Uncertainty Quantification in Prospective and Predictive Patient Specific Cardiac Models

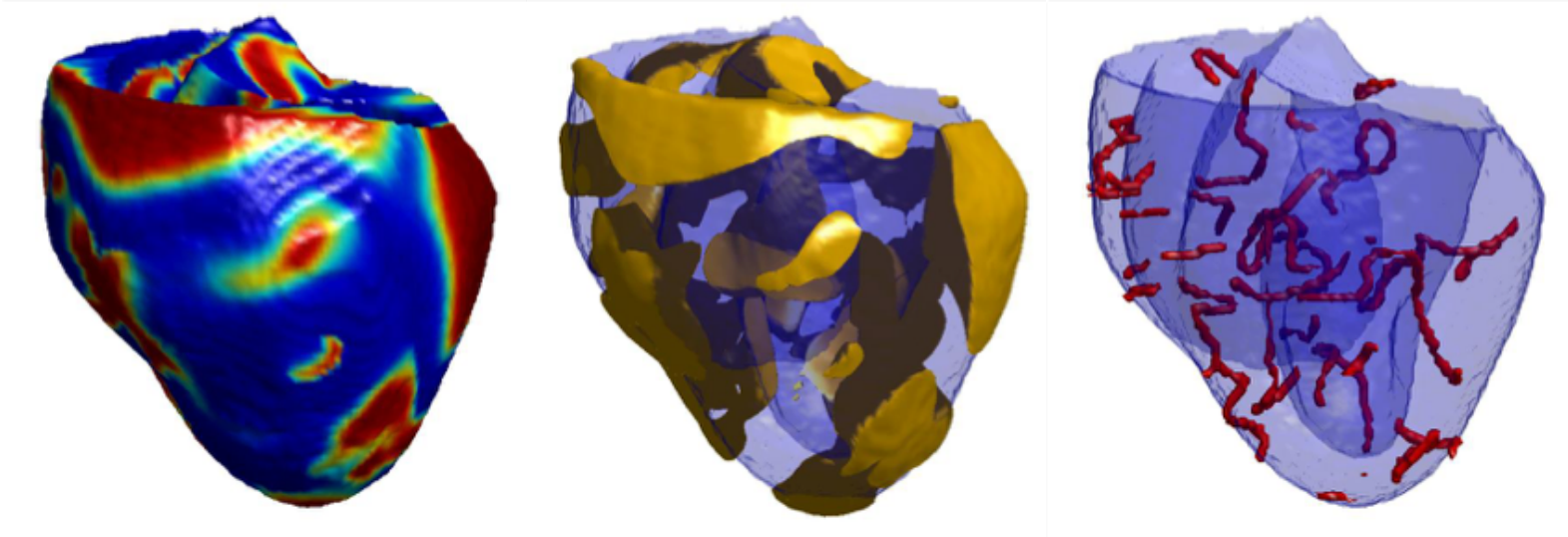
Richard Wilkinson, **Sam Coveney**, Jeremy Oakley, Richard Clayton, **Cesare Corrado**, Caroline Roney, Orod Razeghi, Steven Niederer

University of Sheffield
Kings College London

EPSRC
Engineering and Physical Sciences Research Council

Project overview

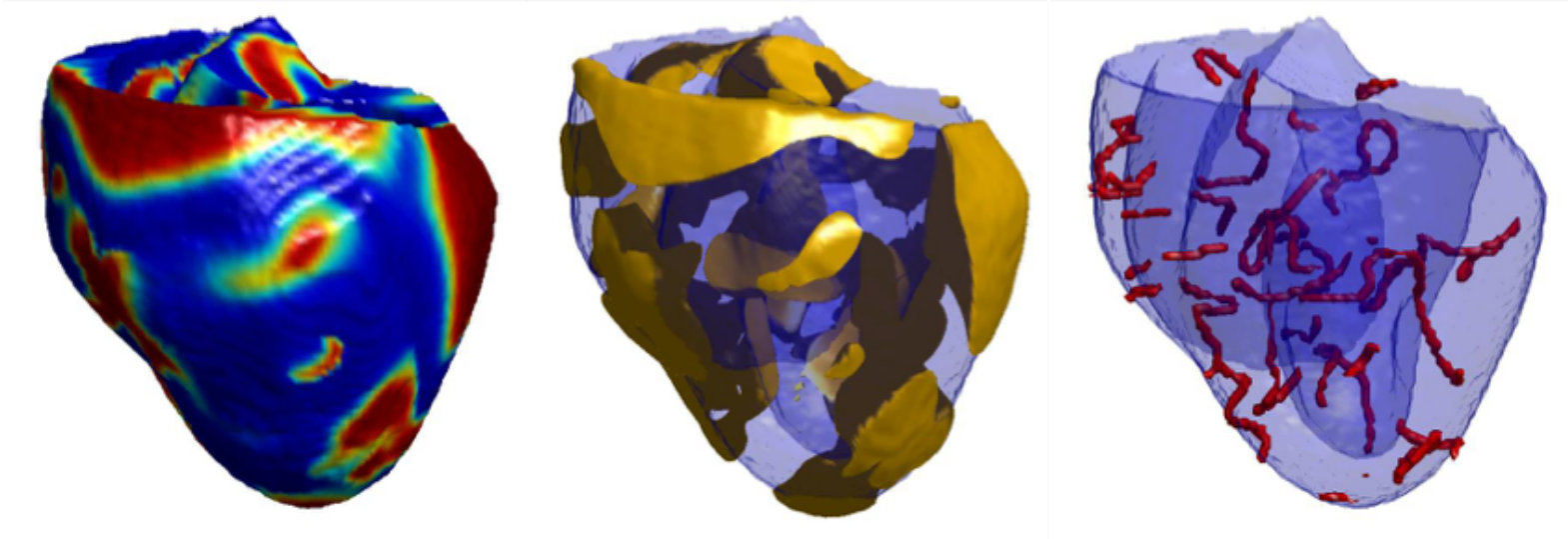
Atrial fibrillation



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

Project overview

Atrial fibrillation



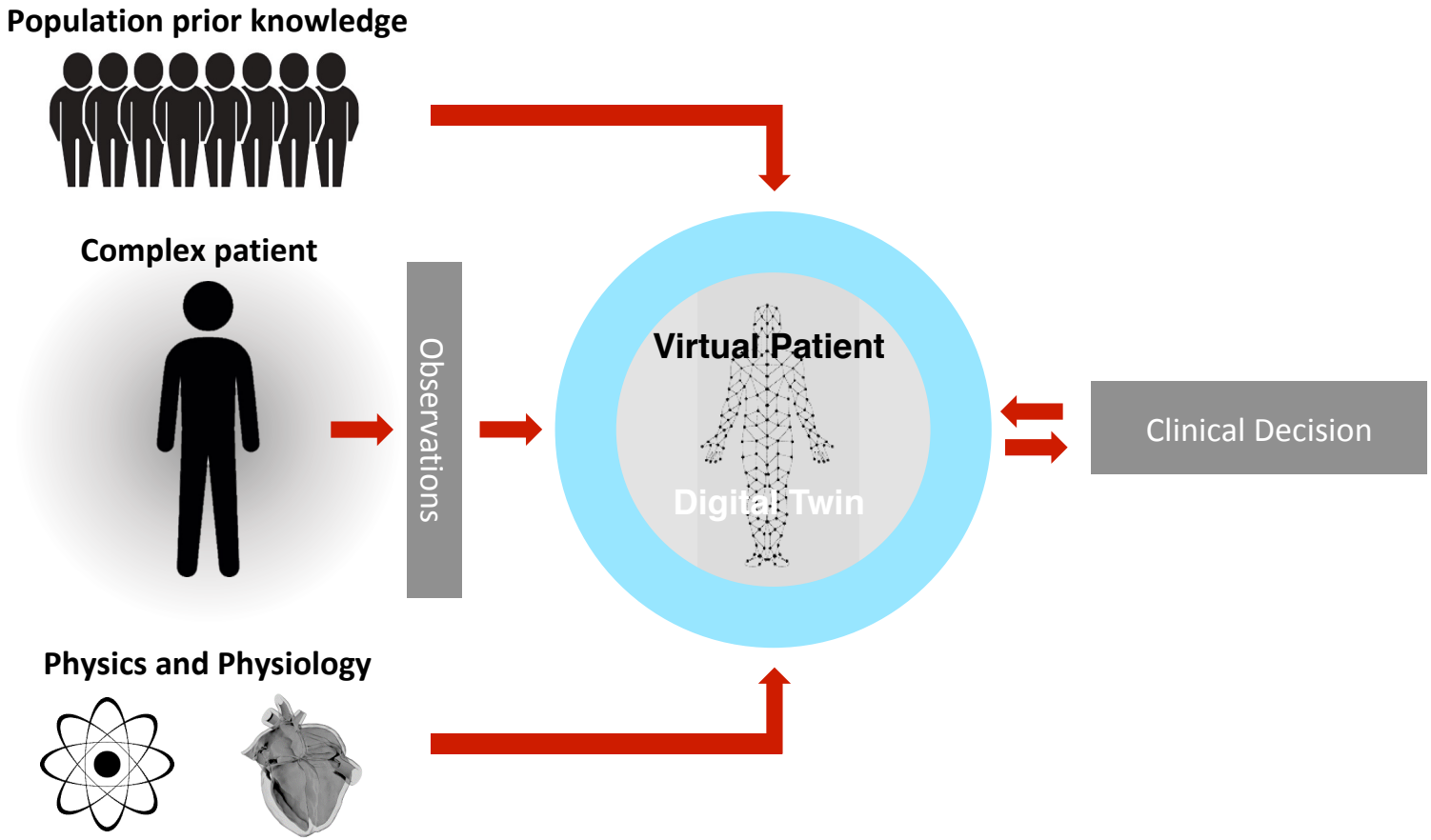
Atrial fibrillation (AF) - rapid and uncoordinated electrical activation (arrhythmia) leading to poor mechanical function.

