

# Dimension Reduction for the Emulation of Cardiac Electrophysiology Models for Single Cells and Tissue

Brodie A J Lawson<sup>1</sup>, Chris C Drovandi<sup>1</sup>, Pamela Burrage<sup>1</sup>, Blanca Rodriguez<sup>2</sup>, Kevin Burrage<sup>1,2</sup>

<sup>1</sup> ARC Centre of Excellence for Mathematical and Statistical Frontiers,  
Queensland University of Technology, Brisbane, Australia

<sup>2</sup> Department of Computer Science, University of Oxford, Oxford City, United Kingdom

## Abstract

*Widespread variability in the electrophysiological behaviour of individual cardiac cells, as well as between the hearts of different members of a population, presents a significant challenge to both the biological and mathematical understanding of cardiology. This variability underpins the differential responses to heterogeneities in pathologies of the heart, and to drug treatments, and so a thorough understanding is critical. A range of techniques exist for both uncertainty quantification and exploration of variability in mathematical models, but these require evaluation of the model at large numbers of points in a parameter space and the complexity of these models can make such analyses prohibitively computationally expensive. We demonstrate the use of dimension reduction to allow Gaussian processes to emulate the complex spatiotemporal outputs of heart models, thus making studies of variability feasible. Significant improvements in computational speed are achieved.*

## 1. Introduction

When electrically stimulated, a cardiac cell undergoes a pattern of depolarisation and repolarisation known as the action potential (AP), encoded by the change in its membrane potential over time. The shape of the AP governs important properties such as conduction velocity and refractory time, but can differ considerably between individual cells even of the same type. This variability makes experimental data much harder to interpret, and raises further questions about the cell properties that might be varying in order to produce these changes. Exploration of this variability using mathematical models is an area of significant recent interest [1–3].

Variability persists up to the tissue and organ level, with differences in heart function due to factors including age, hormone levels and pathologies. This creates important effects like a differential response to drug treatments. The

sources of such variability are difficult to quantify, potentially arising from wide-scale differences in cell properties or from small regions of inhomogeneities that produce knock-on effects, even potential re-entrant currents. Efforts to characterise this variability are also hindered by the computational complexity and the very fine computational meshes required to resolve the strongly nonlinear dynamics of heart models.

Emulation is a technique that allows for uncertainty quantification and exploration of variability in complex systems where model runtime is a limiting factor [4]. The general idea is to treat the model to be emulated as a black box, focusing only on the *response surface* of each of the model outputs in terms of the model's inputs. This black box is termed the *simulator*. Emulation is achieved by constructing an *emulator* (also known as a surrogate model, or meta-model) that approximates the response surfaces of each of the simulator's outputs using a mechanism that is much quicker to evaluate. Techniques for emulation include regression, polynomial chaos expansions, or Gaussian processes (GPs), the technique we focus on in this work. Each approach makes use of limited information about the response surface, namely the outputs of the simulator evaluated at a set of *training points*. Emulation is then about filling out the response surfaces of the outputs between these points.

## 2. Emulation of Heart Models

### 2.1. Gaussian Process Emulation

GPs are a popular technique for emulation that use a generalisation of the multivariate normal distribution with mean and covariance specified as *functions*, allowing for the modelling of continuous processes [5]. Training a GP emulator involves selecting appropriate forms for the mean and covariance function in responses to the training data, typically by specifying parameteric forms for these functions and then determining appropriate values for the hy-

perparameters of these parametric forms.

Setting aside Bayesian approaches that place prior distributions on the hyperparameters and integrate over the entire space of potential GPs, a natural method for choosing hyperparameters is to maximise the *likelihood*. Assuming a zero mean function, this takes the form

$$\log \mathcal{L} = -\frac{1}{2} \mathbf{y}^T \mathbf{K}^{-1} \mathbf{y} - \frac{1}{2} \log \det(\mathbf{K}) - \frac{N}{2} \log 2\pi. \quad (1)$$

Here  $\mathbf{K}$  is the matrix of covariances between all of the  $m$  training points, and  $\mathbf{y}$  the vector of simulator outputs at each of the training points. The hyperparameters change the covariances between the training points and enter into the likelihood (1) via the elements of the  $\mathbf{K}$  matrix. Maximising this likelihood is a continuous optimisation problem to which traditional methods can be applied, with the derivatives of equation (1) with respect to the hyperparameters available at little extra cost [5].

Once the GP has been trained, predictions  $\mathbf{y}^*$  are made using the mean of the GP at the prediction points, with its variance providing a measure of uncertainty in these predictions. The mean and variance are given by

$$\mathbf{y}^* = \mathbf{K}^* \mathbf{K}^{-1} \mathbf{y} \quad \text{Var} = \mathbf{K}^{**} - \mathbf{K}^* \mathbf{K}^{-1} \mathbf{K}^*,$$

with  $\mathbf{K}^*$  the matrix of covariances between training and prediction points and  $\mathbf{K}^{**}$  the vector of variances at the prediction points.

