Supplementary Material for Sampling Methods for Exploring Between Subject Variability in Cardiac Electrophysiology Experiments

C. C. Drovandi[⋆†], N. Cusimano[⋆†], S. Psaltis[⋆†], B. A. J. Lawson[⋆†],
A. N. Pettitt[⋆†], P. Burrage[⋆†] and K. Burrage[⋆†‡]
[⋆]Mathematical Sciences, Queensland University of Technology, Brisbane, Australia 4000
[†]ARC Centre of Excellence for Mathematical & Statistical Frontiers
[‡] Department of Computer Science, University of Oxford, UK
email: c.drovandi@qut.edu.au

July 11, 2016

Appendix A Latin Hypercube Sampling

Introduced by McKay et al. (1979), LHS has been widely used in the electrophysiology literature for investigating the between-subject variability that exists in experimental observations (Marder and Taylor, 2011; Britton et al., 2013; Dutta et al., 2013). LHS is designed to generate parameter combinations such that good coverage of the parameter space is achieved.

Latin squares are commonly used in experimental design (Hinkelmann and Kempthorne, 2008) for allocating treatments, for instance, to experimental units, allowing only one occurrence of each treatment type per experiment. In this work, we consider a Latin square (two-dimensional parameter space) to be one in which no more than one sample is taken from each row and column (or division) for a given parameter combination. Note that a Latin hypercube is the extension of this to higher dimensions. It is possible to form multiple Latin squares by sampling more than once in each division, and this situation is shown in Fig. 1, where three Latin squares are formed by taking three samples per division. Furthermore, it shows that multiple configurations can be formed (represented by each colour), which are essentially new LHS runs. However, Fig. 1 shows an idealised situation in which there is no overlap in the divisions across multiple configurations. In reality, this may not occur as there is no information passed from one configuration to the next. They can be considered as separate instances of LHS.

Typically, LHS is performed using only a single sample per division and a single configuration, and this method will be the focus of this work. The process required for LHS can be described as follows.

Given a d dimensional parameter space, we first specify lower and upper bounds for each component θ_j of $\boldsymbol{\theta}$, such that, $lb_j < \theta_j < ub_j$, $j = 1, \ldots, d$. We divide the range for each θ_j into M equally probable subdivisions and randomly sample each subdivision exactly once.

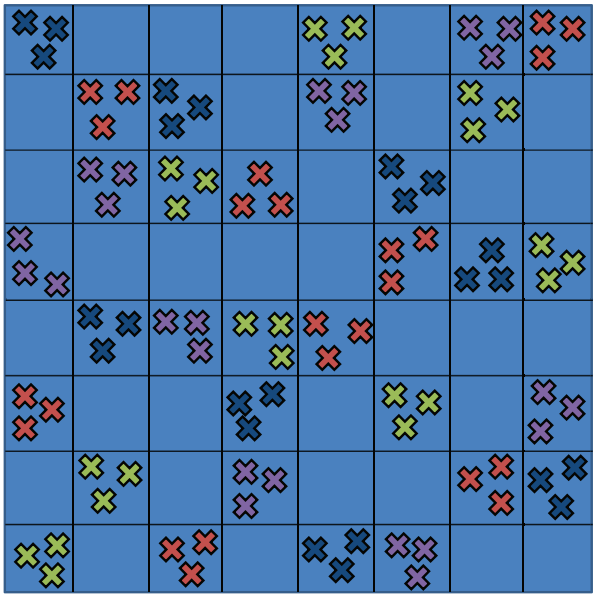

Figure 1: Representation of a two-parameter LHS approach. Each colour represents a different configuration of LHS.

We then build a $d \times M$ matrix $\boldsymbol{P}$ in which each row corresponds to a different component of $\boldsymbol{\theta}$ and contains a random permutation of the M sampled values for that particular θ_j. The M parameters $\boldsymbol{\theta}^1, \ldots, \boldsymbol{\theta}^M$ generated by the LHS strategy are simply given by the M columns of the matrix $\boldsymbol{P}$. The model is then simulated at these values of $\boldsymbol{\theta}$.

As discussed in the main paper, the model outputs are then compared to the observed data by way of summary statistics. If the set of appropriate constraints are satisfied, then the parameter combination is accepted as giving a 'match' to the data.

For the case study in the main paper, we perform two separate cases of LHS: firstly, with 5000 divisions per parameter range, one sample per division and one configuration; and secondly, with 10000 divisions per parameter range, one sample per division and one configuration. We choose to implement LHS in this way as it is the way commonly discussed in the literature. It is unclear at this stage whether more divisions and fewer samples is the most effective approach of implementing LHS.

Appendix B SMC Algorithm

The SMC algorithm developed for the main paper is shown in Algorithm 1.

Appendix C Complete description of the BR model

The Beeler–Reuter (BR) cell model is defined as a system of eight ordinary differential equations governing the temporal evolution of the transmembrane potential v of a single ventricular

Algorithm 1 SMC algorithm with an MCMC kernel for sampling from the sequence of POM probability distributions defined in (2) of the main paper. We present the algorithm as sampling over the space of $\boldsymbol{\theta}$ however we actually sample over the transformed space of $\boldsymbol{\phi}$. Samples from the $\boldsymbol{\theta}$ space can be easily recovered by back-transformation.