- 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).

Project overview II

Patient Specific Cardiac Models

Aim: predict whether an AF patient will develop AT following ablation, infer the reentry pathways, and then guide the surgical ablation to treat for both in a single procedure.



But how confident are we in our prediction

Project overview III

To infer reentry pathways we

- use a complex simulator (encoding scientific knowledge) to see whether AT can be maintained

This requires

- Left atrium geometry, spatially distributed tissue properties, fibre directions, etc for the individual patient

all of which are unknown.

Project overview III

To infer reentry pathways we

- use a complex simulator (encoding scientific knowledge) to see whether AT can be maintained

This requires

- Left atrium geometry, spatially distributed tissue properties, fibre directions, etc for the individual patient

all of which are unknown.

Workflow:

- MRI to build patient specific left atrium mesh, identify fibrosis.
- Electrophysiology study to learn electrical activation map, conduction velocities
- Interpolate to entire LA, allowing conduction velocity and restitution curves to be calculated
- Use these to inform the heterogeneity in our prior distribution of the tissue properties
- Build an emulator of the simulator
- Find our posterior distribution over tissue parameters etc
- Predict AT pathways, make clinical recommendations

Uncertainty quantification

Project aim is characterize and combine the uncertainties to make decisions that take our lack of knowledge into account.

- Noisy data, recorded at a small number of sparse, uncertain locations
- Large number of unknown parameters
- Complex simulator (limited computational resource)
- Misspecification/discrepancy

Uncertainty quantification

Project aim is characterize and combine the uncertainties to make decisions that take our lack of knowledge into account.

- Noisy data, recorded at a small number of sparse, uncertain locations
- Large number of unknown parameters
- Complex simulator (limited computational resource)
- Misspecification/discrepancy

$$\mathbb{P}(E\text{vent}|D\text{ata}) = \int \mathbb{P}(E|\theta, x, f)\pi(\theta, x, f|D)d\theta dxdf$$

where

$$\pi(\theta, x, f|D) \propto \pi(D|\theta, x, f)\pi(\theta)\pi(x)\pi(f)$$

We need to characterize variability at the

- population level $\pi(\theta), \pi(x)$ etc
- individual level $\pi(\theta, x, f, ...|D)$ – may need to be done online
- and the physics/simulator $\pi(D|\theta, x, f)$

Uncertainty quantification

Project aim is characterize and combine the uncertainties to make decisions that take our lack of knowledge into account.

- Noisy data, recorded at a small number of sparse, uncertain locations
- Large number of unknown parameters
- Complex simulator (limited computational resource)
- Misspecification/discrepancy

$$\mathbb{P}(E\text{vent}|D\text{ata}) = \int \mathbb{P}(E|\theta, x, f)\pi(\theta, x, f|D)d\theta dx df$$

where

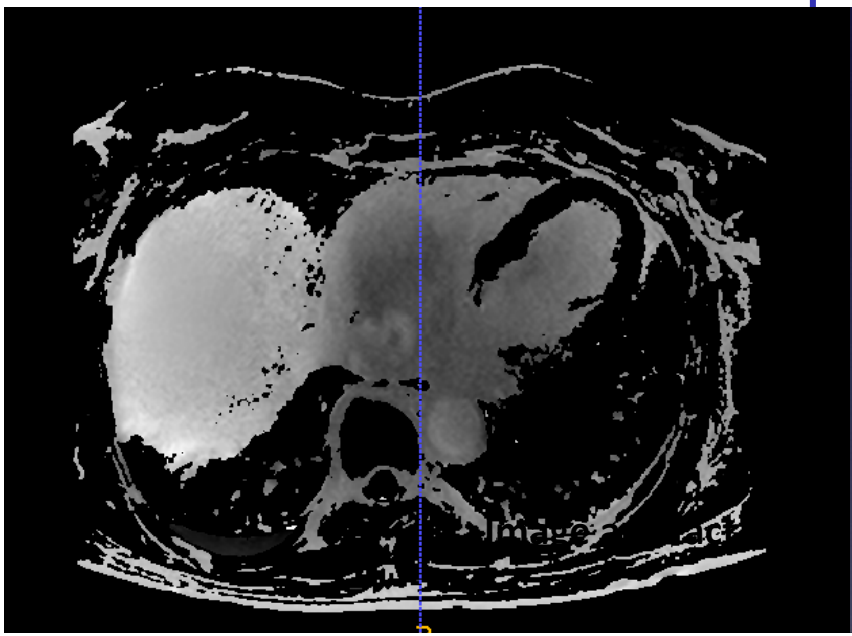
$$\pi(\theta, x, f|D) \propto \pi(D|\theta, x, f)\pi(\theta)\pi(x)\pi(f)$$

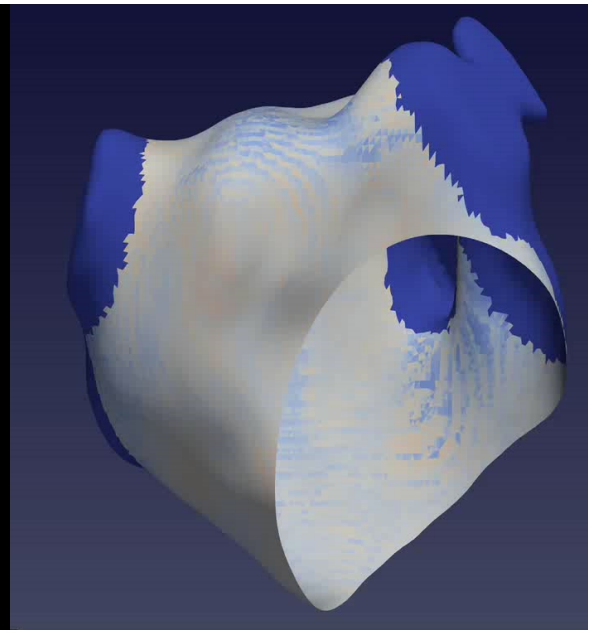
We need to characterize variability at the

- population level $\pi(\theta), \pi(x)$ etc
- individual level $\pi(\theta, x, f, ...|D)$ – may need to be done online
- and the physics/simulator $\pi(D|\theta, x, f)$

Pragmatic approach necessary.

