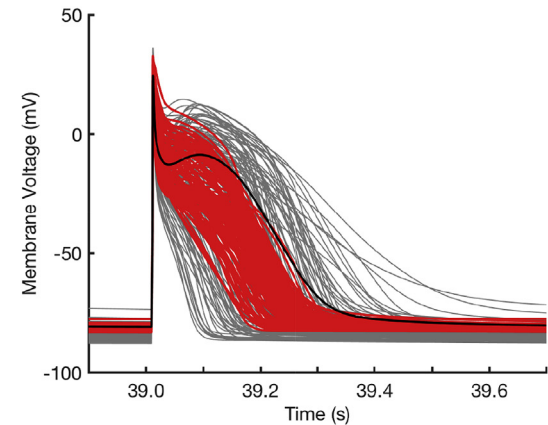


Fig. 8. Most implausible outputs at each wave for the Courtemanche model (a) and Maleckar model (b) during history matching to the SR dataset. Note that V_{amp} was not included in history matching for the Maleckar model.

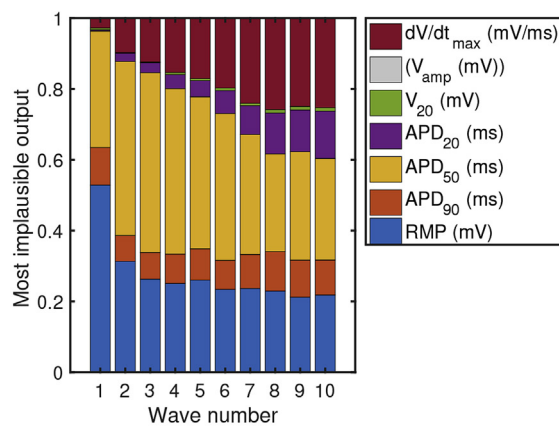
(a) Courtemanche model



(b) Courtemanche model



(c) Maleckar model



(d) Maleckar model

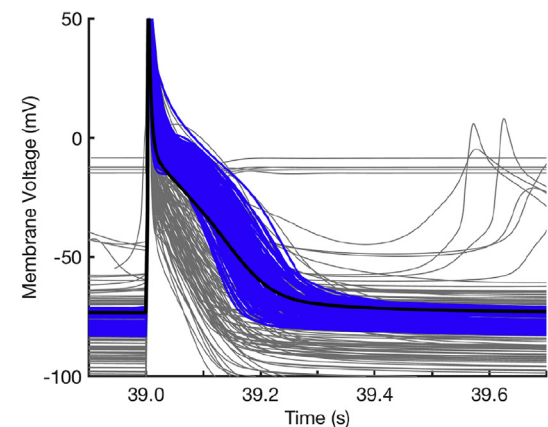


Fig. 9. (a) Most implausible outputs at each wave for the Courtemanche model during history matching to data from cells that have been remodelled in response to chronic AF. (b) Initial design data and action potentials from wave 10. (c) Most implausible outputs at each wave for the Maleckar model during history matching to data from cells that have been remodelled in response to chronic AF. Note that V_{amp} was not included in history matching for the Maleckar model. (d) Initial design data and action potentials from wave 10.

The focus of this study was on the application of history matching to cardiac cell models. However, by history matching to the SR and cAF datasets we were able to compare the range of inputs in each model consistent with normal and remodelled cardiac cells, and these are shown in Fig. 10. We stress that each set of points shown are projections of the NROY regions, and so the plots do not show the co-variation of inputs represented by the complex shape of NROY in input space that are required to match the desired outputs. Because we are not confident that each input was independent, we did not do any statistical tests. However, it is clear from Fig. 10a that for the Courtemanche model G_{K1} tended to be smaller and G_{to} tended to be larger and less constrained with fitting to the SR dataset compared to cAF. Fig. 10b shows that in the Maleckar model a similar trend for G_{K1} was observed, although for both the SR and cAF datasets the selected range for G_t may not have been large enough.

5. Discussion and conclusions

In this paper we report on the application of Bayesian history matching to the problem of selecting a set of inputs for two cardiac cell models that produce outputs consistent with experimental observations. This approach has successfully identified regions of input space that can be sampled to produce sets of inputs that generate model runs consistent with experimental observations. However, this study has raised several new questions and identified limitations. As cardiac cell models are increasingly used for applications such as quantitative safety pharmacology, these new questions are likely to have implications for future work in this area.