Here we use a squared exponential covariance,

$$k(\boldsymbol{\theta}, \boldsymbol{\theta}') = \sigma^2 \exp \left( -\sum_i \frac{(\theta_i - \theta'_i)^2}{2l_i^2} \right) + \sigma_n \delta(\boldsymbol{\theta}, \boldsymbol{\theta}'),$$

with  $\sigma$ ,  $\sigma_n$  and the  $l_i$  the hyperparameters.  $\sigma$  is the variance in the process,  $\sigma_n$  is the noise associated with each of the observations and  $l$  are the “effect lengths” of each of the input parameters, a measure of how sensitive the output is to that parameter. In emulation, the observations are runs of the simulator and so there is no noise associated, but it is still included as a regulariser for the process.

A simulator’s multiple outputs can be emulated separately, using independent GPs for each output. However, this independence creates an issue when correlation between outputs can be expected or when proximal locations in space or time are expected to be similar (correlated). This issue is resolved using dimension reduction techniques that seek to emulate the representative *features* that encode spatial or temporal outputs. These features can be treated as independent and hence emulated individually.

## 2.2. Emulating Single Cell Models

Models for the AP of a single cardiac cell represent the cell as a volume capacitor of capacitance  $C_m$ , and thus

express the change in membrane potential,  $V$  over time as the combined effect of all transmembrane currents,

$$\frac{dV}{dt} = \frac{-1}{C_m} \left( \sum_j I_j + I_{stim} \right). \quad (2)$$

Here  $I_{stim}$  is an external stimulus and the  $I_j$  are the set of model-specific transmembrane currents, which are expressed in terms of a set of gating and other state variables all with their own associated ODEs. Equation (2) is thus a system of ODEs, and is highly non-linear. Following evidence that ion channel density is one of the significantly variable properties between cells, studies in variability often focus on varying constants that express the current density of each individual current [1, 3].

We seek to produce an emulator that replaces numerical integration of the system (2) for specified values of the variable channel conductances. The output of this emulator is the AP curve,  $V(t)$ ,  $t \in [0, t_{end}]$ , that results after pacing to a steady state using a fixed stimulus regime. We represent the  $V(t)$  curve in terms of a set of features that can then be emulated by independent GPs. Previous work has used AP *biomarkers*, measures of the key properties of the AP such as its duration, resting potential and activation amplitude, as these features, but this does not allow re-construction of the original  $V(t)$  curve from the emulated features. Our approach emulates full AP profiles, from which biomarkers can then be easily calculated if required.

We obtain a set of features by decomposing the AP into a truncated infinite series over orthogonal functions,

$$V(t) \approx \sum_{i=1}^{N_f} a_i P_i(\tilde{t}). \quad (3)$$

The features are then the coefficients of this series,  $a_i$ . Predicted APs will always be a sum over smooth functions, and thus smooth themselves, satisfying the requirement of correlation between nearby timepoints. After testing a range of potential functional bases, we selected Legendre polynomials  $L_i(t)$  transformed to be orthogonal over the range  $[0, t_{end}]$ , for which the coefficients of the series are

$$a_i = \frac{2i+1}{2} \int_0^{t_{end}} V(t) L_i \left( 2 \frac{t}{t_{end}} - 1 \right) dt. \quad (4)$$

We calculate the integrals in (4) using Clenshaw–Curtis quadrature, as this naturally places large numbers of quadrature points towards the boundaries of the domain and hence can be expected to better resolve the extremely rapid upstroke in response to stimulus at  $t = 0$ .

### 2.3. Emulating Tissue Models

The spread of excitation in tissue is represented by the monodomain partial differential equation,

$$\frac{\partial V}{\partial t} = \chi C_m \nabla \cdot (D \nabla V) - \frac{1}{C_m} \left( \sum_j I_j + I_{stim} \right). \quad (5)$$

Here  $D$  is the conductivity tensor,  $\chi$  the volume-to-surface ratio for cardiac cells and  $C_m$ ,  $I_j$  and  $I_{stim}$  are all as defined for the single cell AP model (2). Equation (5) is much more costly to simulate than single cell models, as it involves the same strong non-linearities now acting at each point in the spatial domain. This necessitates a fine spatial grid or mesh, and the different timescales of behaviour in different regions of the system make it difficult to use adaptive timestep methods. This increase in model runtime makes emulation much more beneficial, but also increases the complexity of the emulation problem as the outputs to be emulated now occupy a higher number of dimensions. Furthermore, (5) is capable of producing extremely complex patterns of excitation that we expect to pose great challenge to statistical emulators. For this work we use a simple stimulation protocol that results only in radially expanding waves as proof of concept.