for $i = 1, \ldots, N$ **do**
  Simulate $\boldsymbol{\theta}^i \sim p(\cdot)$ and generate the vector of summary statistics from the model $\boldsymbol{S}^{\boldsymbol{\theta}^i}$
end for
for $c = 1, \ldots, K$ **do**
  Set $N_a = N/2$
  Determine the set $A_c^{\max}$ which is the smallest possible range so that all particles satisfy the constraint $\mathbb{1}(S_c^{\boldsymbol{\theta}^i} \in A_c^{\max})$ for $i = 1, \ldots, N$
  while $A_c^{\max} \supset A_c$ **do**
    Determine the set A_c^{next} which is the largest possible range so that N_a particles satisfy the constraint $\mathbb{1}(S_c^{\boldsymbol{\theta}^i} \in A_c^{\text{next}})$
    If $A_c^{\text{next}} \subset A_c$ then set $A_c^{\text{next}} \equiv A_c$ and adjust the value of N_a appropriately
    for $i = N_a + 1$ to N (note if $N_a = N$ then this loop is not required) **do**
      Re-sample $\boldsymbol{\theta}^i \sim \{\boldsymbol{\theta}^j\}_{j=1}^{N_a}$
      for $r = 1$ to R **do**
        Propose move $\boldsymbol{\theta}^* \sim q(\cdot|\boldsymbol{\theta}^i)$
        Compute acceptance ratio $MH = \min\left(1, \frac{p(\boldsymbol{\theta}^*)q(\boldsymbol{\theta}^i|\boldsymbol{\theta}^*)}{p(\boldsymbol{\theta}^i)q(\boldsymbol{\theta}^*|\boldsymbol{\theta}^i)}\right)$
        if $U(0,1) > MH$ **then**
          Reject $\boldsymbol{\theta}^*$ without simulating data and go to next iteration of the for loop
        end if
        Generate the vector of summary statistics from the model $\boldsymbol{S}^{\boldsymbol{\theta}^*}$
        if $S_s^{\boldsymbol{\theta}^*} \in A_s$ for $s = 1, \ldots, c-1$ and $S_c^{\boldsymbol{\theta}^*} \in A_c^{\text{next}}$ **then**
          Set $\boldsymbol{\theta}^i = \boldsymbol{\theta}^*$
        end if
      end for
    end for
    Determine the set $A_c^{\max}$ which is the largest possible range so that all particles satisfy the constraint $\mathbb{1}(S_c^{\boldsymbol{\theta}^i} \in A_c^{\max})$ for $i = 1, \ldots, N$
  end while
end for

cell and the evolution of a vector of secondary variables $\boldsymbol{z} = (m, h, j, d, f, x, c)^T$, used in the description of the dynamics of the various ion channels present in the cellular membrane. The six variables m, h, j, d, f, x are typically called gating variables, while $c = 10^7[\mathrm{Ca}]$ represents the intracellular calcium concentration and has been scaled to simplify the notation. The equation for the transmembrane potential is

$$C_m \frac{dv}{dt} + I_{\mathrm{ion}} = I_{\mathrm{stim}}. \tag{1}$$

In equation (1), C_m is the cell membrane capacitance per unit area, I_{ion} is the ionic current defined as the sum of all transmembrane currents generated by the opening and closing of the ion channels in the cell membrane, and I_{stim} represents an externally applied electrical stimulus.

In the BR model, the ionic current is the sum of four components and can be written as follows:

$$I_{\mathrm{ion}} = I_{\mathrm{Na}} + I_{\mathrm{K}} + I_{\mathrm{x}} + I_{\mathrm{s}}, \tag{2}$$

where I_{Na} is the current carried by sodium

$$I_{\mathrm{Na}} = (g_{\mathrm{Ina}}\, m^3 h j + 0.003)(v - 50),$$

I_{K} and I_{x} are potassium currents, defined respectively by

$$I_{\mathrm{K}} = g_{\mathrm{Ik}} \left\{ \frac{4(\exp(0.04(v+85)) - 1)}{\exp(0.08(v+53)) + \exp(0.04(v+53))} + \frac{0.2(v+23)}{1 - \exp(-0.04(v+23))} \right\},$$

$$I_{\mathrm{x}} = g_{\mathrm{Ix}}\, x\, \frac{\exp(0.04(v+77)) - 1}{\exp(0.04(v+35))},$$

and I_{s} is the calcium current given by

$$I_{\mathrm{s}} = g_{\mathrm{Is}}\, d\, f\, (v + 82.3 + 13.0287 \ln(10^{-7} c)).$$

In the equations above, C_m is in $\mu\mathrm{F}\cdot\mathrm{cm}^{-2}$, all currents are in $\mu\mathrm{A}\cdot\mathrm{cm}^{-2}$, v is in mV, the six gating variables are dimensionless, [Ca] is in moles per litre ($\mathrm{mole}\cdot\mathrm{L}^{-1}$), and time is expressed in ms. The parameters $g_{\mathrm{Ina}}, g_{\mathrm{Ik}}, g_{\mathrm{Ix}}$, and g_{Is} represent the four maximal current densities of the system and have been used in this work for the comparison of the POM methods. In the original formulation of the model, Beeler and Reuter (1977) specify these four parameters as $g_{\mathrm{Ina}} = 4\,\mathrm{mS}\cdot\mathrm{cm}^{-2}$, $g_{\mathrm{Ik}} = 0.35\,\mu\mathrm{A}\cdot\mathrm{cm}^{-2}$, $g_{\mathrm{Ix}} = 0.8\,\mu\mathrm{A}\cdot\mathrm{cm}^{-2}$ and $g_{\mathrm{Is}} = 0.09\,\mathrm{mS}\cdot\mathrm{cm}^{-2}$.

In order to complete the formulation of the BR cell model, we provide here the analytic expression of the seven ODEs governing the secondary variables of the model, as given in the original work by Beeler and Reuter (1977).

Let q represent any of the six gating variables of the model. The corresponding ODE of the BR cell model can be written as follows:

$$\frac{dq}{dt} = \alpha_q(v)(1 - q) - \beta_q(v)\, q, \tag{3}$$

	C_1	C_2	C_3	C_4	C_5	C_6	C_7
ms^{-1}	ms^{-1}	mV^{-1}	mV	$\mathrm{mV}^{-1}\cdot\mathrm{ms}^{-1}$	mV	mV^{-1}	–
α_m	0	0	47	-1	47	-0.1	-1
β_m	40	-0.056	72	0	0	0	0
α_h	0.126	-0.25	77	0	0	0	0
β_h	1.7	0	22.5	0	0	-0.082	1
α_j	0.055	-0.25	78	0	0	-0.2	1
β_j	0.3	0	32	0	0	-0.1	1
α_d	0.095	-0.01	-5	0	0	-0.72	1
β_d	0.07	-0.017	44	0	0	0.05	1
α_f	0.012	-0.008	28	0	0	0.15	1
β_f	0.0065	-0.02	30	0	0	-0.2	1
α_x	0.0005	0.083	50	0	0	0.057	1
β_x	0.0013	-0.06	20	0	0	-0.04	1

Table 1: Parameter values for the BR model as given in Beeler and Reuter (1977).

where α_q and β_q are the channel opening and closing rates associated to the particular q (3) is referring to. For all gating variables, both rates are functions of the transmembrane potential v and have the following form:

$$\frac{C_1 \exp(C_2(v + C_3)) + C_4(v + C_5)}{\exp(C_6(v + C_3)) + C_7}. \tag{4}$$

The values and the units of the constants involved in the expression (4) according to the original formulation of the model are specified as in Table 1.