5.1. Comparison with previous work

In the present study we used experimental data described by Sánchez et al. (2014), which has been previously used to identify populations of both the Courtemanche and Maleckar models consistent with the SR and cAF datasets (Sánchez et al., 2014; Lawson et al., 2018). Although we examined a slightly different set of model inputs, the overall differences in maximal conductances between the SR and cAF datasets are broadly comparable. Earlier studies have examined simulated remodelling in the Courtemanche and Maleckar models, as well as the predecessor of the Maleckar model (Wilhelms et al., 2012; Cherry and Fenton, 2008),

taking into account other features of the cell models including their dynamic behaviour and Ca^{2+} handling. We did not take these features into account, but recognise their importance because differences in simulated intracellular Ca^{2+} can discriminate between models that produce identical action potentials (Sarkar and Sobie, 2010), and dynamic model properties such as restitution can also be used to improve identifiability (Groenendaal et al., 2015).

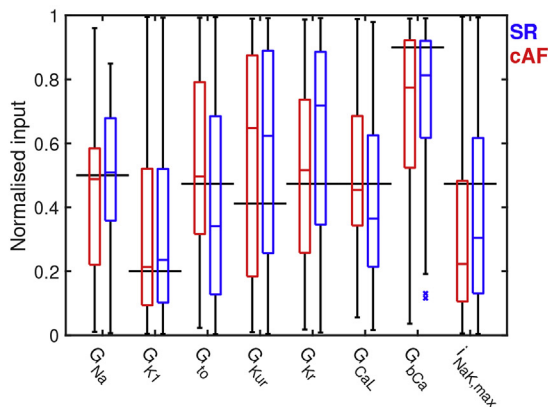
In this study we fixed the intracellular concentrations of Na^+ and K^+ in the Courtemanche model in line with previous work (Wilhelms et al., 2012). Other approaches to stability in this model have been proposed Van Oosterom and Jacquemet (2009), and could be deployed in future work.

5.2. Insights into models

History matching has yielded some insights into the models that we examined. The plots shown in Fig. 7 indicate that the NROY region in the 8-dimensional input space appears to be contiguous. With the default set of inputs, the Courtemanche model produces a prominent spike and dome action potential shape, with an APD of around 300 ms while in contrast the Maleckar action potential with default inputs is more triangular. It is possible to modify the inputs to the Courtemanche model so that it produces a more triangular action potential, and this approach has been used to simulate the effects of remodelling resulting from chronic atrial fibrillation (Cherry and Fenton, 2008). In the present study, history matching to the SR dataset identified a set of inputs that produced a prominent spike and dome configuration in the Maleckar model by increasing I_t . This finding demonstrates the flexibility of cardiac cell models. However, fitting to the cAF dataset did not have the opposite effect of a more triangular action potential in either model, although this may be because of the choice of action potential biomarkers we used, and this is discussed further below.

In order to describe the population observations and to provide a more restrictive test of the usefulness of history matching, we used the median and squared median absolute deviation as measures of the experimental data as opposed to the mean and variance. Median based measures are more robust as they are less susceptible to influence by outliers. Using median based measures, we found that it was not possible to match the Maleckar model to the V_{amp} biomarker, since the model consistently produced V_{amp} values far above the range of observed V_{amp} values. This could be a limitation of the model, or a consequence of gating kinetics that

