Creation of predictive cardiac excitation models at the tissue scale with machine learning in augmented state space

Desmond Kabus[1,2] Antoine A.F. de Vries[2] Daniël A. Pijnappels[2] Hans Dierckx[1,*]

2023-05-16

Abstract

When creating in-silico models of excitable media, often only one spatio-temporal variable can be measured easily. This may for instance be optical voltage mapping data of the electrical waves in cardiac muscle tissue controlling the heart beat. We use exponential moving averages and gradients computed using the least-squares method to extract additional states from this one variable to span a sparse discretised 3D state space. To this augmented state space, we fit a simple polynomial model to predict the evolution of the one state variable. For optical voltage mapping data of human atrial myocyte monolayers stimulated using a stochastic burst pacing protocol, the surrogate model is able to describe the excitation of the system, as well as wave propagation.

[1] Department of Mathematics, KU Leuven Campus Kortrijk (KULAK), Belgium
[2] Laboratory of Experimental Cardiology, Leiden University Medical Center (LUMC), The Netherlands

[*] Correspondence: h.dierckx@kuleuven.be

Contents

1 Introduction **1**

2 Methods **2**
2.1 Data acquisition 2
2.1.1 Optical voltage mapping 2
2.1.2 Numerical simulation 2
2.2 State space augmentation 3
2.2.1 Exponential moving averages 4
2.2.2 Least-squares absolute gradient . . . 4
2.3 State space histograms 4
2.4 Model fitting 5

3 Results **5**
3.1 Surrogate model for synthetic data 5
3.1.1 State space 5
3.1.2 Forward model 6
3.2 Surrogate model for optical mapping data . 6
3.2.1 State space 6
3.2.2 Forward model 6

4 Discussion **6**

5 Conclusion **11**

6 Acknowledgments **11**

1 Introduction

With the goal of building a digital twin of each patient's heart, i.e., a full recreation of it in the computer, the scientific community is creating sophisticated models for different mechanisms in the heart [15, 17]. One of them is the contraction of the heart which is controlled by electrical waves travelling through the myocardium [4]. An *in-silico model of cardiac tissue electrophysiology* is a mathematical model that is able to describe the dynamics of these excitation patterns [4].

Typically, in-silico models for cardiac electrophysiology fall into one of two categories. On the one hand, there are the mathematical models that focus mostly on the overall dynamics in the tissue [1, 3, 6, 14]; and on the other hand there are detailed models that aim to model each current across the cell membrane [5, 13, 16]. Typically, these models are obtained by fitting in-vitro data: patch-clamping of single cells, optical mapping of monolayers or of the epicardium.

A common property in the models focussing on the overall dynamics rather than the individual ion currents is that the state variables used in them often do not have an equivalent in the physical tissue. These internal, hidden states are not fit to actual data, instead they are often chosen as a means to an end.

Capturing cell memory is crucial in the modelling of cardiac tissue. This may for instance be done an exponentially decaying memory variable [7, 18].

A recent step forward in in-vitro models for cardiac electrophysiology was the conditional immortalisation of human atrial myocytes (hiAMs) [9]. These well-differentiated cells address some of the shortcomings of using heart muscle cell cultures of animals for in-vitro modelling: Cells from different species have different electrophysiological behaviour, i.e., they vary in parameters such as the action potential duration (APD), the conduction velocity (CV) or restitution characteristics, resulting in different wave propagation [4, 9, 13]. No in-silico version of the electrophysiological behaviour of this cell line has yet been published.

In this work, we present a general method to fit in-silico tissue models from just one recorded variable over two-dimensional space and time. The method requires only minimal assumptions and works by expanding the data to a multi-dimensional state space. We both re-create an existing in-silico model with our method and use it to create an in-silico tissue model for hiAMs only from optical voltage mapping data. The steps of our method are outlined graphically in Fig. 1. In section 2, we describe how the data are acquired, how they are processed, expanded, and

how a surrogate model is finally fit to them. In section 3, we present preliminary results from the models that are generated with this method, which should be thought of as a proof of concept. Finally, in sections 4–5, we wrap up by summarising the method and discussing its advantages and shortcomings, as well as giving an outlook on how the models fit by our method can be improved.

2 Methods

2.1 Data acquisition

In this work, we study spatio-temporal data from two excitable media: an in-vitro data set from *optical voltage mapping* (OVM) and a in-silico data set from finite-differences simulations solving the mono-domain equations [4]. In this section, we outline how these data were generated and processed.

2.1.1 Optical voltage mapping

To obtain a two-dimensional recording of the electrical activity of tissue as an in-vitro data set, we perform optical mapping using a voltage-sensitive dye of $10\,\mathrm{cm}^2$ hiAM monolayers in a six-well culture plate. We followed the OVM protocols that can be found in the earlier publication introducing hiAMs [9]. The recorded data consist of one variable, i.e., the measured difference in intensity of the emitted light, over space and time. Each of the 100×100 square pixels has a size of $\Delta x = 250\,\mu\mathrm{m}$ and the temporal resolution is $\Delta t = 1\,\mathrm{ms}$ between each of the 61 440 consecutive frames in time. We assume that the monolayer is approximately homogeneous and isotropic.

To excite the tissue, we place a bi-polar stimulation electrode with two poles with a diameter of 0.5 mm and a distance of 1 mm at the well's boundary, and stimulate with $N = 200$ current pulses of random amplitude, random duration, and random timing. We call this procedure the *stochastic burst pacing* (SBP) protocol, inspired by stochastic pacing experiments done by Krogh-Madsen et al. [11]. The amplitudes A_n are sampled from a uniform distribution $\mathcal{U}([-0.8\,\mathrm{mA}, 0.8\,\mathrm{mA}])$. The durations $\tau_n = \tau_\mathrm{u} + \tau_\mathrm{e}$ are sampled from a combination of a uniform and exponential distribution:

$$\tau_\mathrm{u} \sim \mathcal{U}([0\,\mathrm{ms}, 10\,\mathrm{ms}]) \tag{1}$$

$$\tau_\mathrm{e} \sim \mathrm{Exp}\,(1/40\,\mathrm{ms}) \tag{2}$$

and likewise the time $T_n = T_\mathrm{u} + T_\mathrm{e}$ between subsequent pulses:

$$T_\mathrm{u} \sim \mathcal{U}([0\,\mathrm{ms}, 50\,\mathrm{ms}]) \tag{3}$$

$$T_\mathrm{e} \sim \mathrm{Exp}\,(1/200\,\mathrm{ms}) \tag{4}$$

where $\mathcal{U}([a, b])$ denotes the uniform distribution on the interval $[a, b]$ and $\mathrm{Exp}\,(\lambda)$ the exponential distribution with decay rate λ:

$$\mathrm{Exp}\,(t; \lambda) \propto \begin{cases} \mathrm{e}^{-\lambda t} & t \geq 0 \\ 0 & t < 0 \end{cases} \tag{5}$$

The stimulus current hence takes the form:

$$I_\mathrm{stim}(t) = \sum_{n=1}^{N} A_n [\mathrm{H}\,(t - t_n) - \mathrm{H}\,(t - t_n - \tau_n)] \tag{6}$$

with start time of the n-th pulse:

$$t_n = \sum_{m=1}^{n} [T_m + \tau_m] \tag{7}$$

and the Heaviside function:

$$\mathrm{H}\,(t) = \begin{cases} 0 & t < 0 \\ \frac{1}{2} & t = 0 \\ 1 & t > 0 \end{cases} \tag{8}$$

We preprocess these data in a similar way as in a previous publication [10]: We apply Gaussian blurring with a kernel size of three grid points and rescale the data to zero-mean and unit variance.

As a further step, we subtract the baseline of the data, such that the resting state is close to zero. To determine the baseline, we make the assumption that a neighbourhood of 4×4 pixels over a time span of 2 s, the resting state corresponds to the most common value of the recorded intensity.

