

# Uncertainty quantification approaches for patient specific cardiac models

Richard Wilkinson, Sam Coveney, Jeremy Oakley, Steven Niederer,  
Richard Clayton

School of Maths and Statistics  
University of Sheffield

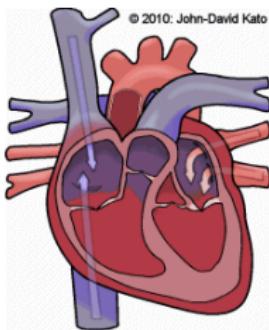
16 January 2019

# The heart

The heart is an electrical-mechanical pump, which contracts under electrical potential.

Focusing on the left atrium

- left atrium receives oxygenated blood from the lungs
- left atrium pumps this blood to the left ventricle
- left ventricle pumps this oxygenated blood to the body

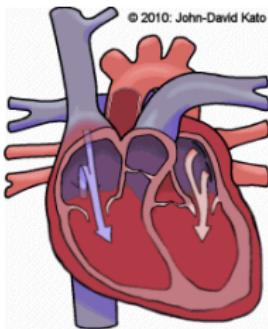


# The heart

The heart is an electrical-mechanical pump, which contracts under electrical potential.

Focusing on the left atrium

- left atrium receives oxygenated blood from the lungs
- left atrium pumps this blood to the left ventricle
- left ventricle pumps this oxygenated blood to the body

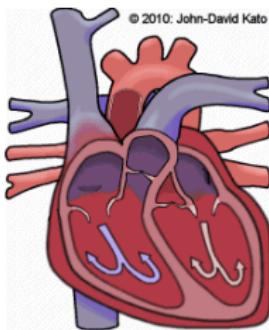


# The heart

The heart is an electrical-mechanical pump, which contracts under electrical potential.

Focusing on the left atrium

- left atrium receives oxygenated blood from the lungs
- left atrium pumps this blood to the left ventricle
- left ventricle pumps this oxygenated blood to the body

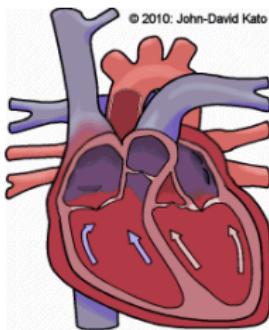


# The heart

The heart is an electrical-mechanical pump, which contracts under electrical potential.

Focusing on the left atrium

- left atrium receives oxygenated blood from the lungs
- left atrium pumps this blood to the left ventricle
- left ventricle pumps this oxygenated blood to the body

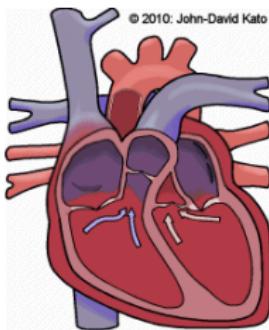


# The heart

The heart is an electrical-mechanical pump, which contracts under electrical potential.

Focusing on the left atrium

- left atrium receives oxygenated blood from the lungs
- left atrium pumps this blood to the left ventricle
- left ventricle pumps this oxygenated blood to the body

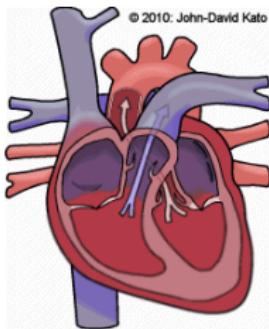


# The heart

The heart is an electrical-mechanical pump, which contracts under electrical potential.

Focusing on the left atrium

- left atrium receives oxygenated blood from the lungs
- left atrium pumps this blood to the left ventricle
- left ventricle pumps this oxygenated blood to the body

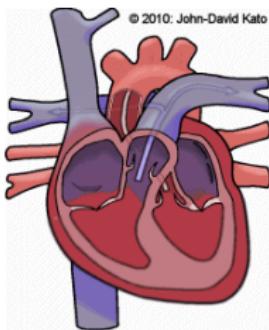


# The heart

The heart is an electrical-mechanical pump, which contracts under electrical potential.

Focusing on the left atrium

- left atrium receives oxygenated blood from the lungs
- left atrium pumps this blood to the left ventricle
- left ventricle pumps this oxygenated blood to the body

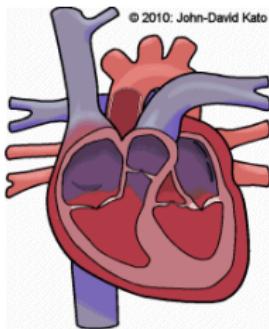


# The heart

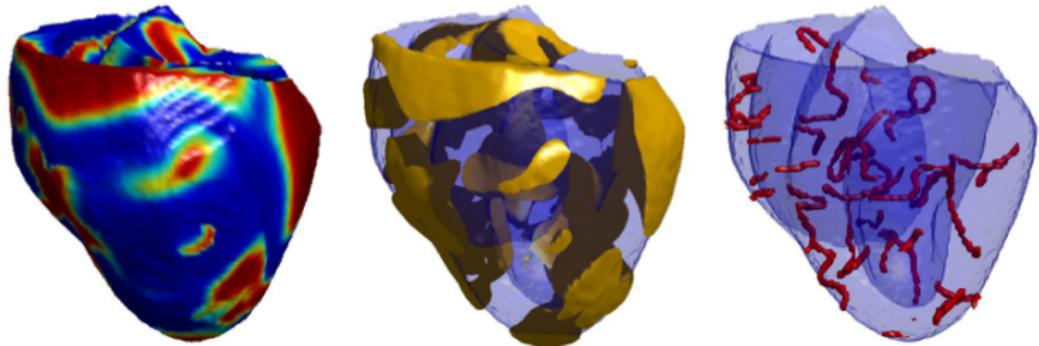
The heart is an electrical-mechanical pump, which contracts under electrical potential.

Focusing on the left atrium

- left atrium receives oxygenated blood from the lungs
- left atrium pumps this blood to the left ventricle
- left ventricle pumps this oxygenated blood to the body

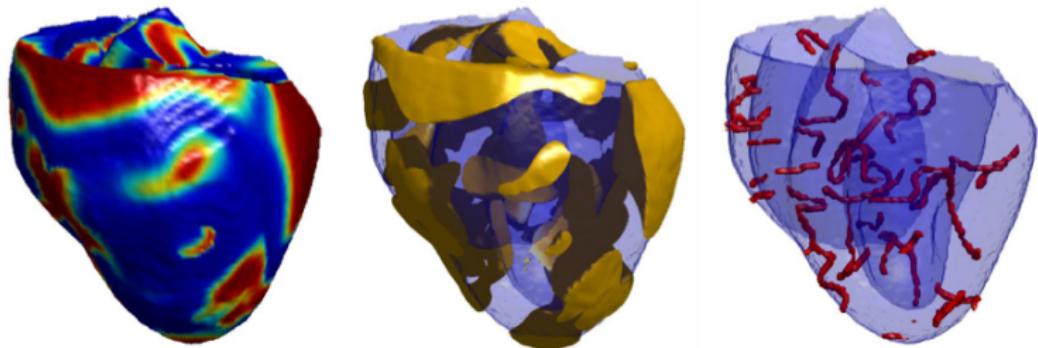


## Atrial fibrillation



Atrial fibrillation (AF) - rapid and uncoordinated electrical activation (arrhythmia) leading to poor mechanical function.

## Atrial fibrillation



Atrial fibrillation (AF) - rapid and uncoordinated electrical activation (arrhythmia) leading to poor mechanical function.

- Affects around 610,000 people in UK.
- Catheter ablation removes/isolates pathological tissue that sustain/initiate AF.
- 40% of patients subsequently experience atrial tachycardia (AT).

## Patient Specific Cardiac Models

Aim: predict the probability an AF patient will develop AT following ablation, and then treat for both in a single procedure.

- Each intervention carries a 6% risk of a major complication, and costs  $\sim \text{£}8000$ .

## Patient Specific Cardiac Models

Aim: predict the probability an AF patient will develop AT following ablation, and then treat for both in a single procedure.

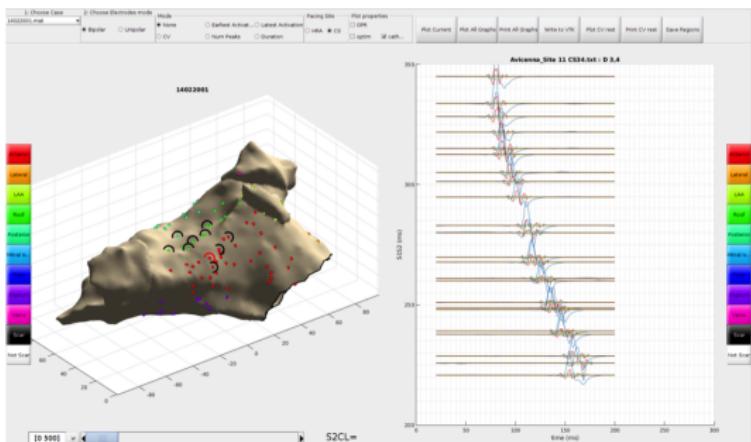
- Each intervention carries a 6% risk of a major complication, and costs ~£8000.

Personalised biophysical models have the ability to predict patient response to treatment

- cardiac models at forefront of personalised modelling
- models are currently deterministic - simulating a single outcome.
- clinical diagnosis is rarely definitive - we need to account for uncertainties.
  - ▶ decisions are not always made on basis of most likely outcome:  
consider costs and benefits across all potential outcomes weighted by their probability.  
e.g. if patient has 30% chance of complication - this should influence decision making.

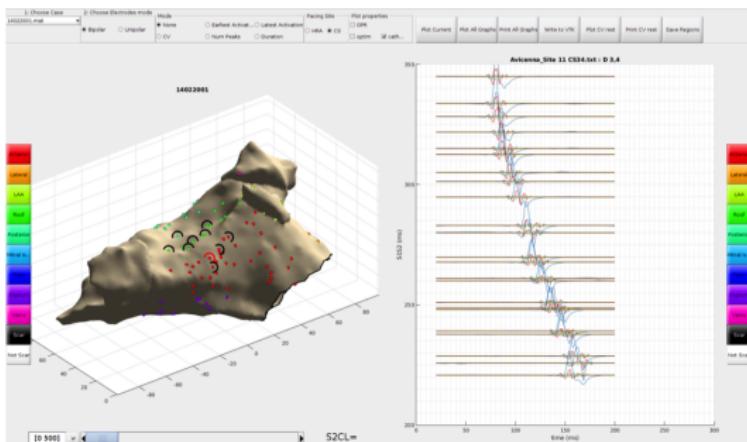
# Tissue electrophysiology simulation

Our primary measurements are surface measurements of electrophysiology.



# Tissue electrophysiology simulation

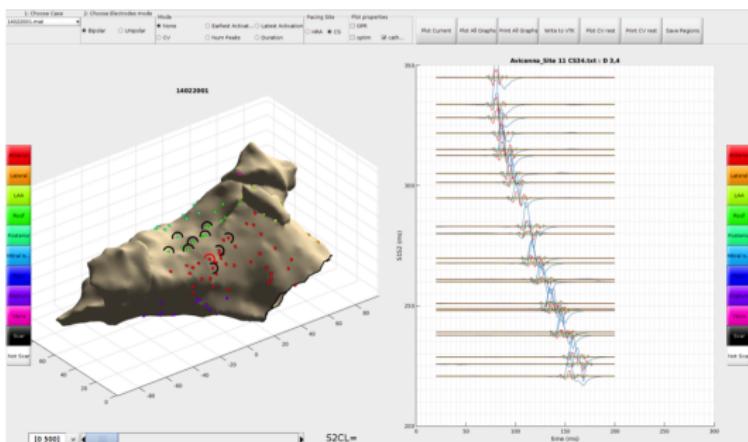
Our primary measurements are surface measurements of electrophysiology.