Problem 1: Uncertain shape - Cesare Corrado



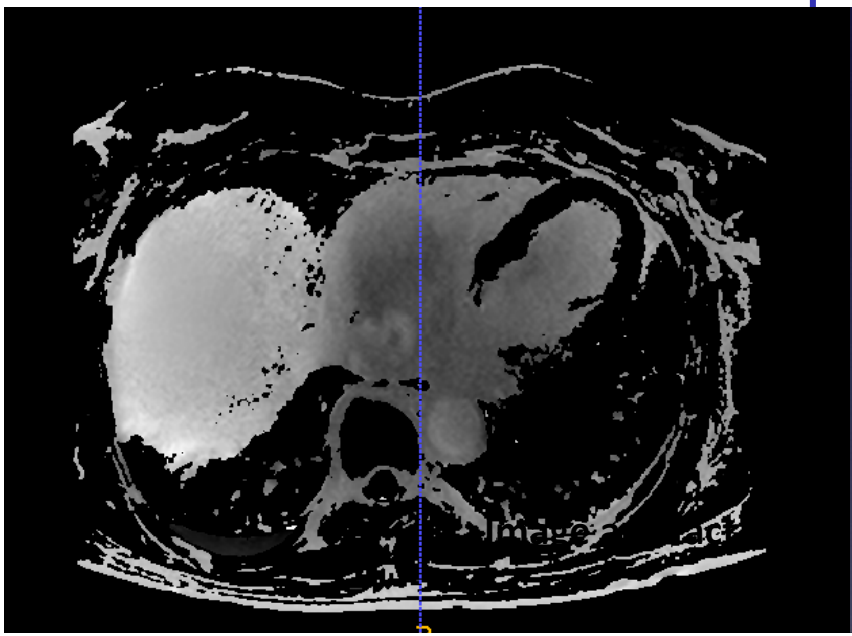


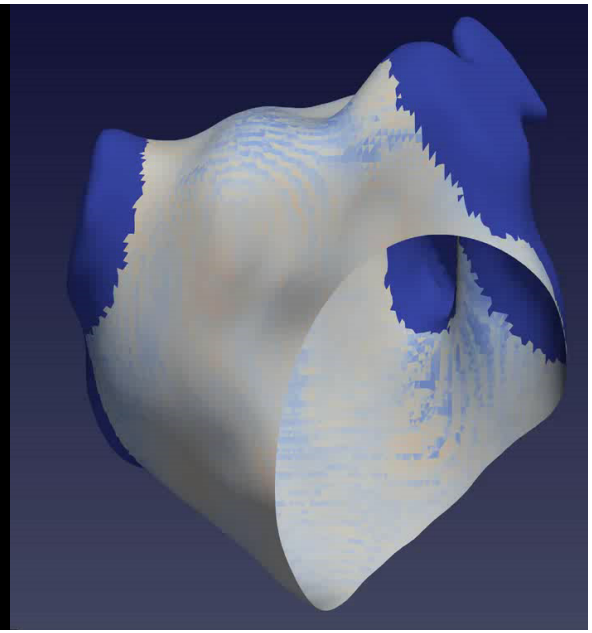
Measure shape $x_{obs} \in \mathbb{R}^D$ where $D \sim 10^5$

$$x_{obs} = x_{true} + e' \quad \text{where} \quad e' \sim N(0, \Sigma')$$

How can we parsimoniously describe the variation in atrial shapes in the populations $x^1_{true}, \ldots, x^n_{true}$?

Problem 1: Uncertain shape - Cesare Corrado





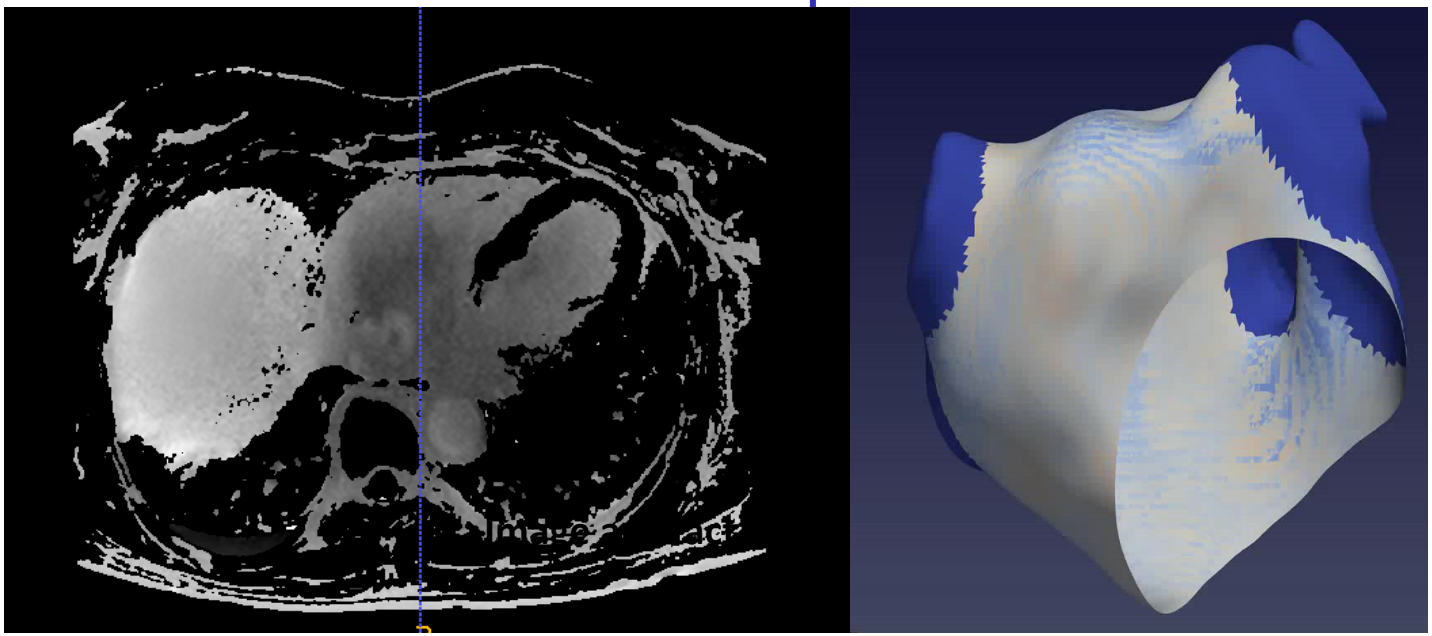
Measure shape $x_{obs} \in \mathbb{R}^D$ where $D \sim 10^5$

$$x_{obs} = x_{true} + e' \quad \text{where} \quad e' \sim N(0, \Sigma')$$

How can we parsimoniously describe the variation in atrial shapes in the populations $x^1_{true}, \ldots, x^n_{true}$?
Working in the standard basis is infeasible

$$x_{obs} = \sum_i x_{obs,i} v_i \quad \text{where} \quad (v_i)_j = \delta_{ij}$$