The final preprocessing step for this data set is to smooth the data with an *exponential moving average* (EMA) with $\alpha = 0.1$, which, for time $t \geq 0$ with frame duration Δt and the signal over time $s(t)$ at each position in space, is iteratively defined as:

$$\tilde{s}(t) = \begin{cases} \alpha s(t) + [1 - \alpha]\tilde{s}(t - \Delta t) & t > 0 \\ s(t = 0) & t = 0 \end{cases} \tag{9}$$

or with $a = \frac{\alpha}{\Delta t} > 0$:

$$\tilde{s}(t) = \mathrm{e}^{-at} s(0) + a \int_0^t \mathrm{e}^{a[t' - t]} s(t')\,\mathrm{d}t' \tag{10}$$

We refer to this recording as the OVM data set and $u(t, \boldsymbol{x})$ denotes the unit-less signal after preprocessing which corresponds to the electrical activity of the cells, i.e., it is linked to the transmembrane voltage over space and time.

2.1.2 Numerical simulation

To obtain a fully controllable testing data set, we have also run a similar experiment in-silico. We use the finite-differences method to solve the reaction-diffusion system using a rescaled version of the minimalist two-variable tissue model by Aliev and Panfilov [1], or in short the AP96 model, whose equations we will briefly summarise shortly (eqs. 15–16).

As the diffusivity scales as the square of the *conduction velocity* (CV), we set it such that the medium has a measured CV around the CV of hiAMs [4, 9]:

$$D = 0.0625\,\mathrm{mm}^2\,\mathrm{ms}^{-1} \tag{11}$$

$$\Rightarrow \mathrm{CV} = 0.17\,\mathrm{m\,s}^{-1} \tag{12}$$

And we rescale the entire model over time such that its *action potential duration* (APD) has a value of roughly the APD of hiAMs [9]:

$$\mathrm{APD} = 100\,\mathrm{ms} \tag{13}$$

$$\Rightarrow r = \frac{25.58}{\mathrm{APD}} = 0.2558\,\mathrm{ms}^{-1} \tag{14}$$

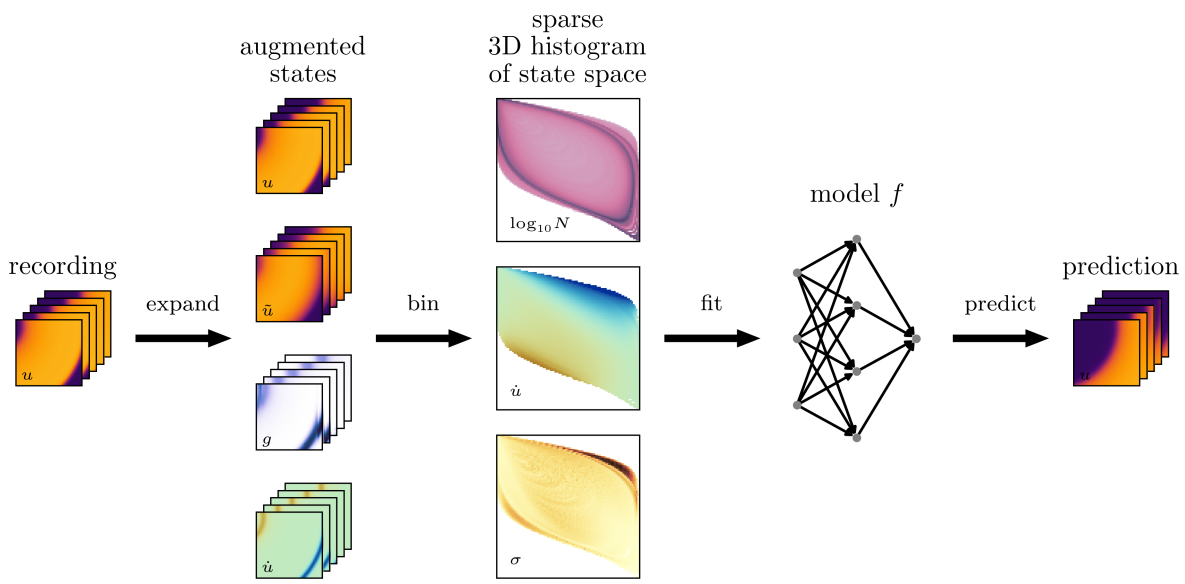

Figure 1: Outline of the surrogate model fitting procedure. In the first step, relevant features are extracted from the spatio-temporal data u. Then the state space is discretised and histogrammed. The value of the time-derivative $\dot{u}$ of the main variable is analysed statistically in this state space. A surrogate model can then be fit to the state space and used to predict the evolution of the medium.

Table 1: Parameters of the rescaled AP96 model.

Parameter	Value
a	0.15
ϵ_0	0.002
k	8
μ_1	0.2
μ_2	0.3
D	$0.0625\,\mathrm{mm}^2\,\mathrm{ms}^{-1}$
r	$0.2558\,\mathrm{ms}$

The time derivatives are rescaled by this time scaling factor r [1]:

$$\partial_t u = \nabla \cdot D\nabla u - rku[u-a][u-1] - ruv + \frac{i_{\mathrm{stim}}}{C} \tag{15}$$

$$\partial_t v = -r\epsilon v - r\epsilon ku[u-a-1] \tag{16}$$

$$\epsilon = \epsilon_0 + \frac{\mu_1 v}{u+\mu_2} \tag{17}$$

We have thereby introduced the millimetre as the spatial unit and the millisecond for time into the AP96 model which uses arbitrary space and time units. The variables $u(t,\boldsymbol{x})$ and $v(t,\boldsymbol{x})$ remain unit-less. The value $u=0$ corresponds to the resting state and when excited, u approaches the value $u=1$.

For all other parameters, a, ϵ_0, k, μ_1, and μ_2, we use the default values, as in the original publication by Aliev and Panfilov [1]. An overview of all parameters can be found in Table 1.

We have run a simulation over 30 000 frames at $\Delta t = 1\,\mathrm{ms}$ per frame at the same resolution as for the in-vitro data, OVM, and with the same SBP. The electrode was placed off-centred at the side of one edge of the medium. For the time integration, we have used a simple Euler-forward step with numerical time step 0.1 ms and a second-order central finite-differences stencil for the Laplacian $\nabla \cdot D\nabla$. The square simulated tissue is homogeneous and isotropic with no-flux boundary conditions.

We refer to this synthetic recording as the AP96 data set. We only include the variable $u(t,\boldsymbol{x})$ in the data set, as in the in-vitro data (OVM). For surrogate model fitting, we hence also only use one variable related to the transmembrane voltage of the tissue and not the internal state of the cell, which in the AP96 model is encapsulated in the second variable $v(t,\boldsymbol{x})$.

2.2 State space augmentation

The multi-dimensional space spanned by all variables describing the state of the modelled medium, $\boldsymbol{u} = [u, v, ...]^{\mathrm{T}}$, is called *state space* in contrast to the physical space spanned by $\boldsymbol{x}$ and time t.

An in-silico tissue model usually prescribes a unique reaction $\boldsymbol{R}$ at each position $\boldsymbol{u}$ in the state space. We write a more generic version of eqs. 15–16, the mono-domain description [4]:

$$\partial_t \boldsymbol{u} = \boldsymbol{P}\nabla \cdot \boldsymbol{D}\nabla \boldsymbol{u} + \boldsymbol{R}(\boldsymbol{u}) \tag{18}$$

with the diffusivity matrix $\boldsymbol{D}$, where $\boldsymbol{P}$ is a diagonal matrix describing which variables in $\boldsymbol{u}$ are to be diffused.

For the two-variable AP96 model, with $\boldsymbol{u} = [u, v]^{\mathrm{T}}$ and $\boldsymbol{P} = \mathrm{diag}\,(1, 0)$, the state space of the AP96 model is shown in Fig. 2, coloured only by the component in u of the reaction term $\boldsymbol{R}$. It can be seen that the second variable v can be used to distinguish cases with positive and negative reaction at the same value of u. A model that can describe both excitation from a resting state and recovery to it, is only possible with at least two variables, such that the inertial manifold can be a loop in the state space.