We model tissue electrical activation using the monodomain equation on a shell model of LA with local activation given by the Mitchell-Schaeffer ionic model.

# Tissue electrophysiology simulation

Our primary measurements are surface measurements of electrophysiology.



We model tissue electrical activation using the monodomain equation on a shell model of LA with local activation given by the Mitchell-Schaeffer ionic model.

Accurate predictions require patient specific models, but clinical data is sparse and noisy.

We need to

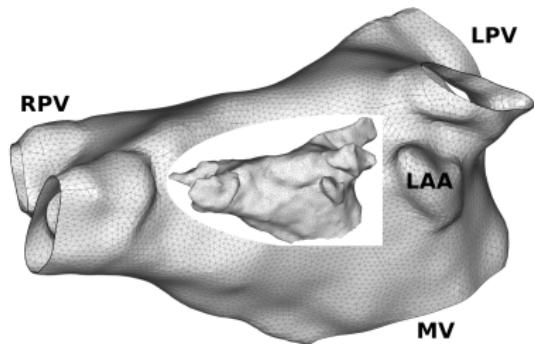
- Estimate conduction velocity on the atrium using ECG measurements
- Infer tissues properties, including regions of fibrotic material
- Predict AT pathways
- Aid clinical decision making (accounting for uncertainty)

# Uncertainty Quantification (UQ) for patient specific models

If personalised cardiac models are to be used to guide clinical decision making, we need quantification of the uncertainties and probabilities of likely outcomes.

Sources of uncertainty:

- Inferring activation time from ECGs
- Shape of left atrium
- Location of electrodes
- Estimation of local tissue properties (5 parameters per mesh point)
- Inference of AT pathways
- Effect of any interventions
- ...

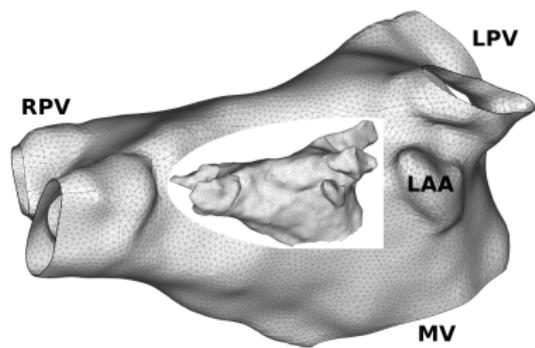


# Uncertainty Quantification (UQ) for patient specific models

If personalised cardiac models are to be used to guide clinical decision making, we need quantification of the uncertainties and probabilities of likely outcomes.

Sources of uncertainty:

- Inferring activation time from ECGs
- Shape of left atrium
- Location of electrodes
- Estimation of local tissue properties (5 parameters per mesh point)
- Inference of AT pathways
- Effect of any interventions
- ...



Do the data contain enough info to allow us to make informative robust predictions? How big an effect does source of uncertainty have, i.e., which matter? Do we need to model them all? Can some be ignored?

# What is Uncertainty Quantification (UQ)

Uncertainty Quantification (UQ)  $\equiv$  statistics with complex models

- A 'complex model' is one that is expensive to evaluate.

# What is Uncertainty Quantification (UQ)

Uncertainty Quantification (UQ)  $\equiv$  statistics with complex models

- A 'complex model' is one that is expensive to evaluate.

Typical tasks

- Uncertainty propagation

$$\theta \sim \pi(\cdot), \text{ and } Y = f(\theta), \quad \text{then } Y \sim ?$$

# What is Uncertainty Quantification (UQ)

Uncertainty Quantification (UQ)  $\equiv$  statistics with complex models

- A 'complex model' is one that is expensive to evaluate.

Typical tasks

- Uncertainty propagation

$$\theta \sim \pi(\cdot), \text{ and } Y = f(\theta), \quad \text{then } Y \sim ?$$

- Parameter estimation

$$\text{Observe } Y = f(\theta) + \epsilon, \text{ then } \theta = ?$$

# What is Uncertainty Quantification (UQ)

Uncertainty Quantification (UQ)  $\equiv$  statistics with complex models

- A 'complex model' is one that is expensive to evaluate.

Typical tasks

- Uncertainty propagation

$$\theta \sim \pi(\cdot), \text{ and } Y = f(\theta), \quad \text{then } Y \sim ?$$

- Parameter estimation

$$\text{Observe } Y = f(\theta) + \epsilon, \text{ then } \theta = ?$$

- Sensitivity analysis:  $\theta = (\theta_1, \dots, \theta_p)^\top \sim \pi(\cdot)$ . If we can measure one component of  $\theta$ , which should we choose to minimize  $\text{Var}(Y)$ ?
- Decision making: my model is uncertain, the parameters are uncertain, the data is noisy, but I need to make a decision...

# What is Uncertainty Quantification (UQ)

Uncertainty Quantification (UQ)  $\equiv$  statistics with complex models

- A 'complex model' is one that is expensive to evaluate.

Typical tasks

- Uncertainty propagation

$$\theta \sim \pi(\cdot), \text{ and } Y = f(\theta), \quad \text{then } Y \sim ?$$

- Parameter estimation

$$\text{Observe } Y = f(\theta) + \epsilon, \text{ then } \theta = ?$$

- Sensitivity analysis:  $\theta = (\theta_1, \dots, \theta_p)^\top \sim \pi(\cdot)$ . If we can measure one component of  $\theta$ , which should we choose to minimize  $\text{Var}(Y)$ ?
- Decision making: my model is uncertain, the parameters are uncertain, the data is noisy, but I need to make a decision...

UQ should be a synergy between statistics, applied mathematics and domain sciences

## Tools

The Bayesian approach to the inverse problem: represent all uncertainties as probability distributions

$$\pi(\theta|D) = \frac{\pi(D|\theta)\pi(\theta)}{\pi(D)}$$

# Tools

The Bayesian approach to the inverse problem: represent all uncertainties as probability distributions

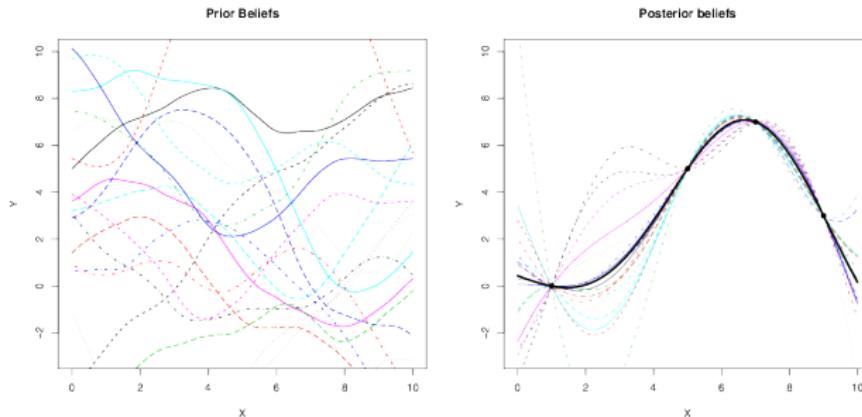
$$\pi(\theta|D) = \frac{\pi(D|\theta)\pi(\theta)}{\pi(D)}$$

Primary methodological tools

- Gaussian process interpolation
- Approximate Bayesian Computation (ABC) - 'likelihood-free' inference
- Machine learning models (deep neural nets etc)
- Basis function expansions, dimension reduction, reduced order models etc.

# Gaussian processes

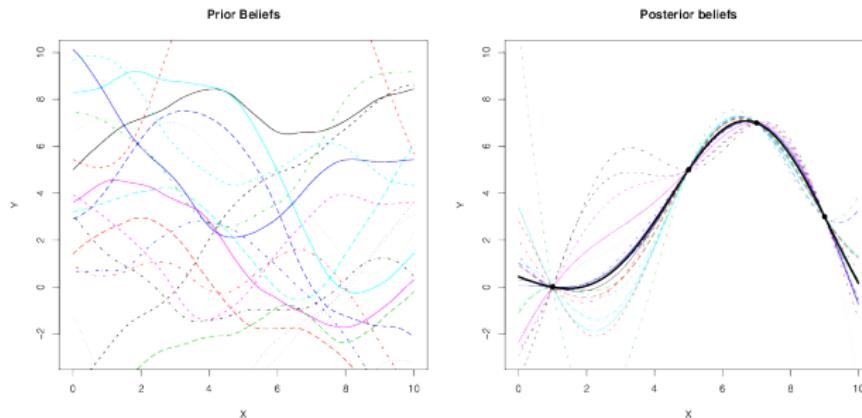
Primarily based around the use of Gaussian processes



Think of these as probabilistic models of functions.

# Gaussian processes

Primarily based around the use of Gaussian processes



Think of these as probabilistic models of functions.

A GP is a random process indexed by  $x \in \mathcal{X}$  say, such that for every finite set of indices,  $x_1, \dots, x_n$ ,

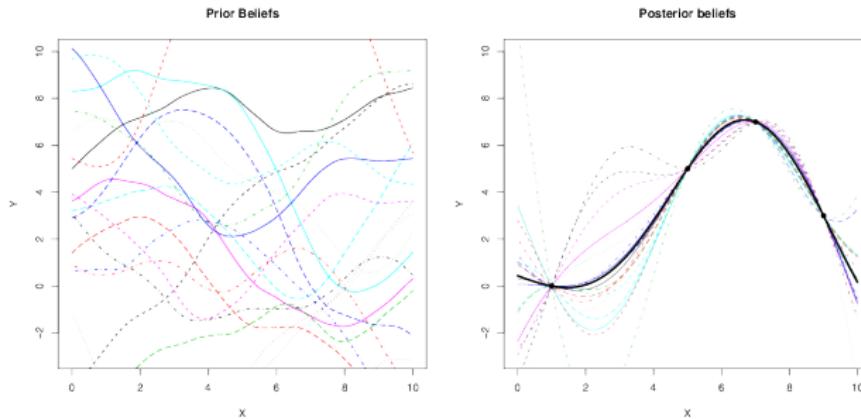
$$\mathbf{f} = (f(x_1), \dots, f(x_n))$$

has a multivariate Gaussian distribution.

Key choice is the covariance/kernel function  $k(x, x') = \text{Cov}(f(x), f(x'))$

# Gaussian processes

Primarily based around the use of Gaussian processes



Think of these as probabilistic models of functions.

A GP is a random process indexed by  $x \in \mathcal{X}$  say, such that for every finite set of indices,  $x_1, \dots, x_n$ ,

$$\mathbf{f} = (f(x_1), \dots, f(x_n))$$

has a multivariate Gaussian distribution.

Key choice is the covariance/kernel function  $k(x, x') = \text{Cov}(f(x), f(x'))$

Why would we want to use this very restricted model?

## Answer 1

Class of models is closed under various operations.

## Answer 1

Class of models is closed under various operations.

- Closed under addition

$$f_1(\cdot), f_2(\cdot) \sim GP \quad \text{then} \quad (f_1 + f_2)(\cdot) \sim GP$$

## Answer 1

Class of models is closed under various operations.

- Closed under addition

$$f_1(\cdot), f_2(\cdot) \sim GP \quad \text{then} \quad (f_1 + f_2)(\cdot) \sim GP$$

- Closed under Bayesian conditioning, i.e., if we observe

$$\mathbf{D} = (f(x_1), \dots, f(x_n))$$

then

$$f | D \sim GP$$

but with updated mean and covariance functions.

## Answer 1

Class of models is closed under various operations.

- Closed under addition

$$f_1(\cdot), f_2(\cdot) \sim GP \quad \text{then} \quad (f_1 + f_2)(\cdot) \sim GP$$

- Closed under Bayesian conditioning, i.e., if we observe

$$\mathbf{D} = (f(x_1), \dots, f(x_n))$$

then

$$f|D \sim GP$$

but with updated mean and covariance functions.