Finally, the evolution in time of the scaled intracellular calcium concentration $c = 10^7[\mathrm{Ca}]$ is governed by the following differential equation:

$$\frac{dc}{dt} = 0.07(1 - c) - I_{\mathrm{s}}, \tag{5}$$

where I_s is the calcium current defined above.

C.1 Initial condition

The initial condition for the eight variables involved in the system is as follows: we assume that the transmembrane potential is initially set equal to $v = -85\,\mathrm{mV}$ and for the vector of secondary variables we define $\boldsymbol{z} = (0, 1, 1, 0, 1, 0, 1)$. We let the simulation run for a certain time interval without the application of any external current, that is, we specify $I_{\mathrm{stim}} = 0$ for $t \in [0, t_1]$, so that all variables in the system are allowed to reach the corresponding steady state.

C.2 Stimulation protocol

We stimulate the cell with a train of five current pulses of 2 ms duration, delivered at a cycle length of 1000 ms, that is, at $t_i = t_1 + (i - 1) \times 1000$ for $i = 1, \ldots, 5$. The amplitude of all five

stimuli is set equal to twice the diastolic threshold of the model at the considered cycle length ($I_{\text{stim}} = 28\,\mu\text{A}\cdot\text{cm}^{-2}$). The diastolic threshold is measured by following the action potential duration definition proposed by Monasterio et al. (2014) for their single-cell simulations.

C.3 Implementation of the solution

The solution of the system of differential equations is computed with the MATLAB® ODE solver **ode15s** with default error tolerances on the interval $[0, t_f]$ (for a suitable $t_f > t_5$ so that the entire action potential triggered by the fifth electrical stimulus could be visualised).

For the comparison of the solution obtained with different parameters, we focus on the last action potential generated by the electrical stimuli, that is, we consider the solution profile of the variable v on the temporal interval $[t_5, t_f]$. In all the model simulations run for the manuscript we choose $t_1 = 500\,\text{ms}$ and $t_f = 5500\,\text{ms}$.

C.4 Characterisation of the solution

In order to characterise the data produced by a given simulation, we compute a set of four biomarkers as explained in Section 2.3 of the main paper. We also consider the value of the solution at a set of unevenly spaced time points, with a higher concentration where the action potential presents rapid variations in order to more accurately capture critical features of the solution profile. Letting $t = 0$ represent the initial time point of the temporal interval of interest (i.e. t_5), we consider the following vector of 28 time points (in ms): (0.1, 0.3, 0.6, 1, 1.10, 1.25, 1.35, 1.5, 1.75, 2, 2.75, 3.5, 4.25, 5, 10, 15, 20, 25, 35, 50, 75, 100, 125, 150, 200, 250, 300, 350). This choice clearly depends on the solution behaviour produced by a given stimulation protocol. However, the same idea can be easily implemented for any other choice of simulation parameters and stimulation protocol by simply identifying regions of rapid/slow change in the generated action potential.

Appendix D Beeler-Reuter Summary statistic data

Shown are the sample mean and standard deviation (Table 2), and the sample minimum and maximum (Table 3) of all the summary statistics considered for parameter sets 1-4.

Table 2: Means and standard deviations for each of the summary statistics.

Summary Statistic	parameter set 1 mean (sd)	parameter set 2 mean (sd)	parameter set 3 mean (sd)	parameter set 4 mean (sd)
t_1 0.1 (ms)	-81.25 (0.05)	-81.35 (0.25)	-81.13 (0.31)	-81.23 (0.23)
t_2 0.3 (ms)	-75.82 (0.05)	-75.93 (0.26)	-75.69 (0.32)	-75.80 (0.23)
t_3 0.6 (ms)	-67.87 (0.05)	-67.99 (0.29)	-67.73 (0.35)	-67.84 (0.26)
t_4 1 (ms)	-56.89 (0.09)	-57.06 (0.39)	-56.71 (0.47)	-56.88 (0.36)
t_5 1.10 (ms)	-53.61 (0.15)	-53.82 (0.46)	-53.40 (0.55)	-53.61 (0.43)
t_6 1.25 (ms)	-47.03 (0.42)	-47.36 (0.76)	-46.71 (0.89)	-47.13 (0.74)
t_7 1.35 (ms)	-39.93 (1.00)	-40.47 (1.33)	-39.46 (1.51)	-40.25 (1.36)
t_8 1.5 (ms)	-19.94 (3.12)	-21.02 (3.26)	-19.27 (3.19)	-21.13 (3.48)
t_9 1.75 (ms)	15.43 (3.38)	14.84 (3.27)	15.66 (2.59)	14.00 (3.47)
t_{10} 2 (ms)	31.74 (2.53)	31.43 (2.46)	31.74 (2.01)	30.62 (2.59)
t_{11} 2.75 (ms)	36.04 (1.71)	35.85 (1.66)	36.00 (1.42)	35.27 (1.74)
t_{12} 3.5 (ms)	34.49 (1.51)	34.25 (1.47)	34.50 (1.26)	33.81 (1.56)
t_{13} 4.25 (ms)	32.06 (1.42)	31.75 (1.40)	32.13 (1.21)	31.43 (1.49)
t_{14} 5 (ms)	29.64 (1.34)	29.24 (1.37)	29.76 (1.20)	29.03 (1.46)
t_{15} 10 (ms)	18.81 (0.95)	18.04 (1.61)	19.21 (1.62)	18.31 (1.45)
t_{16} 15 (ms)	14.55 (0.67)	13.57 (1.96)	15.11 (2.07)	14.06 (1.58)
t_{17} 20 (ms)	13.46 (0.48)	12.35 (2.22)	14.10 (2.38)	12.93 (1.74)
t_{18} 25 (ms)	13.69 (0.35)	12.50 (2.38)	14.38 (2.56)	13.13 (1.87)
t_{19} 35 (ms)	14.96 (0.18)	13.73 (2.49)	15.67 (2.69)	14.38 (2.01)
t_{20} 50 (ms)	15.59 (0.05)	14.39 (2.43)	16.25 (2.63)	15.04 (2.00)
t_{21} 75 (ms)	13.33 (0.01)	12.20 (2.33)	13.94 (2.53)	12.88 (1.92)
t_{22} 100 (ms)	9.37 (0.03)	8.20 (2.36)	9.94 (2.58)	8.95 (2.00)
t_{23} 125 (ms)	4.64 (0.03)	3.36 (2.55)	5.20 (2.77)	4.21 (2.22)
t_{24} 150 (ms)	-0.82 (0.03)	-2.32 (2.92)	-0.26 (3.17)	-1.30 (2.63)
t_{25} 200 (ms)	-15.87 (0.05)	-18.67 (5.02)	-15.25 (5.34)	-16.73 (4.71)
t_{26} 250 (ms)	-43.97 (0.13)	-51.72 (12.22)	-43.62 (12.35)	-46.24 (11.08)
t_{27} 300 (ms)	-81.26 (0.02)	-81.08 (2.56)	-78.44 (4.91)	-80.72 (1.81)
t_{28} 350 (ms)	-82.56 (0.00)	-82.73 (0.41)	-82.39 (0.46)	-82.62 (0.32)
biomarker$_1$ AP Peak	36.06 (1.73)	35.87 (1.68)	36.02 (1.43)	35.28 (1.76)
biomarker$_2$ Dome Peak	15.62 (0.07)	14.41 (2.45)	16.29 (2.65)	15.07 (2.01)
biomarker$_3$ Max dv/dt	176.23 (10.71)	176.28 (10.36)	175.16 (8.78)	171.45 (10.51)
biomarker$_4$ APD$_{90}$	278.37 (0.39)	271.57 (14.54)	281.85 (16.34)	277.38 (12.20)