Problem 1: Uncertain shape - Cesare Corrado

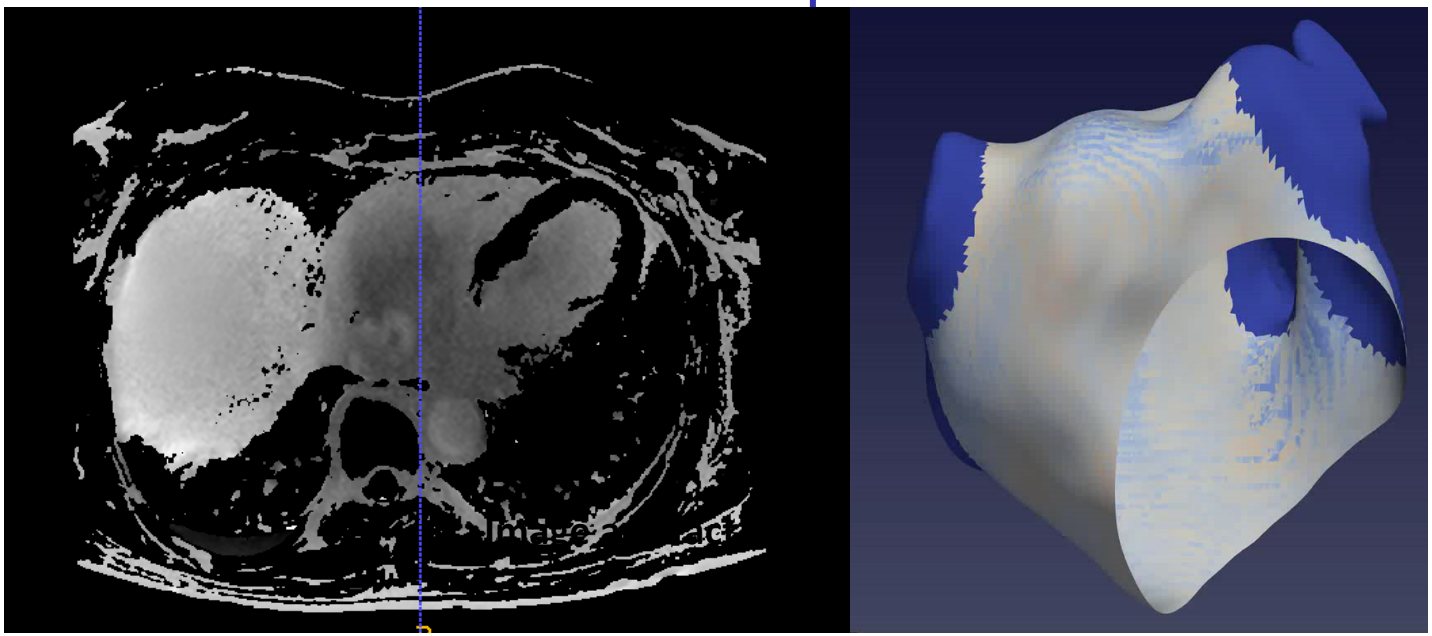


Aim: change into a basis that allows variation to be described in lower dimension

$$x_{obs} = \mu + \sum_{i=1}^{d} \lambda_i u_i + e \quad \text{where} \quad e \sim N(0, \Sigma)$$

where $\lambda = (\lambda_1, \ldots, \lambda_d)^\top$ is the new coordinate describing variation in $\mathbb{R}^d$ (where $d \ll D$) for the orthonormal basis $\{u_1, \ldots, u_d\}$.

Problem 1: Uncertain shape - Cesare Corrado



Aim: change into a basis that allows variation to be described in lower dimension

$$x_{obs} = \mu + \sum_{i=1}^{d} \lambda_i u_i + e \quad \text{where} \quad e \sim N(0, \Sigma)$$

where $\lambda = (\lambda_1, \ldots, \lambda_d)^\top$ is the new coordinate describing variation in $\mathbb{R}^d$ (where $d \ll D$) for the orthonormal basis $\{u_1, \ldots, u_d\}$.
Determine the reduced basis, error variance Σ and prior $\lambda \sim N(0, \Sigma_\lambda)$ from the population. cf Andrew McCulloch's talk

Use Bayesian approach to characterize uncertainty about individual anatomy via

$$\pi(\lambda|x_{obs}) \propto \pi(x_{obs}|\lambda, \hat{\Sigma})\pi(\lambda)$$

where now typically $\lambda \in \mathbb{R}^{10}$.

Use Bayesian approach to characterize uncertainty about individual anatomy via

$$\pi(\lambda|x_{obs}) \propto \pi(x_{obs}|\lambda, \hat{\Sigma})\pi(\lambda)$$

where now typically $\lambda \in \mathbb{R}^{10}$.

Choice of basis?

Aim: characterize the population variability, i.e., $p(x)$ in a lower dimensional space

Use Bayesian approach to characterize uncertainty about individual anatomy via

$$\pi(\lambda|x_{obs}) \propto \pi(x_{obs}|\lambda, \hat{\Sigma})\pi(\lambda)$$

where now typically $\lambda \in \mathbb{R}^{10}$.

Choice of basis?

Aim: characterize the population variability, i.e., $p(x)$ in a lower dimensional space

- PCA basis is optimal
  - ie let $U = [u_1, \ldots, u_d] \in \mathbb{R}^{D \times d}$ be the first d eigenvectors of $Var_p(x)$. Then $\lambda = U^\top x = (\langle u_1, X \rangle, \ldots, \langle u_d, X \rangle)^\top$.
  - u_i maximizes $u_i^\top Var_p(x) u_i$ s.t. $\langle u_i, u_j \rangle = \delta_{ij}$
  - Equivalently, $UU^\top X$ is the best rank d approximation to X (in the Frobenius norm $\sim L_2$).

Use Bayesian approach to characterize uncertainty about individual anatomy via

$$\pi(\lambda|x_{obs}) \propto \pi(x_{obs}|\lambda, \hat{\Sigma})\pi(\lambda)$$

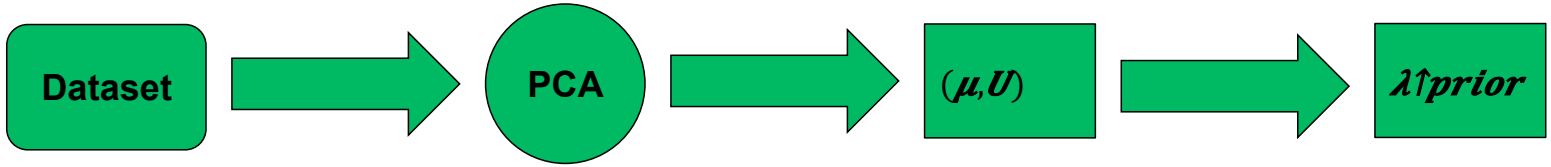
where now typically $\lambda \in \mathbb{R}^{10}$.

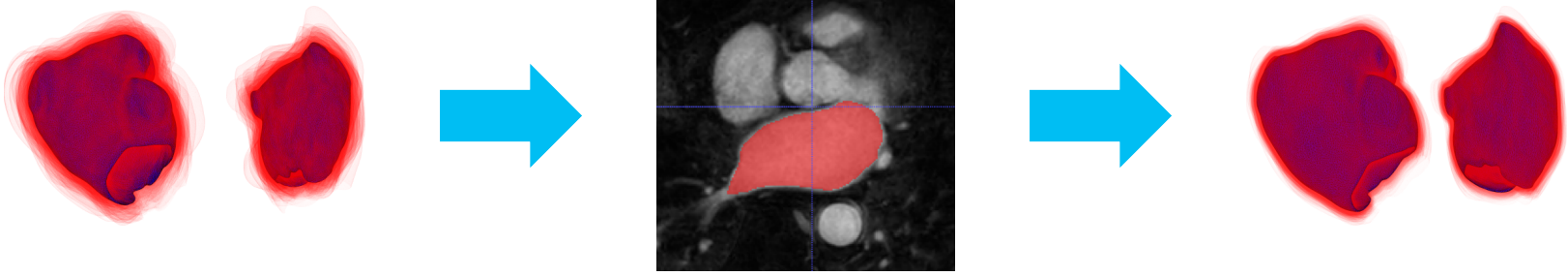
Choice of basis?

Aim: characterize the population variability, i.e., $p(x)$ in a lower dimensional space

- PCA basis is optimal
  - ie let $U = [u_1, \ldots, u_d] \in \mathbb{R}^{D \times d}$ be the first d eigenvectors of $Var_p(x)$. Then $\lambda = U^\top x = (\langle u_1, X \rangle, \ldots, \langle u_d, X \rangle)^\top$.
  - u_i maximizes $u_i^\top Var_p(x) u_i$ s.t. $\langle u_i, u_j \rangle = \delta_{ij}$
  - Equivalently, $UU^\top X$ is the best rank d approximation to X (in the Frobenius norm $\sim L_2$).

But for other purposes (eg in supervised problems) PCA can give poor dimension reduction.