We seek to emulate the solution of (5) on a 2D domain at a small number of  $p$  ‘snapshots’ in time, using  $m$  training points. The number of finite differencing nodepoints used is  $N^2$ , in this case  $401 \times 401 = 160,801$ , creating a problem too large to emulate directly. The orthogonal function decomposition technique used for dimension reduction in Section 2.2 would require  $(N_f + 1)(N_f + 2)/2$  terms, also a prohibitive number of features to emulate. Instead, we use a kernel isomap transform, a form of dimension reduction along the lines of multidimensional scaling that uses geodesic, instead of Euclidean, distances to better capture manifold structures in the high-dimensional space [6].

Kernel isomap projects the original  $m \times N^2$  matrix of training data onto a  $k$ -dimensional feature space, with the number of features free to be chosen. This produces an  $m \times k$  matrix at each of the  $p$  snapshots. The features are the  $k$  most dominant eigenvectors of the kernel matrix, in the same fashion as kernel principal component analysis. Predictions for  $V(\mathbf{x}, t)$  are formed by using GP emulation to predict the co-ordinates in feature space and then transforming back into the original space by undoing the isomap transform via local linear interpolation [7].

## 3. Results and Conclusions

### 3.1. Single Cell Model Emulation

We emulate the Grandi *et al.* [8] model for the human atrial AP. We take as 11 variable parameters the conduc-

tances of all currents except for background and chlorine currents, including the rate of  $\text{Ca}^{2+}$  uptake and release by the sarcoplasmic reticulum. Each parameter is allowed to vary over the range  $\pm 30\%$  of the values given in the original work, a choice that reflects a physiologically reasonable level of variability [3].

Training and test data were produced by simulating the ODE system (2) using MATLAB’s *ode15s* routine, then calculating the corresponding coefficients for the Legendre polynomial series for each of the APs in the training set with  $N_f = 500$ . 550 points were selected randomly throughout the parameter space, with 500 used for training the remaining 50 reserved for testing. GPs were trained, and subsequent predictions made using the GP package for MATLAB available from <http://gaussianprocess.org>.

Use of a trained GP emulator instead of the simulator reduced computational time by a factor of 48, even for this comparatively simple problem of ODE integration. The majority of APs are emulated very well (Figure 1a), with occasional discrepancies in the resting potential and during the initial upstroke. Representative AP curves are compared with those produced by the simulator in Figure 1b, including one where the emulator performs poorly (dashed line). This AP is atypical, featuring an early afterdepolarisation (EAD), and without training points close to this location in the parameter space the GP emulator cannot predict this unexpected behaviour. However, the overall time to repolarisation is still well estimated by the emulated AP.

The biomarkers associated with the emulated curves (Figure 1c) on the whole perform about equally as well as GPs based on direct emulation of the biomarkers [9]. We suspect that the approximation of biomarkers using emulated AP curves would be significantly improved by using more training points in order to increase the chances of detecting ‘outlier’ behaviours in the simulator, though of course direct emulation of the biomarkers is also expected to improve to some extent when more training data is provided.

### 3.2. Tissue Model Emulation

As a preliminary demonstration of spatiotemporal emulation, we simulate (5) coupled to the Beeler-Reuter model for ventricular APs. This model uses four currents, and hence four variable parameters, and again we use a variability of  $\pm 30\%$ . We choose  $k = 10$  isomap features to emulate, and increasing the number of features was found not to improve results. Reflecting the increased cost of running the monodomain equation simulator, we use only 100 training points. Emulation of the monodomain problem proved to be more than four orders of magnitude faster than running the simulator.

Figure 2 shows an example simulation and the corresponding emulator output at different snapshots in time.