Table 3: Minimums and maximums for each of the summary statistics.

Summary Statistic	parameter set 1 min / max	parameter set 2 min / max	parameter set 3 min / max	parameter set 4 min / max
t_1 0.1 (ms)	-81.32 / -81.19	-81.60 / -80.91	-81.56 / -80.80	-81.50 / -80.81
t_2 0.3 (ms)	-75.88 / -75.76	-76.19 / -75.47	-76.15 / -75.35	-76.07 / -75.36
t_3 0.6 (ms)	-67.93 / -67.81	-68.28 / -67.48	-68.24 / -67.36	-68.16 / -67.37
t_4 1 (ms)	-57.03 / -56.79	-57.51 / -56.35	-57.39 / -56.15	-57.35 / -56.27
t_5 1.10 (ms)	-53.85 / -53.45	-54.41 / -52.94	-54.21 / -52.66	-54.20 / -52.83
t_6 1.25 (ms)	-47.72 / -46.58	-48.55 / -45.82	-47.97 / -45.25	-48.15 / -45.60
t_7 1.35 (ms)	-41.52 / -38.91	-42.77 / -37.79	-41.41 / -36.64	-42.04 / -37.40
t_8 1.5 (ms)	-24.79 / -16.72	-27.18 / -15.30	-22.78 / -12.95	-25.36 / -14.53
t_9 1.75 (ms)	10.12 / 18.75	8.60 / 18.78	11.94 / 20.31	9.95 / 19.25
t_{10} 2 (ms)	27.72 / 34.12	27.02 / 33.97	28.45 / 35.00	27.77 / 34.20
t_{11} 2.75 (ms)	33.31 / 37.55	32.98 / 37.47	33.58 / 38.17	33.48 / 37.61
t_{12} 3.5 (ms)	32.08 / 35.83	31.58 / 35.91	32.49 / 36.56	32.14 / 36.09
t_{13} 4.25 (ms)	29.80 / 33.32	29.08 / 33.59	30.45 / 34.27	29.74 / 33.80
t_{14} 5 (ms)	27.49 / 30.83	26.55 / 31.29	28.37 / 31.99	27.29 / 31.52
t_{15} 10 (ms)	17.30 / 19.65	15.24 / 21.05	16.72 / 21.84	16.10 / 21.05
t_{16} 15 (ms)	13.48 / 15.15	10.75 / 17.09	11.80 / 17.88	11.33 / 16.59
t_{17} 20 (ms)	12.68 / 13.88	9.54 / 16.14	10.29 / 16.91	9.75 / 15.13
t_{18} 25 (ms)	13.13 / 14.00	9.74 / 16.43	10.27 / 17.16	9.62 / 15.47
t_{19} 35 (ms)	14.68 / 15.12	11.12 / 17.66	11.36 / 18.33	10.59 / 16.90
t_{20} 50 (ms)	15.50 / 15.63	12.06 / 18.13	12.03 / 19.06	11.31 / 17.45
t_{21} 75 (ms)	13.32 / 13.35	10.01 / 15.84	9.84 / 16.80	9.31 / 15.15
t_{22} 100 (ms)	9.34 / 9.41	5.96 / 11.92	5.72 / 12.95	5.26 / 11.38
t_{23} 125 (ms)	4.61 / 4.68	0.93 / 7.35	0.57 / 8.49	0.13 / 6.94
t_{24} 150 (ms)	-0.85 / -0.77	-5.13 / 2.22	-5.66 / 3.61	-6.10 / 1.93
t_{25} 200 (ms)	-15.92 / -15.79	-23.63 / -11.05	-24.85 / -8.65	-25.35 / -11.16
t_{26} 250 (ms)	-44.09 / -43.72	-64.21 / -33.89	-67.10 / -29.11	-67.00 / -33.89
t_{27} 300 (ms)	-81.28 / -81.23	-82.82 / -76.16	-82.80 / -69.38	-82.78 / -76.60
t_{28} 350 (ms)	-82.56 / -82.55	-83.10 / -82.09	-83.06 / -81.91	-83.04 / -81.99
biomarker$_1$ AP Peak	33.32 / 37.60	32.99 / 37.50	33.59 / 38.21	33.48 / 37.63
biomarker$_2$ Dome Peak	15.52 / 15.68	12.06 / 18.18	12.04 / 19.09	11.32 / 17.51
biomarker$_3$ Max dv/dt	159.89 / 186.73	160.76 / 186.71	158.88 / 185.55	159.01 / 186.76
biomarker$_4$ APD$_{90}$	278.02 / 279.02	257.81 / 294.92	255.44 / 303.69	255.95 / 294.24

Appendix E Case Study: Courtemanche-Ramirez-Nattel Model

A natural question is how the POM approaches perform when applied to a more complicated problem involving a higher-dimensional parameter space and a larger number of summary statistics output by the model. RND, LHS and SMC methods are applied to a more sophisticated ODE model for action potentials in cardiac cells (atrial in this case), the Courtemanche-Ramirez-Nattel model (Courtemanche et al., 1998). This model uses 21 coupled ODEs to model a larger number of currents which all contribute to the overall membrane potential and how it behaves in response to an external stimulus, though the general pattern of response does still match the action potential presented in Fig. 1 of the main paper, albeit potentially without the subsequent dome peak after initial depolarisation.