(a) Courtemanche model



(b) Maleckar model

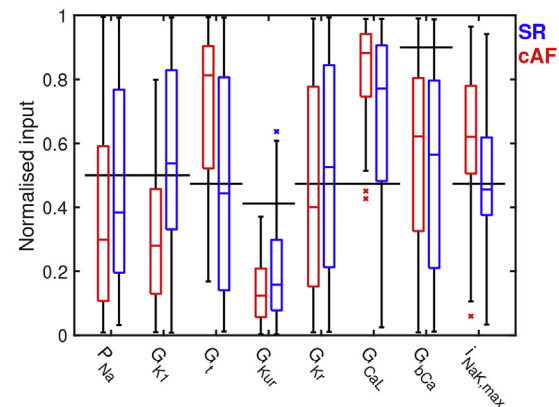
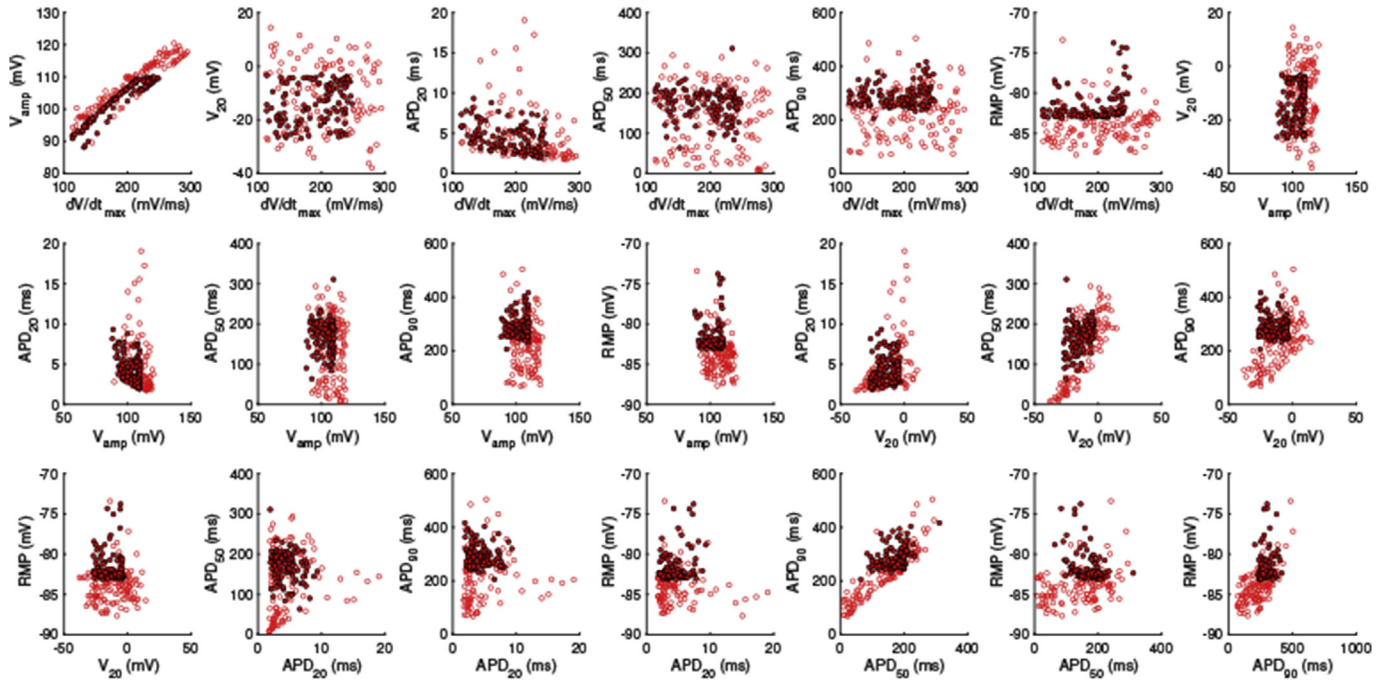


Fig. 10. Comparison of inputs in wave 10 for the Courtemanche (a) and Maleckar (b) models when history matched to data from normal atrial cells (SR, blue) and cells that have been remodelled in response to chronic AF (cAF, red). Horizontal lines indicate the default values of each input.