- Closed under any linear operation. If  $\mathcal{L}$  is a linear operator, then

$$\mathcal{L}f \sim GP(\mathcal{L}m, \mathcal{L}k\mathcal{L}^\top)$$

e.g.  $\frac{df}{dx}$ ,  $\int f(x)dx$ ,  $Af$  are all GPs. Can also analytically condition on  $\mathcal{L}f = 0$ , e.g. incompressible flow  $\nabla \cdot \nabla f = 0$

## Answer 2: non-parametric/kernel regression

- Linear regression  $y = x^\top \beta + \epsilon$  can be written solely in terms of inner products  $x^\top x$ .

$$\begin{aligned}\hat{\beta} &= \arg \min ||y - X\beta||_2^2 + \sigma^2 ||\beta||_2^2 \\ &= (X^\top X + \sigma^2 I) X^\top y \\ &= X^\top (X X^\top + \sigma^2 I)^{-1} y \quad (\text{the dual form})\end{aligned}$$

## Answer 2: non-parametric/kernel regression

- Linear regression  $y = x^\top \beta + \epsilon$  can be written solely in terms of inner products  $x^\top x$ .

$$\hat{\beta} = \arg \min ||y - X\beta||_2^2 + \sigma^2 ||\beta||_2^2$$

$$= (X^\top X + \sigma^2 I) X^\top y$$

$$= X^\top (X X^\top + \sigma^2 I)^{-1} y \quad (\text{the dual form})$$

- We know that we can replace  $x$  by a feature vector in linear regression, e.g.,  $\phi(x) = (1 \ x \ x^2 \ \cos(x))^\top$  etc.

## Answer 2: non-parametric/kernel regression

- Linear regression  $y = x^\top \beta + \epsilon$  can be written solely in terms of inner products  $x^\top x$ .

$$\hat{\beta} = \arg \min ||y - X\beta||_2^2 + \sigma^2 ||\beta||_2^2$$

$$= (X^\top X + \sigma^2 I) X^\top y$$

$$= X^\top (X X^\top + \sigma^2 I)^{-1} y \quad (\text{the dual form})$$

- We know that we can replace  $x$  by a feature vector in linear regression, e.g.,  $\phi(x) = (1 \ x \ x^2 \ \cos(x))^\top$  etc.
- For some features, inner product is equivalent to evaluating a kernel

$$\phi(x)^\top \phi(x') \equiv k(x, x')$$

where  $k : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$  is a semi-positive definite function.

## Answer 2: non-parametric/kernel regression

- Linear regression  $y = x^\top \beta + \epsilon$  can be written solely in terms of inner products  $x^\top x$ .

$$\hat{\beta} = \arg \min ||y - X\beta||_2^2 + \sigma^2 ||\beta||_2^2$$

$$= (X^\top X + \sigma^2 I) X^\top y$$

$$= X^\top (X X^\top + \sigma^2 I)^{-1} y \quad (\text{the dual form})$$

- We know that we can replace  $x$  by a feature vector in linear regression, e.g.,  $\phi(x) = (1 \ x \ x^2 \ \cos(x))^\top$  etc.
- For some features, inner product is equivalent to evaluating a kernel

$$\phi(x)^\top \phi(x') \equiv k(x, x')$$

where  $k : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$  is a semi-positive definite function.

**Kernel trick:** lift  $x$  into infinite dimensional feature space by replacing inner products  $x^\top x'$  by  $k(x, x')$ , but never evaluate these features, only the  $n \times n$  kernel matrix.

$$\hat{y}' = m(x') = \sum_{i=1}^n \alpha_i k(x, x_i)$$

Generally, we don't think about features, we just choose a kernel. But choosing a kernel is implicitly choosing features, and our model only includes functions that are linear combinations of this set of features (the Reproducing Kernel Hilbert Space (RKHS) of  $k$ ).

Generally, we don't think about features, we just choose a kernel. But choosing a kernel is implicitly choosing features, and our model only includes functions that are linear combinations of this set of features (the Reproducing Kernel Hilbert Space (RKHS) of  $k$ ).

**Example:** If (modulo some detail)

$$\phi(x) = \left( e^{-\frac{(x-c_1)^2}{2\lambda^2}}, \dots, e^{-\frac{(x-c_N)^2}{2\lambda^2}} \right)$$

then as  $N \rightarrow \infty$  then

$$\phi(x)^\top \phi(x) = \exp\left(-\frac{(x-x')^2}{2\lambda^2}\right)$$

Generally, we don't think about features, we just choose a kernel. But choosing a kernel is implicitly choosing features, and our model only includes functions that are linear combinations of this set of features (the Reproducing Kernel Hilbert Space (RKHS) of  $k$ ).

**Example:** If (modulo some detail)

$$\phi(x) = \left( e^{-\frac{(x-c_1)^2}{2\lambda^2}}, \dots, e^{-\frac{(x-c_N)^2}{2\lambda^2}} \right)$$

then as  $N \rightarrow \infty$  then

$$\phi(x)^\top \phi(x) = \exp\left(-\frac{(x-x')^2}{2\lambda^2}\right)$$

Although our simulator may not lie in the RKHS defined by  $k$ , this space is much richer than any parametric regression model (and can be dense in some sets of continuous bounded functions), and is thus more likely to contain an element close to the simulator than any class of models that contains only a finite number of features.

## Answer 3: Naturalness of GP framework

Why use **Gaussian** processes as non-parametric models?

## Answer 3: Naturalness of GP framework

Why use **Gaussian** processes as non-parametric models?

One answer might come from Bayes linear methods<sup>1</sup>.

If we only knew the expectation and variance of some random variables,  $X$  and  $Y$ , then how should we best do statistics?

---

<sup>1</sup>statistics without probability

## Answer 3: Naturalness of GP framework

Why use **Gaussian** processes as non-parametric models?

One answer might come from Bayes linear methods<sup>1</sup>.

If we only knew the expectation and variance of some random variables,  $X$  and  $Y$ , then how should we best do statistics?

It can been shown, that the best second-order inference we can do to update our beliefs about  $X$  given  $Y$  is

$$\mathbb{E}(X|Y) = \mathbb{E}(X) + \text{Cov}(X, Y)\text{Var}(Y)^{-1}(Y - \mathbb{E}(Y))$$

which is exactly the Gaussian process update for the posterior mean.

So GPs are in some sense very natural approaches.

---

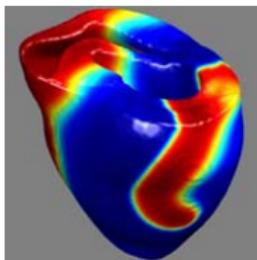
<sup>1</sup>statistics without probability

## Work flow

- MRI to build patient specific left atrium mesh, segmented to identify regions of fibrosis.
- Electrophysiology study to learn local activation map,
- Interpolate to entire LA, allowing conduction velocity and restitution curves to be calculated
- Local tissue properties estimated giving personalized probabilistic model
- Predict AT pathways
- Clinical decision support

## Local activation time

Picture of electrical wave spiraling around ventricle (fibrillation).  
Red: 'active' cardiac tissue Blue: 'inactive' cardiac tissue



We want to know the time of arrival of the 'electrical wave front'.  
We call this the Local Activation Time (LAT).

- An electrophysiology (EP) study performed by inserting catheters and electrodes on LA surface, to measure electrical activity.

## Local activation time

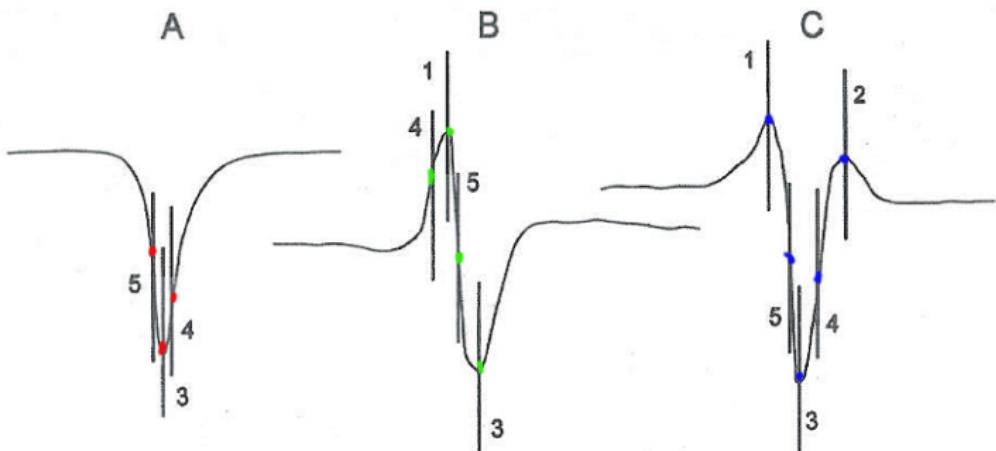


Figure 4. Example of bipolar electrogram configurations and associated landmarks used for activation time analysis. A: uniphasic; B: biphasic; C: triphasic. 1: peak positive voltage; 2: secondary peak positive voltage; 3: peak negative voltage; 4: maximum positive derivative; 5: maximum negative derivative. See text for details.

The literature is unclear about how LAT should be inferred from a bipolar electrogram.

Some methods more robust than others, and few include uncertainty.

## Local activation time

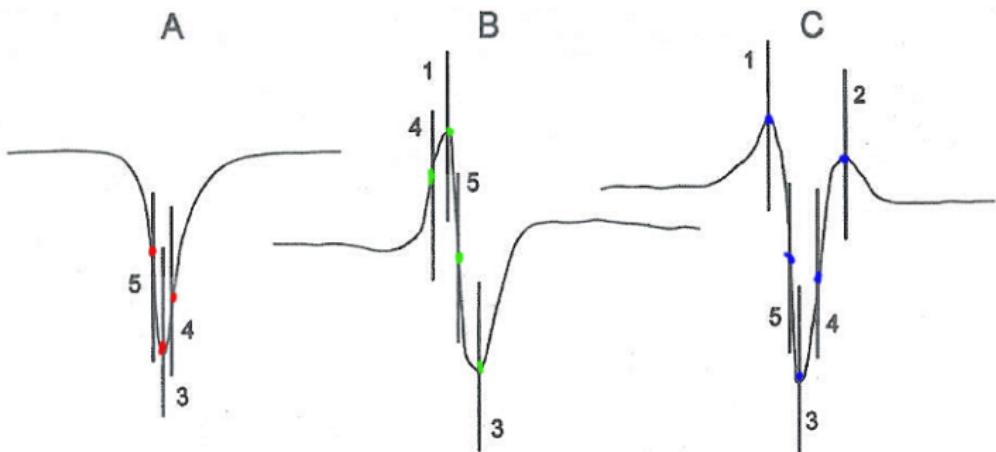
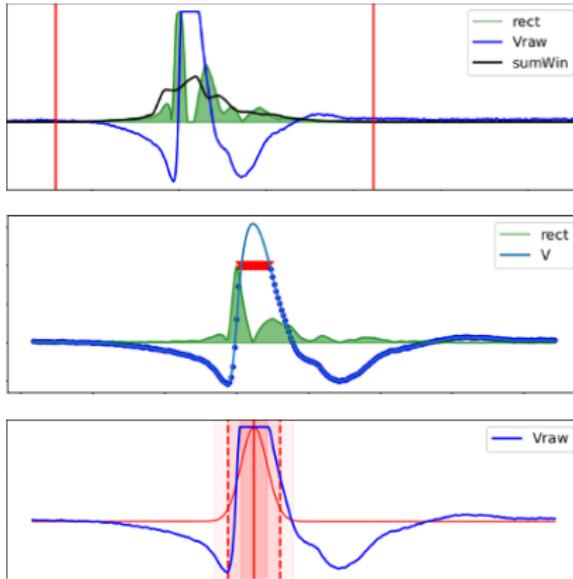


Figure 4. Example of bipolar electrogram configurations and associated landmarks used for activation time analysis. A: uniphasic; B: biphasic; C: triphasic. 1: peak positive voltage; 2: secondary peak positive voltage; 3: peak negative voltage; 4: maximum positive derivative; 5: maximum negative derivative. See text for details.

