Unlocking datasets by calibrating populations of models to data density: a study in atrial electrophysiology

Brodie A. J. Lawson[1*], Christopher C. Drovandi[1], Nicole Cusimano[2],
Pamela Burrage[1], Blanca Rodriguez[3], Kevin Burrage[1,3†]

[1]Australian Research Council Centre of Excellence for Mathematical and Statistical Frontiers,
School of Mathematical Sciences, Queensland University of Technology, Brisbane, Australia
[2]Basque Center for Applied Mathematics, Bilbao, Spain
[3]Department of Computer Science, University of Oxford, Oxford, United Kingdom

[*]To whom correspondence should be addressed; E-mail: b.lawson@qut.edu.au.
[†] Visiting Professor

The understanding of complex physical or biological systems nearly always requires a characterisation of the variability that underpins these processes. In addition, the data used to calibrate such models may also often exhibit considerable variability. A recent approach to deal with these issues has been to calibrate populations of models (POMs), multiple copies of a single mathematical model but with different parameter values, in response to experimental data. To date this calibration has been largely limited to selecting models that produce outputs that fall within the ranges of the dataset, ignoring any trends that might be present in the data. We present here a novel and general methodology for calibrating POMs to the distributions of a set of measured values in a dataset. We demonstrate our technique using a dataset from a cardiac elec-

trophysiology study based on the differences in atrial action potential readings between patients exhibiting sinus rhythm (SR) or chronic atrial fibrillation (cAF), and the Courtemanche-Ramirez-Nattel model for human atrial action potentials. Not only does our approach accurately capture the variability inherent in the experimental population, we also demonstrate how the POMs that it produces may be used to extract additional information from the data used for calibration, including improved identification of the differences underlying stratified data. We also show how our approach allows for different hypotheses regarding the variability in complex systems to be quantitatively compared.

Introduction

Mathematical modelling is vital for the understanding of complex phenomena, but the use of mathematical models requires careful specification of their parameter values against available data. In many applications, model predictions can vary sharply in response to even small changes in the values of their parameters, and yet experimental efforts to determine these values are invariably associated with either some kind of uncertainty or inherent variability underlying the processes that are being measured. In biological and physiological contexts, for example, not only are these uncertainties typically very large, the values of representative parameters also exhibit considerable variation between different members of a population, due to differences in physiology and genetics. Properly accounting for this variability using mathematical models is critical to furthering understanding in such fields (*1*).

With regards to uncertainty quantification (UQ), techniques such as Monte Carlo sampling (*2*), polynomial chaos expansions (*3*), and Bayesian approaches including Gaussian processes (*4*), allow for the impacts of uncertainty in parameter values upon model outputs (predictions)

to be quantified, or for parameter values and their uncertainties to be determined in response to data collected for model outputs. However, each works from the perspective of a single, immutable model with some fixed uncertainties in its inputs and corresponding uncertainties in its outputs. This becomes an issue when one wishes for example to determine which features (parameter values) in a population predict different classes of outputs, or to consider the impacts of changes to the underlying model itself.

On the other hand, a very natural approach for modelling and understanding the variability within populations is the recent technique known as populations of models (POMs) (*5–7*). In this approach, a collection of varying individuals is represented in kind by a collection of individual models, with the idea that the collection of models exhibits the same variability as the population being modelled. Although each individual model typically differs only in terms of the values of its parameters, each remains a model in its own right, allowing for subpopulations within the POM to be identified and analysed, and for the underlying model to be easily adjusted once a POM has been constructed. Somewhat related are genetic algorithms that use multiple copies of a model with differing parameter values as their organisms (*8*), although there the focus is on breeding a single model that best fits data for a single individual, and not on characterising variability in a population.

In this paper our motivating application is that of cardiac electrophysiology, although the POMs technique has also been used in biomechanics (*9*), cholesterol pharmacology (*10, 11*) and is highly relevant to systems biology in general (*12*). The action of the heart depends on the excitable, highly nonlinear (*13*) nature of cardiac cells, which undergo a carefully-controlled process of ion uptake and release in response to electrical stimulus. In addition to the temporary intake of Ca^{2+} ions that produces the cellular contraction associated with the heartbeat, control of the potential difference across the cell's membrane also prevents it from being re-stimulated too quickly. The time course of the membrane potential in response to stimulus is known as

the action potential (AP), and it is the AP or the important features of it (biomarkers) that are commonly recorded in single-cell experiments.

POMs research has been very active in this setting, inspired by developments in neuro-science (*5, 6*) and beginning with the sensitivity analysis studies of Sobie (*14*). An important advancement was then to *calibrate* these POMs in response to experimental or clinical data, so that the models in these *in silico* populations generated outputs that were physiologically reasonable (*7*). Such calibrated POMs have been used to great effect, including suggesting modifications to existing models of rabbit ventricular cells (*15*) and human atrial cells (*16*) required to reproduce specific data, determining the electrophysiological properties that lead to the dangerous phenomena of alternans (*17, 18*), repolarisation abnormalities (*19*) and atrial fibrillation (*20*), and characterising the sources of the differing function of failing hearts (*21*). The technique has also been used to explore the variable response of a population to anti-arrhythmic drug treatments (*7, 22, 23*), to the onset of ischemia in rabbits (*24*), and the effects of hypertrophic cardiomyopathy (*25*). Most relevant to our work, calibrated POMs were used by Sanchez *et al.* (*26*) to explore the differences between patients exhibiting sinus rhythm (SR) or chronic atrial fibrillation (cAF).

To date, calibration of POMs has been achieved almost exclusively by rejecting any trialled models that produce outputs that correspond to measurable quantities falling outside the ranges of observations for those same quantities in the dataset (*27*). This prevents any obviously un-physical models from being accepted into the population, but does not necessarily guarantee a good correspondence with the biomarkers' distribution in the experimental data. Selection of models according to the ranges of values observed in the data creates a feasible region that is necessarily hyperrectangular, whereas the actual multidimensional spread of experimental measurements may be a much more complex shape. Additionally, it may be desirable that the selected models are not only *feasible*, but also that together they exhibit key features of

4

the distributions of experimental data such as regions of high or low density and correlations between measured quantities. Of course, these considerations are only relevant when there is sufficient data to reasonably estimate the distributions of the measured quantities, and otherwise range-based calibration is highly appropriate. Two recent works in cardiac electrophysiology did select parameter values by ensuring that the models together exhibited appropriate mean and standard deviation for the experimental measures of interest (*28, 29*), a step towards the distribution-driven calibration technique we introduce in this publication.

We extend a recent statistically-informed sampling technique for POM construction (*22*), sequential Monte Carlo (SMC), in order to propose a new method that produces POMs that are directly calibrated to data distributions. As compared to Latin hypercube sampling (LHS), the uniform sampling approach typically used for range-based calibration (*27*), SMC identifies regions of the parameter space that correspond to a higher density of data and generates proportionally more samples in such locations. It also allows the complexity of the sampling problem to be introduced in a gradual fashion, relevant also when calibrating to ranges (*22*). We demonstrate not only how our calibration technique generates POMs that reproduce the distributions in experimental data, but more importantly how this type of calibration allows for additional insights to be gained from this data. Specifically, we demonstrate how biomarker data from myocytes taken from SR and cAF atria encodes almost all of the differences in ion channel expression that characterise the cAF pathology, but that distributions of the biomarkers must be used for calibration in order to recover them. We also show that our distribution-calibrated POMs serve as a method for parameter selection in response to data, and as a tool to explore different hypotheses regarding the sources and extent of variability in unobserved properties that manifest experimentally-observed variability. We finally conclude by discussing our new approach, when and how it should be used, and the implications for modelling and understanding variability in all its manifestations both within and outside of cardiac electrophysiology.

Experimental Data and Model Choice

The experimental data we use to demonstrate our techniques for data-calibrated POMs is that presented by Sanchez *et al.* (*26*), and consists of biomarker values measured from recorded APs for 469 cells taken from the right atrial appendages of 363 patients. These patients belonged to either one of two groups, those exhibiting standard sinus rhythm (SR) and those exhibiting chronic atrial fibrillation (cAF). Thus, not only are there two datasets for which we demonstrate our calibration process, but we are also able to explore how the *in silico* populations calibrated to the two datasets compare. The large number of recordings in the dataset makes it particularly suited to the idea of distributional calibration, because the distribution implied by the data can be meaningfully estimated.

The biomarker values used to quantify the recorded APs are the action potential duration computed at 20, 50 and 90% repolarisation (APD_{20}, APD_{50}, APD_{90}), the action potential amplitude (APA), resting membrane potential (RMP), the membrane potential at 20% of APD_{90} (V_{20}) and the maximum upstroke velocity (dV/dt_{max}). More information regarding the experimental conditions under which the data was collected is available in (*26*).

Figure 1 visualises the spread of biomarker values for both SR and cAF APs. Clear differences in the distributions of SR and cAF biomarkers are immediately apparent, and indeed when the biomarkers are considered individually all of them except for dV/dt_{max} show a statistically significant difference ($p < 0.001$) between SR and cAF myocytes (*26*). It is thus reasonable to expect that calibration to these two datasets will be indicative of the differences that underlie the cAF pathology, as compared to SR. We also note that the values of dV/dt_{max} range as high as 434V/s, which is much larger than the typical values of 150-300V/s for this biomarker in atrial myocytes (*30, 31*).

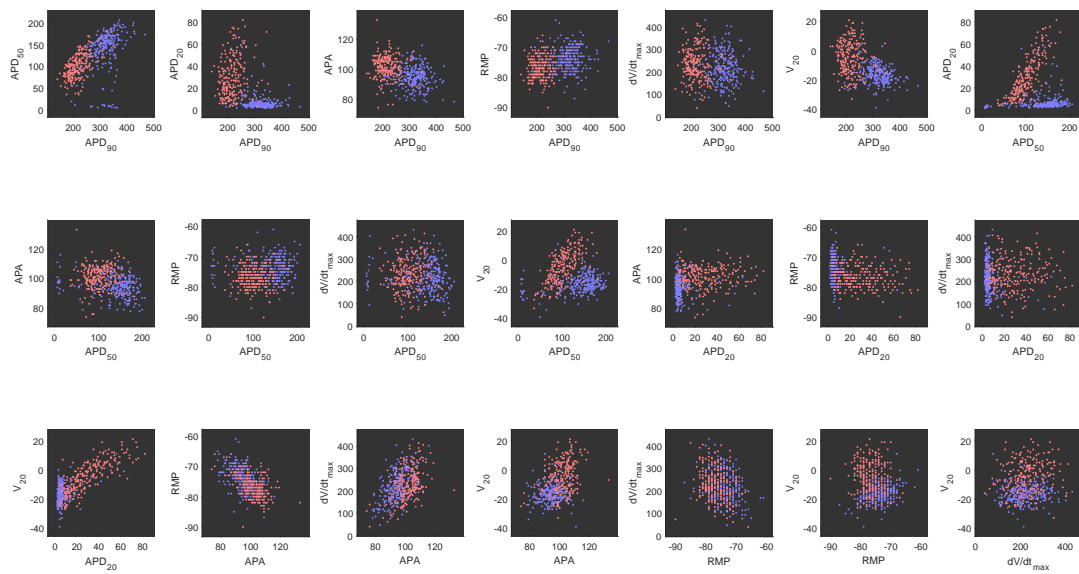To construct our *in silico* POMs, we selected the Courtemanche-Ramirez-Nattel (CRN)

Figure 1: **Experimental recordings of AP biomarkers show significant differences between SR and cAF.** Pairwise scatterplots of each unique pair of biomarkers in the SR dataset (blue) and the cAF dataset (red). Clear differences in how biomarker values are distributed between the two populations are seen, especially for the APD biomarkers. On the whole, the data exhibits a large amount of variability, highlighting the importance of characterising this variability in understanding cardiac electrophysiology.

model (*32*) following preliminary studies that suggested that out of three potential models it was most able to capture biomarker values seen in the Sanchez *et al.* dataset. A separate benchmarking study also suggested that this model, despite being one of the first developed for human atria, predicted APDs very well for data from both SR and cAF patients (*33*). POMs were then constructed by allowing the conductances of the model's most important currents to vary (see Materials and Methods for further details and justification). However, as will be seen later our approach initially struggled when attempting to replicate the larger values of dV/dt_{max} in the dataset, and the relationship between this biomarker and APA. Modifying the search parameters to additionally include a time constant associated with the inwards Na^+ current, and allowing the conductance of this current to vary more significantly, allowed these features of the dataset to be much better captured.

Additionally, we later make use of data regarding ion channel remodelling under cAF from a variety of individual studies. This data is used not for calibration, but solely for the comparison between what has been observed in the literature and what is predicted by our calibrated POMs.

Results

SMC Significantly Improves Calibration of Populations of Models to Data

We used 2000 particles to initialise our SMC sampling process (see Materials and Methods for details of the algorithm), obtaining POMs composed of 1938 unique models for the SR data, and 1931 unique models for the cAF data. Using 10000 trialled models generated using LHS (10 samples with 1000 divisions in each parameter dimension) produced 1319 accepted models for the SR dataset and 1338 models for the cAF dataset after calibration to biomarker ranges. We note that these numbers should not be compared as a measure of efficiency, for the SMC algorithm involves multiple model runs for each particle and is solving a more difficult sampling problem that takes the distributions of the data into account.