(a) Courtemanche model



(b) Maleckar model

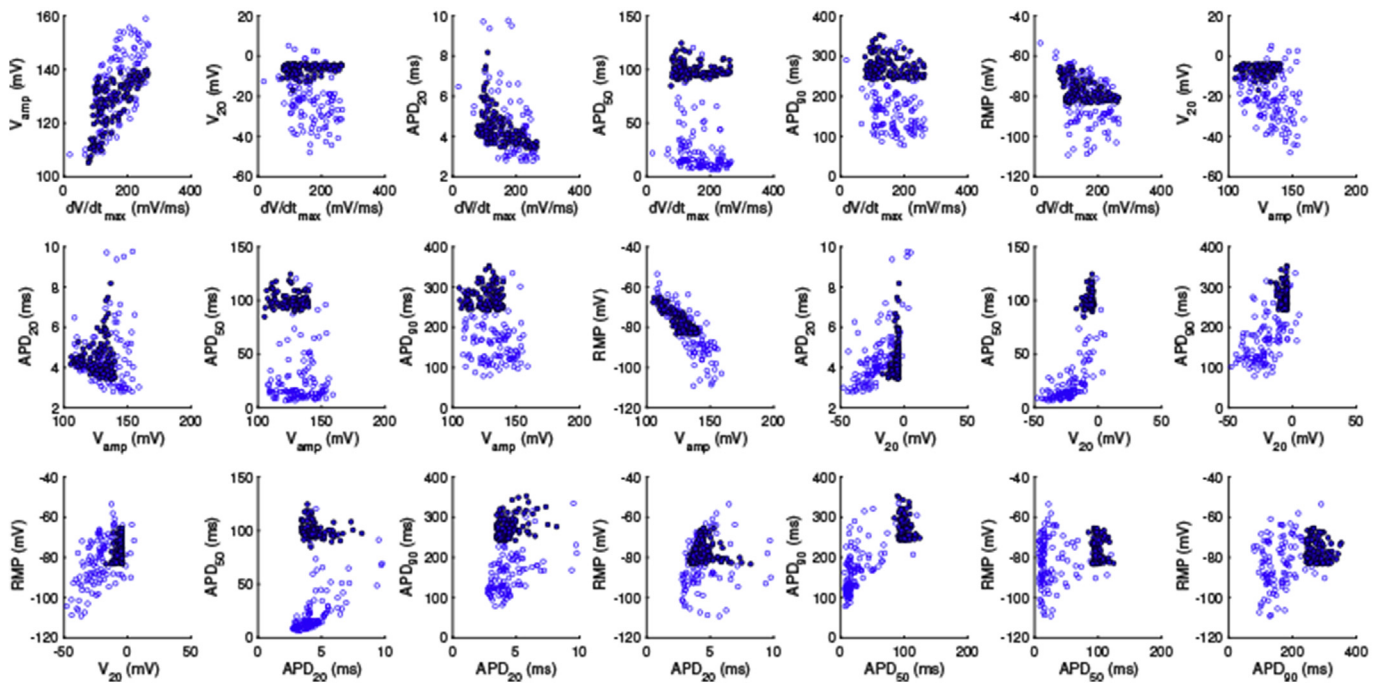


Fig. 11. : Correlations between each pair of design data outputs in (a) the Courtemanche model, and (b) the Maleckar model arising from simulator runs in wave 1 and wave 11 of history matching to the SR dataset. Each point shows one pair of outputs from the design data; points from wave 1 are shown in blue, and points from wave 11 in red.

were not varied in the present study. Matches to the mean and variance were possible, but only because an inflated variance allowed matches far outside any actual observations in the population.

It should be noted that our approach described here did not aim to find the probability distribution of inputs which reproduce an observed probability distribution on the model outputs. However, given an appropriate sampling technique in the identified input space, this would be possible to do efficiently by using emulators in place of simulators. We observed a complex configuration of inputs in the NROY space, and this is illustrated in the projections of sampled points shown in Fig. 7. Some of the samples lie close to the upper or lower bound, in particular the lower bound of G_{Kur} and the upper bound of $G_{Ca,L}$ in the Maleckar model. However, reducing G_{Kur} below the lower bound, and increasing $G_{Ca,L}$ above the upper bound tended to result in abnormal action potentials.

The potential set of inputs for cardiac cell models that could be investigated using a history matching approach is very large. In this study we chose a subset of model inputs that have a physiological interpretation, since maximum conductances reflect the density of ion channels in the membrane. History matching other model inputs, such as dynamic gating parameters (Pathmanathan et al., 2015) or Ca^{2+} handling parameters (Cherry and Fenton, 2008) would of course be possible, but suitable outputs to match against would need to be chosen.