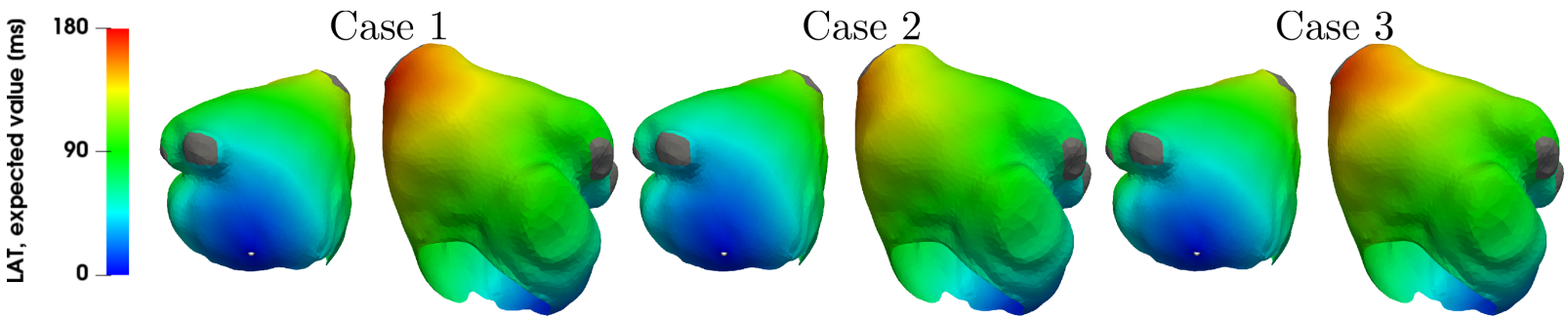




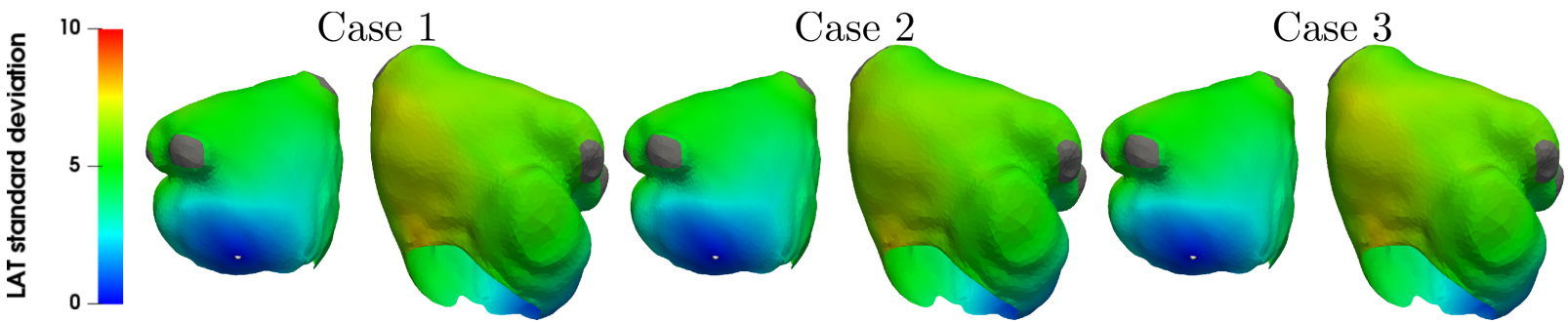
Population Mean and shape (λ) Priors

New Measure

Update Shape (λ) Posterior Distribution



Expected value



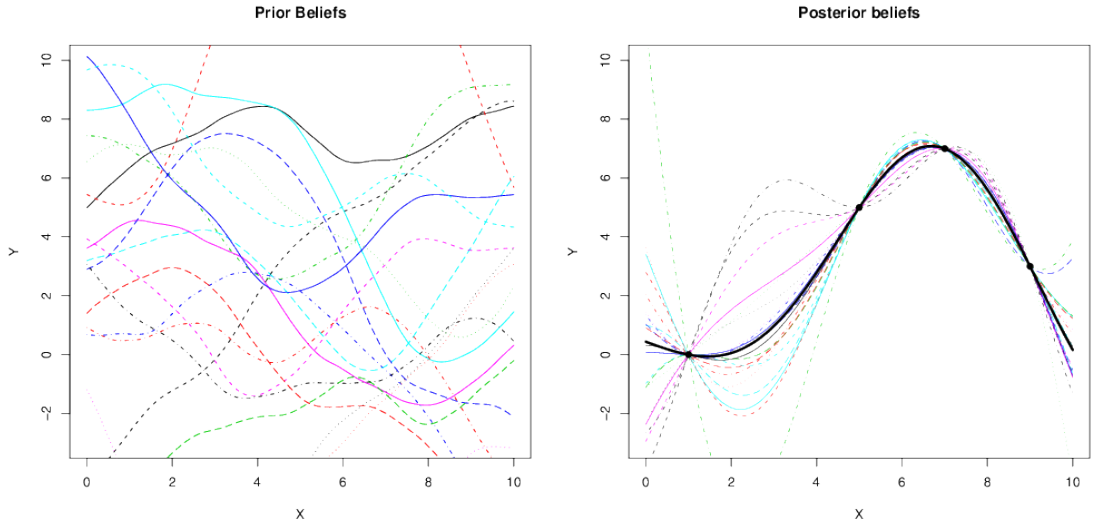
Standard deviation

Problem 2: Interpolation of LAT - Sam Coveney

Aside: Gaussian processes (GP)

Regression: given data $\{x_i, y_i = f(x_i)\}_{i=1}^n$ learn f.

- x is location on the atrium, $f(x)$ is activation time
- x is a simulator parameter, $f(x)$ a complex simulator prediction.



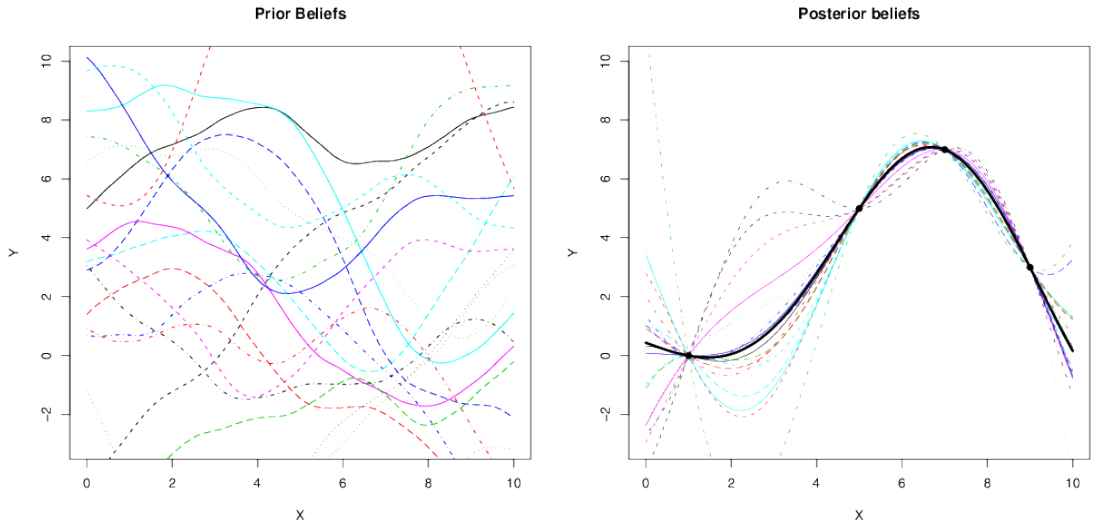
GPs can be thought of as probabilistic models of functions.

Problem 2: Interpolation of LAT - Sam Coveney

Aside: Gaussian processes (GP)

Regression: given data $\{x_i, y_i = f(x_i)\}_{i=1}^n$ learn f.

- x is location on the atrium, $f(x)$ is activation time
- x is a simulator parameter, $f(x)$ a complex simulator prediction.



GPs can be thought of 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, $x_1, \ldots, x_n$,

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

has a multivariate Gaussian distribution.

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

Why use GPs? Answer 1

Class of models is closed under various operations.

Why use GPs? Answer 1

Class of models is closed under various operations.

- Closed under addition

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

Why use GPs? 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

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

then

$$f|D \sim GP$$

but with updated mean and covariance functions.

Why use GPs? 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

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

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 $\nabla \cdot \nabla f = 0$

Why use GPs? Answer 2: non-parametric/kernel regression

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

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

Why use GPs? Answer 2: non-parametric/kernel regression

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

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

- 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.

Why use GPs? Answer 2: non-parametric/kernel regression

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

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

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

Why use GPs? Answer 2: non-parametric/kernel regression

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

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

- 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 $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is a semi-positive definite function.
Kernel trick: lift x into infinite dimensional feature space by replacing inner products $x^\top x'$ by $k(x, x')$.
Never evaluate the 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.