For the SR dataset the SMC-calibrated POMs show a significantly better degree of localisation to data-dense regions in the biomarker space, when compared to POMs calibrated to biomarker ranges (Figure 2). Over the selected range of parameters, the CRN model tends to produce APD_{90} and RMP values lower than the majority of the data, and APA values higher than the majority of the data, and so such models are then over-represented in the LHS POM, which only rejects models that fall outside of the range of the data. Our POM calibrated to data density using SMC greatly reduces the impacts of this 'model bias', and also captures very well the extent of variance in the V_{20} values in the dataset. However, there are still clear discrepancies in the distributions for RMP and dV/dt_{max}, especially the latter where neither the distribution-calibrated or range-calibrated POMs produce any models with dV/dt_{max} at the upper end of the range in the data. We specifically address this issue later in the paper. Figure 2b shows the constructed POMs and the data in terms of each of the different pairs of biomarkers, making it clearer how our SMC sampling approach is selecting models more concentrated in regions of higher data density. However, it is evident that a further removal of some models of the population would improve the concordance between POM and data, highlighting the importance of our refinement process that we describe subsequently. The results for the cAF dataset are very similar (see Figure S1).

Refinement via Selection of Optimal Subpopulations Captures Well the Variability in Datasets

In order to further improve the fit between constructed POMs and data, we include a second phase of our calibration process that selectively removes models from the SMC-constructed POMs, which we term "refinement". Our easily-approximated divergence measure ρ provides a means for selecting this subpopulation, using a simulated annealing-type algorithm (see Materials and Methods, "Further POMs Refinement").
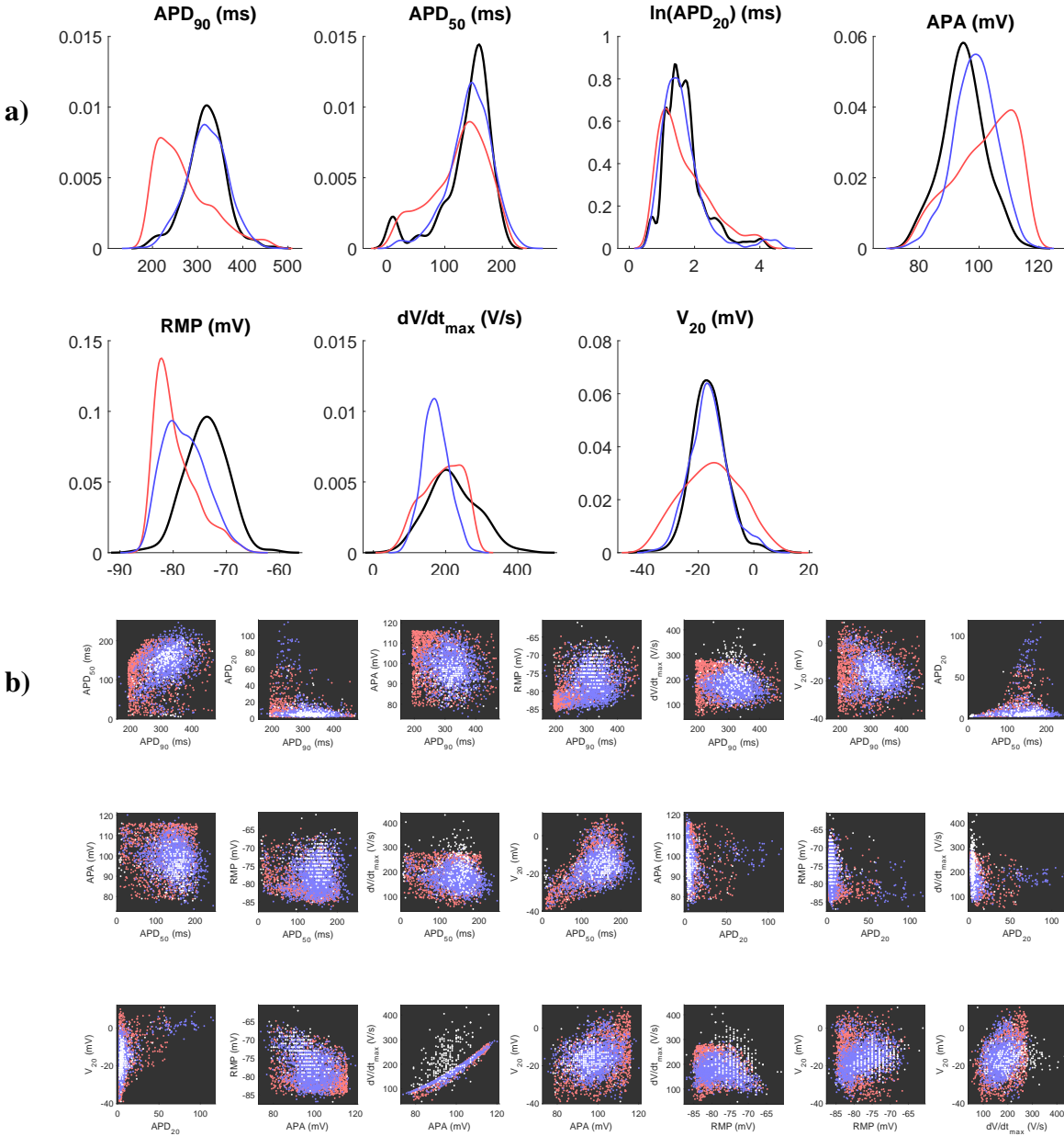
Figure 2: **Calibration to biomarker distributions as opposed to their ranges reduces model bias. a)** Marginal distributions of the biomarkers in the SR dataset (black) and POMs calibrated to biomarker distributions using the SMC algorithm (blue) or calibrated to biomarker ranges using LHS (red). The natural logarithm of APD_{20} values is used to better display their distribution. SMC for distributional calibration is seen to provide a significant improvement in agreement with the data. **b)** Pairwise scatterplots of each unique pair of biomarkers in the SR dataset (white) and the POMs constructed using SMC matched to distributions (blue) and LHS matched to ranges (red). The SMC-generated POM demonstrates good localisation to the dense regions in the data, but clearly requires further calibration. An obvious correlation between APA and dV/dt_{max} is exhibited by the model, regardless of the sampling method used, but this correlation is not present in the data.

For the SR dataset, minimising ρ produced a refined POM of 275 models with very good representation of variability in the dataset (red line in Figure 3), with the marginal distributions of the biomarkers matching very well apart from APD_{20}, APA and dV/dt_{max}. The inability of the calibration process to capture the distributions of these biomarkers is predominantly explained by the strong correlation in the model between APA and dV/dt_{max}, as clearly seen in Figure 2b (third row, third column). This correlation inescapably ties these two biomarkers together, and prevents the refinement process from selecting models that are appropriately spread across both at once.

The inability to disentangle dV/dt_{max} from APA in the model, along with the model's consistent underestimation of this biomarker as compared to the data, motivates the use of a second divergence measure $\hat{\rho}$ (see Materials and Methods) that reduces the emphasis on this biomarker in the calibration process. When a subpopulation that minimises $\hat{\rho}$ is selected, the result is a POM composed of 327 models that shows slight improvements in the marginal distributions of the other biomarkers and a more significant improvement in APD_{20} (blue line in Figure 3). As might be expected, the distribution of APA values is now exceedingly well-represented in the POM, as the algorithm's efforts to minimise the divergence from the maximum upstroke velocity no longer hamper its ability to fit the distribution of the highly correlated APA. Table S1 also shows how the key statistical properties of the biomarker data are captured by our refined POM. Similarly good performance is also achieved by refined POMs calibrated to the cAF dataset (Figure S2 and Table S1).

Figure 3b provides a visual demonstration of the two-phase calibration process, showing the selection of a subpopulation of SMC-selected models that corresponds well to the density of data across the biomarker space by minimising $\hat{\rho}$. Although a large proportion of models are rejected by this process (83% for the SR dataset), larger POMs can be generated as desired by using additional particles in the SMC algorithm, or by modifying the 'energy' minimised
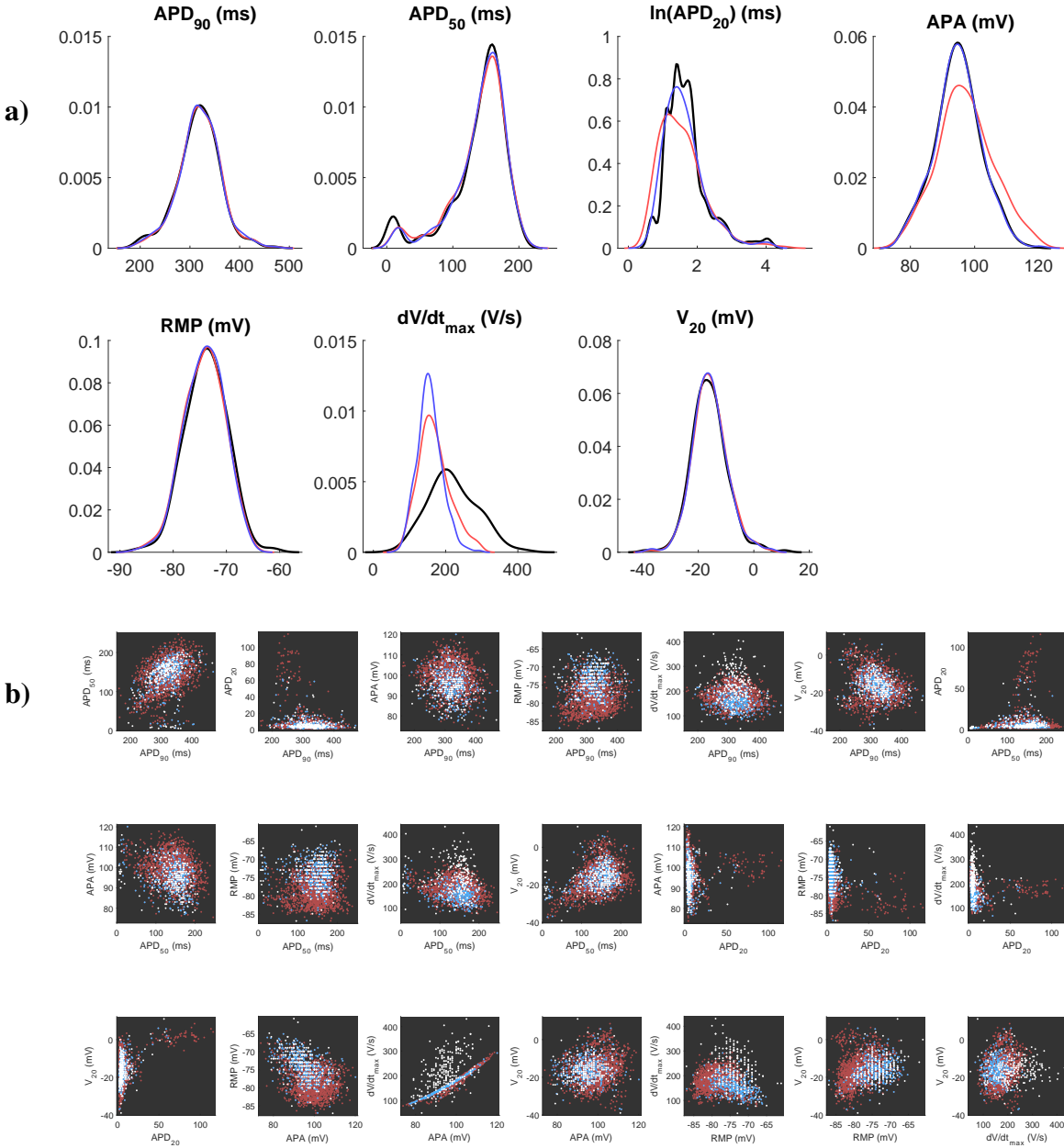
Figure 3: **Selection of an optimal subpopulation almost fully captures biomarker variability.** **a)** Marginal distributions of the biomarkers in the SR dataset (black) and POMs selected as subpopulations of the SMC-generated POM that minimised ρ (red) or $\hat{\rho}$ (blue). The simulated annealing algorithm clearly succeeds at selecting a representative subpopulation, but the distributions of the APA and dV/dt$_{max}$ are not quite captured. De-emphasising dV/dt$_{max}$ in the calibration process provides very good capture of variability in all other biomarkers. **b)** Pairwise scatterplots of each unique pair of biomarkers in the SR dataset (white) and the models from the SMC-generated POM that were accepted (light blue) or rejected (dark red) in the process of minimising $\hat{\rho}$. Outside of dV/dt$_{max}$, especially its relationship with APA, the features of the data are very well represented by the final POM.

in the refinement process so that POMs of a larger size are encouraged. We also note that the improved performance of calibrating to our modified divergence measure $\hat{\rho}$ comes at little cost to the value of the original divergence measure ρ, as shown by Table S2. For this reason we favour POMs calibrated by minimising $\hat{\rho}$ in the remainder of this work.