5.3. Experimental data and action potential biomarkers

Our study made use of experimental datasets, using a set of biomarkers designed to capture the action potential shape (Sánchez et al., 2014). While these biomarkers provide a useful starting point, the experimental variability was large with a coefficient of variation of up to 30%, which is larger than in some less comprehensive datasets for similar cells (Wilhelms et al., 2012). Another limitation of the experimental data is the extent of correlations among the biomarkers, illustrated by plotting pairs of design data outputs in Fig. 11. It is likely that the biomarkers used in the present study are not an optimal description of action potential shape, and more robust descriptors may be needed.

5.4. Future directions and open questions

In this study we have demonstrated that history matching can be used to identify sets of cardiac model inputs that can reproduce a set of uncertain experimental observations. We have not been able to identify distributions of parameter values because of compensation mechanisms within the models as well as degeneracy. Nevertheless, despite history matching to population statistics, in which correlation between observed outputs in individuals will be blurred out, we have still managed to cut-down the plausible space significantly. Further work could attempt to specify correlations in the observed outputs and use a multivariate form of the implausibility criterion to reduce the non-implausible space further. This may assist in the degeneracy problem as the correlations between outputs would place stricter conditions on which inputs are non-implausible.

Model discrepancy aims to quantify the difference between a model and the real world system that it represents. It is a crucial missing ingredient for model calibration, but at present it is not clear how model discrepancy should be expressed or even bounded. Progress in this area is important, because understanding the systematic mismatch between a model and the real system will enable experimental data to be properly used in model calibration. The history matching approach allows plausible inputs to be captured even in the case of model discrepancy, provided this is

defined well. With careful use of suitable biomarkers it may be possible to learn the model discrepancy by adjusting it to ensure that we can match to known inputs. In the case of V_{amp} for the Maleckar model in this study, it was clear that there was significant model discrepancy present since a match to the population was simply not possible.

If progress can be made in these areas, then we can expect to see a new generation of cardiac cell models (or improved parameterisation of existing models), which take into account the variability of real cardiac cells, and can be matched with known confidence to specific datasets. This type of model can not only be expected to provide new insights into physiological mechanisms, but also to provide robust and reliable tools for quantitative safety pharmacology.

Acknowledgements

We are very grateful to Ian Vernon and Michael Goldstein at the University of Durham, as well as Richard Wilkinson and Jeremy Oakley at the University of Sheffield, for valuable discussions about history matching and Gaussian processes. This work was funded by EPSRC grant EP/K037145/1.

Appendix A. Supplementary data

Supplementary data related to this article can be found at <https://doi.org/10.1016/j.pbiomolbio.2018.08.001>.