In this case a larger number of biomarkers are generated from each output action potential, seven in total to roughly match published biomarker data from atrial cells (Sánchez et al., 2014). Artificial data with between subject variability is generated by varying twelve input parameters, still corresponding to the conductances (or maximum rates) of the different current flows in and out of the cell which control its membrane potential. The range of possible values for these parameters is chosen such that means and variances of the biomarker values are reasonable with respect to the published data (in this case atrial cells from patients with chronic atrial fibrillation). Further details on the model and its simulation are given below.

E.1 Model and Data

The Courtemanche-Ramirez-Nattel model (hereafter Courtemanche model for brevity) simulates membrane potentials in atrial cells by tracking current flow created by the transfer of Na^+, K^+ and Ca^{2+} ions through a number of channels. The strengths of these currents are the varied parameters, $\boldsymbol{\theta} = (g_{\text{Ina}}, g_{\text{Ibna}}, g_{\text{Ik1}}, g_{\text{Ito}}, g_{\text{Ikr}}, g_{\text{Iks}}, g_{\text{Ikur}}, g_{\text{Ica,L}}, g_{\text{Ibca}}, I_{\text{naca}-\text{max}}, I_{\text{nak}-\text{max}}, g_{\text{Icap}-\text{max}})$. Respectively these are the conductances of the fast Na^+ current, the background Na^+ current, the inward rectifier K^+ current, the transit outward K^+ current, the rapid and slow delayed rectifier K^+ currents, the ultrarapid rectifier K^+ current, the L-type Ca^{2+} current, the background Ca^{2+} current and the maximal values of the Na^+/Ca^{2+} pump current, the Na^+/K^+ pump current, and the sarcolemmal Ca^{2+} pump current.

Output action potentials are generated by stimulating the cell ten times at a pacing rate of 1 Hz, with a stimulus amplitude of 2000 pA applied for 2 ms. From these output action potentials, seven biomarkers are recorded. These include three of the four used for the Beeler-Reuter case study (action potential peak, the APD_{90} and the maximum upstroke velocity) along with the APD_{50} and APD_{20}, the resting membrane potential, and the membrane potential at a time 20% of APD_{90} (v_{20}). The biomarkers regarding the height and location of the dome peak are not used, because the Courtemanche model is intended for atrial cells, which do not necessarily exhibit dome peaks in their action potentials. In the case of time series data, the times for which the transmembrane potential is recorded are $t = (0.5, 1, 2.5, 5, 7.5, 10, 15, 25, 40, 50, 75, 100, 125, 150, 200, 250, 300)$ ms.

Artificial data is generated again by selecting ten values of $\boldsymbol{\theta}$, using a base value and a window of variability which provided a reasonable match with published data for atrial cells, $\boldsymbol{\theta_T} = (0.4, 0.1, 0.07, 0.33, 0.15, 1600, 4, 0.003, 0.0013, 1.5, 0.55, 7.5)$ and $\boldsymbol{\theta} = \boldsymbol{\theta_T} \pm 30\%$. Using

this spread of theta values to represent between-subject variability corresponds to published means and standard deviations of biomarker values for atrial cells from patients with chronic atrial fibrillation Sánchez et al. (2014) reasonably well, making the Courtemanche model case study both an example of a more difficult problem and also a more clinically realistic one. The action potentials for the dataset, along with a subset of the population of models found using SMC, is shown in Fig. 2.

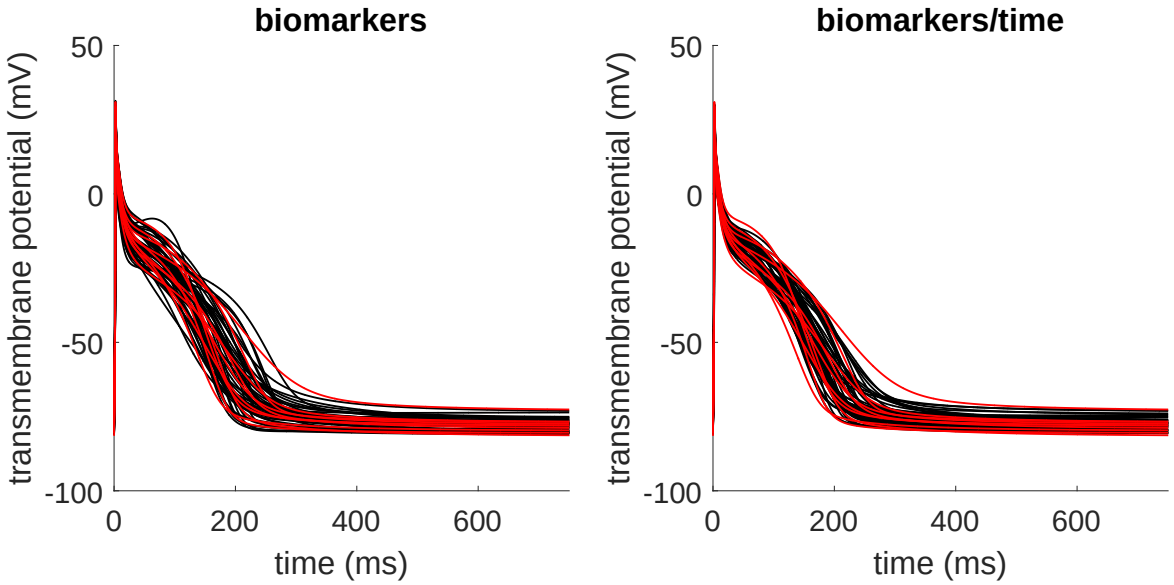

Figure 2: Solution trajectories corresponding to the artificial data for the Courtemanche model (red). A subset of 50 POMs found using the SMC algorithm is also visualised (black).

The methods as stated in the main paper are applied to the Courtemanche model. Here we use the minimum and maximum values from the artificial data to form the constraints on the biomarker and time series data. LHS is performed using 5,000 divisions in each dimension, with one sample taken to provide 5,000 potential models tested for satisfaction of the constraints on the summary statistics. Preliminary simulations suggest that RND and LHS perform similarly. SMC is applied in a similar way as described in Appendix B. For the perturbation step we use the resampled particles to fit a mixture of two beta distributions to each margin (after transforming the parameters to the unit interval (0,1)) and apply a Gaussian copula in an attempt to account for the correlation between parameters. We use $R = 7$ MCMC steps after each resampling operation for the perturbation step.