Our SMC sampling process with subsequent refinements has been seen to successfully produce POMs that accurately reflect the distributions of outputs in the data, for two (SR and cAF) datasets that show significant variation between individual samples. This was achieved using a strongly nonlinear model and moderately high numbers of both variable parameters and observed biomarkers. However, the issue remains that our POMs show a strong correlation between the biomarkers dV/dt$_{\text{max}}$ and APA, despite very little relationship between these two biomarkers in the data. We now explore whether further variability in additional parameters can explain this discrepancy.

Variability in Na$^+$ Ion Channel Inactivation Further Explains the Trends in Upstroke Velocity

The upstroke portion of the AP is controlled primarily by the inwards sodium current I_{Na}, and so we focus our exploration of additional variability on this current. One hypothesis is that variation in the time constant of the "h" gate, which controls the initial inactivation of I_{Na}, could explain the lack of correlation between upstroke velocity and APA in the data, yet implicit in the calibrated POMs. If the rate at which I_{Na} switches off during the upstroke is varied, then this serves to decouple these two biomarkers. For example, a fast upstroke need not necessarily also lead to a high APA if I_{Na} deactivates more quickly. Variability in the rate of this inactivation has been experimentally observed in human atrial myocytes (*34*).

Introducing a scaling parameter for this time constant, $\tau_{h(scale)}$, we performed a new calibration but with this additional parameter allowed to vary $\pm 100\%$ along with the channel conduc-

tances. Additionally, we increased the range of g_{Na} so that it could take values up to $+300\%$, in order to give the CRN model the best chance at realising the very high upstroke velocities seen in the dataset. Refinement for this study minimised ρ, not $\hat{\rho}$, as our goal is to specifically explore whether or not the distribution of dV/dt$_{\text{max}}$ in the data can be captured by our calibration process.

The new distribution-calibrated POMs show less correlation between dV/dt$_{\text{max}}$ and APA, although the two biomarkers still remain more strongly correlated than in the data. The result is best demonstrated for the cAF dataset, shown in Figure 4a. Furthermore, the marginal distribution of dV/dt$_{\text{max}}$ is significantly improved (Figure 4b), with little lost in the fit to distributions of the other biomarkers (see Table S2). The results for SR are very similar (Figure S3). This suggests that variance in the inactivation rate of I_{Na} may be an important contributor to the variability seen in AP upstrokes in the data.

Distributional Calibration Produces Populations of Models that Capture Important Data Features

Previous POM studies in cardiac electrophysiology have been almost entirely limited to varying only ion channel conductances, given their established importance to variability in cardiomyocyte behaviour (*27*). This reflects the large variation of ion channel numbers across individual myocytes. Furthermore, there are less clear mechanisms for incorporating the variability associated with additional parameters controlling the time and voltage dependencies in ion channel activation/inactivation. For these reasons, we return to the original POMs where the only variable parameters are the channel conductances, and refinement is performed by minimising $\hat{\rho}$, reducing the emphasis on dV/dt$_{\text{max}}$. We now investigate how these distribution-calibrated POMs can be used to inform cardiac electrophysiology, and hence demonstrate the advantages offered by our calibration technique.
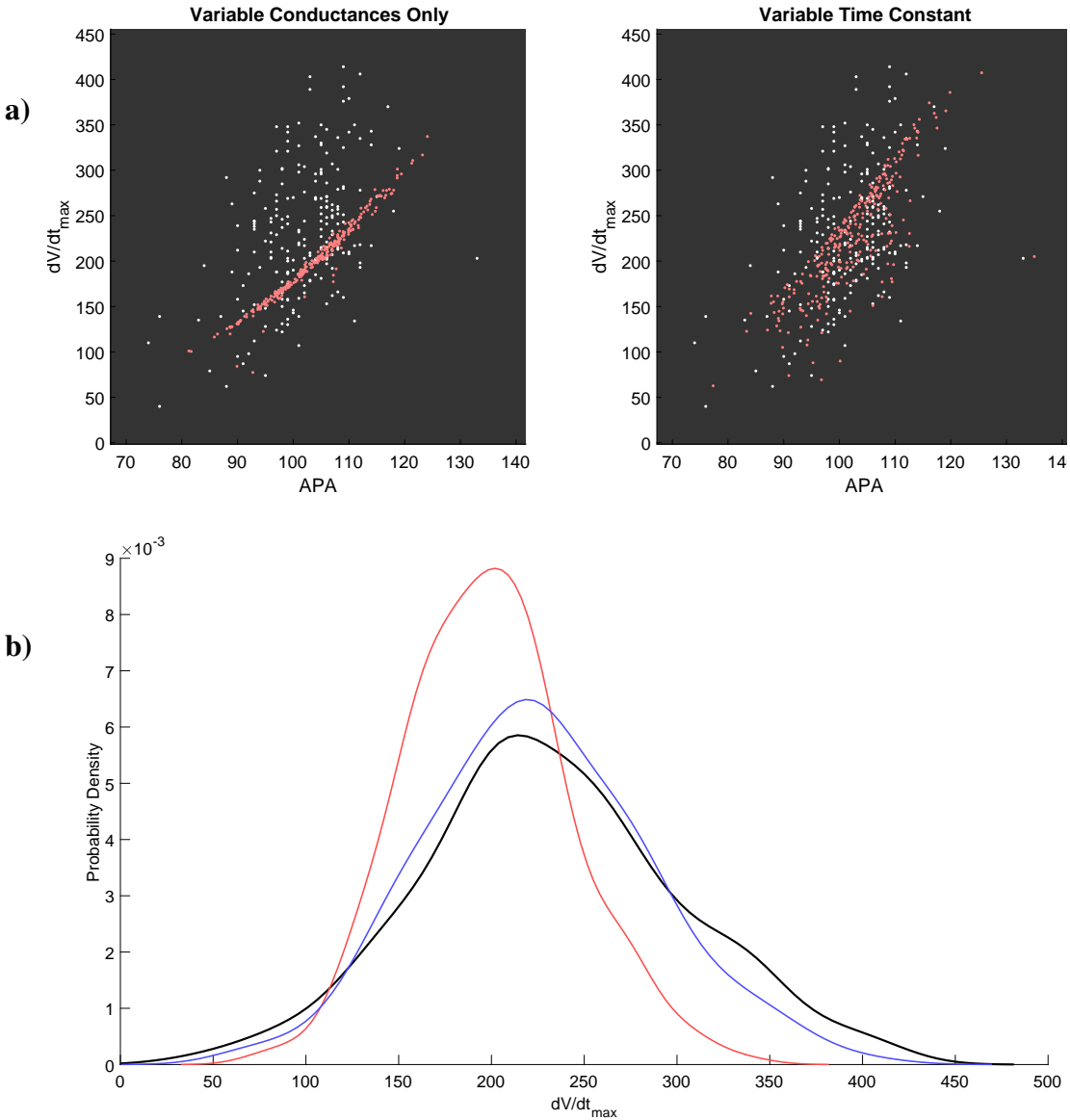
Figure 4: **Further variance in I_{Na} improves the realisation of dV/dt$_{max}$ values in the cAF dataset. a)** Pairwise scatterplot of APA and dV/dt$_{max}$ values in the cAF dataset (white) and those accepted by distribution-calibrated POMs minimising ρ (red). Allowing variance in the time constant significantly reduces the correlation between these two biomarkers in the POM, better realising the spread of the data. **b)** Marginal distribution of dV/dt$_{max}$ values in the cAF dataset (black) and the calibrated POMs varying only current conductances (red) or with additional variance in I_{Na} conductance and inactivation time (blue). This additional variance allows our calibrated POM to almost capture the marginal distribution of dV/dt$_{max}$ values were the original POM fails.
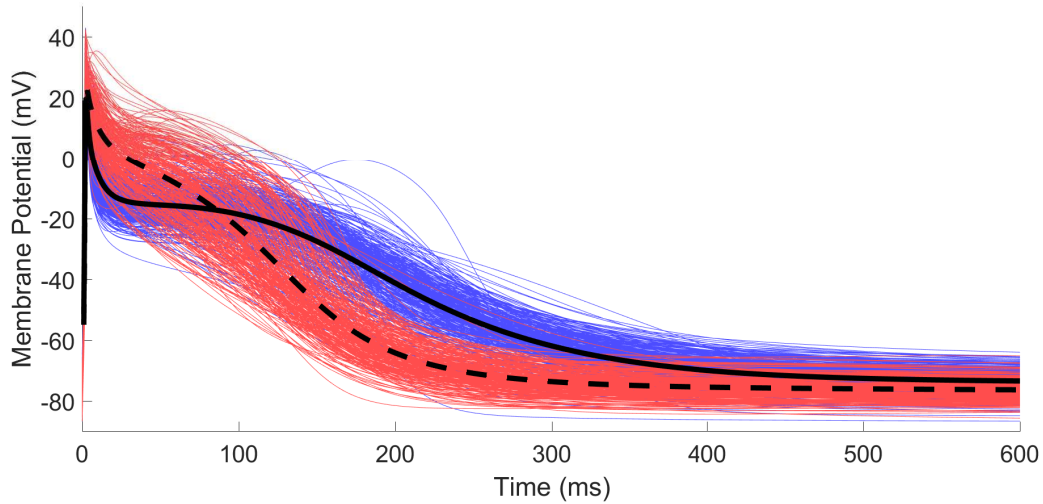
Figure 5: **Distributional calibration captures the morphological differences between SR and cAF atrial APs.** Atrial APs produced by simulation of the populations of CRN models calibrated to biomarker data for patients exhibiting sinus rhythm (blue) and chronic atrial fibrillation (red). Also displayed are the average of all traces for the sinus rhythm (solid) and atrial fibrillation (dashed) populations. The increase in AP triangulation and reduced refractory period associated with chronic atrial fibrillation is clearly demonstrated, especially by the averaged traces.

We begin by directly considering the most important output of the populations of CRN models that we have produced, the APs, shown in Figure 5. Significant differences between the SR and cAF populations are immediately observed. SR APs demonstrate an initial period of very rapid repolarisation after the AP peak, then an extended plateau phase that eventually resumes gradual repolarisation back to the resting potential. In contrast, cAF APs show much less significant initial repolarisation, but their lack of any significant plateau phase and overall faster repolarisation produces significantly lower APDs (and hence a decreased refractory period).

These features are well known to be associated with cAF, which is characterised by far more triangular APs that lack a noticeable plateau phase and return to resting potential more rapidly than APs in healthy sinus rhythm (*35*). The morphological differences between SR and cAF APs are seen in the data, with the rapid repolarisation followed by plateau in the SR population

16

implied by very small APD_{20} values and larger APD_{50} values, while the AP triangulation in the cAF population is seen in larger APD_{20} but smaller APD_{50} values. Calibrating to distributions naturally takes these features of the data into account, successfully selecting models that predict the appropriate morphologies. When calibrating to ranges for this dataset, the models selected do successfully produce APs that show the reduced APD as associated with cAF, but are less successful in predicting the accompanying differences in AP morphology (see Figure S4, also Figures 2 and 3 in (*26*)). We note that previous studies have created additional biomarkers that can be expressed in terms of the original biomarkers, such as measures of triangulation based on combinations of different APD values, in order to allow more effective calibration by capturing these additional features — see (*21*). Our method, however, does not depend upon identifying the important trends in data or designing additional outputs to capture them, making it generally applicable.

Accurate prediction of the specific shapes of the APs for SR and cAF patients is important as it suggests that the differential actions of the many ionic currents that together produce the AP are being well captured by the POMs calibrated to distributions. This is critical when it comes to using these POMs for further analysis, such as considering the response of the different members of the population to drug treatments that act on specific membrane currents (*7, 22*), or identifying the differences in underlying electrophysiology that characterise the two populations. We demonstrate these aspects in the following subsection.

Distributional Calibration Produces POMs that Capture Key Atrial Electrophysiological Aspects

(i) Impacts of cAF-induced Remodelling

We have constructed POMs calibrated to the SR and cAF datasets by varying the relative strengths of the different currents that contribute to the human atrial AP, and thus any sig-

nificant differences in parameter values selected for the two datasets suggest that it is changes in these currents that produce the modified APs associated with the cAF pathology. Indeed, electrical remodelling of atrial myocytes that changes the densities of their different ion channels is a well-known feature of cAF and contributes to the persistence of the condition (*36*). Dobrev and Ravens (*37*) provide a review of the experimental evidence for the changes in current density associated with cAF, although further remodelling has since been experimentally identified (*31*).