In both types of data sets we consider in this work, we only have data for one variable available, namely u. To still be able to differentiate between excitation and recovery,

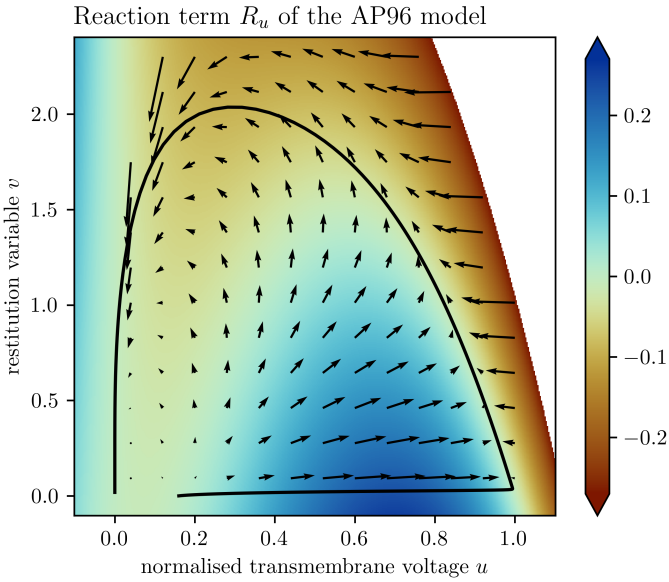

Figure 2: Reaction in state space of the AP96 model. The reaction term $\boldsymbol{R}$ is a vector field defining a reaction at each point $\boldsymbol{u}$ in the state space. Colouring is according to the u-component of the reaction term R_u. The trajectory through the state space from excitation of one point is shown by example.

we extract additional information from u over space and time to use as additional states for the state space, so-called *augmentations*. A well-designed *augmented state space* unfolds the data such that cells in the tissue with different histories are identified by different state vectors in it.

As we want to find surrogate models that can be used as a forward model in a similar way as in eq. 18, we impose two more requirements on the states with which we augment the state space:

1. Augmentations should be local in space, i.e., they should only use states within a neighbourhood around each point.
2. Causality must be respected, i.e., only data from previous points in time and the current point in time may be used to compute each augmentation.

While other augmentations were tried in the development of this method, such as a pseudo-ECG or the last APD and DI at each point over time, we are here presenting two augmentations we found particularly useful.

2.2.1 Exponential moving averages

Information that we want to capture with a state space augmentation is that at the same value of u, the model can evolve differently. It may increase at the front of waves, during excitation, or decrease at the wave backs, during recovery.

One way to do sense excitation and recovery is by expanding the state space with augmented states that capture the passage of time. A naïve way to do so may be to just include a time-delayed signal, which is a useful augmentation in phase mapping [8]. This is however not very robust to noise, as noise in the past may directly impact the update. Another option may be to use Hilbert transforms, but as their computation involves an integral over all of time, they violate causality and are therefore not meeting our requirements [12].

The EMA (eq. 9) is robust to noise and typically lags behind the signal by an amount of time inversely proportional to the constant α. It can intuitively be used to differentiate between excitation and recovery for signals u in-between the resting state and excited state: If an EMA with a suitable scale α is lower than u, it is likely that the medium is currently getting excited; if it is higher, it is likely recovering.

Multiple EMAs can be used to be able to capture more advanced behaviour: With an appropriately chosen α, we can differentiate between states where the tissue is activated a lot and where it is mostly resting. With such an EMA, restitution characteristics can be captured.

We augment both data sets (OVM and AP96) with an EMA with $\alpha = 0.1$, denoted by $\tilde{u}$.

2.2.2 Least-squares absolute gradient

We also conjecture that information about the slopes and other derivatives in space of u is useful for a well fitting surrogate model. Aiming for high robustness with respect to noise, we compute the average absolute value of the gradient in a neighbourhood by fitting planes using the least-squares algorithm. We call this quantity GLSq. It is computed in these steps:

1. At each point $\boldsymbol{x}$ on the rectangular lattice, fit a plane to a selection of nearby grid points.
2. The absolute value of the slope of this plane is the value GLSq.

We select the points in the neighbourhood of each point $\boldsymbol{x}$ based on two criteria:

a. Only consider points in a given radius r_{GLSq} around $\boldsymbol{x}$.
b. Of these points, only include those whose grid index relative to $\boldsymbol{x}$ is divisible by a given integer N_{GLSq}.
c. Do not fit to points with invalid values or outside the medium.

Different values of r_{GLSq} and N_{GLSq} can be used to tweak the desired scale over which the gradient is computed while only including a reasonable number of points in the fit.

We augment the data sets (OVM and AP96) with a GLSq variable with $r_{\mathrm{GLSq}} = 5$ and $N_{\mathrm{GLSq}} = 5$, denoted by g.

2.3 State space histograms

With this choice of state space expansion, the state vector is $\boldsymbol{u} = [u, \tilde{u}, g]^{\mathrm{T}}$.

In the usual dynamics observed in the excitable systems we study in this work, not all points $\boldsymbol{u}$ in the state space are equally likely. Some state vectors $\boldsymbol{u}$ may be quite common as being part of the inertial manifold, while others might be completely absent in the data.

A model that is fit directly to the data may be heavily biased towards treating the common state vectors with high accuracy, neglecting the less common state vectors. We therefore discretise the state space with three variables to

100 bins in each dimension and collect the data in a multi-dimensional histogram. By fitting to the average value inside each bin instead of all data points, we purposefully treat each state vector in the observed state space with the same weight. For better memory usage, these histograms are computationally implemented using multi-dimensional sparse arrays. This efficient memory management enables building higher-dimensional state spaces, though in this work we require only three variables for now. State vectors which are completely absent in the data do not take up memory this way.

By example for the AP96 model, it can be seen in Fig. 3 that a part of the state space is not visited and most of the samples are on the inertial manifold, a closed approximately one-dimensional loop. We conjecture that in higher dimensions the state space is even more sparse, e.g. with detailed ionic models containing more than 10 state variables.

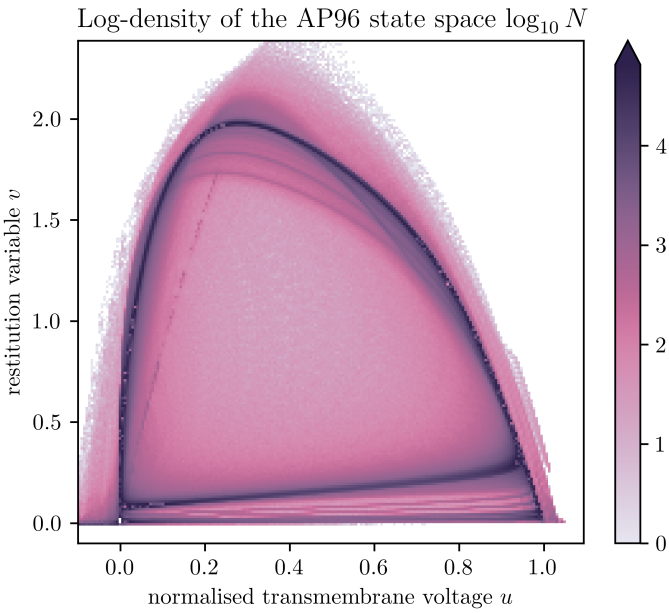

Figure 3: Likelihood of each point in the state space of the AP96 model for the numerical simulation generating AP96.

In the forward models we develop in this work, we only use one variable u over space and time from which all other variables in the state vector $\boldsymbol{u}$ can be derived. The time-derivative $\dot{u}$ fully defines the reaction of the surrogate model.

In a well chosen state space, for each state vector $\boldsymbol{u}$, the reaction $\dot{u}$ is the same, i.e., the states are well separated. Therefore, we calculate the mean value of $\dot{u}$ computed using finite-differences at each bin in the discretised state space, as well as its standard deviation. This mean value of $\dot{u}$ takes a similar role as R_u for the AP96 model and *"colours"* the state space in a comparable way to R_u in Fig. 2.

2.4 Model fitting

A surrogate model will be fit to the data in state space as a scalar-valued function f:

$$f : \boldsymbol{u} \mapsto \dot{u} \tag{19}$$

such that the model consists of the state space expansion $\boldsymbol{u}$ using eq. 9 and the algorithm for GLSq outlined in section 2.2.2, as well as the Euler-forward step:

$$u(t + \Delta t, \boldsymbol{x}) = u(t, \boldsymbol{x}) + \Delta t f(\boldsymbol{u}(t, \boldsymbol{x})) \tag{20}$$

A simple choice for f is to use the least-squares algorithm to find the polynomial of a given order which minimises the loss:

$$\mathcal{L}(f) = \sum_{\boldsymbol{u} \in \mathcal{U}} \left[\dot{u}(\boldsymbol{u}) - f(\boldsymbol{u})\right]^2 \tag{21}$$

where $\mathcal{U}$ denotes the centres of all bins in the discretised state space containing at least one data point.