The literature is unclear about how LAT should be inferred from a bipolar electrogram.

Some methods more robust than others, and few include uncertainty.

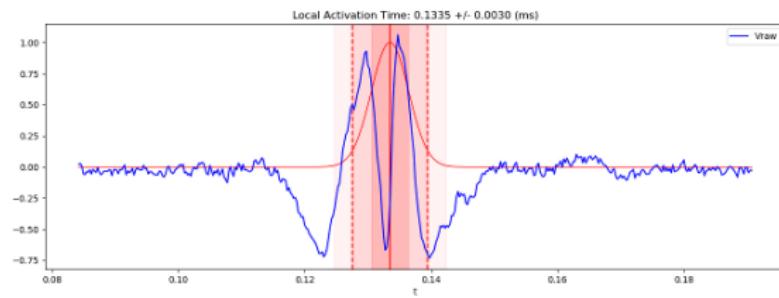
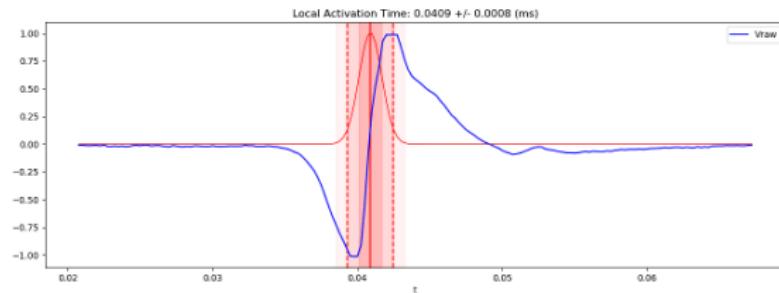
## LAT uncertainty from ECG



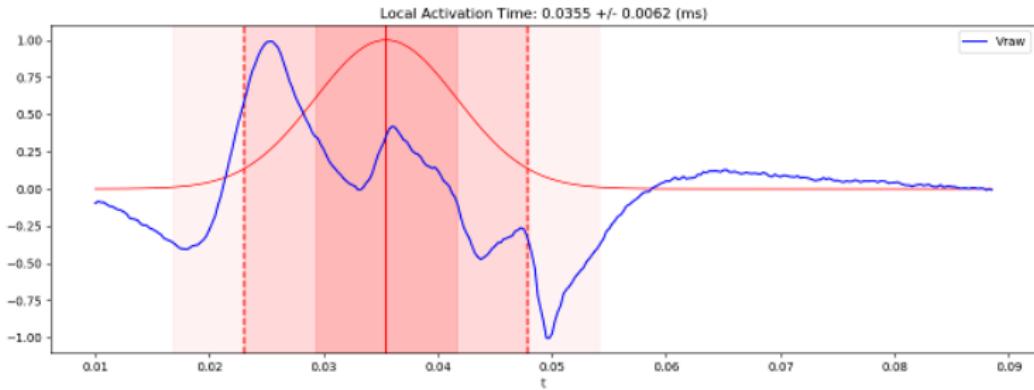
LAT:  $t_{50}$  : 50% of cumulative area under rectified (green) curve

- fit GP
- Impute missing values
- Estimate  $t_{50}$ , or maximum, or steepest gradient ( $f \sim GP$  then  $f' \sim GP$ )....

# Estimate uncertainty



$$2\sigma_{EGM} = (t_{75} - t_{25})/2$$

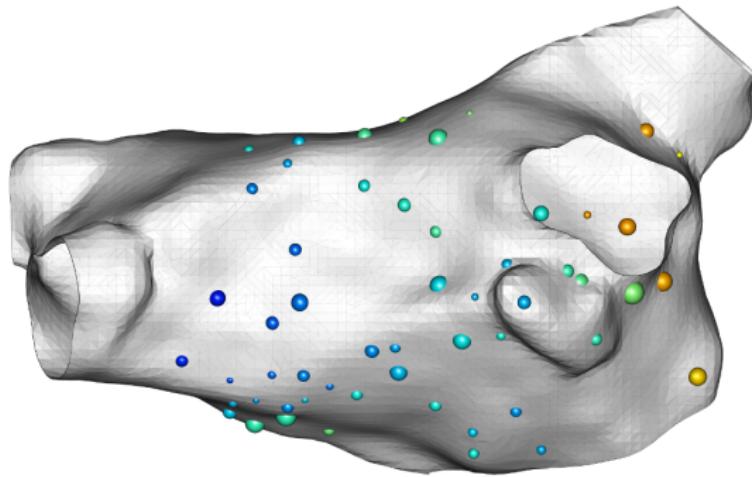


This electrogram is noisy and the signal unclear, but it is often all we have at some location.

- It tells us something, but precision is low.

## Interpolation

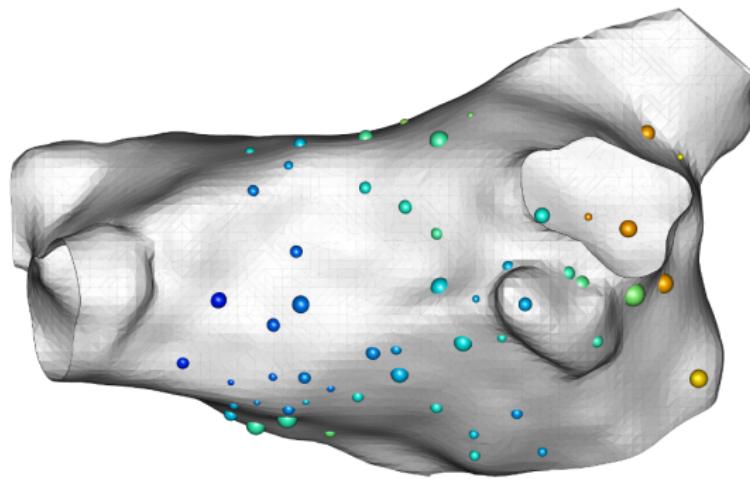
Typically, only able to measure LAT a small number ( $\sim 10s$ ) of locations on the atrium.



How can we interpolate to other locations?

## Interpolation

Typically, only able to measure LAT a small number ( $\sim 10s$ ) of locations on the atrium.



How can we interpolate to other locations?

$$t_{50}(x) = \text{LAT}(x) + \epsilon_{EGM} + \epsilon_{position}$$

$$\epsilon_{EGM} \sim \mathcal{N}(0, \sigma_{EGM}^2)$$

$$\epsilon_{position} \sim \mathcal{N}(0, \sigma_{position}^2)$$

# GP interpolations

We want to model

$$LAT(x) \sim GP(m(x), k(x, x'))$$

but standard approaches to Gaussian process interpolation won't work on complex manifolds such as the atrium

- Typically we define covariance function to be a function of the Euclidean distance between two points i.e.  $k(x, x') \equiv k(\|x - x'\|_2)$ ,

We want the interpolation to take into account distance on the manifold travelled by electrical wave.

- Defining a valid positive definite covariance function on the manifold is hard!

## INLA-SPDE approach: Lindgren, Rue, Lindstrom (2011)

Coveney et al 2019

Instead of a GP model (formulated in terms of a covariance function) we use a Gaussian Markov random field (GMRF) which allows the model to be specified via a sparse precision matrix (allows solution in  $O(n^{3/2})$  instead of  $O(n^3)$ ).

## INLA-SPDE approach: Lindgren, Rue, Lindstrom (2011)

Coveney et al 2019

Instead of a GP model (formulated in terms of a covariance function) we use a Gaussian Markov random field (GMRF) which allows the model to be specified via a sparse precision matrix (allows solution in  $O(n^{3/2})$  instead of  $O(n^3)$ ).

- GPs with a Matern covariance function are stationary solutions to explicit linear stochastic partial differential equation.
- Allows us to solve GPs using the machinery of finite element methods
- Makes it easy to work on irregular domains - we now only need to solve a SPDE on a particular triangular mesh for any given problem.

## INLA-SPDE approach: Lindgren, Rue, Lindstrom (2011)

Coveney et al 2019

Instead of a GP model (formulated in terms of a covariance function) we use a Gaussian Markov random field (GMRF) which allows the model to be specified via a sparse precision matrix (allows solution in  $O(n^{3/2})$  instead of  $O(n^3)$ ).

- GPs with a Matern covariance function are stationary solutions to explicit linear stochastic partial differential equation.
- Allows us to solve GPs using the machinery of finite element methods
- Makes it easy to work on irregular domains - we now only need to solve a SPDE on a particular triangular mesh for any given problem.