Sanchez *et al.* (*26*) also compared the POMs generated for the CRN model and the models of Maleckar *et al.* (*38*) and Grandi *et al.* (*39*) when calibrated to the ranges of the SR and the cAF data, by varying the six currents identified as most important to AP properties. They identified a statistically significant upregulation of I_{K1} in all three models, with changes in other currents found to be model-dependent. In the case of the CRN model, Sanchez *et al.* also found statistically significant decreases in I_{CaL} and I_{to} in accordance with experimental observation (*36*). However, the observed decreases in I_{to} and I_{CaL} were quite small and the POMs constructed failed to identify I_{Kur} as a significantly downregulated current in cAF. Furthermore, I_{NaCa} showed a statistically significant decrease, despite Na^+/Ca^{2+} exchanger action being known to *increase* in cAF-afflicted atria (*40*).

In contrast, our POMs calibrated to data density show many significant differences in current strengths between the SR and cAF populations (Figure 6), and do very well at identifying the currents that are known to be remodelled in response to cAF (Table 1). The distribution-calibrated POMs correctly identify the primary changes in current activity that characterise cAF (decreased I_{to}, I_{Kur} and I_{CaL}, and increased I_{K1}), and in the case of I_{to} and I_{Kur} also does very well in identifying the specific extent of these changes. Our POMs also correctly identify the upregulation of I_{NaCa} and its extent, and only miss the upregulation of I_{Ks}. Although upregulation of I_{Ks} in cAF has been experimentally shown to be very strong ($+100\%$) (*41*), the rather
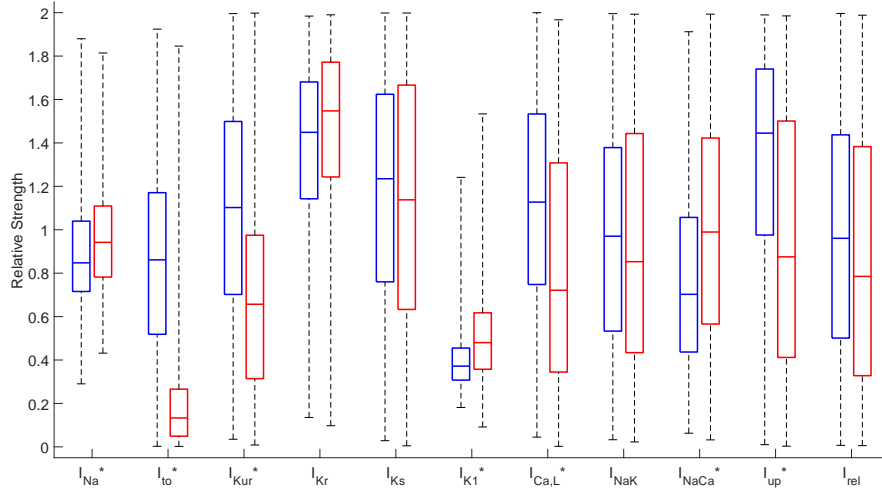
Figure 6: **Accepted parameter values for the SR and cAF populations predict well the changes in ionic behaviour associated with the cAF pathology.** Boxplot of θ values composing the POMs calibrated to the SR (blue) and cAF (red) datasets. Values are expressed in relation to the base parameter values for the CRN model. Current densities that show statistically significant differences ($p < 0.001$ from the Mann-Whitney U test) are indicated with a *. The currents most well-known as remodelled in cAF (I_{to}, I_{Kur}, I_{K1} and I_{CaL}) all show significant differences in the correct directions.

minor role that this current plays in normal repolarisation makes any changes in it much less significant and thus difficult to detect through biomarker recordings. Lastly, our POMs suggest that the net change in atrial SERCA function (Ca^{2+} uptake) under cAF is a decrease, interesting in light of the combination of positive and negative impacts on SERCA function that have been observed experimentally (*42, 43*). Our prediction of a net decrease agrees with the modification used by Grandi *et al.* (*39*) to represent the cAF case in their AP model.

When we compared the SR and cAF POMs constructed using LHS calibrated to biomarker ranges, the results were similar to those seen in the Sanchez *et al.* study, even though our POMs allowed more currents to vary in strength. Listed in Table 1 and visualised in Figure S5, the range-calibrated POMs underestimate the decreases in I_{to} and especially I_{Kur}, and do not identify I_{CaL} as an important current in electrical remodelling under cAF. I_{NaCa} is also suggested to

Parameter	Exp.	POMs (dist.)	POMs (ranges)
g_{Na}	$\updownarrow^1$ (*44*)	$+11\%$	$\leftrightarrow$
g_{to}	$\sim -70\%$ (*37*)	-85%	-51%
g_{Kur}	$\sim -50\%$ (*37*)	-40%	-6%
g_{Kr}	$\leftrightarrow^2$ (*31*)	$\leftrightarrow$	$+10\%$
g_{Ks}	$\sim +100\%$ (*41*)	$\leftrightarrow$	$\leftrightarrow$
g_{K1}	$\sim +100\%$ (*37*)	$+29\%$	$+33\%$
g_{CaL}	$\sim -70\%$ (*37, 45*)	-36%	$\leftrightarrow$
$I_{NaK(\text{max})}$	$\leftrightarrow$ (*46*)	$\leftrightarrow$	$+10\%$
$I_{NaCa(\text{max})}$	$\sim +40\%$ (*40*)	$+41\%$	-18%
$I_{up(\text{max})}$	$\updownarrow^3$ (*42, 43*)	-39%	$\leftrightarrow$
k_{rel}	$\uparrow^4$ (*42*)	$\leftrightarrow$	$\leftrightarrow$

Table 1: **Experimentally observed changes in current density associated with cAF are well predicted by POMs calibrated to distributions.** Changes in median current activities between the POMs calibrated to either the distributions, or the ranges of the SR and cAF datasets, as compared with experimentally observed (Exp.) measurements of changes in current densities associated with this pathology. Experimental figures are taken from the specified references and rounded to the closest 10% to reflect the general uncertainty in their measurements, and in some cases represent the combined result of multiple studies. The $\leftrightarrow$ symbol indicates no significant change observed ($p \geq 0.01$ from the Mann-Whitney U test for POMs), and other symbols used are explained by the following notes: [1]Peak I_{Na} current is reduced, but sustained I_{Na} increased. [2]Decreases in mRNA levels have been observed (*37*), but no direct experimental evidence for I_{Kr} change in cAF has yet been provided. [3]Ca^{2+} uptake is reduced by decreased Serca2a levels, but increased by enhanced phosphorylation of SERCA inhibitors. [4]Ca^{2+} release is increased, but in a 'leaky' fashion not necessarily best represented by changes to k_{rel} in the CRN model. Distribution-calibrated POMs detect more of the differences in current densities that underlie the cAF pathology, correlating well with experimentally-observed changes in the greatest majority of current densities.

be downregulated, and small but significant differences in I_{Kr} and I_{NaK} are identified despite these not agreeing with experimental evidence.

What we have seen is that, surprisingly, the AP biomarker data alone is enough to quantify the majority of impacts on ion channel conductances caused by electrical remodelling under cAF. However, in order to successfully extract this information from the dataset using POMs, it was required that we make use of the full amount of information contained within the data by calibrating to the distributions of these biomarker values. Our calibration technique thus serves as a means of detecting more subtle differences between multiple datasets (or stratified data), and the calibrated POMs that reflect these differences may then inform what changes in model parameters underlie them.

(ii) Response to Anti-arrhythmic Treatment

Arrhythmias in the heart are typically treated by drugs that block specific ion channels, reducing the impact of the corresponding current(s) on the AP. A common target is the rapid component of the delayed K^+ rectifier current (I_{Kr}), which activates comparatively late in the AP and is a primary contributor to repolarisation in this phase. Reducing flow due to this current hence prolongs the AP, and can restore SR in patients with cAF (*47*). I_{Kr} was also the current chosen to explore the differential response of a variable population to drug treatment using POMs in a previous study (*7*).

In our POMs calibrated to the SR and cAF datasets, the cAF models show significantly larger I_{Kr} that also activates slightly earlier (Figure 7a), contributing to the more rapid repolarisation and lack of a plateau phase in the cAF APs. However, the maximum conductance of the channel, g_{Kr}, shows no significant difference between the SR and cAF POMs (see Table 1). This indicates that the increase in I_{Kr} activity is symptomatic of the changed AP morphology in cAF, which impacts the voltage-dependent gating behaviour of these ion channels.
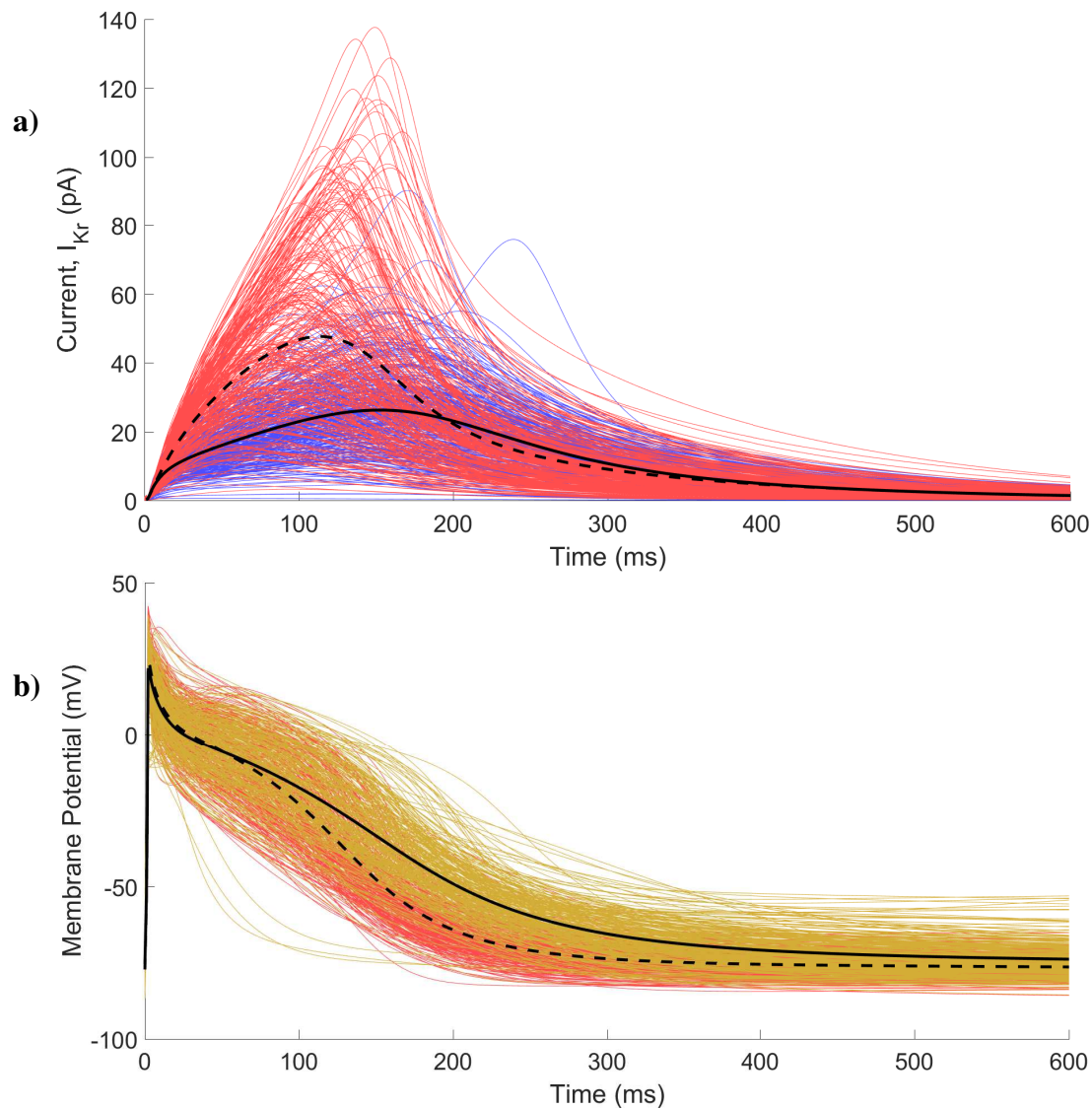
Figure 7: **Calibration to distributions produces models that respond appropriately to anti-arrhythmic treatment via I_{Kr} block. a)** Traces of the rapid component of the delayed rectifier K$^+$ current for the models calibrated to the SR dataset (blue, solid) and the cAF dataset (red, dashed). The cAF population demonstrates an almost twofold increase in the activity of this current. **b)** APs after treatment by 50% I_{Kr} block (gold) show significant AP prolongation compared to the same models without I_{Kr} block (red). This is also demonstrated by the averaged traces for both (black lines, treated – solid, untreated cAF – dashed). The restoration of atrial refractoriness in patients with cAF is clearly demonstrated.

We explore I_{Kr} block via drug treatment by first pacing the models in the cAF POM until steady state (see Materials and Methods), then reducing g_{Kr} by 50% and repeating the full stimulus protocol. In correspondence with the observed effects of such drug treatments, the APDs of almost all models (95%) are restored to values associated with healthy SR (Figure 7b). This demonstrates the ability of our calibration to AP biomarker measurements to produce models that exhibit appropriate behaviours even in situations outside of those to which they were calibrated. However, there are a small number of models that predict *decreased* APD in response to treatment, as shown by the post-drugblock APs (gold) that fall to the left of the untreated cAF APs (red), as well as a few models that repolarise to unreaslitically high resting potentials ($> -60\,\text{mV}$). One advantage of the POMs framework is that these models that show unexpected behaviour can be directly examined in order to determine the underlying causes, potentially identifying risk factors for these adverse reactions.