- choice of kernel implicitly chooses features
- model only includes functions that are linear combinations of the features (the RKHS of k)

Generally, we don't think about features, we just choose a kernel.

- choice of kernel implicitly chooses features
- model only includes functions that are linear combinations of the features (the RKHS of k)

Example: If (modulo some detail)

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

then as $N \to \infty$ 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.

- choice of kernel implicitly chooses features
- model only includes functions that are linear combinations of the features (the RKHS of k)

Example: If (modulo some detail)

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

then as $N \to \infty$ then

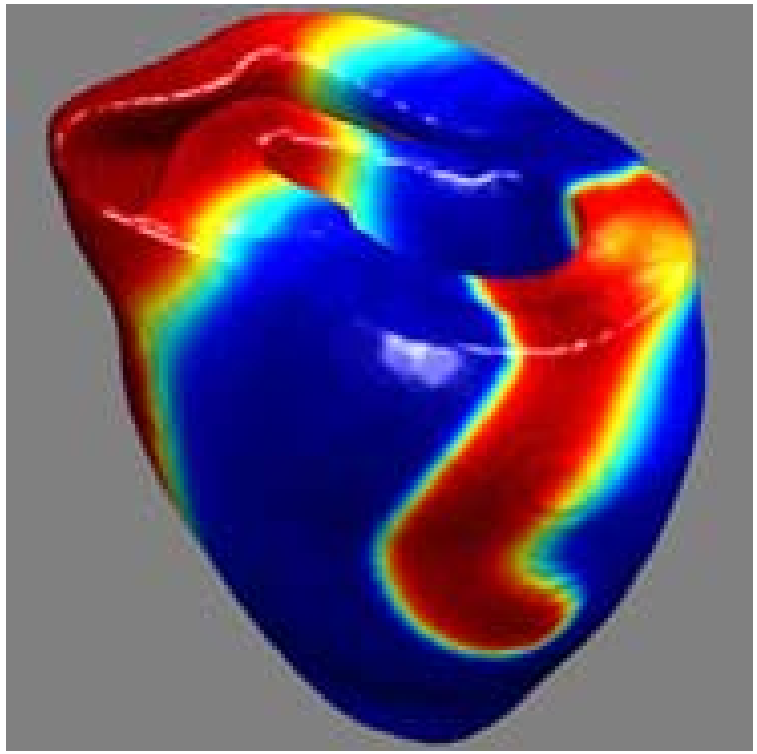
$$\phi(x)^\top \phi(x) = \exp\left(-\frac{(x-x')^2}{2\lambda^2}\right)$$

Although our simulator may not lie in the RKHS defined by k, this space is much richer than any parametric regression model (possibly dense in some set of continuous functions)

- more likely to contain an element close to the simulator than any finite class of models

Local activation time

Red: 'active' cardiac tissue Blue: 'inactive' cardiac tissue



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

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

Interpolation

The LAT map tells us conduction velocities.

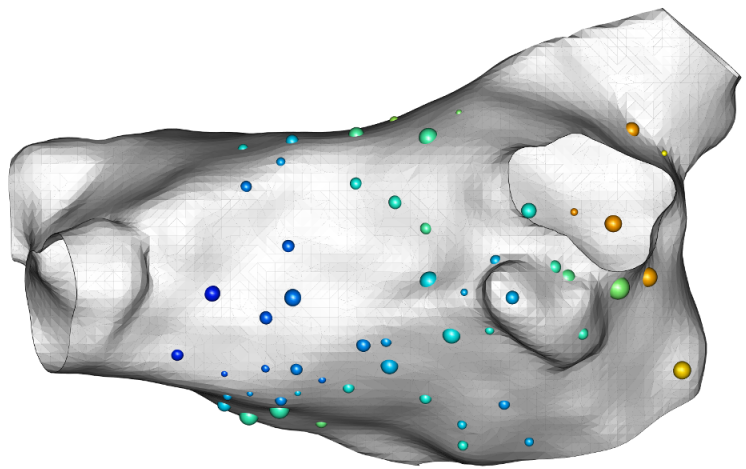
- Heterogeneity in the conduction velocity tells us something about heterogeneity in the tissue properties.

Interpolation

The LAT map tells us conduction velocities.

- Heterogeneity in the conduction velocity tells us something about heterogeneity in the tissue properties.

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



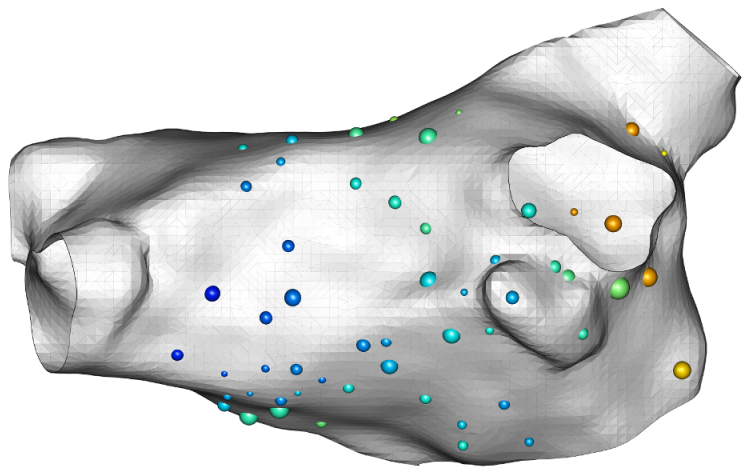
Idea: Interpolate the LAT map, use this to guide our prior distribution for tissue properties for the simulator

Interpolation

The LAT map tells us conduction velocities.

- Heterogeneity in the conduction velocity tells us something about heterogeneity in the tissue properties.

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



Idea: Interpolate the LAT map, use this to guide our prior distribution for tissue properties for the simulator

How can we interpolate to other locations?

$$LAT_{obs}(x) = \mathrm{LAT}_{true}(x) + \epsilon_{EGM} + \epsilon_{position}$$

GP interpolation

We want to model

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

but standard approaches won't work on complex atrial manifolds

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

GP interpolation

We want to model

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

but standard approaches won't work on complex atrial manifolds

- Typically covariance is 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 formulated in terms of a covariance function, for Matern covariance functions Whittle showed we can represent the GP as a stochastic partial differential equation (SPDE):

$$(\kappa^2 - \Delta)^{\alpha/2} LAT(x) = W(x)$$

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

Coveney *et al.* 2019

Instead of a GP formulated in terms of a covariance function, for Matern covariance functions Whittle showed we can represent the GP as a stochastic partial differential equation (SPDE):

$$(\kappa^2 - \Delta)^{\alpha/2} LAT(x) = W(x)$$

- Allows us to fit GPs using the machinery of finite element methods (allows solution in $O(n^{3/2})$ instead of $O(n^3)$).
- Makes it easy to work on irregular domains.

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

Coveney *et al.* 2019

Instead of a GP formulated in terms of a covariance function, for Matern covariance functions Whittle showed we can represent the GP as a stochastic partial differential equation (SPDE):

$$(\kappa^2 - \Delta)^{\alpha/2} LAT(x) = W(x)$$

- Allows us to fit GPs using the machinery of finite element methods (allows solution in $O(n^{3/2})$ instead of $O(n^3)$).
- Makes it easy to work on irregular domains.