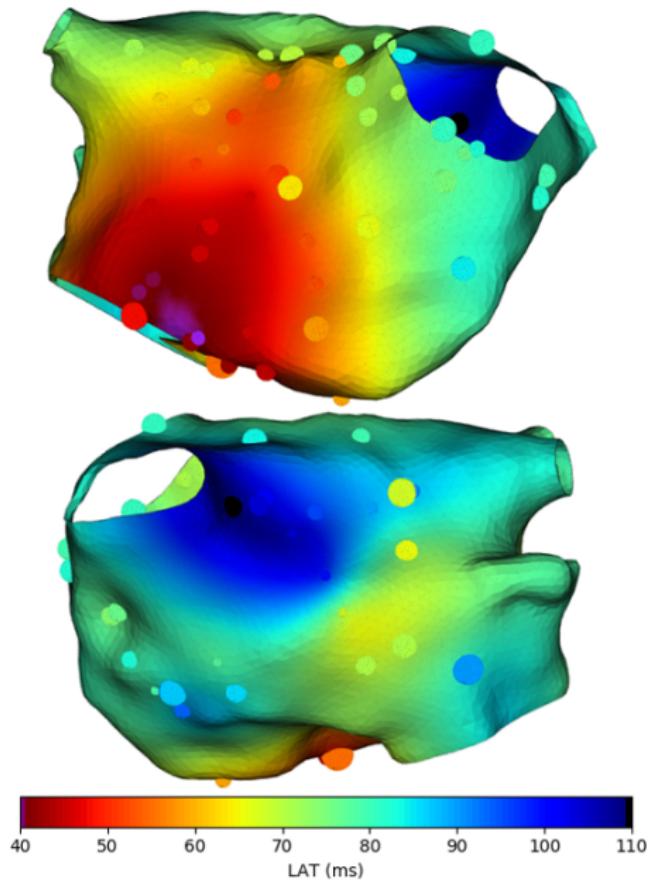
$$LAT(x) = \sum_{k=1}^n w_k \psi_k(x)$$

with  $w_k \sim N(0, \tilde{Q}^{-1})$  where  $\tilde{Q}$  is sparse. Note

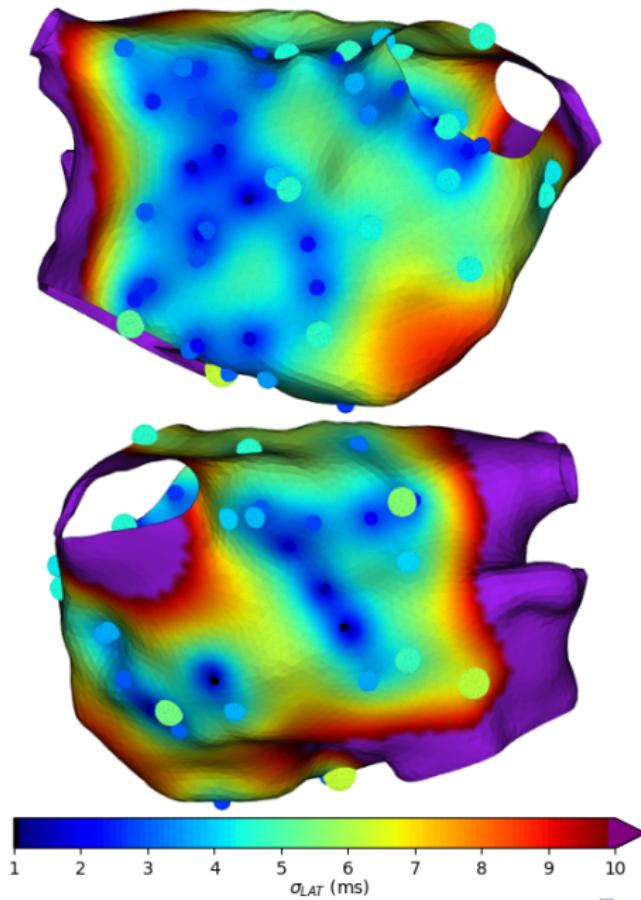
$$LAT(\cdot) \sim GP(0, Q^{-1})$$

for some  $Q$

## Results - mean



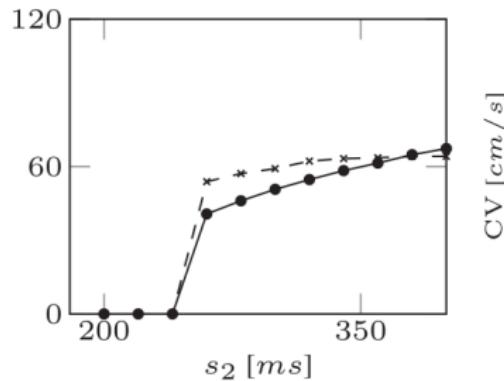
## Results - standard deviation



## S1-S2 interpolation

The **electrical restitution curve** describes the recovery of action potential duration as a function of the interbeat interval (the diastolic interval).

- During an EP study the heart is 'paced' at a regular S1 interval.
- Premature interbeats introduced at interval S2
- As the S2 interval shortens the heart tissue will eventually cease to recover in time to activate for both beats



## S1-S2 interpolation

The EP study measures activation time at  $\sim 30$  locations and  $\sim 10$  S2 intervals. Can use INLA-SPDE approach to interpolate LAT at the locations for a given S2 value.

How can we borrow strength from different S2 intervals to improve the interpolation?

## S1-S2 interpolation

The EP study measures activation time at  $\sim 30$  locations and  $\sim 10$  S2 intervals. Can use INLA-SPDE approach to interpolate LAT at the locations for a given S2 value.

How can we borrow strength from different S2 intervals to improve the interpolation?

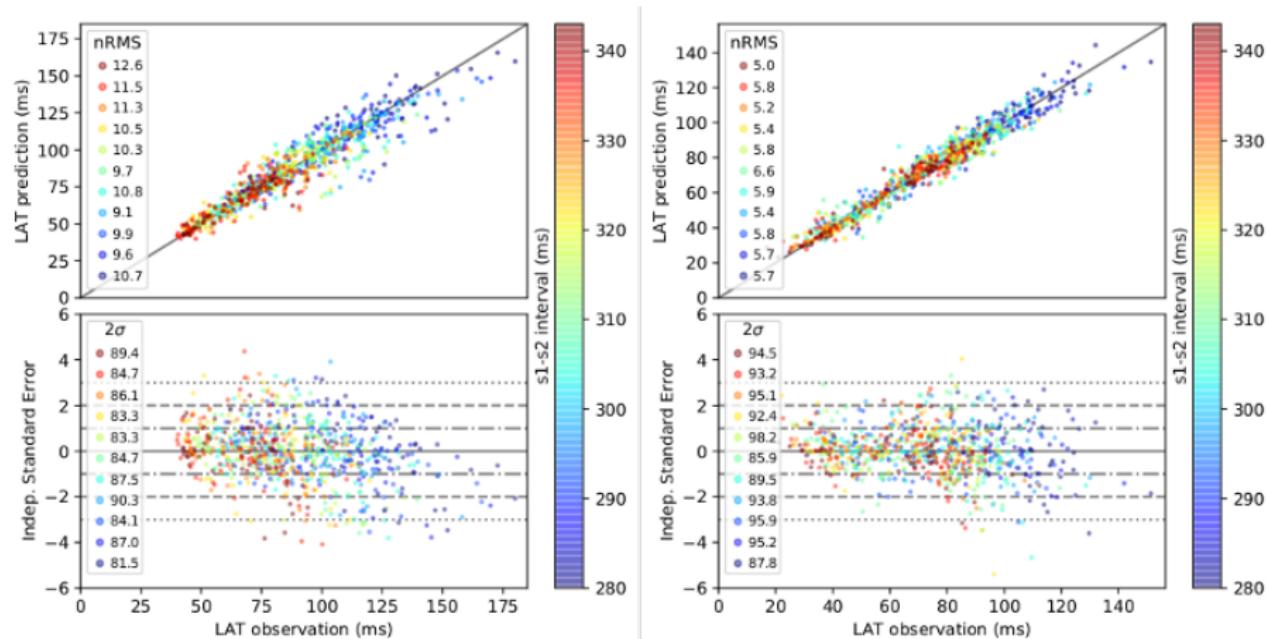
Simplest way is to add S2 as an input, and assume an AR(1) relationship between  $LAT(x, S2_{i+1})$  and  $LAT(x, S2_i)$