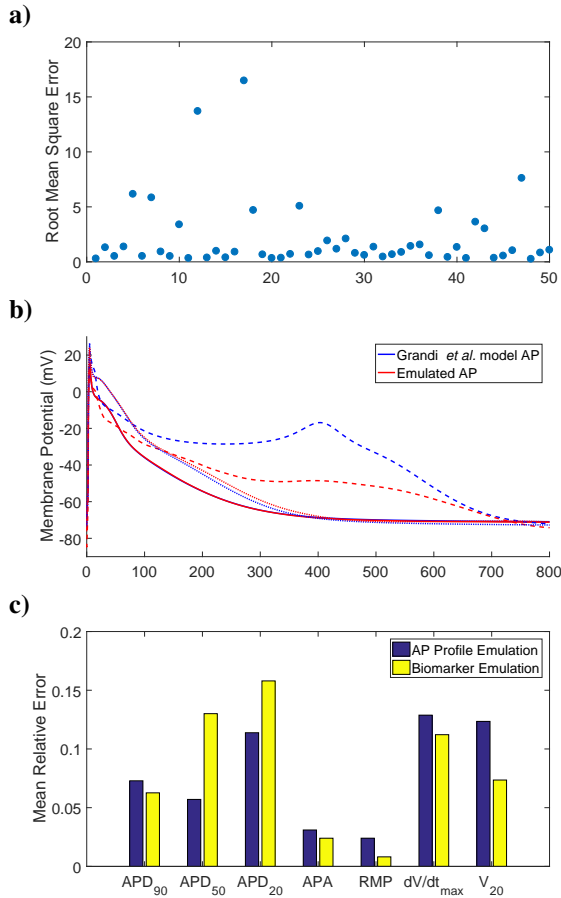


Figure 1. **a)** The root mean square errors along the curve obtained for the 50 AP profiles in the test dataset. **b)** Emulated and simulated APs at three test points, demonstrating the range of emulator behaviours. The line styles (solid, dotted, dashed) delineate the different test cases. **c)** Comparison between direct emulation of biomarkers and biomarkers calculated from emulated AP profiles.

The spatial spread of the travelling wave of excitation is well predicted in terms of front shape, speed of propagation and refractory time of cells. Minor smoothing of the profile is seen, but this does not impact on the important properties of the wavefront. However, a wide range of simulator behaviours are seen across the parameter space, and not all test points show the same fidelity in emulator predictions. Further analysis and additional training points, selected by a rigorous experimental design, are required.

Our initial results demonstrate the potential for massive speed increases in simulating spatiotemporal heart models in two or three dimensions. However, the complexity of these models makes emulation a challenging problem that requires further investigation. Of particular interest is the emulation of heterogeneous structures in the heart, which generate far more complex excitation patterns.

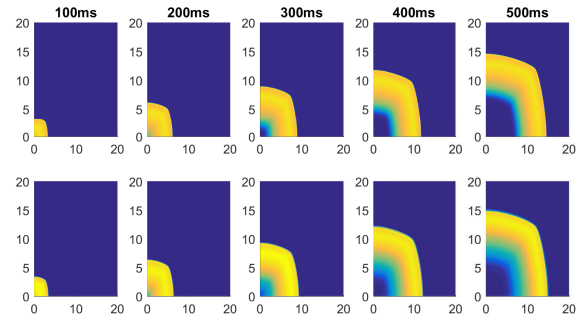


Figure 2. Simulator output (top row) and the corresponding emulator prediction (bottom row) for a series of temporal snapshots. The extent of excitation (depolarisation) is marked by the brightness of the image.

## References

- [1] Sarkar AX, Christini DJ, Sobie EA. Exploiting mathematical models to illuminate electrophysiological variability between individuals. *J Physiol* 2012;590:2555–2567.
- [2] Britton OJ, Bueno-Orovio A, Van Ammel K, Lu HR, Towart R, Gallacher DJ, Rodriguez B. Experimentally calibrated population of models predicts and explains intersubject variability in cardiac cellular electrophysiology. *Proc Natl Acad Sci U S A* 2013;110:E2098–2105.
- [3] Muszkiewicz A, Britton OJ, Gemmel P, Passini E, Sanchez C, Zhou X, Carusi A, Quinn TA, Burrage K, Bueno-Orovio A, Rodriguez B. Variability in cardiac electrophysiology: Using experimentally-calibrated populations of models to move beyond the single virtual physiological human paradigm. *Prog Biophys Mol Biol* 2016;120:115–127.
- [4] Sacks J, Welch WJ, Mitchell TJ, Wynn HP. Design and analysis of computer experiments. *Stat Sci* 1989;4:409–423.
- [5] Rasmussen CE, Williams CKI. *Gaussian Processes for Machine Learning*. MIT Press, 2006.
- [6] Tenenbaum JB, de Silva V, Langford JC. A global geometric framework for nonlinear dimensionality reduction. *Science* 2000;290:2319–2323.
- [7] Xing W, Shah AA, Nair PB. Reduced dimensional Gaussian process emulators of parameterized partial differential equations based on isomap. *Proc R Soc A* 2015;471:20140697.
- [8] Grandi E, Pandit SV, Voigt N, Workman AJ, Dobrev D, Jalife J, Bers DM. Human atrial action potential and  $Ca^{2+}$  model: Sinus rhythm and chronic atrial fibrillation. *Int Physiol Circ. Res.*;109:1055–1066.
- [9] Chang ETY, Strong M, Clayton RH. Bayesian sensitivity analysis of a cardiac cell model using a gaussian process emulator. *PLoS One* 2015;10:e0137004.

Address for correspondence:

Brodie A. J. Lawson  
 Queensland University of Technology, 2 George Street,  
 BRISBANE, Queensland, Australia 4000  
 b.lawson@qut.edu.au