E.2 Summary statistic data

Shown are the sample mean and standard deviation (Table 2), and the sample minimum and maximum (Table 3) of all the summary statistics considered for the Courtemanche model.

Table 4: Means, standard deviations and minimum and maximum values for each of the summary statistics in the Courtemanche model.

Summary Statistic	mean (sd)	min/max
t_1 0.5 (ms)	-68.26 (2.40)	-71.65 / -63.00
t_2 1 (ms)	-58.40 (2.54)	-61.97 / -52.83
t_3 2.5 (ms)	22.99 (5.12)	15.80 / 30.11
t_4 5 (ms)	8.52 (2.97)	3.71 / 14.06
t_5 7.5 (ms)	3.41 (3.15)	-1.86 / 9.38
t_6 10 (ms)	-0.47 (3.03)	-5.74 / 5.25
t_7 15 (ms)	-6.56 (2.95)	-11.90 / -1.10
t_8 25 (ms)	-12.70 (3.43)	-18.85 / -6.48
t_9 40 (ms)	-16.65 (4.31)	-23.71 / -8.84
t_{10} 50 (ms)	-18.39 (4.74)	-25.64 / -9.83
t_{11} 75 (ms)	-22.56 (5.43)	-29.35 / -13.23
t_{12} 100 (ms)	-27.81 (5.90)	-36.14 / -19.43
t_{13} 125 (ms)	-35.06 (6.80)	-47.47 / -25.07
t_{14} 150 (ms)	-44.37 (8.90)	-60.24 / -30.86
t_{15} 200 (ms)	-62.36 (10.75)	-75.14 / -42.92
t_{16} 250 (ms)	-71.68 (6.83)	-78.48 / -55.09
t_{17} 300 (ms)	-74.90 (4.36)	-79.24 / -63.88
biomarker$_1$ $v_{peak} - v_{rest}$	101.40 (6.60)	88.79 / 109.30
biomarker$_2$ v_{rest}	-78.14 (2.37)	-81.49 / -72.95
biomarker$_3$ Max dv/dt	198.04 (36.99)	138.06 / 246.78
biomarker$_4$ APD_{90}	217.95 (38.16)	168.08 / 299.45
biomarker$_5$ APD_{50}	95.42 (29.58)	52.86 / 130.00
biomarker$_6$ APD_{20}	6.28 (2.64)	2.95 / 9.60
biomarker$_7$ v_{20}	-17.45 (4.76)	-25.66 / -9.03

Table 5: Comparison of LHS and SMC efficiency for the Courtmanche model, with efficiency calculated as the mean number of unique models found divided by the mean number of simulations required. Means and standard deviations use a sample size of ten full simulations of each method.

	biomarkers			biomarkers/times		
Method	mean sims (sd)	mean models (sd)	eff (%)	mean sims (sd)	mean models (sd)	eff (%)
LHS	5000	36 (5)	0.7	5000	0.7 (0.7)	0.01
SMC	8962 (160)	394 (15)	4.4	10561 (340)	393 (20)	3.7

E.3 Results

Performance results of the different approaches are listed in Table 5. All approaches are successfully able to find sets of input parameters with which to build a population of models that fall within the range of between-subject variability of the biomarkers. However, as expected for a type of rejection sampling, due to the curse of dimensionality the LHS approach requires a larger number of trial models to form even small populations of models. Indeed, when constraints derived from time series data are also incorporated, the LHS method quite often fails to find even a single viable model using 5000 trial points. SMC shows no such issue, retaining similar numbers of unique models from the same number of starting particles even for the more difficult problem with the additional constraints. However, the Courtemanche model has highlighted the increased difficulty of maintaining a fully diversified POM in the SMC approach, as seen by the reduction in the number of unique models found by the method in comparison to the results for the Beeler-Reuter model. We will consider alternative perturbation kernels to improve this diversity in further research.

Appendix F Results for one parameter

Histograms from the SMC POM for when only one parameter is varied is shown in Fig. 3.