$$LAT(x, S2_{i+1}) \sim N(\rho LAT(x, S2_i), (1 - \rho^2)Q^{-1})$$

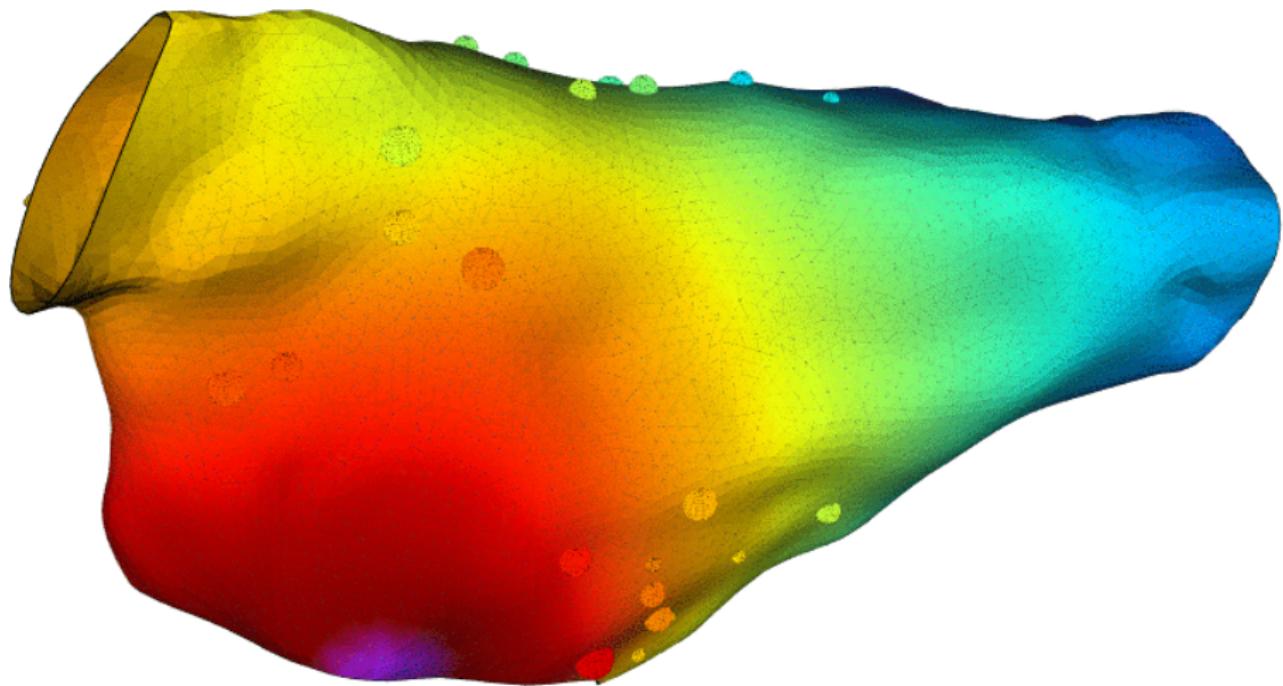
or more precisely

$$LAT(x, S2) \sim GP(0, Q_{S2}^{-1} \otimes Q^{-1})$$

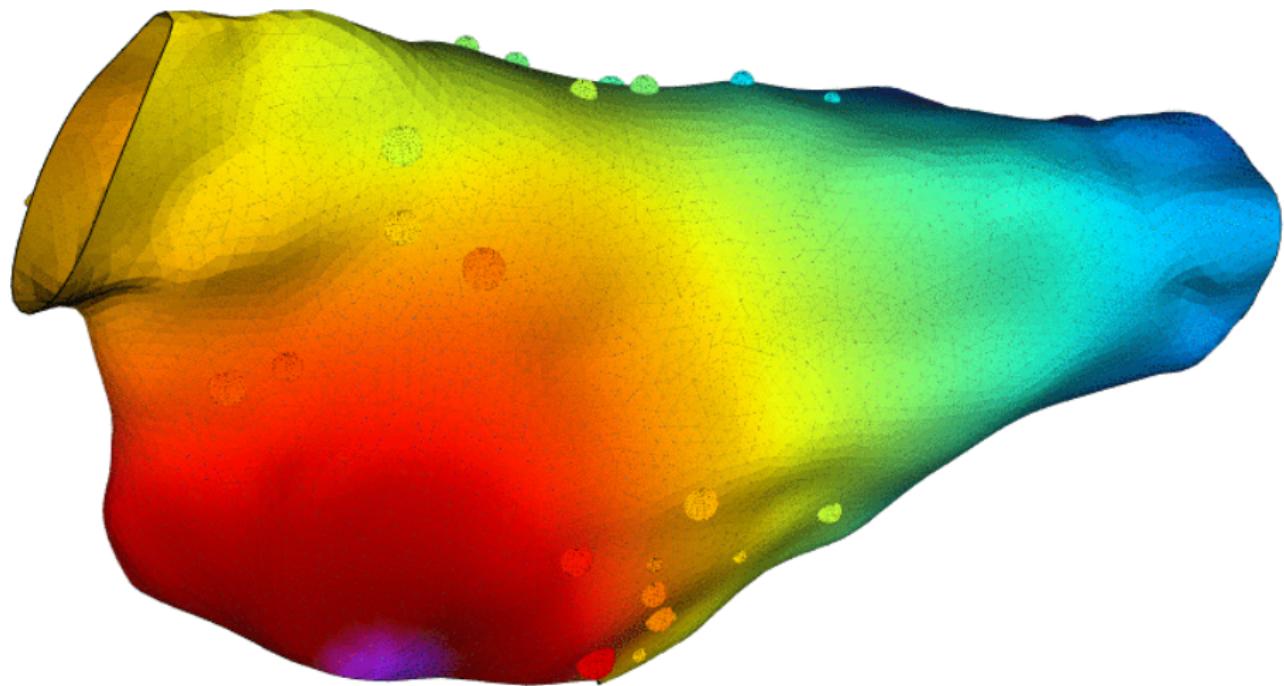
# Results: Cross validation



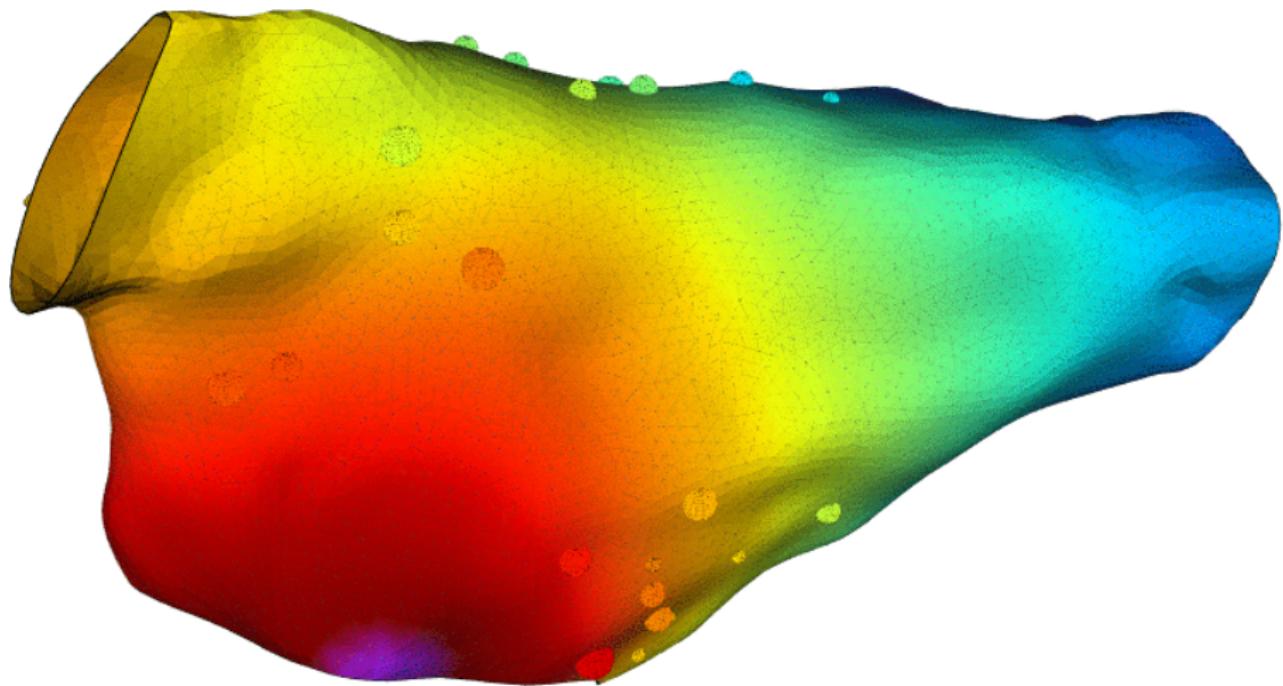
## Random samples



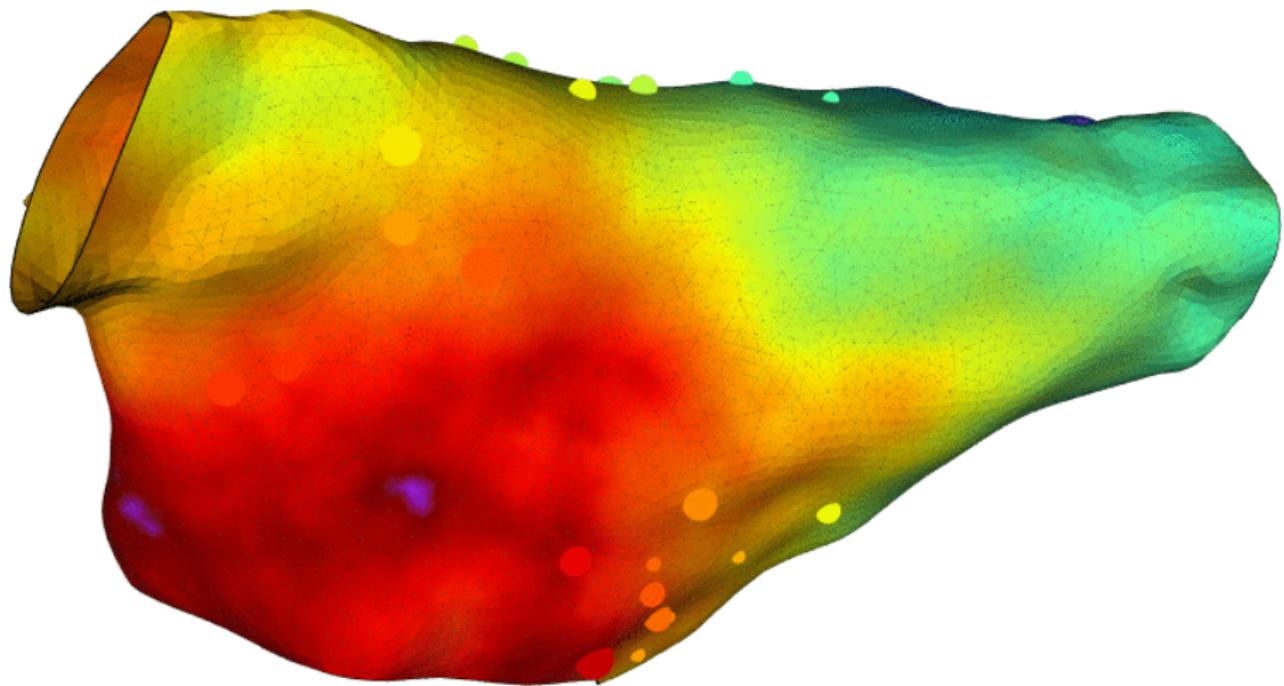
## Random samples



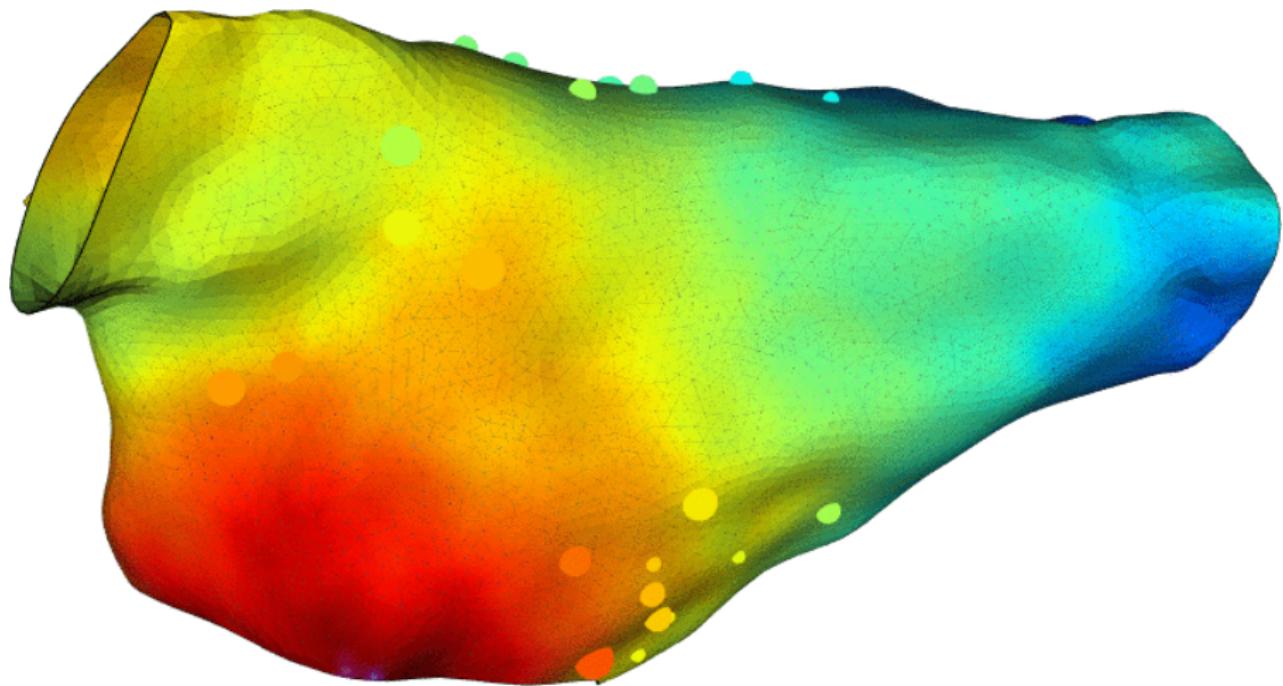
## Random samples



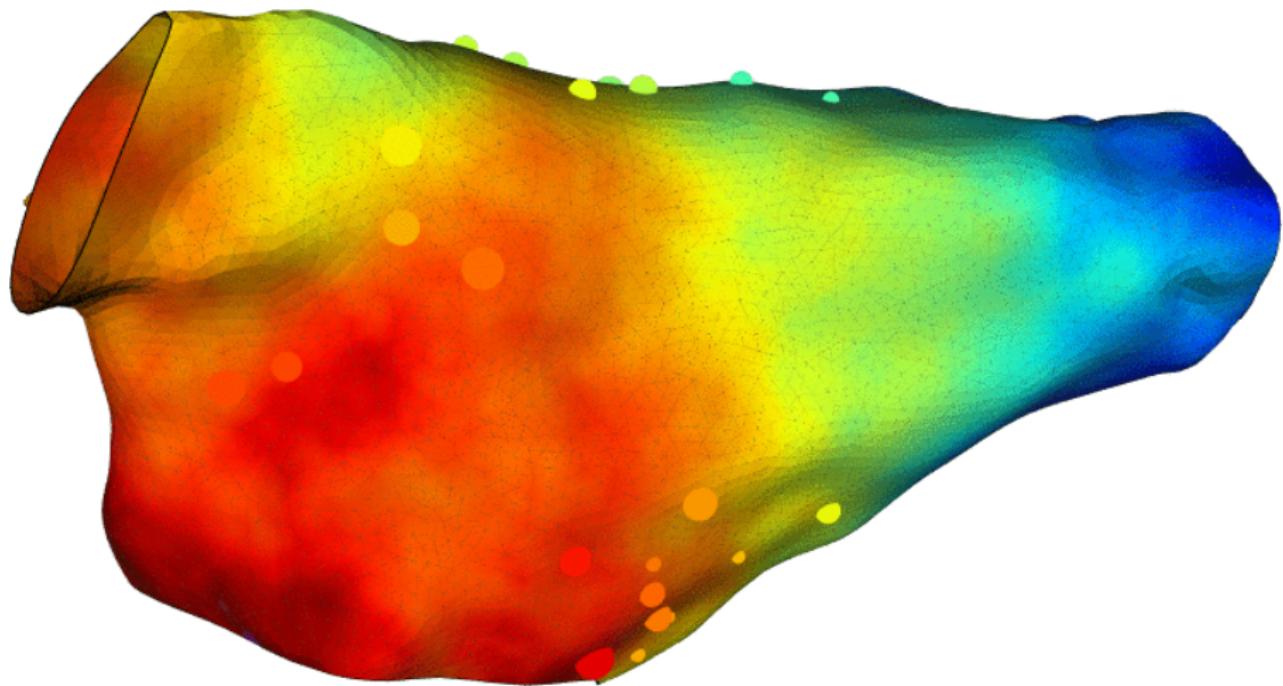
## Random samples



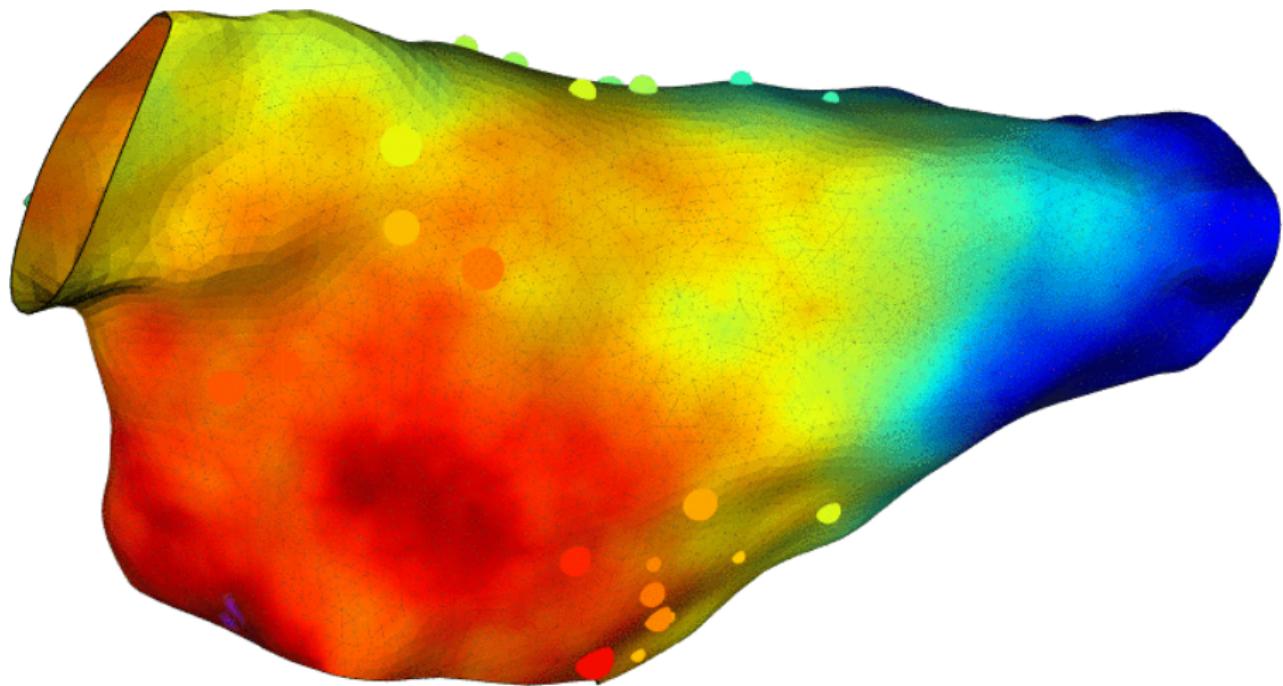
## Random samples



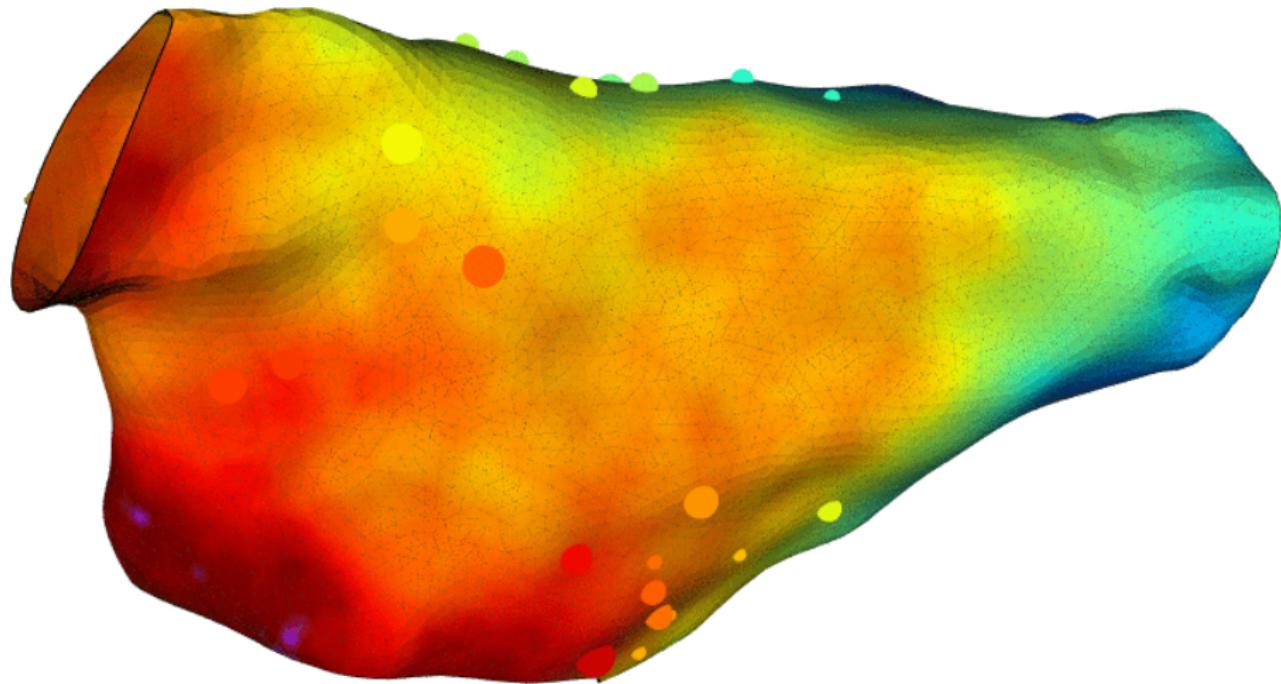
## Random samples



## Random samples



## Random samples



Unfortunately random samples produce unphysical (non-monotonic) patterns...

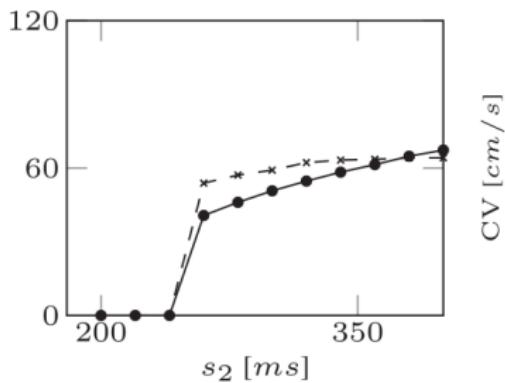
# Manifold uncertainty

## Model data comparison - decorrelating results

## Model calibration

Mitchell-Schaeffer model has 5 parameters,  $\theta \in \mathbb{R}^5$ , at every node on mesh ( $\sim 10000$  locations).

Want to infer these by comparison of measured restitution curve to simulated value:



# Bayesian inference of parameters

The Bayesian approach is to find the posterior distribution

$$\pi(\theta|D) \propto \pi(\theta)\pi(D|\theta)$$

posterior  $\propto$  prior  $\times$  likelihood

- usual intractability in Bayesian inference is not knowing  $\pi(D)$ .
- a problem is doubly intractable if  $\pi(D|\theta) = c_\theta p(D|\theta)$  with  $c_\theta$  unknown (cf Murray, Ghahramani and MacKay 2006)
- a problem is completely intractable if  $\pi(D|\theta)$  is unknown and can't be evaluated (unknown is subjective). I.e., if the analytic distribution of the simulator,  $f(\theta)$ , run at  $\theta$  is unknown.

Completely intractable models are where we need to resort to ABC methods