References

- Andrianakis, I., Vernon, I.R., McCreesh, N., McKinley, T.J., Oakley, J.E., Nsubuga, R.N., Goldstein, M., White, R.G., 2015. Bayesian history matching of complex infectious disease models using emulation: a tutorial and a case study on HIV in Uganda. *PLoS Comput. Biol.* 11 (1), e1003968.
- Balse, E., Boycott, H.E., 2017. Ion channel trafficking: control of ion channel density as a target for arrhythmias? *Front. Physiol.* 8 (OCT), 1–6.
- Bingham, D., Sitter, R.R., Tang, B., 2009. Orthogonal and nearly orthogonal designs for computer experiments. *Biometrika* 96 (1), 51–65.
- Britton, O.J., Bueno-Orovio, A., Van Ammel, K., Lu, H.R., Towart, R., Gallacher, D.J., Rodríguez, B., May, 2013. Experimentally calibrated population of models predicts and explains intersubject variability in cardiac cellular electrophysiology. *Proc. Natl. Acad. Sci. U.S.A.* 110 (23), E2098–E2105. <http://www.ncbi.nlm.nih.gov/pubmed/23690584>.
- Cairns, D.I., Fenton, F.H., Cherry, E.M., 2017. Efficient parameterization of cardiac action potential models using a genetic algorithm. *Chaos: An Interdisciplinary Journal of Nonlinear Science* 27 (9), 093922. <http://aip.scitation.org/doi/10.1063/1.5000354>.
- Chang, E.T.Y., Strong, M., Clayton, R., 2015. Bayesian sensitivity analysis of a cardiac cell model using a Gaussian process emulator. *PLoS One* 10 (6), e0130252. <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0130252>.
- Cherry, E.M., Fenton, F.H., 2008. Visualization of spiral and scroll waves in simulated and experimental cardiac tissue. *N. J. Phys.* 10, 125016.
- Colatsky, T., Fermini, B., Gintant, G., Pierson, J.B., Sager, P., Sekino, Y., Strauss, D.G., Stockbridge, N., 2016. The comprehensive in vitro proarrhythmia assay (CiPA) initiative update on progress. *J. Pharmacol. Toxicol. Meth.* 81, 15–20. <https://doi.org/10.1016/j.vascn.2016.06.002>.
- Courtemanche, M., Ramirez, R.J., Nattel, S., 1998. Ionic mechanisms underlying human atrial action potential properties: insights from a mathematical model. *AJP (Am. J. Physiol.)* 275, H301–H321.
- Dokos, S., Lovell, N.H., 2004. Parameter estimation in cardiac ionic models. *Prog. Biophys. Mol. Biol.* 85 (2–3), 407–431. <http://www.ncbi.nlm.nih.gov/pubmed/15142755>.
- Fink, M., Niederer, S.A., Cherry, E.M., Fenton, F.H., Koivumäki, J.T., Seemann, G., Thul, R., Zhang, H., Sachse, F.B., Crampin, E.J., Smith, N.P., 2011. Cardiac cell modelling: observations from the heart of the cardiac physiome project. *Prog. Biophys. Mol. Biol.* 104, 2–21.
- Groenendaal, W., Ortega, F.A., Kherlopian, A.R., Zygmunt, A.C., Krogh-Madsen, T., Christini, D.J., 2015. Cell-specific cardiac electrophysiology models. *PLoS Comput. Biol.* 11 (4), 1–22.
- Johnstone, R.H., Chang, E.T.Y., Bardenet, R., de Boer, T.P., Gavaghan, D.J., Pathmanathan, P., Clayton, R., Mirams, G.R., 2015. Uncertainty and variability in models of the cardiac action potential: can we build trustworthy models? *J. Mol. Cell. Cardiol.* 96, 49–62. <https://doi.org/10.1016/j.yjmcc.2015.11.018>.
- Kennedy, M.C., O'Hagan, A., 2000. Predicting the output from a complex computer code when fast approximations are available. *Biometrika* 87, 1–13.