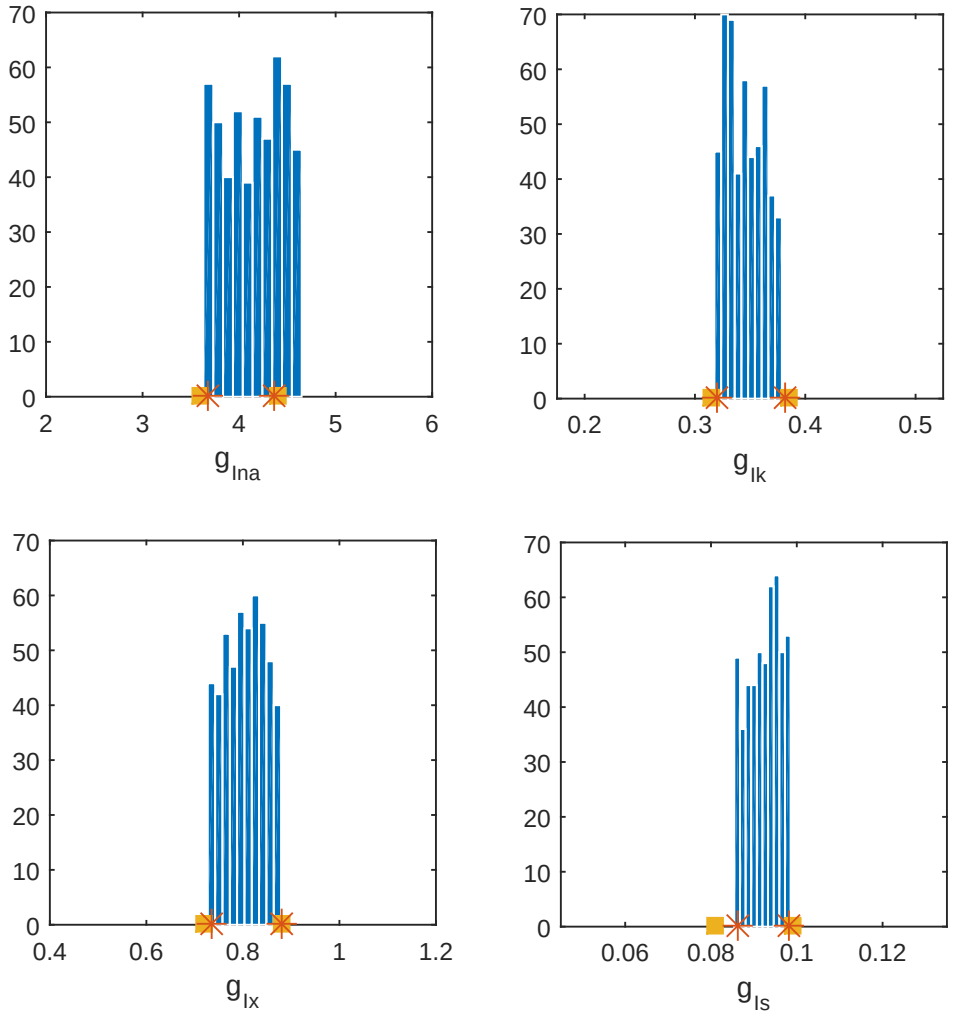

Figure 3: Histograms from the SMC POM for when only one parameter is varied. In all cases the biomarker/time statistics are used. The yellow squares denote the limits of the space that the parameter values used to generate the data are sampled from and the orange stars denote the minimum and maximum value of the parameter used to generate the data. The range of the x-axis gives the range of the parameter space explored.

Appendix G Before and After Drug Block Analysis - The 95% confidence interval case

Solution trajectories before and after drug block for the original data set and the POMs calibrated with the 95% confidence interval constraints are shown in Fig. 4.

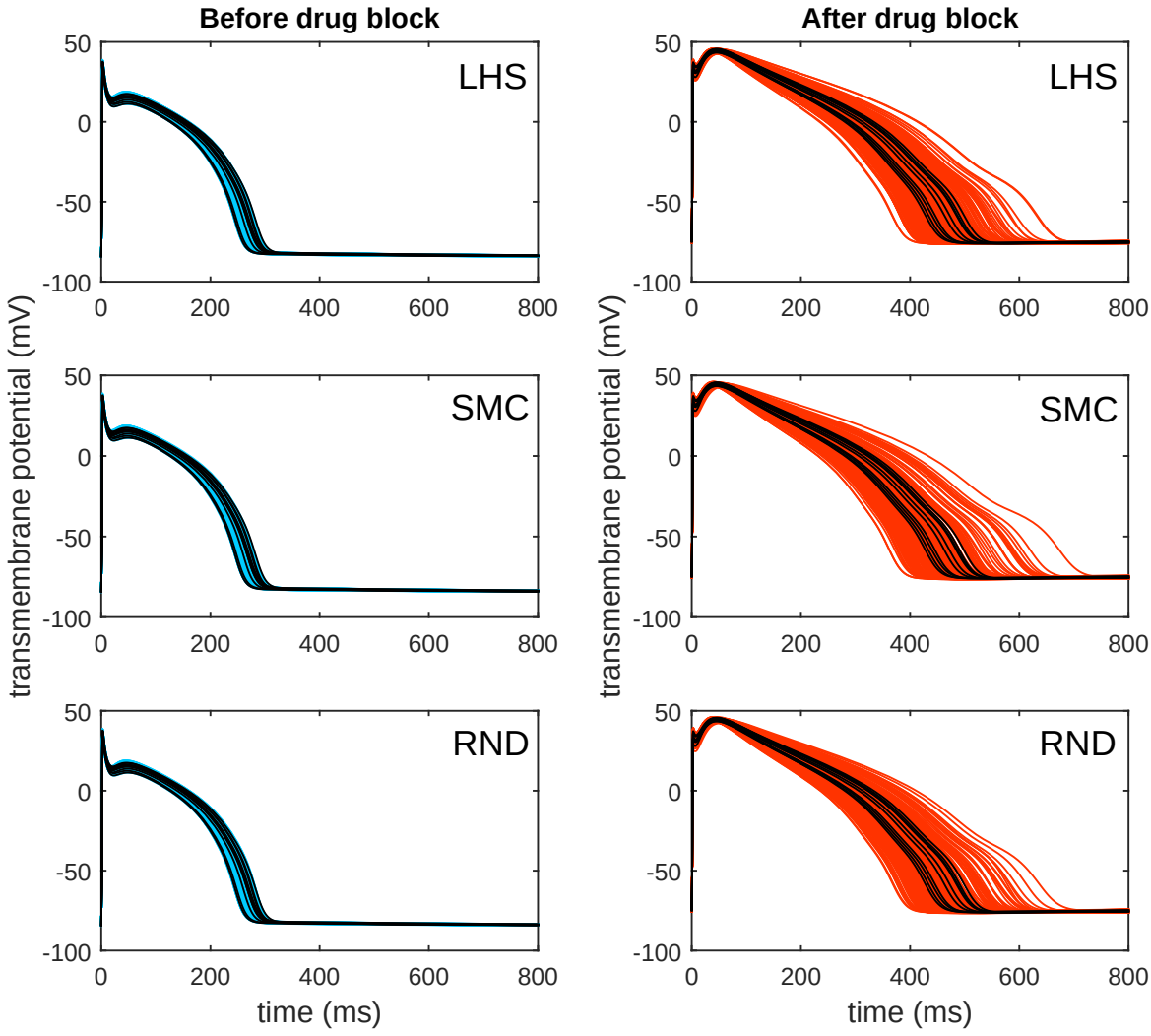

Figure 4: Solution trajectories corresponding to the original data set and the three calibrated populations of 200 models obtained with LHS, SMC and RND methods. The left hand column represents the trajectories corresponding to the set of parameters $\boldsymbol{\theta}$ (without drug block), while the right hand column represents the trajectories corresponding to the set of parameters $\boldsymbol{\theta}_{\text{mod}}$ (with drug block).

Histograms of the APD_{90} distribution before and after drug block, and scatter plots of $\boldsymbol{\theta}_{\mathbf{mod}}$ for the POMs calibrated with the 95% confidence interval constraints are shown in Fig. 5.