As a proof of concept, we choose to fit a third order polynomial which is linear with respect to each of the components:

$$f(\boldsymbol{u}) = \begin{array}{l} w_{\tilde{u}gu}\tilde{u}gu + w_{\tilde{u}g}\tilde{u}g + w_{\tilde{u}u}\tilde{u}u + w_{\tilde{u}}\tilde{u} + \\ w_{gu}gu + w_g g + w_u u + w \end{array} \tag{22}$$

To prevent the surrogate model from leaving the valid state space, after each Euler-forward step, we clip the predicted u to the valid range for which we have data. Note that doing so also limits the EMA $\tilde{u}$ to this range, while the GLSq g can still leave the range observed in the data, i.e., for larger than typical slopes.

To span the state space $\mathcal{U}$ that the surrogate model is fit to, we use points up to a specific point in time and at least 20 grid points away from the boundary to minimise boundary effects on the used data. The evolution of the system after this point in time is used for testing the model by comparing prediction to reference.

3 Results

In this section, we present how the state space expansion and surrogate model that we chose in section 2, deal with the two data sets. As the results for the in-silico synthetic data are easier to interpret, we begin with those data.

3.1 Surrogate model for synthetic data

3.1.1 State space

For the AP96 data set from numerical simulation of the AP96 model (cf. section 2.1.2), the 3D state space spanned by $\boldsymbol{u} = [u, \tilde{u}, g]^{\mathrm{T}}$ is quite sparse. 276 757 bins in this space contain at least one data point, only 28 % of all bins. Of these bins, 30 % contain at least 10 data points. For the bins with at least one point, a bin contains 102 data points on average.

In Figs. 4 & 5, we show two different projections to 2D of the state space; to u and $\tilde{u}$, and to u and g, respectively. In Fig. 6, we present a three-dimensional view of the state space. In each of these figures, in panel A, it can be seen that most data points lay on a loop-like inertial manifold, while a large volume of the state space is still covered by data points (panel B). Colouring by the mean of $\dot{u}$, we can see that the two halfes of the loop nicely distinguish between excitation and recovery, in blue and red shades in panel C. In panel D, the state space is coloured by the

standard deviation of $\dot{u}$ for the data points in the bins. In large parts of these projections, the standard deviation is much lower than the mean of $\dot{u}$, showing that the model's possible states are well-separated in the chosen surrogate state space $\boldsymbol{u}$.

3.1.2 Forward model

A model in the context of this work consists of a *"colouring"* of the entire 3D state space representing the $\dot{u}$ at each possible state $\boldsymbol{u}$.

The surrogate model f is fit to the 3D state space as outlined in section 2.4 resulting in:

$$f(\boldsymbol{u}) = \begin{array}{l} -0.231\,30\,\tilde{u}gu - 0.589\,65\,\tilde{u}g + \\ 0.016\,863\,\tilde{u}u + 0.005\,441\,0\,\tilde{u} + \\ 0.482\,58\,gu + 0.099\,841\,g - \\ 0.012\,559\,u - 0.005\,277\,2 \end{array} \tag{23}$$

Subsequently, a forward simulation has been performed with the model leading to a prediction of u based on the history until that point which can be compared to the actual evolution of u in Fig. 7. This simple model displays excitation and recovery, as well as propagation of the wave. At the boundary of the simulated medium, build-up of u leads to the state vector hitting the hard limit of $u = 1$ and remains stationary there. Due to this effect, the behaviour is unphysical where the wave hits the boundary. Inside of the medium however, the predicted evolution is comparable to the reference. A ripple effect then spreads out from the wave which is also an unphysical effect from the numerical integration.

For the prediction, an action potential duration of APD $= (36.0 \pm 1.0)$ ms can be measured for the polarisation threshold of $u = 0.7$, which is shorter than the reference APD $= (34.0 \pm 1.0)$ ms.

Using the method by Bayly et al. [2], conduction velocity has been measured to be CV $= (0.132 \pm 0.024)\,\mathrm{m\,s^{-1}}$ for the prediction, and CV $= (0.1535 \pm 0.0031)\,\mathrm{m\,s^{-1}}$ for the reference.

3.2 Surrogate model for optical mapping data

3.2.1 State space

For the OVM data set from optical voltage mapping of hiAM cells (cf. section 2.1.1), the 3D state space is sparse, too. Only 13 % of the bins contain at least one data point. Of these bins, 44 % contain at least 10 data points. For the bins with at least one point, a bin contains 117 data points on average.

As can be seen in Figs. 8, 9, & 10, for these data we are also observing a loop-like inertial manifold with two halfes as in the other data set.

3.2.2 Forward model

For this data set, the fit polynomial for the surrogate model is:

$$f(\boldsymbol{u}) = \begin{array}{l} -0.016\,190\,\tilde{u}gu - 0.128\,61\,\tilde{u}g + \\ 0.003\,260\,5\,\tilde{u}u - 0.064\,610\,\tilde{u} + \\ 0.112\,37\,gu + 0.074\,417\,g + \\ 0.058\,239\,u - 0.004\,102\,7 \end{array} \tag{24}$$

The prediction by the forward model for the OVM data set can be seen in Fig. 11. It captures wave propagation and excitation. However, before the system is able to recover, this prediction suffers from the same numerical instabilities as the model fit to the synthetic data.

For these data, we measure CV $= (0.129 \pm 0.059)\,\mathrm{m\,s^{-1}}$ for the prediction, and CV $= (0.220 \pm 0.043)\,\mathrm{m\,s^{-1}}$ for the reference.

4 Discussion

In this paper, we have shown that enough information can be extracted from just one spatio-temporal variable to build a state space which enables the creation of in-silico models that are able to predict the further dynamics of the system in medium-term time scales of around 50 ms. Only the information from optical voltage mapping is already enough to fit a surrogate model to the dynamics. The stochastic burst pacing protocol ensures that even less likely stimulation patterns are observed in the data.

While the variables with which we augment the state space should be seen as preliminary and not optimal choices, the models created with these variables already are able to capture crucial behaviour of the tissue, namely excitation, recovery and wave propagation. It remains to be seen if more or other state space augmentations can be used to be able to replicate higher-order effects.

The final polynomial surrogate model f is fit using very a basic supervised machine learning method. This is however just a small part of the method presented in this work. We have shown that using unsupervised learning, i.e., data transformations, much insight can be gained about the structure of the data. The insights from presenting the data sets in the chosen 3D state spaces as done in this work make it almost trivial to fit the surrogate model to the data. It is remarkable, that such a simple polynomial is able to capture the behaviour of the model sufficiently with the proper state space expansion.

We have not yet optimised the model performance at boundaries of the medium. Also, boundary layer effects on the dynamics may manifest, which need also be learned in a future version of the surrogate model.

The procedure to generate in-silico tissue models however is designed to work well in a setting where not much is known about the wave propagation in an in-vitro tissue model. With it, a surrogate model can be fit with relative ease that then can be used to predict the further evolution of the tissue. Possible applications include an online-learning approach, where the surrogate model's predictive power is continuously improved while new data comes in.