The models that repolarise extremely rapidly following treatment with I_{Kr} blocker are seen to be associated with very small values of g_{NaK}, and thus the unexpected behaviour of these models likely stems from the selection of values for this parameter which are too small to be physiologically realistic. This is a risk of choosing such a large ($\pm100\%$) extent of variation of our parameter values, further motivating our exploration of a lower extent of variability in the following section. Nevertheless, these models do identify a potential risk factor for the use of I_{Kr}-blocking treatments, namely that insufficient I_{NaK} activity can result in dangerous further reduction of the refractory period. Reduced I_{NaK} under I_{Kr} block conditions has also been implicated as a risk factor for repolarisation abnormality in a population of human ventricular models (*19*). In this case, examining the current activity in the low-g_{NaK} models reveals significant Na$^+$ ion accumulation that occurs due to the reduced action of I_{NaK}, which is further hampered by the reduced expulsion of K$^+$ ions through I_{Kr}. This then triggers extreme currents outward through the Na$^+$/Ca^{2+} exchanger, resulting in the extra-rapid repolarisation that

23

is observed.

The models repolarising to resting potentials that are unrealistically high are seen to be associated with incomplete deactivation of the L-type calcium channels and lower values of g_{K1}, resulting in an imbalance of inward and outward current in the unexcited cell that gradually pushes up its membrane potential at rest. Eventually, an alternative steady state is reached where the elevated resting potential largely prevents the cell's sodium channels from opening and the AP is severely disrupted. Parameter values that lead to this behaviour will never be selected by the calibration process, because their APA and RMP values fall outside of the data. However, when outward current due to I_{Kr} is reduced by drug treatment and the balance of ion flow is changed, a few of the models then fail to achieve correct homeostatic balance and instead end up at this alternative steady state. The questions of whether other AP models also predict such alternative steady states, and whether drug treatments can cause individual atrial cells to develop disrupted balances of ion flow at rest (compensated for by their neighbours), are beyond the scope of this paper.

The drug block case study can thus be seen as a means of further calibrating the generated POMs, by testing the ability of all of the models selected to continue predicting reasonable AP curves when subject to established treatments. This is important given the tendency for currents to compensate for one another, resulting in model behaviours that only become manifest subject to this type of further interrogation. In the previous study of Britton *et al.* (*7*), albeit using a different AP model, calibrating to biomarkers recorded for multiple different pacing frequencies was seen to be sufficient for avoiding the selection of models that exhibit unphysical responses to drug block.

(iii) Parameter Selection in Response to Variable Data

For building or modifying AP models, experimental recordings of current actions and the APs themselves are of course essential for the appropriate selection of parameter values. However, the variability present in these recordings makes the selection of model parameters in response to experimental data difficult. Typically such recordings are averaged, providing single traces for currents and APs that can then be calibrated to by inspection or with curve-fitting algorithms. However, this ignores a lot of information present in the full dataset, and can raise the question of which parameter values to use when AP models are known to permit multiple sets of parameter values that all lead to an equivalent AP trace (*48*).

POMs calibrated successfully to data density represent a set of models that all correspond in a sense to some portion of the data. Thus the entirety of the data is reflected in the selection of parameter values across a distribution-calibrated POM, providing a potentially more robust means of selecting an individual set of parameter values for an AP model in response to that data. This is achieved by averaging in the space of parameter values, as opposed to in the space of measured outputs. In the case of this work, the set of parameter values selected for both the SR and cAF POMs have a rather regular distribution (no obvious bimodal behaviour or obvious correlation structures, see Figure S6), and so we consider each variable parameter individually and take the median of its value across all models in the POM, producing two modified CRN models, one each for the SR and cAF datasets.

Figure 8a compares our "median" CRN models to the baseline CRN model, and the baseline CRN model modified for cAF using the experimentally-informed adjustments to its parameters listed in Table 1. The original CRN model is seen to significantly underestimate the RMP, predicting a value of -85.7mV for SR, as compared to -73.8mV for the mean of the SR data (-88.7mV versus -78.4mV for cAF). The APD is similarly underestimated, especially in the case of cAF where the original CRN model (with cAF-adjusted parameters) predicts an APD_{90}

of only 121ms, as compared to a mean value of 216ms in the data. In contrast, our median CRN models for SR and cAF do much better at reflecting the values for these biomarkers seen in this particular dataset (RMP: -74.7mV for SR and -78.4mV for cAF, APD$_{90}$: 335ms for SR and 223 for cAF).

These differences are explained by the parameter values selected for our POMs, which featured many models with much lower values of g_{K1}, and increased values of g_{Kr} as compared to the original published values for the CRN model. These are both late-stage repolarisation currents, and so can somewhat compensate for one another in terms of APD, but I_{K1} is a much more significant controller of RMP, and thus its reduction is necessary to allow the CRN model (which as stated, tended to underestimate RMP) to produce APs that corresponded with the data. As a result of the specific balance of these two currents, APD$_{90}$ values were also increased.

The increased importance of I_{Kr} predicted by our median CRN model (for both SR and cAF) also makes it more predictive of the effects of drug block of this current. Figure 8b compares our median model with the original CRN model, where it can be seen that our calibrated model demonstrates a restoration of APD to SR levels, while the original CRN model shows very little APD prolongation at all. In cAF, the original CRN model predicts a further decrease in relevance of I_{Kr} for repolarisation (*49*), while the models selected for our calibrated populations show an increased magnitude of this current in cAF (Figure 7a). This agrees with the known efficacy of treatments targeting this current for the restoration of healthy SR in patients with cAF (*47*).

The ability of our median model to capture well the results of drug block treatment, despite no data regarding drug block used in the calibration process, leads us to suggest that our technique not only allows the selection of parameter values that better match provided data, but can result in models that are more generally predictive. This is of particular importance in fields such as cardiac electrophysiology, where very different sets of parameter values can produce

very similar APs and further measures may be necessary to differentiate between these multiple sets of parameter values.

Distributional Calibration Can Inform the Extent of Variability in Parameter Values

Calibrating to distributions means that the variability in a supplied dataset can be expected to be explicitly captured by constructed POMs, so long as the model across the specified parameter space is capable of generating outputs that match the data. This means that by using divergence measures such as ρ and $\hat{\rho}$, the differential ability of various parameter spaces to capture variable data with POMs can be considered. The most obvious application of this technique is to explore the level of variability in parameter values needed to explain the data.

Our studies on the Sanchez *et al.* dataset have used a large variation in ion channel conductance values ($\pm 100\%$), following previous studies (*7, 17, 26*). Working with such a variation in parameter values carries the risk of selecting extreme values that are not physiologically sound (for example values close to -100% that essentially switch off an entire current), and there is some suggestion that a level of variability such as $\pm 30\%$ is more appropriate for cardiac ion channel conductances (*26, 27, 50*). We therefore seek to answer the question of whether 30% variability sufficiently explains the variation in the dataset we calibrated to.

In order to select the most relevant portion of the parameter space, we take parameter values $\pm 30\%$ around the values selected for the median CRN models in the previous section. Applying our SMC sampling algorithm and subsequent refinement, we obtained 254 models calibrated to the SR dataset, and 215 models calibrated to the cAF dataset, the minimum allowable number of models in both cases. Neither of these POMs succeeded in fully capturing the variability in the dataset, resulting in divergence measures that compare unfavourably to those obtained using the full $\pm 100\%$ variation in cell properties (Table S2). Examination of the marginal distributions
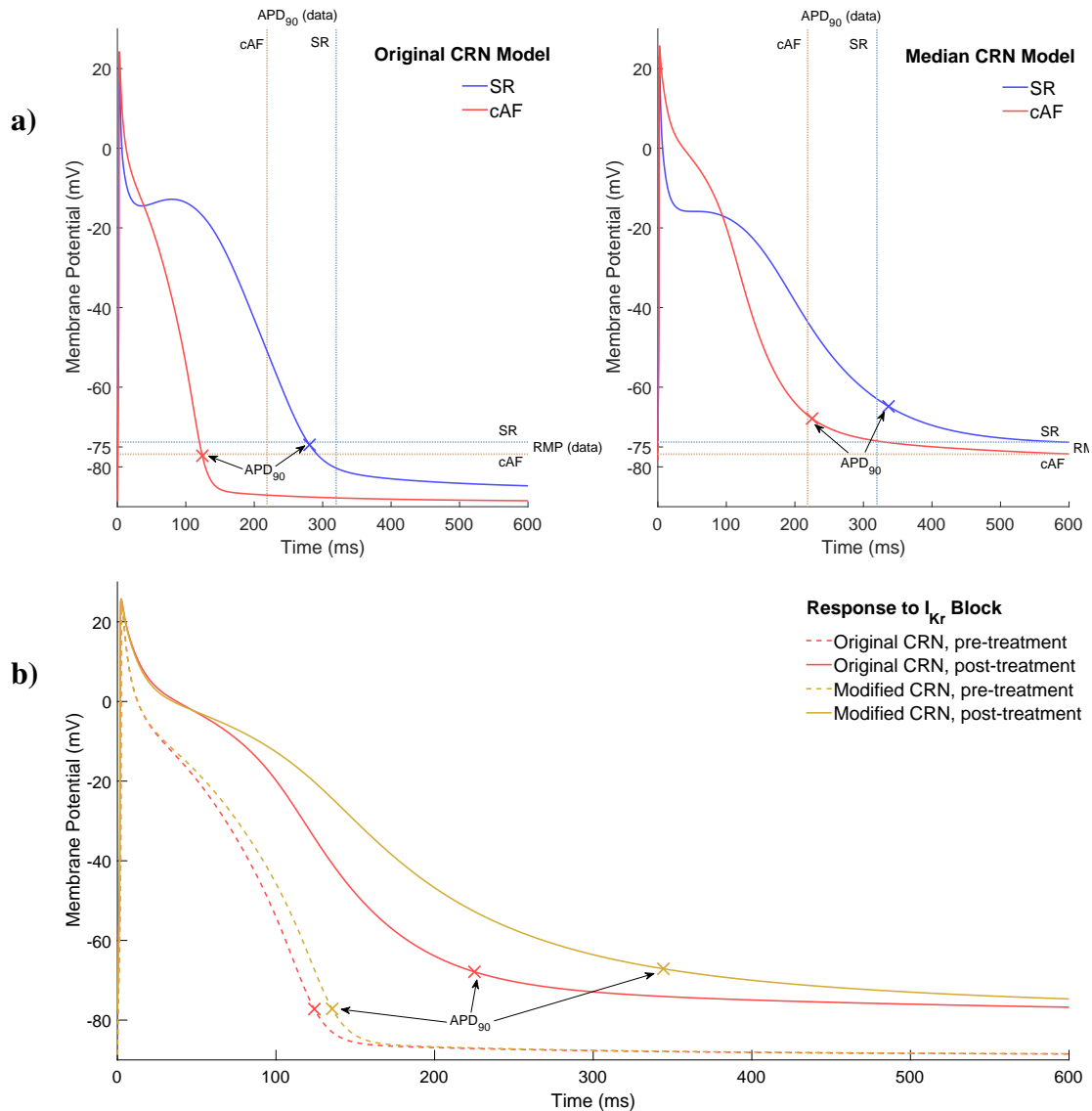
27

Figure 8: **Selection of parameter values using distribution-calibrated POMs produces updated models that correspond to provided data. a)** AP curves for SR (blue) and cAF (red) as predicted by the original, and median CRN model. Experimental calibration via POMs produces a model that fits very well the general trend of the data (mean values of biomarkers from the data indicated by dashed lines). **b)** cAF APs before (red) and after (gold) treatment via I_{Kr} blocker, for the original (dashed) and modified (solid) CRN model. Our median CRN model predicts the antiarrhythmic effects of I_{Kr} block (via APD prolongation), whereas the original CRN model predicts little response.

of the biomarkers for the SR $\pm30\%$ POM reveals that it does successfully capture the general distributions of the data, but fails to show the same extent of variance (Figure 9). Similar results are also seen for the cAF $\pm30\%$ POM (Figure S7).

Furthermore, the secondary mode in the distribution of SR APD_{50} values is completely unrepresented by the POM constructed with reduced variability. This peak most likely corresponds to cells that repolarise more than 50% during the phase of immediate repolarisation, driven primarily by the rapidly activated outward currents I_{to} and I_{Kur}. This results in a cluster of very small APD_{50} values that are separate from the majority of cells, which only reach 50% repolarisation after the plateau phase. General underestimation of biomarker variability in the POM could potentially be explained by measurement error associated with the biomarker data, but the inability to produce models that populate this peak implies that additional variance in cell properties, at least those relevant to I_{to} and/or I_{Kur}, are required to produce models that exhibit sufficient early repolarisation. Coupled with the general underestimation of biomarker variance in the POM, these results certainly imply that additional variability in ion channel conductances beyond $\pm30\%$ underlies the variability in biomarkers that is seen in these datasets.