# 'Likelihood-Free' Inference

## Rejection Algorithm

- Draw  $\theta$  from prior  $\pi(\cdot)$
- Accept  $\theta$  with probability  $\pi(D | \theta)$

Accepted  $\theta$  are independent draws from the posterior distribution,  
 $\pi(\theta | D)$ .

# 'Likelihood-Free' Inference

## Rejection Algorithm

- Draw  $\theta$  from prior  $\pi(\cdot)$
- Accept  $\theta$  with probability  $\pi(D | \theta)$

Accepted  $\theta$  are independent draws from the posterior distribution,  $\pi(\theta | D)$ .

If the likelihood,  $\pi(D|\theta)$ , is unknown:

## 'Mechanical' Rejection Algorithm

- Draw  $\theta$  from  $\pi(\cdot)$
- Simulate  $X \sim f(\theta)$  from the computer model
- Accept  $\theta$  if  $D = X$ , i.e., if computer output equals observation

The acceptance rate is  $\int \mathbb{P}(D|\theta)\pi(\theta)d\theta = \mathbb{P}(D)$ .

# Rejection ABC

If  $\mathbb{P}(D)$  is small (or  $D$  continuous), we will rarely accept any  $\theta$ . Instead, there is an approximate version:

## Uniform Rejection Algorithm

- Draw  $\theta$  from  $\pi(\theta)$
- Simulate  $X \sim f(\theta)$
- Accept  $\theta$  if  $\rho(D, X) \leq \epsilon$

# Rejection ABC

If  $\mathbb{P}(D)$  is small (or  $D$  continuous), we will rarely accept any  $\theta$ . Instead, there is an approximate version:

## Uniform Rejection Algorithm

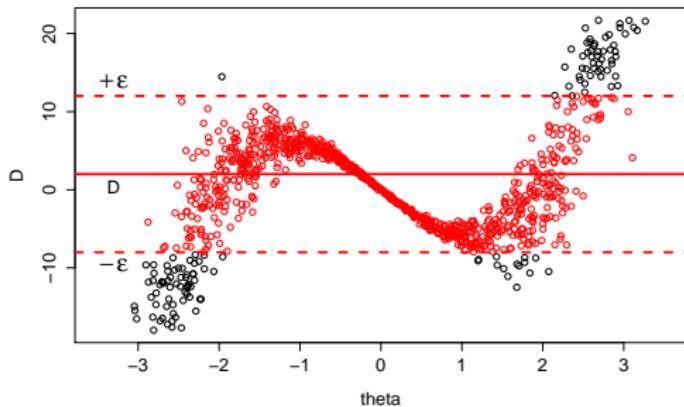
- Draw  $\theta$  from  $\pi(\theta)$
- Simulate  $X \sim f(\theta)$
- Accept  $\theta$  if  $\rho(D, X) \leq \epsilon$

$\epsilon$  reflects the tension between computability and accuracy.

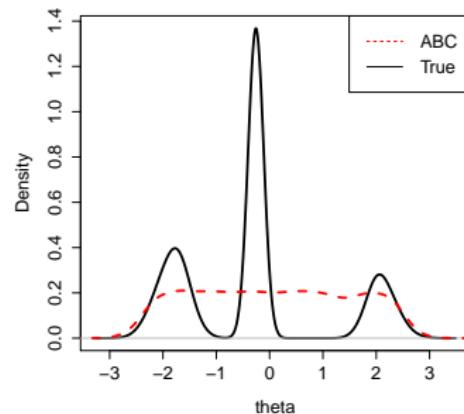
- As  $\epsilon \rightarrow \infty$ , we get observations from the prior,  $\pi(\theta)$ .
- If  $\epsilon = 0$ , we generate observations from  $\pi(\theta | D)$ .