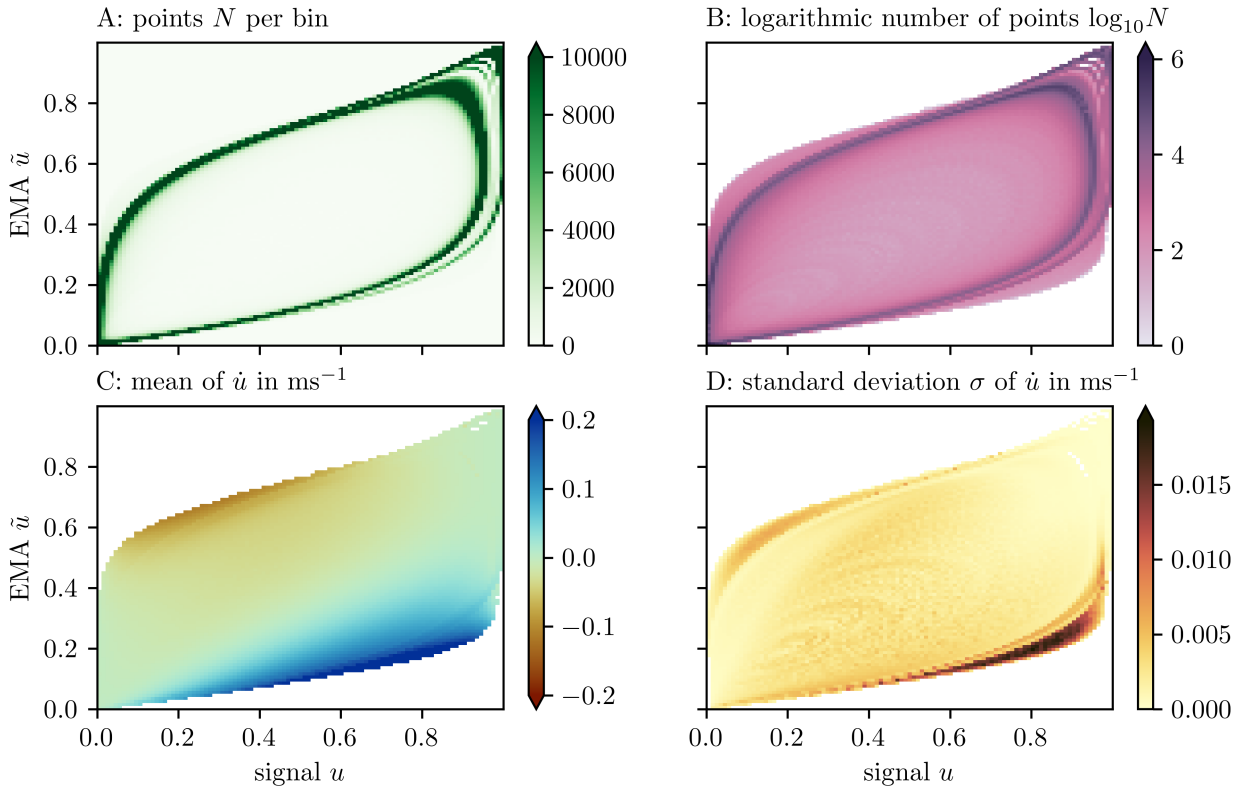

Figure 4: Projection of the 3D state space for the AP96 data set onto two dimensions. (A): By colouring the space by the number N of points in each bin, it can be seen that most points lay on a loop. (B): Colouring by $\log_{10} N$, we can see that still a large volume of the space is covered by data. (C): Looking at the mean of $\dot{u}$ at each bin in this projection, we can see that in one side of the loop, u increases and while it decreases in the other. (D): The standard deviation of $\dot{u}$ gives us an intuition if the states have been chosen well.

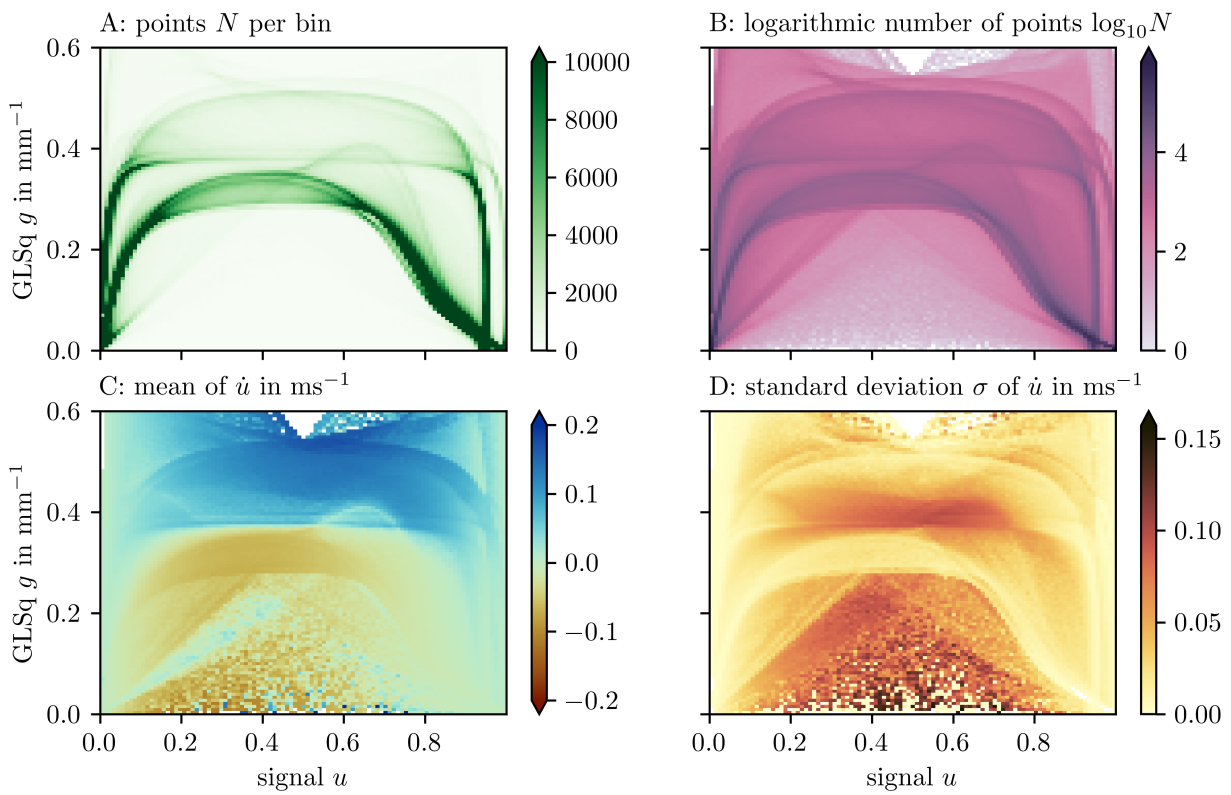

Figure 5: Projection of the 3D state space for the AP96 data set onto two dimensions in the same style as in Fig. 4.