Discussion

Population-based modelling is a powerful technique for allowing deterministic mathematical models to explore and characterise the variability inherent in complex systems (*1*). This includes the use of multiple regression techniques performed on synthetic populations (*14, 51, 52*), and when data is explicitly available, calibrated POMs (*27*). Previously, calibration of POMs has typically been performed by ensuring that all relevant model outputs fall within the ranges of the data (*7, 24, 26*). While this is a perfectly reasonable approach, especially when a low number of experimental samples is available, it does ignore other information inherent in the data and does not strictly guarantee that the models selected will produce outputs that correspond to those seen
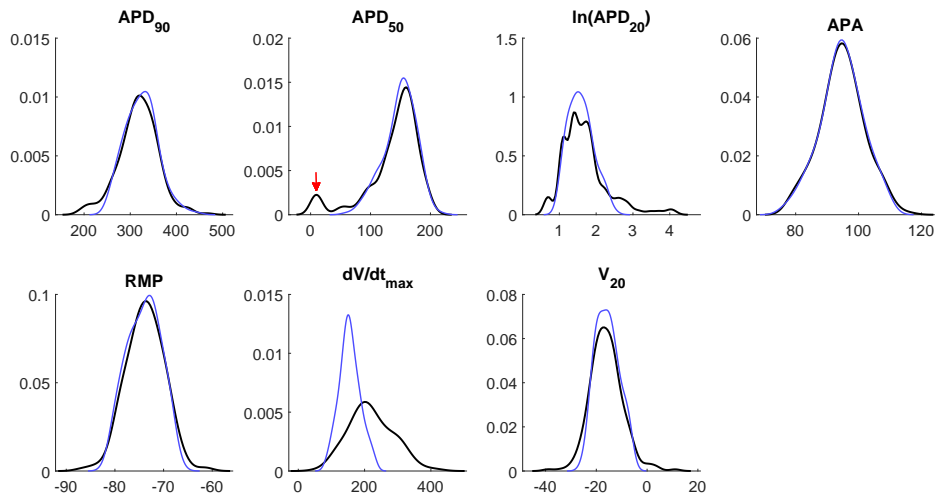
Figure 9: **Variation of** $\pm 30\%$ **in current densities underestimates biomarker variance in the SR dataset.** Marginal distributions of the biomarkers in the SR dataset (black) and distribution-calibrated POM using $\pm 30\%$ variance in ion channel conductances (blue). A reduced search space is still able to recover the general distributions of all biomarkers except for dV/dt_{max}, but the full extent of variation in the APD biomarkers and V_{20} is not present in the calibrated POM. Notably, the very low APD_{50} values recorded for some patients are completely unrepresented in the POM (arrow).

in the data. Our presented calibration technique estimates the underlying distribution of outputs represented by the data and calibrates to this distribution using a combination of SMC (*22*) and a simulated annealing-type algorithm.

We have demonstrated the efficacy of our technique on a pair of datasets from cardiac electrophysiology, a field that has seen a great deal of POMs research following Prinz, Bucher and Marder's (*5*) and Marder and Taylor's (*6*) pioneering studies in neuroscience. The data used was comprised of biomarker values collected for atrial myocytes from hearts exhibiting healthy SR or the cAF pathology (*26*), offering two sets of data to test the calibration technique and the opportunity to compare the models selected to represent healthy and afflicted hearts. Use of SMC to sample according to the distribution of data was found to improve the correspondence between POMs and experimental observations over LHS matched to ranges, but the approximate nature of the technique left room for further improvement. A subsequent refinement technique to select optimal subpopulations of the SMC-constructed POMs was able to produce *in silico* populations that reflected extremely well the variability inherent in the data distributions.

Our demonstration of our calibration technique operated from the assumption that variations in the ion channel conductances were capable of explaining the variability seen in experimental studies, consistent with previous works in the POMs literature (*27*). However, under this assumption we found that the POMs produced by our calibration technique could not match the distribution of dV/dt$_{\mathrm{max}}$ values as well as other biomarkers, and especially the relationship between this biomarker and APA. In order to further address this issue, we considered the effects of incorporating further variability into the behaviour of I_{Na}. We did this by introducing an additional variable parameter, namely the time constant of the gating variable that controls the initial inactivation of this current. This further improved the concordance between the data and our calibrated POMs in terms of these biomarkers, though could not completely capture the nature of their relationship. This suggests that variance in I_{Na} inactivation rates may con-

31

tribute a not insignificant portion of the variability seen in overall excitation behaviour of atrial myocytes, but is insufficient to wholly explain it. Moreover, it demonstrates that the inability of our original calibrated POMs to capture the dV/dt$_{max}$ distribution is most likely the result of our assumptions regarding the sources of variability in the problem being too restrictive, and not due to any issue with the calibration process itself. Indeed, this shows how our calibration technique can be seen as a tool for investigating different assumptions regarding where variability lies in a complex system, or the relative importance of variability in the different components that make up such systems.

Our approach generates models that predict very well the morphologies of different types of APs associated with SR or cAF. The cAF models also demonstrate a more realistic response to antiarrhythmic treatment (AP prolongation in response to I_{Kr} block (*47*)), as compared to the original CRN model. These benefits remained even when the population of models was averaged to produce a single model, demonstrating the efficacy of the technique for the more general problem of selecting parameter values in response to variable data. Taking into account the variability in data in this fashion, as opposed to simply averaging it and fitting parameters to the result, is of particular importance when a system (and its associated mathematical model) is complex, such as in our example of cardiac electrophysiology where the ability of currents to compensate for one another allows for highly similar APs to be produced by very different balances of constituent currents (*7, 53*).

Furthermore, calibration that takes more of the features of data into account may be able to resolve more subtle differences between datasets, making it a powerful approach for identifying the potential causative factors that produce these differences. We have demonstrated this usage of our technique by calibrating POMs to a pair of AP biomarker datasets. Although these biomarkers summarise the characteristic features of the AP, they do not provide much in the way of direct information regarding the activities of the different currents that constitute the AP.

Nevertheless, using only this biomarker data (but only by calibrating to distributions), we found that our resultant POMs for SR and cAF were unexpectedly good at identifying the changes in ion channel activity that characterise the cAF pathology. This was validated by comparison with a suite of studies in the literature specifically quantifying the effects of electrical remodelling on individual currents in cAF (*31, 37*). Interestingly, differences in the uptake of Ca^{2+} by the sarcoplasmic reticulum were also identified, despite this being an internal current that does not directly contribute to the AP (*32*). We suspect that the contribution of the internal calcium dynamics to the current through the Na^+/Ca^{2+} exchanger allows them to be partially identified, even when calibrating only to AP biomarkers, but this point requires further investigation beyond the scope of this paper.

With the ability of our distribution-calibrated POMs to identify the trends in model parameters that underpin specific sets of observable data, we suggest that POMs calibrated by our technique could also be used to identify differences in underlying behaviour that explain for example pathological conditions that are less understood, and not limited to cardiac electrophysiology. This same thought also extends to improved identification of any differences that underlie artificial stratifications of a dataset, for example differences in physiology that correspond to factors such as gender or age (*54, 55*).

When constructing POMs without experimental evidence for the extent of variability in parameters, the choice of parameter space to sample is open. Our primary study used a large space of parameter values ($\pm 100\%$) to give the model the best opportunity to simultaneously fit the two different sets of data and hence demonstrate our technique. However, we also considered the effects of choosing a smaller parameter space ($\pm 30\%$, (*21, 26, 50*)), attaining results that imply that $\pm 30\%$ variance in primary current conductances is insufficient to explain the experimental data, and thus that further variation in these cell properties or others is expected. This sort of exploration into appropriate levels of variability is particularly important in fields such

as cardiac electrophysiology, where the variable cell properties are difficult to directly measure and the extent of variability is not well established (*56*). Such analysis requires the ability to calibrate POMs to distributions, so that the comparisons between parameter spaces of different sizes are unaffected by model bias.

There are several circumstances under which our calibration process might fail or be considered inappropriate. Firstly, calibrating to distributions requires sufficient data in the dataset to form a reasonable approximation to the underlying distribution. When the sample size is insufficient compared to the variance to suggest a specific distribution with any real confidence, obviously distributional calibration is inappropriate and calibration based on range statistics is very reasonable. However, even in these cases a benefit can potentially be gained by enforcing that data is spread evenly across its range, although we have not investigated this in this paper. Secondly, successful calibration can only be achieved when the model used to construct POMs is capable of producing outputs that correspond to the range of values seen in the data. This depends strongly on the variability assumed to be present in the model, as we have seen here with the inability to capture the distribution of dV/dt_{max} values when varying only ion channel conductances. This does however demonstrate the usefulness in considering the distribution of the data and how well constructed POMs fit to it using measures such as ρ, in that it suggests when assumptions regarding variability (or even the models themselves) might need to be reconsidered. Indeed the varying ability of different models or patterns of variability to successfully produce POMs calibrated to a given dataset also serves as a means of comparing and benchmarking them. Lastly, the reliance upon a transformed target density in the SMC algorithm might potentially cause it to select models that do not accurately capture the distribution of outputs in the data, even though on our test problem the method has been seen to perform very well. In the worst case, 'naive' POMs composed of large numbers of models can be constructed by sampling the search space uniformly (such as through LHS) and then refined by our

34

simulated annealing process to select the subpopulation of these that best matches the data.

In conclusion, our presented calibration technique allows datasets to be thoroughly mined in order to produce informative and predictive POMs that capture the variability between individual members of a population. The benefits of successfully accounting for this variability are well established (*1*). We have shown how our technique allows for different assumptions regarding the sources (and extents) of variability in a system to be explored, even when some sources of this variability are indirect or have not been experimentally quantified. We have also demonstrated how calibrating specifically to data distributions allows for much better differentiation between multiple sets of data, shedding light on the changes in underlying properties that explain differences in higher level observables that can be experimentally measured. In this case, differences in AP biomarkers proved sufficient for predicting the majority of changes in ion channel activity that characterise the cAF pathology, but only when calibrating to the distributions of these biomarkers. Finally, we also demonstrated how the models selected by our calibration technique not only agree with the data used for calibration, but are more generally predictive, as shown in this case by a much better representation of the SR and cAF morphologies, and the effects of anti-arrhythmic treatment. We recommend our technique for any systems were variability is expected to be present, a mathematical model is available, and there is sufficient data to make calibration to distributions feasible.

Materials and Methods

Atrial AP Model

The CRN model uses twenty-one coupled ordinary differential equations to simulate among other things the activation and inactivation of nine different sarcolemmal ion channels, as well as the actions of the sarcolemmal Ca^{2+} pump, Na^+/K^+ pump and the Na^+/Ca^{2+} exchanger, which all contribute to the flow of ions in or out of an atrial cell and thus to the changes in mem-

brane potential that create the AP. Ion channels are modelled using a Hodgkin-Huxley (*57*) type formulation, with combinations of gating variables representing the presence/absence of activators/inhibitors that determine channel availability. Ca^{2+} uptake into the network sarcoplasmic reticulum, release from the junctional sarcoplasmic reticulum and the transfer (active or leak) between these two compartments are also represented. For full details of the model, including the specific forms of each of its differential equations see (*32*).

We simulated the CRN model using MATLAB's *ode15s* routine, with a maximum step size of $\Delta t = 1$ms. Biomarkers were measured from output $V(t)$ curves, after discarding any action potentials that failed to excite above -30mV, failed to repolarise to a value of RMP+0.1APA, or exhibited spontaneous depolarisations (judged as subsequent peaks after repolarisation to the aforementioned value). Following Sanchez *et al.*, the model was paced until it reached steady state ($\leq 1\%$ change in all state variables) and then 90 more times, using a stimulus of 2ms of -2210 pA, approximately twice the diastolic threshold for the base model. Temperature and external ion concentrations were adjusted to match the experimental conditions ($T = 309.65$K, $[Na^+]_o = 149.42$mM, $[K^+]_o = 4.5$mM, $[Ca^{2+}]_o = 4.5$mM). No other parameters were modified from their values in the originally published version of the model.

The process of simulating the CRN model (with input parameters $\boldsymbol{\theta}$) and the subsequent calculation of biomarkers from the resulting action potential will be denoted $\mathcal{M}$, and the output biomarkers denoted $\mathbf{y}$, such that

$$\mathbf{y} = \mathcal{M}(\boldsymbol{\theta}). \tag{1}$$

Populations of CRN Models

Populations of models were constructed here by varying a set of inputs to the CRN model, namely the current densities of the fast Na^+ current, the five outward K^+ currents (transit outward, ultrarapid delayed rectifier, rapid and slow delayed rectifiers and the inward recti-

fier), the L-type inward Ca^{2+} current, the Na^+/K^+ pump and Na^+/Ca^{2+} exchanger, and the maximal rates of uptake and release of Ca^{2+} inside the cell by the sarcoplasmic reticulum. Following the notation used in the original CRN paper, the set of inputs is here denoted $\boldsymbol{\theta} = (g_{Na}, g_{to}, g_{Kur}, g_{Kr}, g_{Ks}, g_{K1}, g_{CaL}, I_{NaK(max)}, I_{NaCa(max)}, I_{up(max)}, k_{rel})$, with g denoting the maximal conductances of the different currents, $I_{(max)}$ denoting the maximal actions of pumps and exchangers and k_{rel} being the conductance of the ryanodine receptors that release $Ca2^+$ from the sarcoplasmic reticulum. These were the same currents varied by Muszkiewicz *et al.* in their construction of POMs for human atria (*16*). Sanchez *et al.*'s study (*26*) varied only the six currents they identified as having a significant impact on biomarker values (*58*), and thus did not include g_{Na}, g_{Kr}, g_{Ks}, $I_{up(max)}$ or k_{rel}.