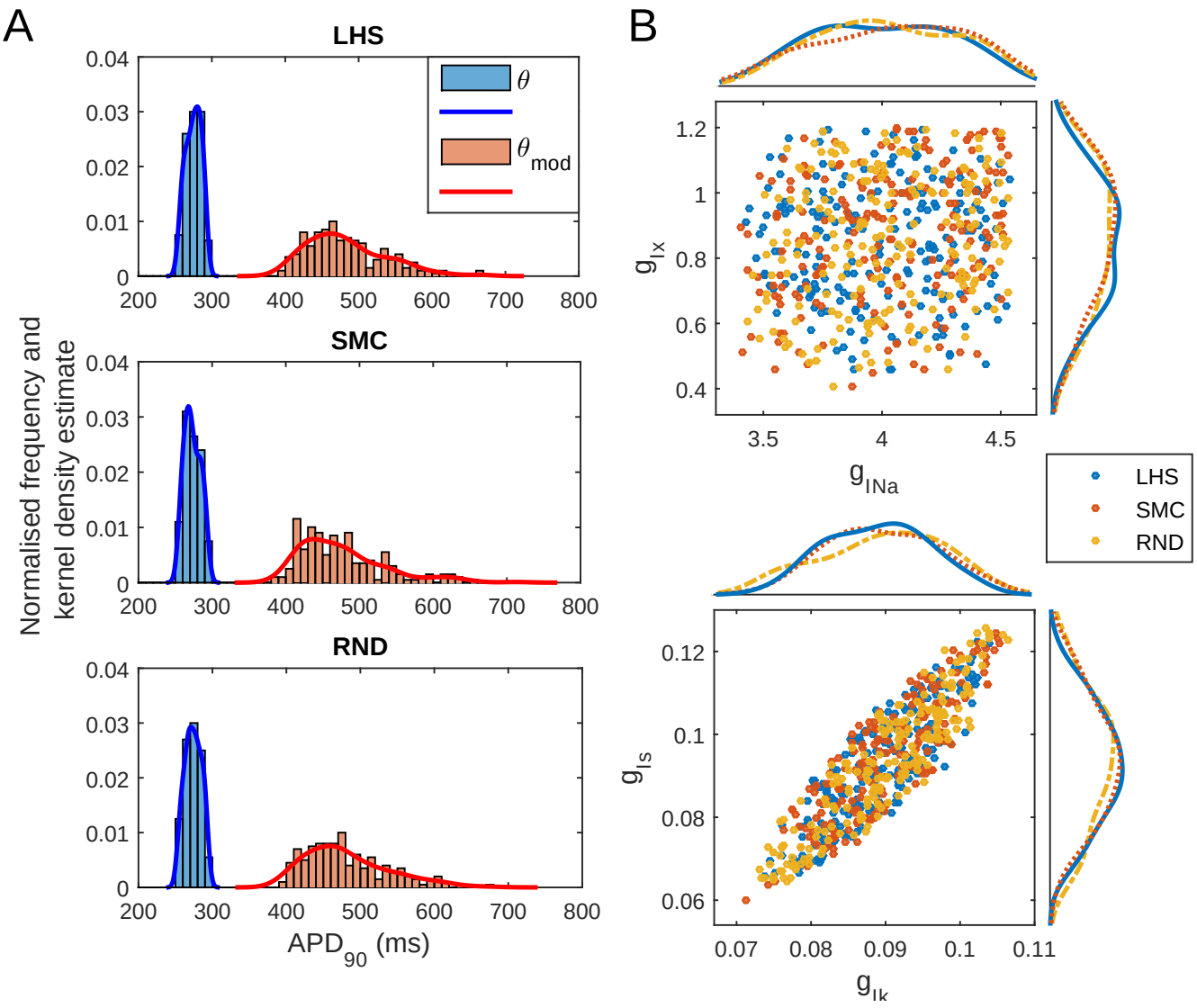

Figure 5: APD_{90} distribution and scatter plot of parameter values for the three POMs considered in the study. (A) APD_{90} distribution for the three POMs considered, before (blue) and after (red) drug block. In each plot we superimpose a kernel density estimate to the normalised histograms of the APD_{90} obtained from the solutions trajectories corresponding to the set of parameters $\boldsymbol{\theta}$ and the set of modified parameters $\boldsymbol{\theta}_{\mathbf{mod}}$. (B) Scatter plots and corresponding kernel density estimates for the location of the four components $(g_{\mathrm{INa}}, g_{\mathrm{Ik}}, g_{\mathrm{Ix}}, g_{\mathrm{Is}})$ of the 200 parameters $\boldsymbol{\theta}_{\mathbf{mod}}$ obtained with the three POM strategies.

References

Beeler, G. and Reuter, H. (1977). Reconstruction of the action potential of ventricular myocardial fibres. *Journal of Physiology*, 268:177–210.

Britton, O. J., Bueno-Orovio, A., Van Ammel, K., Lu, H. R., Towart, R., Gallacher, D. J., and Rodriguez, B. (2013). Experimentally calibrated population of models predicts and explains intersubject variability in cardiac cellular electrophysiology. *Proceedings of the National Academy of Sciences*, 110(23):E2098–E2105.

Courtemanche, M., Ramirez, R. J., and Nattel, S. (1998). Ionic mechanisms underlying human atrial action potential properties: insights from a mathematical model. *American Journal of Physiology-Heart and Circulatory Physiology*, 275(1):H301–H321.

Dutta, A. M. S., Walmsley, J., and Rodríguez, B. (2013). Ionic mechanisms of variability in electrophysiological properties in Ischemia: A population-based study. *Proceedings of the Computing in Cardiology Conference (CINC)*, 40:691–694.

Hinkelmann, K. and Kempthorne, O. (2008). *Design and Analysis of Experiments*, volume 1: Introduction to Experimental Design. John Wiley & Sons, Inc, Hoboken, New Jersey, 2 edition.

Marder, E. and Taylor, A. L. (2011). Multiple models to capture the variability in biological neurons and networks. *Nature Neuroscience*, 14(2):133–138.

McKay, M. D., Beckman, R. J., and Conover, W. J. (1979). A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 21(2):239–245.

Monasterio, V., Carro, J., and Pueyo, E. (2014). Repolarization alternans in human ventricular hyperkalaemic tissue: dependence on current stimulation. *Proceedings XXXII Congreso Anual de la Sociedad Española de Ingeniería Biomédica*, pages 1–4.

Sánchez, C., Bueno-Orovio, A., Wettwer, E., Loose, S., Simon, J., Ravens, U., Pueyo, E., and Rodriguez, B. (2014). Inter-subject variability in human atrial action potential in sinus rhythm versus chronic atrial fibrillation. *PLOS ONE*, 9(8):e105897.