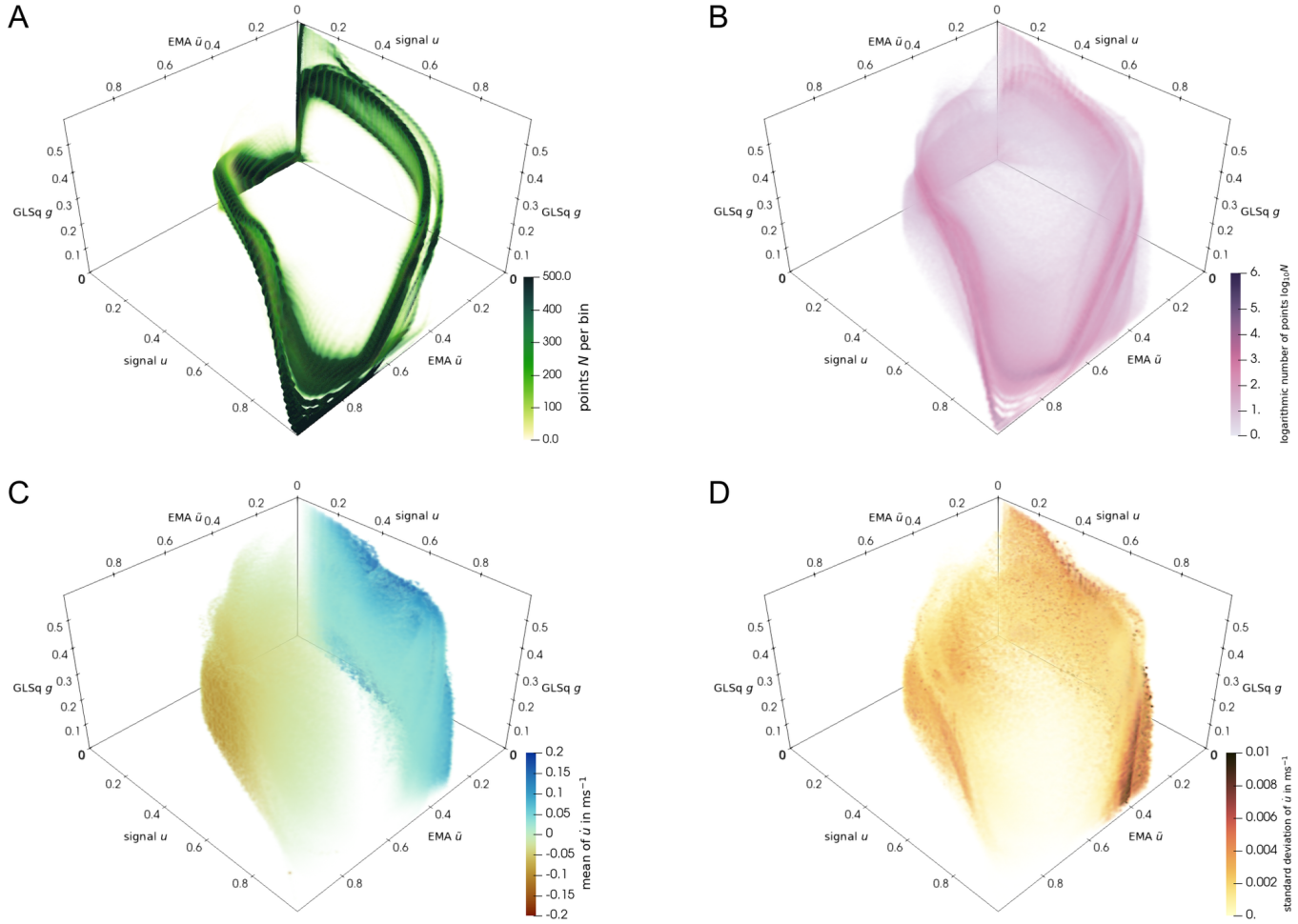

Figure 6: Visualisation of the 3D state space for the AP96 data set in the same style as in Fig. 4.

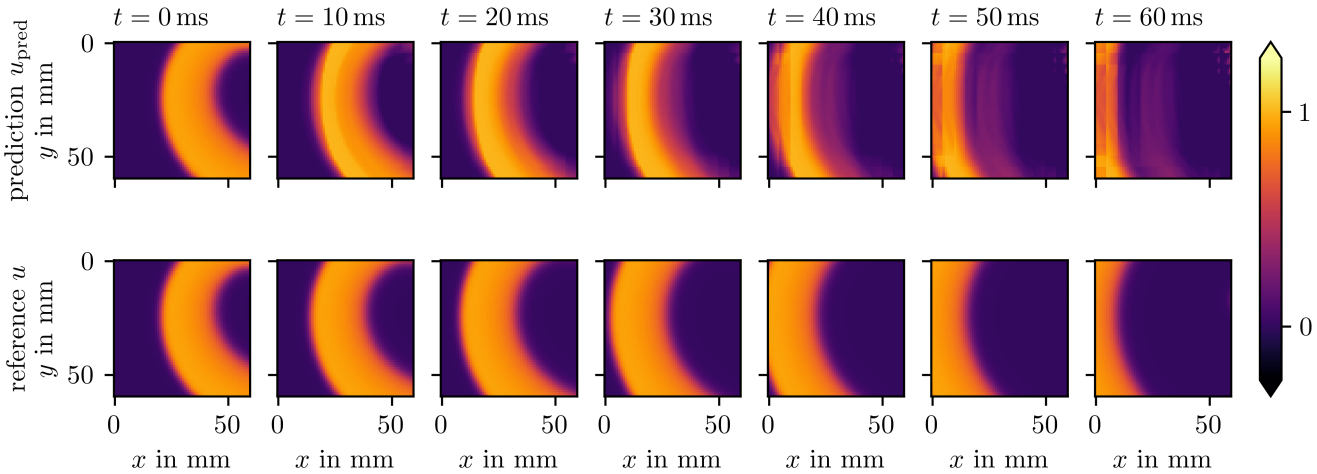

Figure 7: Prediction of the surrogate model with respect to each state over time (top row) compared to the reference solution (bottom row) for the AP96 data set.

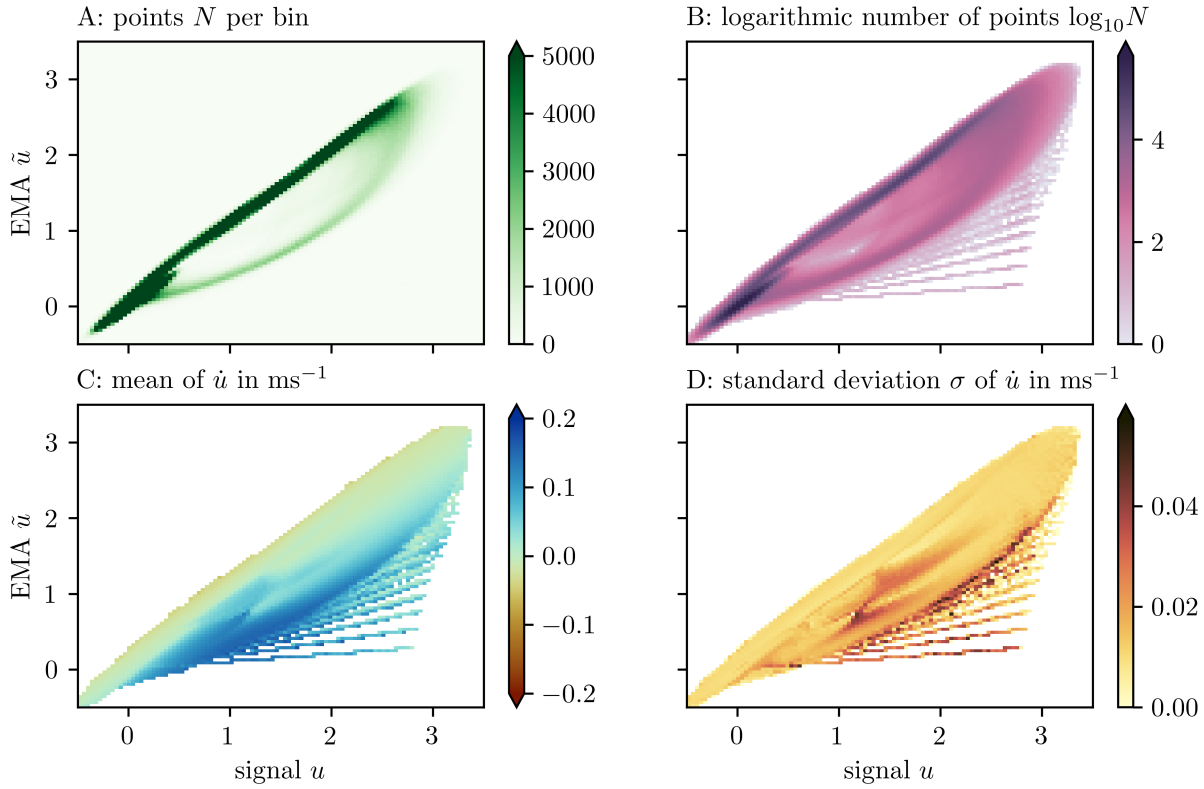

Figure 8: Projection of the 3D state space for the OVM data set onto two dimensions in the same style as in Fig. 4.