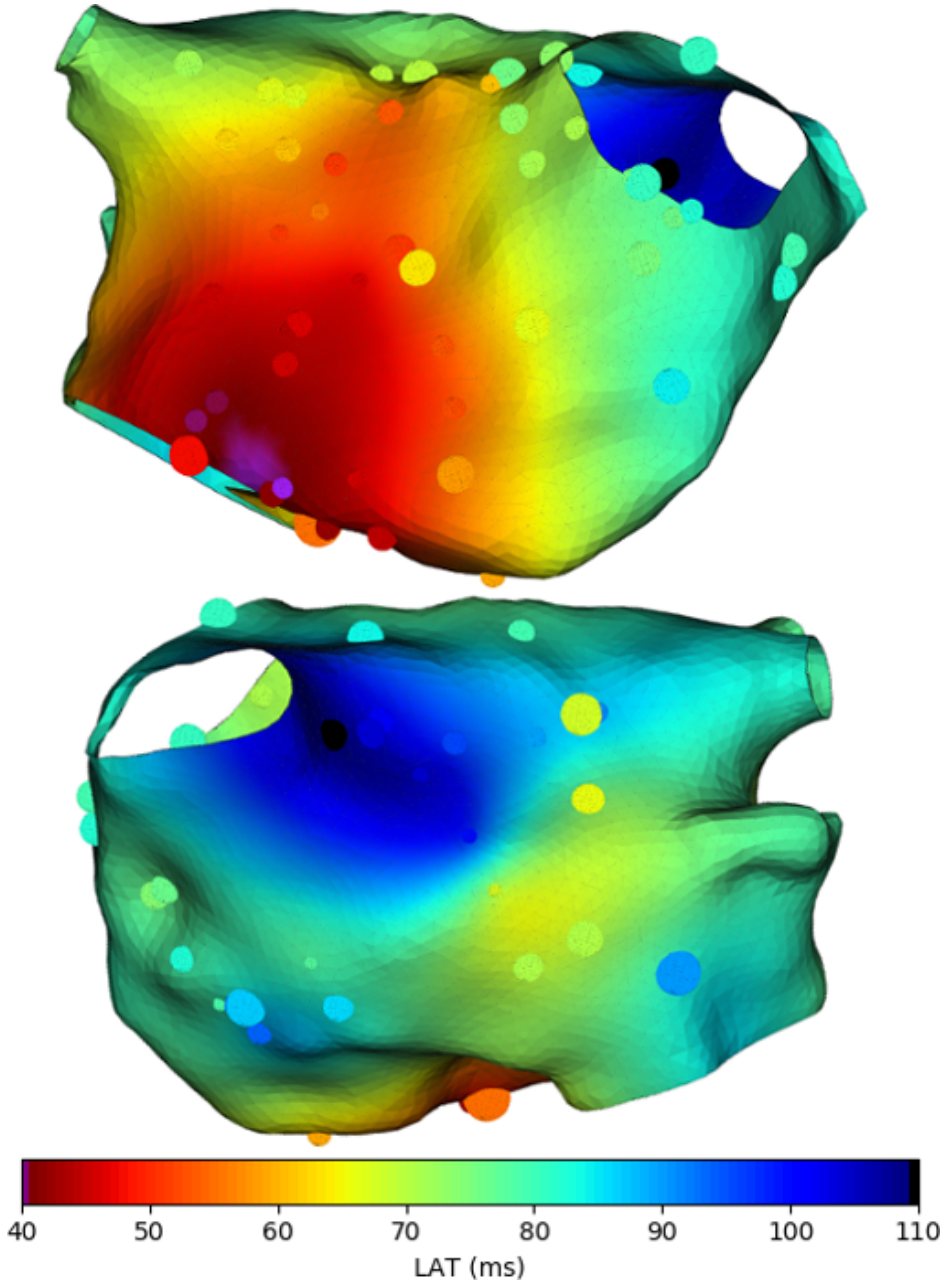
$$LAT(x) = \sum_{k=1}^{n} w_k \psi_k(x)$$

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

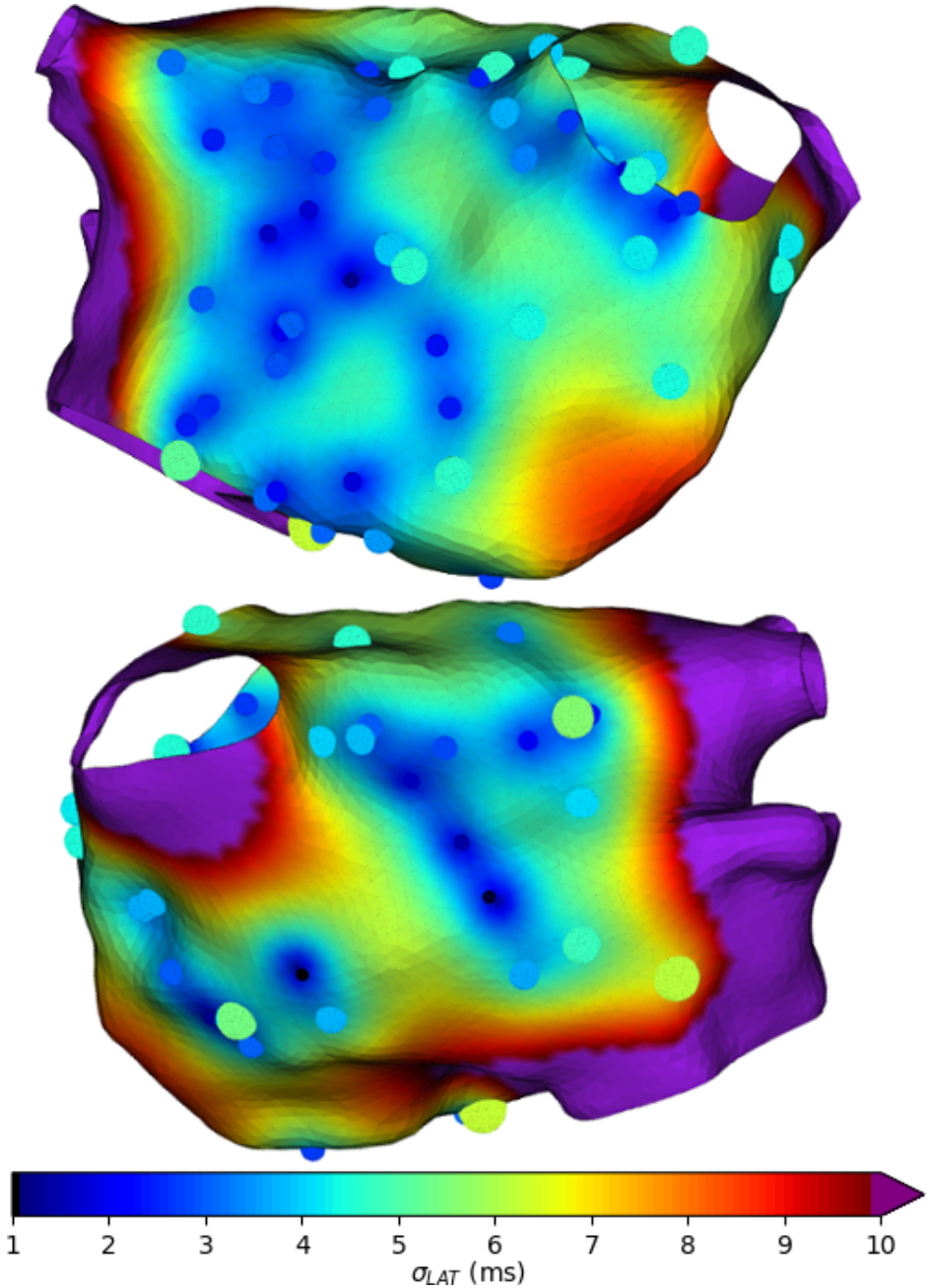
$$LAT(\cdot) \sim GP(0, Q^{-1})$$

for some Q

Results - mean



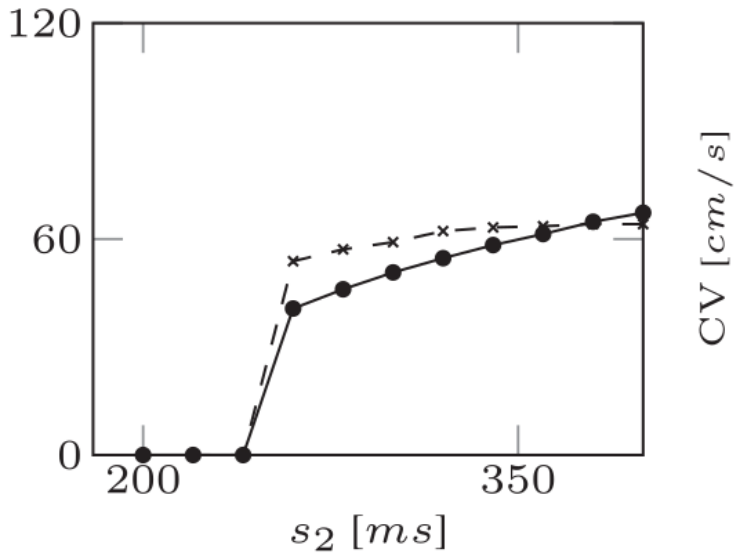
Results - standard deviation



S1-S2 interpolation

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

- 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. We use INLA-SPDE approach to interpolate LAT at the locations for a given S2 value.