Following Sanchez *et al.*, POMs were constructed using a search space of $\pm 100\%$ from the base parameter values for the CRN model. However, whereas their work selected trial points using the sampling method underlying Fourier amplitude sensitivity testing (*59*) and then rejected those that did not produce APs with biomarkers falling within experimentally observed ranges, we used the method described subsequently to produce POMs that not only corresponded to the spread of the experimental data, but also reproduced its distributional features.

Biomarker Joint Distribution Estimation

Distributional calibration first requires estimating the distribution represented by the data, $p(\mathbf{y})$. This was achieved here by multivariate kernel density estimation, which creates a smooth distribution by summing over a series of multivariate Gaussians centred at each of the N individual datapoints,

$$p(\mathbf{y}) \approx \frac{1}{\det(H)^{-1/2} N (2\pi)^{-N_b/2}} \sum_{i=1}^{N} e^{-1/2(\mathbf{y}-\widetilde{\mathbf{y}}_i)^T \mathbf{H}^{-1}(\mathbf{y}-\widetilde{\mathbf{y}}_i)}. \tag{2}$$

Here N_b is the number of biomarkers (seven in this case), $\widetilde{\mathbf{y}}_i$ are the individual points of biomarker data and $\mathbf{H}$ is the bandwidth matrix, a parameter of the density estimator that con-

trols the extent and direction of smoothing. When the distribution to be estimated is normal with unit variance, the optimal bandwidth can be shown (*60*) to be

$$h_{\text{opt}} = \left(\frac{4}{N(N_b + 2)} \right)^{\frac{2}{N_b + 4}},$$

(3)

which motivates a choice of bandwidth matrix

$$
\begin{aligned}
\mathbf{H}_{ij} &= h_{\text{opt}} \, \sigma_i^2 && \text{if } i = j \\
&= 0 && \text{if } i \neq j
\end{aligned}
$$

(4)

Thus the extent of smoothing is weighted in each dimension in terms of the variance in that biomarker observed in the dataset, but correlations between biomarkers are ignored in the choice of $\mathbf{H}$. Note that the estimated density itself still attempts to account for dependencies between biomarkers. The choice to use a diagonal bandwidth matrix tends to be sufficient in practice (*61*).

Another alternative to multivariate kernel density estimation is to approximate $p(\mathbf{y})$ by combining the marginal distributions of each biomarker with a Gaussian copula to approximate their interdependencies. However, for the atrial datasets we use to demonstrate our calibration technique, this approach was found to be less effective and so is not discussed further here.

APD$_{20}$ readings in the SR dataset were predominantly clustered at low values, but with a considerable range. To improve the performance of the kernel density estimation (recalling the bandwidth was selected as optimal for normally distributed data), the APD$_{20}$ values were first logarithmically transformed to make their distribution more regular before use in equation (2).

SMC for POM Calibration to Distributions

Constructing a population of $\boldsymbol{\theta}$ values that exhibits the estimated distribution $p(\mathbf{y})$ is not trivial, given the complex relationship between the two encoded by equation (1). We define $g(\boldsymbol{\theta})$ to

be a probability density over the space $\boldsymbol{\theta}$, and the population distribution of the model output $\mathcal{M}(\boldsymbol{\theta})$ when $\boldsymbol{\theta}$ is drawn according to this distribution we denote $h(\mathbf{y}|g(\boldsymbol{\theta}))$, with the vertical bar | denoting conditioning. That is, $h(\mathbf{y}|g(\boldsymbol{\theta}))$ is the density of $\mathcal{M}(\boldsymbol{\theta})$ when $\boldsymbol{\theta} \sim g(\boldsymbol{\theta})$:

$$h(\mathbf{y}|g(\boldsymbol{\theta})) = \lim_{\Delta_{\mathbf{y}} \to \mathbf{0}} \frac{1}{\prod_{k=1}^{N_b} \Delta_{b_k}} \int_{\mathcal{M}(\boldsymbol{\theta}) \in (\mathbf{y}, \mathbf{y} + \Delta_{\mathbf{y}})} g(\boldsymbol{\theta}) d\boldsymbol{\theta}, \tag{5}$$

where $\Delta_{\mathbf{y}} = (\Delta_{b_1}, \ldots, \Delta_{b_{N_b}})$. Our problem is thus recast as finding the distribution $g(\boldsymbol{\theta})$ that produces a $h(\mathbf{y}|g(\boldsymbol{\theta}))$ that is as close as possible to $p(\mathbf{y})$. Models sampled according to this optimal $g(\boldsymbol{\theta})$ will then exhibit outputs that reproduce the estimated distribution of outputs in the data.

If the "closeness" of $h(\mathbf{y}|g(\boldsymbol{\theta}))$ and $p(\mathbf{y})$ is measured in terms of the Kullback-Leibler divergence between the two distributions, the problem is an optimisation problem

$$g^*(\boldsymbol{\theta}) = \arg\min_{g(\boldsymbol{\theta})} \int_{\mathbf{y}} \ln\left(\frac{h(\mathbf{y}|g(\boldsymbol{\theta}))}{p(\mathbf{y})}\right) h(\mathbf{y}|g(\boldsymbol{\theta})) d\mathbf{y}. \tag{6}$$

The standard method for solving this type of problem is to follow a variational Bayes (VB) type of approach. VB is commonly used to produce parameteric approximations of posterior distributions in Bayesian statistics (*62, chap. 11*). Applying VB to our problem would involve specifying some parametric distribution for $g(\boldsymbol{\theta}) \equiv g(\boldsymbol{\theta}|\phi)$, where ϕ are the parameters of the distribution. For example, if a multivariate normal distribution was adopted for g, ϕ would consist of a mean vector and covariance matrix. Using VB, the problem (6) then reduces to finding the optimal parameters $\phi^* = \arg\min_{\phi} f(\phi)$, with the integral in equation (6) approximated by Monte Carlo integration taking K independent draws from $g(\boldsymbol{\theta}|\phi)$,

$$\int_{\mathbf{y}} \ln\left(\frac{h(\mathbf{y}|g(\boldsymbol{\theta}|\phi))}{p(\mathbf{y})}\right) h(\mathbf{y}|g(\boldsymbol{\theta}|\phi)) d\mathbf{y} \approx \frac{1}{K} \sum_{k=1}^{K} \ln\left(\frac{\hat{h}(\mathbf{y}_k|g(\boldsymbol{\theta}|\phi))}{p(\mathbf{y}_k)}\right) = f(\phi), \tag{7}$$

where $\mathbf{y}_k = \mathcal{M}(\boldsymbol{\theta}_k)$ and $\boldsymbol{\theta}_k \sim g(\boldsymbol{\theta}|\phi)$ for $k = 1, \ldots, K$. For a particular ϕ, an estimate of $h(\mathbf{y}|g(\boldsymbol{\theta}|\phi))$, which we denote as $\hat{h}(\mathbf{y}_k|g(\boldsymbol{\theta}|\phi))$, could be obtained using a kernel density estimate as in equation (2) based on the set of simulated biomarker values $\{\mathbf{y}_k\}_{k=1}^{K}$. There are

several reasons why we did not adopt this approach: firstly it requires us to specify a parametric form for $g(\boldsymbol{\theta})$, secondly evaluating $f(\boldsymbol{\phi})$ is very expensive as it involves solving the model K times and thirdly $\boldsymbol{\phi}$ will be high-dimensional, leading to a difficult optimisation problem.

Instead, we used an approach that is more pragmatic and effective in this application. First, we determined a collection of $\boldsymbol{\theta}$ values (or models) that produced biomarker values $\mathbf{y}$ that have relatively high density with respect to the data density $p(\mathbf{y})$. Then, we removed models from this collection in an iterative fashion so that the distribution of corresponding biomarker values that remained was even closer to $p(\mathbf{y})$ (see Section "Further POMs Refinement").

For the first step we used sequential Monte Carlo (SMC, (63)), following the use of the technique to construct POMs calibrated to ranges in data (22). This technique begins with a set of N particles and traverses them through a sequence of probability distributions by iteratively applying importance sampling, resampling and move steps. We achieve this behaviour by sampling from the sequence of distributions $h(\boldsymbol{\theta}) \propto p(\mathcal{M}(\boldsymbol{\theta}))^{\gamma}$, with $\gamma \in [0, 1]$. It can immediately be seen that $\gamma = 0$ corresponds to the uniform distribution, which is very easy to sample if we specify some lower and upper limits for each component of $\boldsymbol{\theta}$, and that $\gamma = 1$ corresponds to a distribution proportional to $p(\mathcal{M}(\boldsymbol{\theta}))$, which is potentially difficult to sample. Successively incrementing γ after each resample and move step until γ reaches one allows for the complexity of the sampling problem to be introduced gradually. An important aspect of SMC is that it does not require the distributions in the sequence to be properly normalised. The full SMC algorithm is laid out in the supplementary material.

The algorithm requires the use of traditional Markov chain Monte Carlo (MCMC, (64)) steps in order to find unique locations for particles after each resampling step, and the particles must still represent samples from the current target distribution. This is achieved using the Metropolis-Hastings algorithm, which, in our set-up, accepted or rejected any proposed particle

40

moves according to

$$\text{Pr(accept)} = \min\left(1, \frac{[p(\mathcal{M}(\boldsymbol{\theta}_{\text{new}}))]^{\gamma}\mathcal{J}(\boldsymbol{\theta}_{\text{old}}|\boldsymbol{\theta}_{\text{new}})}{[p(\mathcal{M}(\boldsymbol{\theta}_{\text{old}}))]^{\gamma}\mathcal{J}(\boldsymbol{\theta}_{\text{new}}|\boldsymbol{\theta}_{\text{old}})}\right). \tag{8}$$

Here $\mathcal{J}$ is the jumping distribution that generates proposed moves of particles, and in our case does not depend on the previous particle location (that is, $\mathcal{J}(\boldsymbol{\theta}_{\text{new}}|\boldsymbol{\theta}_{\text{old}}) = \mathcal{J}(\boldsymbol{\theta}_{\text{new}})$). The jumping distribution used is a Gaussian mixture model built for a regularised version of the current of locations of all particles after a resampling step (see SMC Algorithm in the supplementary material). The number of MCMC steps performed after each resampling step is chosen adaptively (*65*).

The output of the SMC algorithm is a collection of samples $\{\boldsymbol{\theta}_i\}_{i=1}^{N}$ from a distribution proportional to $p(\mathcal{M}(\boldsymbol{\theta}))$. We note that the corresponding collection of output values $\{\mathbf{y}_i\}_{i=1}^{N}$, where $\mathbf{y}_i = \mathcal{M}(\boldsymbol{\theta}_i)$, is not a sample from the distribution of interest, $p(\mathbf{y})$. The reason for this is that we have not accounted for the nonlinear transformation, $\mathbf{y} = \mathcal{M}(\boldsymbol{\theta})$ to correctly convert the target distribution over the output space to the corresponding target distribution over the parameter space. However, given the fact that the transformation $\mathbf{y} = \mathcal{M}(\boldsymbol{\theta})$ is not analytic and not a one to one function, we suggest that it is not tractable to properly account for it and indeed it may not even be possible to find a distribution over $\boldsymbol{\theta}$ that leads to a distribution of biomarker values consistent with $p(\mathbf{y})$.

Nonetheless, we found that this approach led to a collection of parameter values that generated biomarker values with relatively high density under $p(\mathbf{y})$. We used this collection as the starting point for our subsequent refinement process.

As an alternative to SMC, an MCMC approach could also be used directly to produce samples from a distribution proportional to $p(\mathcal{M}(\boldsymbol{\theta}))$, using the same density (2) and acceptance algorithm (8) with $\gamma = 1$. Indeed we used a modern state-of-the-art MCMC sampler, DiffeRential Evolutionary Adaptive Metropolis (DREAM) (*66*) to verify our SMC algorithm and

found that the two produced comparable results in terms of the distribution of the values of $\boldsymbol{\theta}$ produced. The primary benefit of the SMC algorithm, when it came to the construction of our initial POM, was that the output of the algorithm is a set of unique samples from the distribution almost the same size as the number of particles, N, which is specified by the user. In contrast, MCMC approaches must be run an indefinite amount of time until the chain has been judged to have converged, producing a long chain of samples of initially unknown length. Moreover, these chains contain many repeated samples, and filtering out these repeats will also destroy the desired distribution.