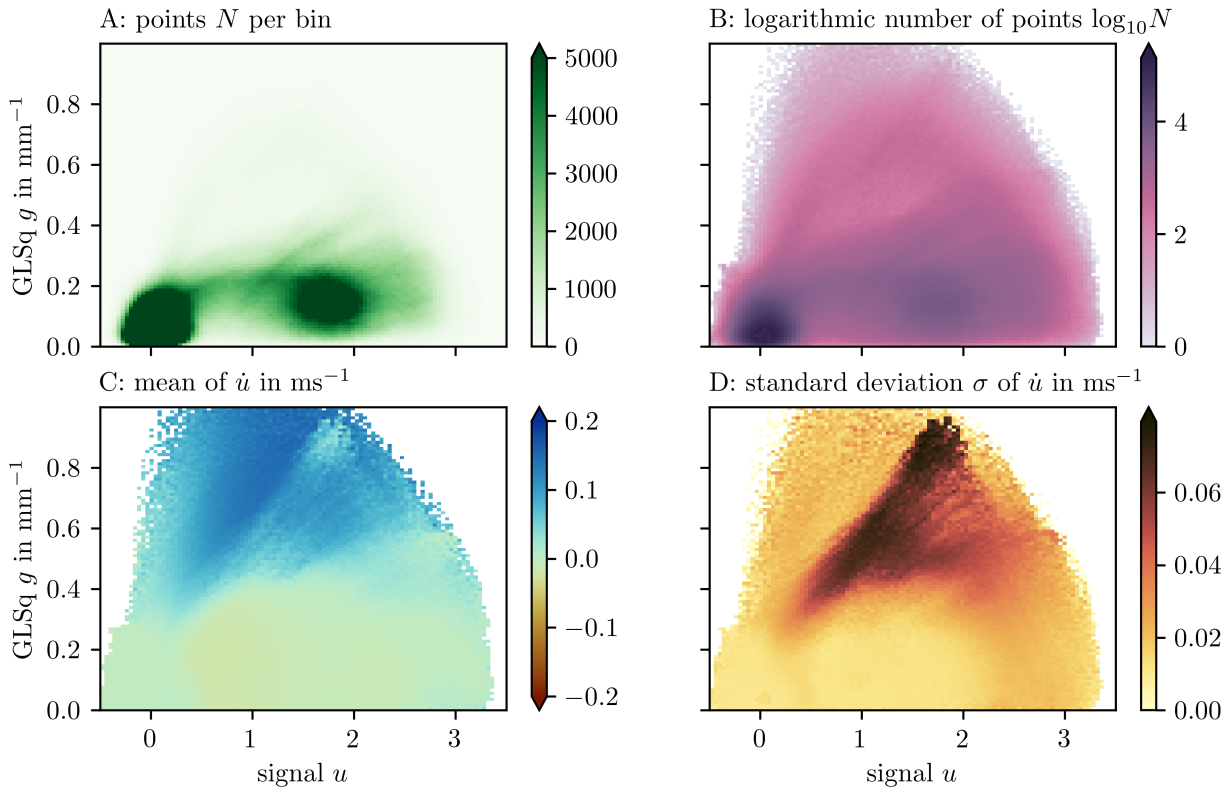

Figure 9: Projection of the 3D state space for the OVM data set onto two dimensions in the same style as in Fig. 4.

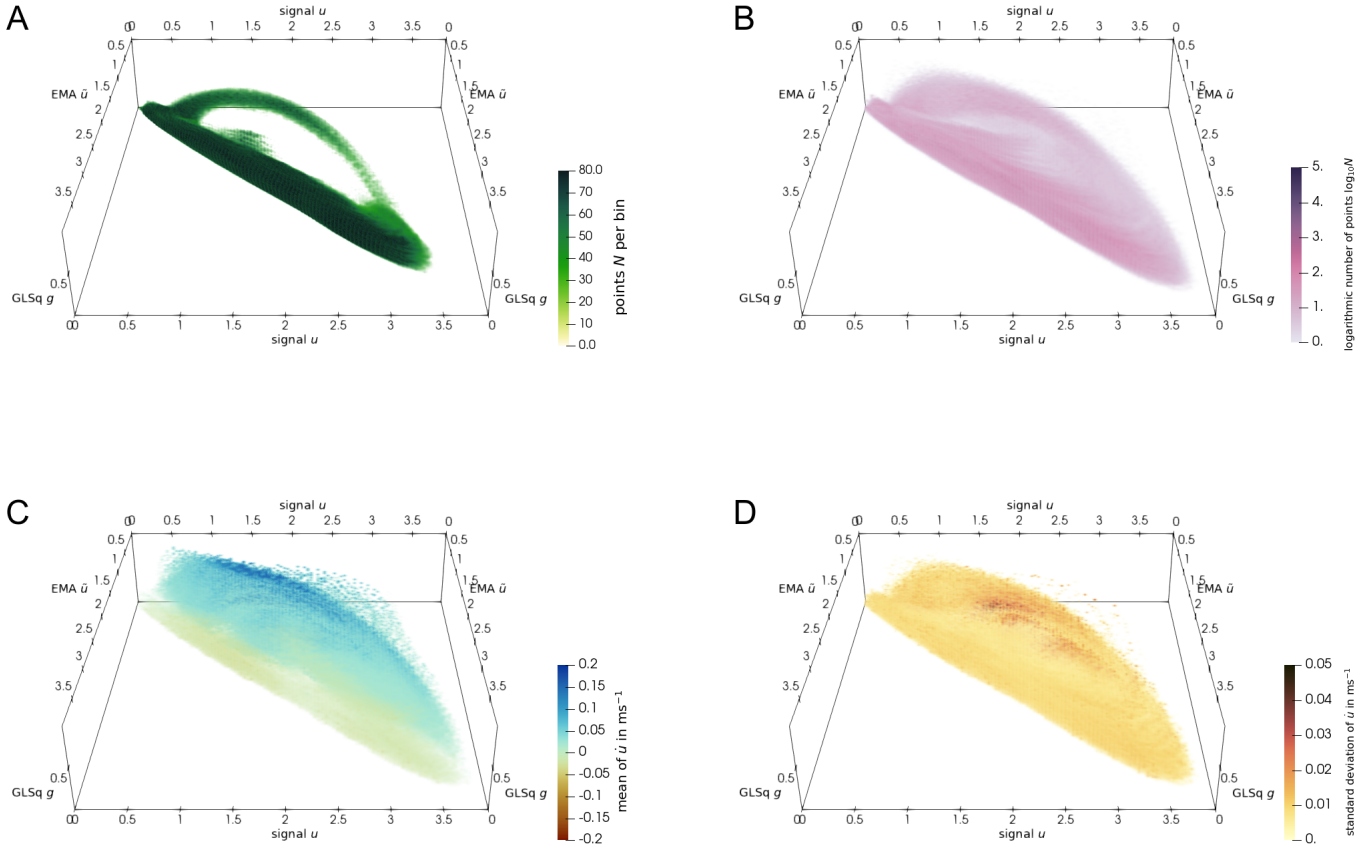

Figure 10: Visualisation of the 3D state space for the OVM data set, in the same style as in Fig. 4.

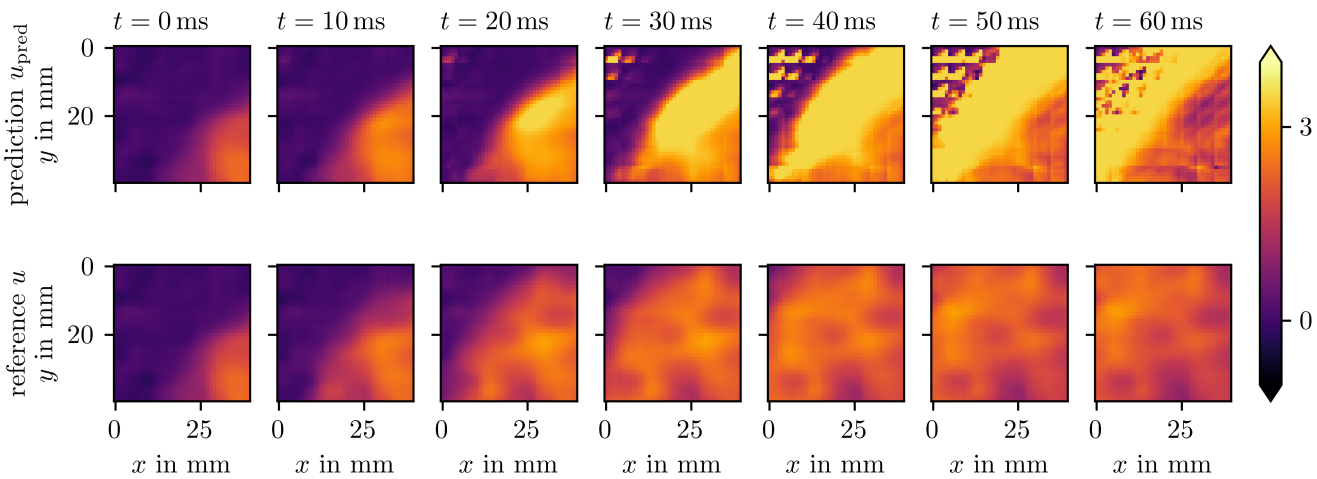

Figure 11: Prediction of the surrogate model with respect to each state over time (top row) compared to the reference solution (bottom row) for the OVM data set.