- Krogh-Madsen, T., Sobie, E.A., Christini, D.J., 2015. Improving cardiomyocyte model fidelity and utility via dynamic electrophysiology protocols and optimization algorithms. *J. Physiol.* 00, 1–12. <http://www.ncbi.nlm.nih.gov/pubmed/26661516>.
- Lawson, B.A.J., Drovandi, C.C., Cusimano, N., Burrage, P., Rodriguez, B., Burrage, K., 2018. Unlocking datasets by calibrating populations of models to data density: a study in atrial electrophysiology. *Science Advances* 4 (January), e1701676. <http://arxiv.org/abs/1706.06763>.
- Loeppky, J.L., Sacks, J., Welch, W.J., 2009. Choosing the sample size of a computer experiment: a practical guide. *Technometrics* 51 (4), 366–376.
- Loewe, A., Wilhelms, M., Schmid, J., Krause, M.J., Fischer, F., Thomas, D., Scholz, E.P., Dössel, O., Seemann, G., 2015. Parameter estimation of ion current formulations requires hybrid optimization approach to be both accurate and reliable. *Frontiers in Bioengineering and Biotechnology* 3, 209. <http://journal.frontiersin.org/Article/10.3389/fbioe.2015.00209/abstract>. <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4710757&tool=pmcentrez&rendertype=abstract>.
- Maleckar, M.M., Greenstein, J.L., Giles, W.R., Trayanova, N. a., 2009. K⁺ current changes account for the rate dependence of the action potential in the human atrial myocyte. *Am. J. Physiol. Heart Circ. Physiol.* 297 (4), H1398–H1410.
- Mirams, G.R., Davies, M.R., Cui, Y., Kohl, P., Noble, D., 2012. Application of cardiac electrophysiology simulations to pro-arrhythmic safety testing. *Br. J. Pharmacol.* 167 (5), 932–945. <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3492977&tool=pmcentrez&rendertype=abstract>.
- Muszkiewicz, A., Britton, O.J., Gemmell, P., Passini, E., Carlos, S., Zhou, X., Carusi, A., Quinn, T.A., Burrage, K., Bueno-orovio, A., Rodriguez, B., 2015. Variability in cardiac electrophysiology: using experimentally-calibrated populations of models to move beyond the single virtual physiological human paradigm. *Prog. Biophys. Mol. Biol.* 120 (2016), 115–127.
- Niederer, S.A., Fink, M., Noble, D., Smith, N.P., 2009. A meta-analysis of cardiac electrophysiology computational models. *Exp. Physiol.* 94 (5), 486–495. <http://www.ncbi.nlm.nih.gov/pubmed/19139063>.
- Oakley, J.E., O'Hagan, A., 2004. Probabilistic sensitivity analysis of complex models: a Bayesian approach. *J. Roy. Stat. Soc. B* 66 (3), 751–769. <http://doi.wiley.com/10.1111/j.1467-9868.2004.05304.x>.
- Pathmanathan, P., Shotwell, M.S., Gavaghan, D.J., Cordeiro, J.M., Gray, R. a., 2015. Uncertainty quantification of fast sodium current steady-state inactivation for multi-scale models of cardiac electrophysiology. *Prog. Biophys. Mol. Biol.* 117 (1), 1–15. <http://linkinghub.elsevier.com/retrieve/pii/S0079610715000097>.
- Pukelsheim, F., 1994. The three sigma rule. *Am. Statistician* 48 (2), 88–91.
- Rasmussen, C.E., Williams, C.K.I., 2006. *Gaussian Processes for Machine Learning*, first ed. MIT press, Cambridge.
- Sánchez, C., Bueno-Orovio, A., Wettwer, E., Loose, S., Simon, J., Ravens, U., Pueyo, E., Rodriguez, B., 2014. Inter-subject variability in human atrial action potential in sinus rhythm versus chronic atrial fibrillation. *PLoS One* 9 (8), e105897. <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4144914&tool=pmcentrez&rendertype=abstract>.
- Sarkar, A.X., Sobie, E. a., 2010. Regression analysis for constraining free parameters in electrophysiological models of cardiac cells. *PLoS Comput. Biol.* 6 (9), e1000914. <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2932676&tool=pmcentrez&rendertype=abstract>.
- Tixier, E., Lombardi, D., Rodriguez, B., Gerbeau, J.-F., 2017. Modelling variability in cardiac electrophysiology: a moment-matching approach. *J. R. Soc. Interface* 14 (133), 20170238. <http://www.ncbi.nlm.nih.gov/pubmed/28835541>. <http://rsif.royalsocietypublishing.org/lookup/doi/10.1098/rsif.2017.0238>.
- Van Oosterom, A., Jacquemet, V., 2009. Ensuring stability in models of atrial kinetics. *Comput. Cardiol.* 36, 69–72.
- Vernon, I., Goldstein, M., Bower, R.G., 2010. Galaxy formation: a Bayesian uncertainty analysis. *Bayesian Analysis* 5 (4), 619–669. <http://projecteuclid.org/euclid.ba/1340110846>.
- Vernon, I., Liu, J., Goldstein, M., Rowe, J., Topping, J., Lindsey, K., 2018. Bayesian uncertainty analysis for complex systems biology models: emulation, global parameter searches and evaluation of gene functions. *BMC Syst. Biol.* 12 (1), 1–29. <https://bmcsystbiol.biomedcentral.com/articles/10.1186/s12918-017-0484-3>.
- Wilhelms, M., Hettmann, H., Maleckar, M.M., Koivumäki, J.T., Dössel, O., Seemann, G., 2012. Benchmarking electrophysiological models of human atrial myocytes. *Front. Physiol.* 3, 487. <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3539682&tool=pmcentrez&rendertype=abstract>.
- Zaniboni, M., Pollard, a. E., Yang, L., Spitzer, K.W., 2000. Beat-to-beat repolarization variability in ventricular myocytes and its suppression by electrical coupling. *Am. J. Physiol. Heart Circ. Physiol.* 278 (3), H677–H687.
- Zaniboni, M., Riva, I., Cacciani, F., Groppi, M., 2010. How different two almost identical action potentials can be: a model study on cardiac repolarization. *Math. Biosci.* 228 (1), 56–70. <http://www.ncbi.nlm.nih.gov/pubmed/20801131>.