$$\epsilon = 10$$

theta vs D



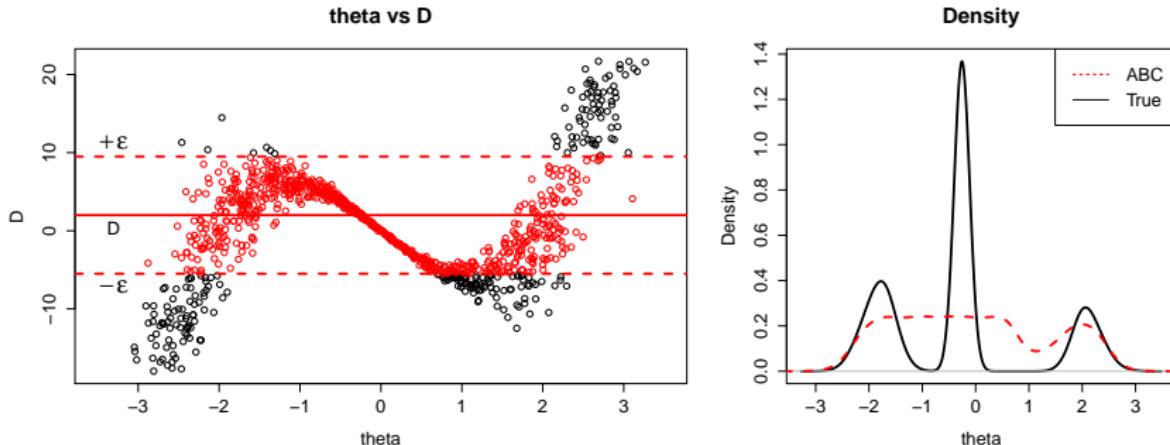
Density



$$\theta \sim U[-10, 10], \quad X \sim N(2(\theta + 2)\theta(\theta - 2), 0.1 + \theta^2)$$

$$\rho(D, X) = |D - X|, \quad D = 2$$

$$\epsilon = 7.5$$

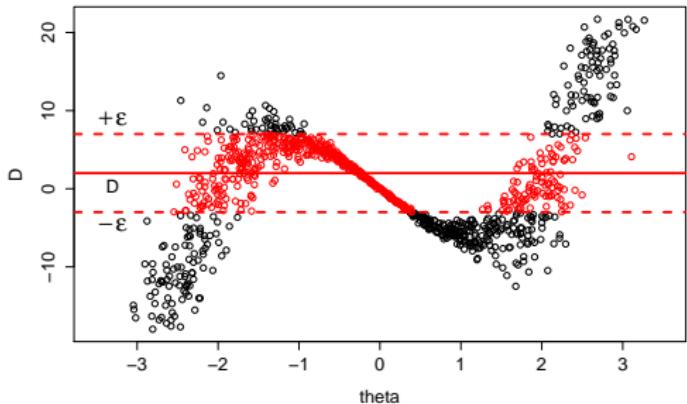


$$\theta \sim U[-10, 10], \quad X \sim N(2(\theta + 2)\theta(\theta - 2), 0.1 + \theta^2)$$

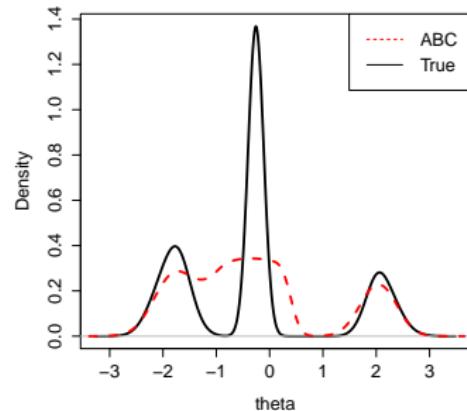
$$\rho(D, X) = |D - X|, \quad D = 2$$

$$\epsilon = 5$$

theta vs D



Density

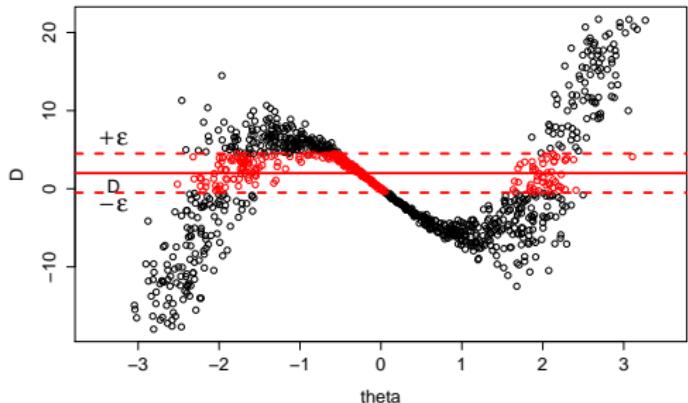


$$\theta \sim U[-10, 10], \quad X \sim N(2(\theta + 2)\theta(\theta - 2), 0.1 + \theta^2)$$

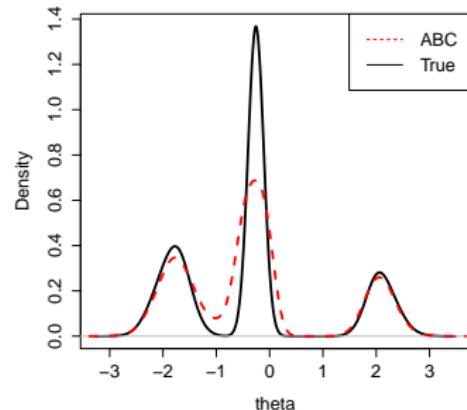
$$\rho(D, X) = |D - X|, \quad D = 2$$

$$\epsilon = 2.5$$

theta vs D



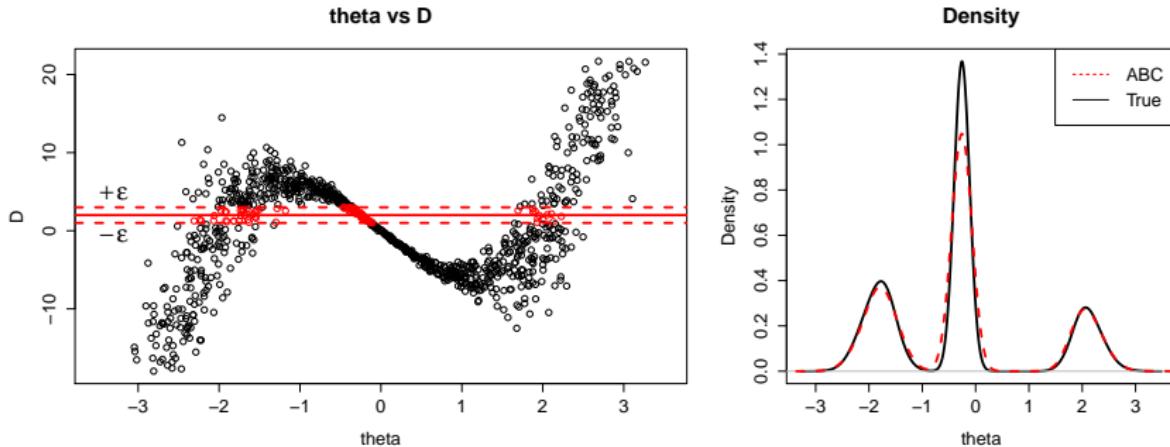
Density



$$\theta \sim U[-10, 10], \quad X \sim N(2(\theta + 2)\theta(\theta - 2), 0.1 + \theta^2)$$

$$\rho(D, X) = |D - X|, \quad D = 2$$

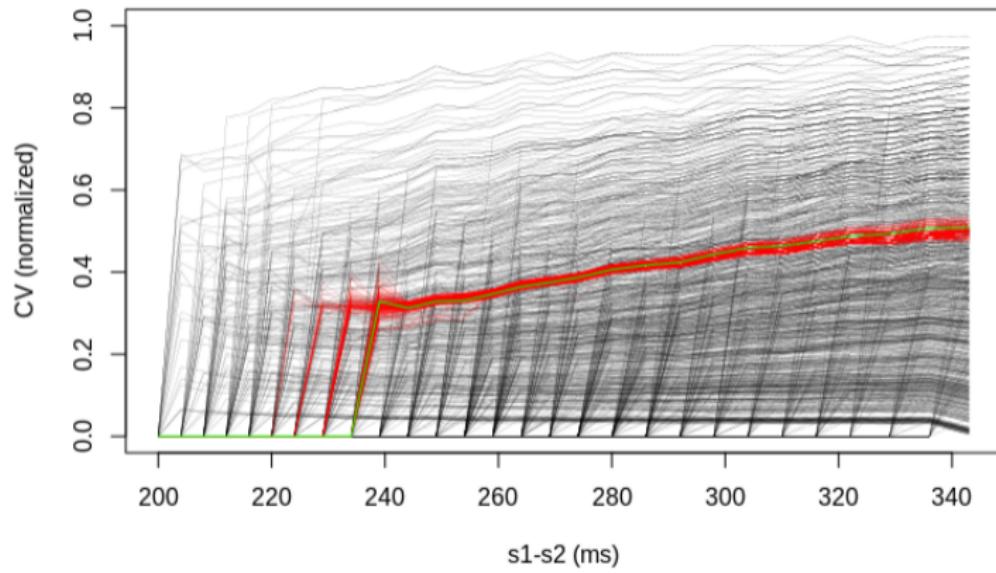
$$\epsilon = 1$$



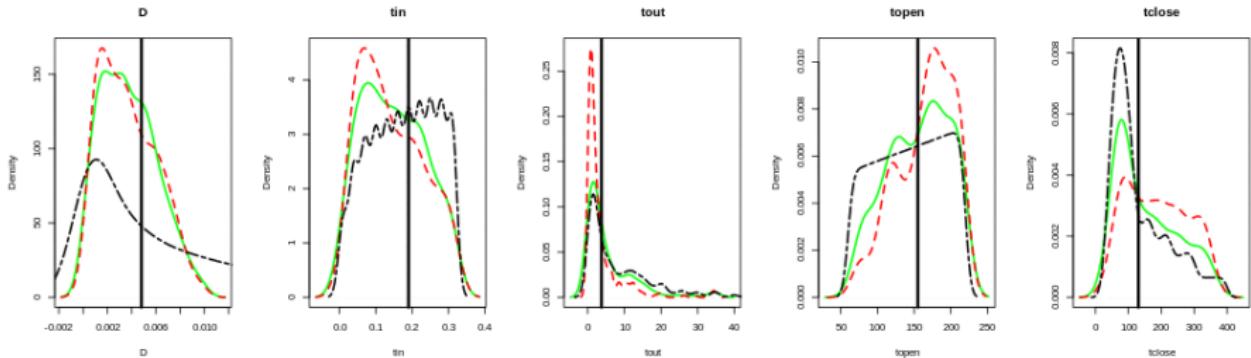
$$\theta \sim U[-10, 10], \quad X \sim N(2(\theta + 2)\theta(\theta - 2), 0.1 + \theta^2)$$

$$\rho(D, X) = |D - X|, \quad D = 2$$

We need to do this for every location ( $\sim 10000$ ) on our mesh. We create a large look up table of  $\{\theta, f(\theta)\}$  pairs allowing rapid inference of approximate posteriors at every location.



Posterior Parameters Density:: Green: with ERP info, Red: without ERP info, Black: prior



Use these posteriors to produce stochastic predictions of atrium wide activation patterns and estimate probability of atrial fibrillation.

# Conclusions

- Huge number of uncertain quantities. We need to find some regularity in the problem to allow us to reduce dimension sufficiently in order to make inference possible

Unclear what best approach is?

- ▶ Interpolate LAT smoothly and fit simulator to the interpolation?
- ▶ Assume smoothness (perhaps with irregularities) in the parameter distribution?

$$\theta(x) = \sum_{i=1}^{O(10)} \lambda_i \phi_i(x)$$

and learn simulator response wrt  $\lambda_i$ ?

- Can we guide data collection protocols?
  - ▶ More measurement locations and fewer S1-S2 intervals, or fewer locations and more S1-S2
- Can we learn with sufficient confidence to improve clinical procedures?

# Conclusions

- Huge number of uncertain quantities. We need to find some regularity in the problem to allow us to reduce dimension sufficiently in order to make inference possible

Unclear what best approach is?

- ▶ Interpolate LAT smoothly and fit simulator to the interpolation?
- ▶ Assume smoothness (perhaps with irregularities) in the parameter distribution?

$$\theta(x) = \sum_{i=1}^{O(10)} \lambda_i \phi_i(x)$$

and learn simulator response wrt  $\lambda_i$ ?

- Can we guide data collection protocols?
  - ▶ More measurement locations and fewer S1-S2 intervals, or fewer locations and more S1-S2
- Can we learn with sufficient confidence to improve clinical procedures?

Thank you for listening!