5 Conclusion

In this work, we presented a method to create very simple in-silico tissue models from in-vitro optical voltage mapping data and synthetic data that capture the basic behaviour of the excitable medium sufficiently. The main novelty of this method is its ability to extract all relevant information from just one state variable over space and time, such as a transmembrane voltage observed via optical voltage mapping.

6 Acknowledgments

We are grateful to Sven O. Dekker, Juan Zhang, and Antoine A.F. de Vries at LUMC for providing access to hiAM monolayers and the optical voltage mapping setup. DK is supported by KU Leuven grant GPUL/20/012. HD is supported by KU Leuven grant STG/19/007.

References

[1] Rubin R. Aliev and Alexander V. Panfilov. A simple two-variable model of cardiac excitation. *Chaos, Solitons & Fractals*, 7(3):293–301, March 1996. ISSN 09600779. doi: 10.1016/0960-0779(95)00089-5. URL https://linkinghub.elsevier.com/retrieve/pii/0960077995000895.

[2] Philip V Bayly, Bruce H KenKnight, Jack M Rogers, Russel E Hillsley, Raymond E Ideker, and William M Smith. Estimation of conduction velocity vector fields from epicardial mapping data. *IEEE transactions on biomedical engineering*, 45(5):563–571, 1998.

[3] Alfonso Bueno-Orovio, Elizabeth M. Cherry, and Flavio H. Fenton. Minimal model for human ventricular action potentials in tissue. *Journal of Theoretical Biology*, 253(3):544–560, aug 2008. doi: 10.1016/j.jtbi.2008.03.029. URL https://doi.org/10.1016%2Fj.jtbi.2008.03.029.

[4] R.H. Clayton, O. Bernus, E.M. Cherry, H. Dierckx, F.H. Fenton, L. Mirabella, A.V. Panfilov, F.B. Sachse, G. Seemann, and H. Zhang. Models of cardiac tissue electrophysiology: Progress, challenges and open questions. *Progress in Biophysics and Molecular Biology*, 104 (1-3):22–48, jan 2011. doi: 10.1016/j.pbiomolbio.2010.05.008. URL https://doi.org/10.1016%2Fj.pbiomolbio.2010.05.008.

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

[6] Flavio Fenton and Alain Karma. Vortex dynamics in three-dimensional continuous myocardium with fiber rotation: Filament instability and fibrillation. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 8(1):20–47, mar 1998. doi: 10.1063/1.166311. URL https://doi.org/10.1063%2F1.166311.

[7] Jeffrey J. Fox, Eberhard Bodenschatz, and Robert F. Gilmour. Period-Doubling Instability and Memory in Cardiac Tissue. *Physical Review Letters*, 89(13):138101, September 2002. ISSN 0031-9007, 1079-7114. doi: 10.1103/PhysRevLett.89.138101. URL https://link.aps.org/doi/10.1103/PhysRevLett.89.138101.

[8] Richard A. Gray, José Jalife, Alexandre V. Panfilov, William T. Baxter, Cándido Cabo, Jorge M. Davidenko, and Arkady M. Pertsov. Mechanisms of Cardiac Fibrillation. *Science*, 270(5239):1222–1223, November 1995. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.270.5239.1222. URL https://www.science.org/doi/10.1126/science.270.5239.1222.

[9] Niels Harlaar, Sven O. Dekker, Juan Zhang, Rebecca R. Snabel, Marieke W. Veldkamp, Arie O. Verkerk, Carla Cofiño Fabres, Verena Schwach, Lente J. S. Lerink, Mathilde R. Rivaud, Aat A. Mulder, Willem E. Corver, Marie José T. H. Goumans, Dobromir Dobrev, Robert J. M. Klautz, Martin J. Schalij, Gert Jan C. Veenstra, Robert Passier, Thomas J. van Brakel, Daniël A. Pijnappels, and Antoine A. F. de Vries. Conditional immortalization of human atrial myocytes for the generation of in vitro models of atrial fibrillation. *Nature Biomedical Engineering*, 6(4):389–402, jan 2022. doi: 10.1038/s41551-021-00827-5. URL https://doi.org/10.1038%2Fs41551-021-00827-5.

[10] Desmond Kabus, Louise Arno, Lore Leenknegt, Alexander V. Panfilov, and Hans Dierckx. Numerical methods for the detection of phase defect structures in excitable media. *PLOS ONE*, 17 (7):1–31, 07 2022. doi: 10.1371/journal.pone.0271351. URL https://doi.org/10.1371/journal.pone.0271351.

[11] Trine Krogh-Madsen, Peter Schaffer, Anne D. Skriver, Louise Kold Taylor, Brigitte Pelzmann, Bernd Koidl, and Michael R. Guevara. An ionic model for rhythmic activity in small clusters of embryonic chick ventricular cells. *American Journal of Physiology-Heart and Circulatory Physiology*, 289(1):H398–H413, July 2005. ISSN 0363-6135, 1522-1539. doi: 10.1152/ajpheart.00683.2004. URL https://www.physiology.org/doi/10.1152/ajpheart.00683.2004.

[12] Pawel Kuklik, Stef Zeemering, Bart Maesen, Jos Maessen, Harry J. Crijns, Sander Verheule, Anand N. Ganesan, and Ulrich Schotten. Reconstruction of Instantaneous Phase of Unipolar Atrial Contact Electrogram Using a Concept of Sinusoidal Recomposition and Hilbert Transform. *IEEE Transactions on Biomedical Engineering*, 62(1):296–302, January 2015. ISSN 0018-9294, 1558-2531. doi: 10.1109/TBME.2014.2350029. URL http://ieeexplore.ieee.org/document/6880777/.

[13] Rupamanjari Majumder, Wanchana Jangsangthong, Iolanda Feola, Dirk L Ypey, Daniël A Pijnappels, and Alexander V Panfilov. A mathematical model of neonatal rat atrial monolayers with constitutively active acetylcholine-mediated k+ current. *PLoS computational biology*, 12(6):e1004946, 2016.

[14] Christopher D. Marcotte and Roman O. Grigoriev. Dynamical mechanism of atrial fibrillation: A topological approach. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 27(9):093936, sep 2017. doi: 10.1063/1.5003259. URL https://doi.org/10.1063%2F1.5003259.

[15] Steven A. Niederer, Joost Lumens, and Natalia A. Trayanova. Computational models in cardiology. *NATURE REVIEWS CARDIOLOGY*, 16(2):100–111, 2019. ISSN 1759-5002. doi: 10.1038/s41569-018-0104-y.

[16] Michelangelo Paci, Jari Hyttinen, Katriina Aalto-Setälä, and Stefano Severi. Computational models of ventricular-and atrial-like human induced pluripotent stem cell derived cardiomyocytes. *Annals of biomedical engineering*, 41(11):2334–2348, 2013.

[17] Natalia A. Trayanova, Ashish N. Doshi, and Adityo Prakosa. How personalized heart modeling can help treatment of lethal arrhythmias: A focus on ventricular tachycardia ablation strategies in post-infarction patients. *WIREs Systems Biology and Medicine*, 12(3), May 2020. ISSN 1939-5094, 1939-005X. doi: 10.1002/wsbm.1477. URL https://onlinelibrary.wiley.com/doi/10.1002/wsbm.1477.

[18] Ning Wei, Yoichiro Mori, and Elena G. Tolkacheva. The role of short term memory and conduction velocity restitution in alternans formation. *Journal of Theoretical Biology*, 367:21–28, February 2015. ISSN 00225193. doi: 10.1016/j.jtbi.2014.11.014. URL https://linkinghub.elsevier.com/retrieve/pii/S0022519314006663.