Further POMs Refinement

The approximate nature of the SMC calibration process encouraged further refinement of the constructed POMs in order to fully capture the statistical distributions seen in the data. This was achieved by selecting a subset of the population such that the new smaller set of models better exhibited the biomarker distributions seen in the data. In order to do this, first a quantitative measure of how well a POM captured the distributions observed in the data was constructed. We used the Jensen-Shannon distance (JSD), a symmetric and finite version of the Kullback-Leibler divergence that remains a measure of the 'distance' between two probability distributions. Labelling the two distributions $p(\mathbf{y})$ and $q(\mathbf{y})$, the Jensen-Shannon distance is given by

$$JSD = \left[\frac{1}{2} \int_{\mathbf{y}} p(\mathbf{y}) \ln\left(\frac{p(\mathbf{y})}{\frac{1}{2}p(\mathbf{y}) + \frac{1}{2}q(\mathbf{y})} \right) d\mathbf{y} + \frac{1}{2} \int_{\mathbf{y}} q(\mathbf{y}) \ln\left(\frac{\mathbf{y})}{\frac{1}{2}p(\mathbf{y}) + \frac{1}{2}q(\mathbf{y})} \right) d\mathbf{y} \right]^{1/2}, \quad (9)$$

with the square root used to make the divergence measure a metric (*67*).

When the data is high-dimensional (say $N \geq 5$), the 'full' JSD between the multivariate joint distributions of the observations in the dataset and those generated by a given POM is difficult to calculate accurately. Therefore, we instead used the distances between the marginal distributions for each of the biomarkers, JSD_i, along with the distances between the bivariate

distributions between all possible pairs of observation variables, JSD_{ij} to create a matrix, the norm of which serves as an approximate measure of fit, namely

$$\rho \;=\; ||\mathbf{P}||_2 , \tag{10}$$

$$\mathbf{P}_{ij} \;=\; JSD_i \qquad \text{for } i = j$$

$$\;=\; JSD_{ij} \qquad \text{for } i \neq j \qquad\qquad (i, j = 1, ..., N_B).$$

The measure ρ takes into account how well the individual distributions of each observed variable are represented by a POM, along with some measure of how well it captures the dependency between these variables. We note that this is not necessarily the best measure of fit, but uses more easily calculated divergences to produce a single value, allowing the use of the technique described below. JSD values used in the calculation of ρ, as integrals, were approximated using standard Riemann integration.

Our refinement process seeks to minimise ρ, using a supplied POM. In this work we use POMs selected using SMC to improve fit with the data, though we note that POMs constructed using typical Monte Carlo sampling techniques (such as LHS and calibrated to the ranges of the data) could also be used as starting points for our refinement procedure. Minimisation is achieved by trialling removal of individual models from the population (or re-introduction of removed models) and then accepting or rejecting them according to the Metropolis probability,

$$\Pr(\text{accept}) = e^{-\Delta\rho/T}. \tag{11}$$

Here $\Delta\rho$ is the change in the overall divergence measure (10) associated with the trialled removal/re-introduction and T is a parameter of the process that controls the likelihood of accepting unfavourable trial updates. This approach is very similar to the approach of simulated annealing (*68*), although we use a fixed value of $T = 0.2$ instead of gradually decreasing it. Every 1000 trial steps, we judge if the choice of subpopulation is wandering too far away from

the optimum by checking if ρ is more than 1% larger than the current best ρ value found, and if so, restart the process back to the configuration corresponding to the best ρ value.

The only additional condition we use is that the size of the subpopulation of models cannot fall below the number of datapoints, ensuring that the resulting population will not become small enough to lose meaning. If a larger population of models is desired, ρ (representing the 'energy' of the system that is minimised over the course of the annealing process) can be replaced by a new expression in equation (11) that penalises both higher values of ρ and small numbers of models in the population.

We saw (see Results) that the data for one biomarker, namely the maximum upstroke velocity, took values not predicted by the CRN model in the search space, and strong correlations exhibited by the model were not seen in the data. This made it appealing to de-emphasise the contributions of this biomarker to the POM refinement process. This was achieved by creating a second divergence measure, $\hat{\rho}$, that is the 2-norm of a modified version of the performance matrix $\mathbf{P}$ with the row and column corresponding to the maximum upstroke velocity overwritten with zeroes, except for the diagonal element. Minimising $\hat{\rho}$ instead allowed the distributions of the other biomarkers to be better fit by the refinement process, at the cost of producing POMs that did not strongly reflect the distribution of maximum upstroke velocities in the data. Given that we expected the maximum upstroke velocity to be the biomarker most subject to experimental measurement error in the data, we consider this a reasonable decision.

References and Notes

1. A. X. Sarkar, D. J. Christini, E. A. Sobie, *J. Physiol.* **590**, 2555 (2012).

2. J. F. Ogilvie, *Comput. Chem.* **8**, 205 (1984).

3. D. Xiu, G. M. Karniadakis, *SIAM J.* **24**, 619 (2002).

4. A. O'Hagan, M. C. Kennedy, J. E. Oakley, *Uncertainty analysis and other inference tools for complex computer codes* (Oxford University Press, 1999), vol. 6, pp. 503–524.

5. A. Prinz, D. Bucher, E. Marder, *Nat. Neruosci.* **7**, 1345 (2004).

6. E. Marder, A. L. Taylor, *Nat. Neurosci.* **14**, 133 (2011).

7. O. J. Britton, *et al.*, *Proc. Natl. Acad. Sci. U. S. A.* **110**, E2098 (2013).

8. Z. Syed, E. Vigmond, S. Nattel, L. J. Leon, *Med. Biol. Eng. Comput.* **43**, 561 (2005).

9. M. Sierra, *et al.*, *Ann. Biomed. Eng.* **43**, 2444 (2015).

10. N. C. A. van de Pas, *et al.*, *J. Pharmacokinet. Pharmacodyn.* **41**, 351 (2014).

11. R. J. Allen, T. R. Rieger, C. J. Musante, *CPT Pharmacometrics Syst. Pharmacol.* **5**, 140 (2016).

12. A. Carusi, *Stud. Hist. Philos. Sci. C* **48**, 28 (2014).

13. Z. Qu, G. Hu, A. Garfinkel, J. N. Weiss, *Phys. Rep.* **543**, 61 (2014).

14. E. A. Sobie, *Biophys. J.* **96**, 1264 (2009).

15. P. Gemmell, K. Burrage, B. Rodriguez, T. A. Quinn, *PLoS One* **9**, e90112 (2014).

16. A. Muszkiewicz, A. Bueno-Orovio, X. Liu, B. Casadei, B. Rodriguez, *Comput. Cardiol.* **41**, 761 (2014).

17. X. Zhou, *et al.*, *Comput. Cardiol.* **40**, 855 (2013).

18. X. Zhou, *et al.*, *Circ. Res.* **118**, 266 (2016).

19. O. J. Britton, A. Bueno-Orovio, L. Virág, Varró, B. Rodriguez, *Front. Physiol.* **8**, 278 (2017).

20. A. Liberos, *et al.*, *Heart Rhythm* **13**, 2358 (2016).

21. J. Walmsley, *et al.*, *PLoS One* **8**, e56359 (2013).

22. C. C. Drovandi, *et al.*, *J. R. Soc. Interface* **13**, 20160214 (2016).

23. E. Passini, *et al.*, *Front. Physiol.* **8**, 668 (2017).

24. P. Gemmell, K. Burrage, B. Rodriguez, T. A. Quinn, *Prog. Biolphys. Mol. Biol.* **121**, 169 (2016).

25. E. Passini, *et al.*, *J. Mol. Cell. Cardiol.* **96**, 72 (2016).

26. C. Sánchez, *et al.*, *PLoS One* **9**, e105897 (2014).

27. A. Muszkiewicz, *et al.*, *Prog. Biophys. Mol. Boil.* **120**, 115 (2016).

28. M. C. Lancaster, E. A. Sobie, *Clin. Pharmacol. Ther.* **100**, 371 (2016).

29. E. Tixier, D. Lombardi, B. Rodriguez, J. F. Gerbeau, *J. R. Soc. Interface* **14**, 20170238 (2017).

30. H. Gelband, H. L. Bush, M. R. Rosen, R. J. Myerburg, B. F. Hoffman, *Circ. Res.* **30**, 293 (1972).

31. S. V. Pandit, *Cardiac Electrophysiology: From Cell to Bedside*, D. Zipes, J. Jalife, eds. (Elsevier, 2013), pp. 309–318, 6th edn.

32. M. Courtemanche, R. J. Ramirez, S. Nattel, *Am. J. Physiol.* **275**, H301 (1998).

33. M. Wilhelms, *et al.*, *Front. Physiol.* **3** (2013).

34. M. Schneider, T. Proebstle, V. Hombach, A. Hannekum, R. Rüdel, *Pflugers Arch.* **428**, 84 (1994).

35. M. Boutjdir, *et al.*, *Pacing Clin. Electrophysiol.* **9**, 1095 (1986).

36. S. Nattel, B. Burstein, D. Dobrev, *Circ. Arrhythm. Electrophysiol.* **1**, 62 (2008).

37. D. Dobrev, U. Ravens, *Basic. Res. Cardiol.* **98**, 137 (2003).

38. M. M. Maleckar, J. L. Greenstein, W. R. Giles, N. A. Trayanova, *Am. J. Physiol. Heart. Circ. Physiol.* **297**, H1398 (2009).

39. E. Grandi, *et al.*, *Circ. Res.* **109**, 1055 (2011).

40. S. Neef, *et al.*, *Circ. Res.* **106**, 1134 (2010).

41. R. Caballero, *et al.*, *J. Am. Coll. Cardiol.* **55**, 2346 (2010).

42. A. El-Armouche, *et al.*, *Circulation* **114**, 670 (2006).

43. N. Voigt, A. W. Trafford, U. Ravens, D. Dobrev, *Circulation* **120**, Suppl. 18 (2009).

44. S. Sossalla, *et al.*, *J. Am. Coll. Cardiol.* **55**, 2330 (2010).

45. N. Gaborit, *et al.*, *Circulation* **112**, 471 (2005).

46. A. J. Workman, K. A. Kane, A. C. Rankin, *Cardiovasc. Res.* **59**, 593 (2003).

47. T. J. Colatsky, T. M. Argentieri, *Drug Dev. Res.* **33**, 235 (1994).

48. M. Zaniboni, I. Riva, F. Cacciani, M. Groppi, *Math. Biosci.* **228**, 56 (2010).

49. M. Courtemanche, R. J. Ramirez, S. Nattel, *Cardiovasc. Res.* **42** (1999).

50. L. Romero, E. Pueyo, M. Fink, B. Rodríguez, *Am. J. Physiol. Heart Circ. Physiol.* **297**, H1436 (2009).

51. A. X. Sarkar, E. A. Sobie, *PLoS. Comput. Biol.* **6**, e1000914 (2010).

52. A. X. Sarkar, E. A. Sobie, *Heart Rhythm* **8**, 1749 (2011).

53. S. A. Mann, *et al.*, *J. Mol. Cell. Cardiol.* **100**, 25 (2016).

54. T. Taneja, B. W. Manhert, R. Passman, J. Goldberger, A. Kadish, *Pacing Clin. Electrophysiol.* **24**, 16 (2001).

55. K. Sakabe, *et al.*, *Pacing Clin. Electrophysiol.* **26**, 1238 (2003).

56. G. R. Mirams, P. Pathmanathan, R. A. Gray, R. H. Clayton, *J. Physiol.* **594**, 6833 (2016).

57. A. L. Hodgkin, A. F. Huxley, *J. Physiol.* **117**, 500 (1952).

58. C. Sanchez, *et al.*, *Am. J. Physiol. Heart. Circ.* **302**, H1146 (2012).

59. S. Marino, I. B. Hogue, C. J. Ray, D. E. Kirschner, *J. Theor. Biol.* **254**, 178 (2008).

60. B. W. Silverman, *Density Estimation for Statistics and Data Analysis* (Chapman and Hall, 1986).

61. M. P. Wand, M. C. Jones, *Kernel Smoothing*, vol. 60 of *Monographs on Statistics and Applied Probability* (Chapman & Hall, 1995).

62. C. M. Bishop, *Pattern recognition and machine learning*, vol. 128 (Springer, 2006).

63. P. Del Moral, A. Doucet, A. Jasra, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **68**, 411 (2006).

64. S. Brooks, A. Gelman, G. L. Jones, X. Meng, *Handbook of Markov Chain Monte Carlo* (CRC press, 2011).

65. C. C. Drovandi, A. N. Pettit, *Biometrics* **67**, 225 (2011).

66. J. A. Vrugt, *Environ. Modell. Softw.* **75**, 273 (2016).

67. D. M. Endres, J. E. Schindelin, *IEEE Trans. Inf. Theory* **49**, 1858 (2003).

68. S. Kirkpatrick, C. D. Gelatt, M. P. Vecchi, *Science* **220**, 671 (1983).