- allows us to 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. We use INLA-SPDE approach to interpolate LAT at the locations for a given S2 value.

- allows us to 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 $LAT(x, S2_i)$

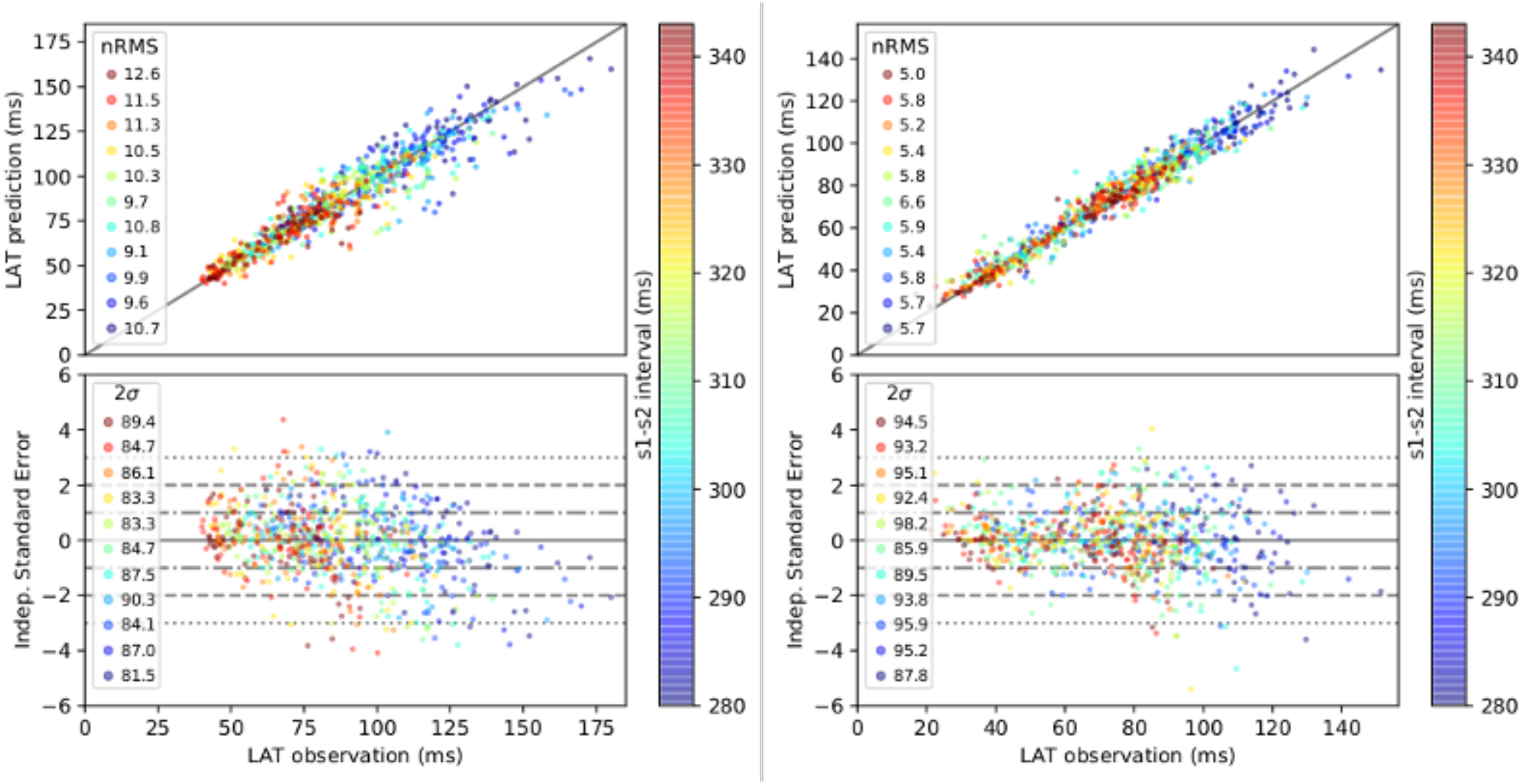
$$LAT(x, S2_{i+1}) \sim N(\rho LAT(x, S2_i), (1 - \rho^2)Q^{-1})$$

or more precisely

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

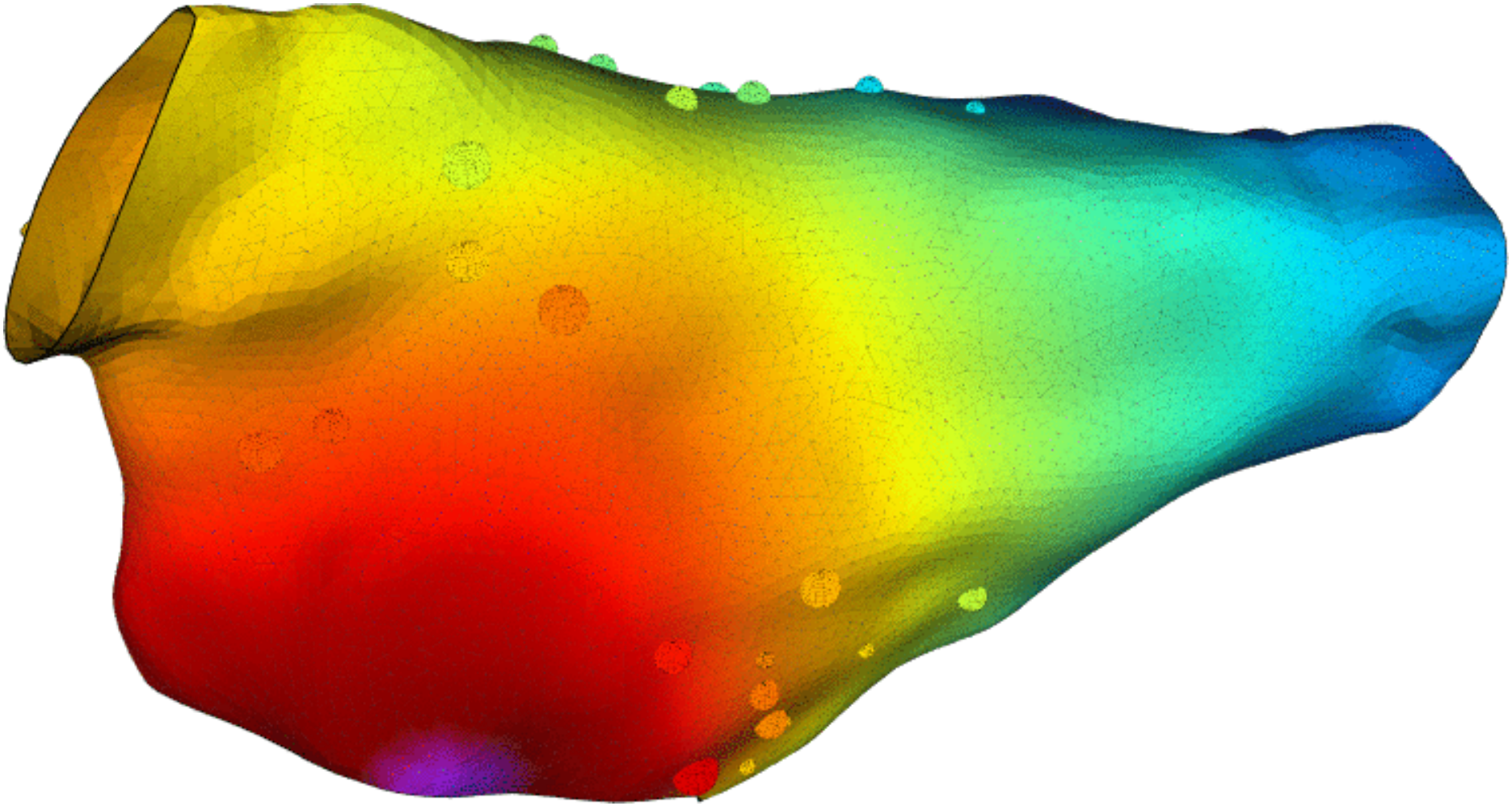
Cf Dirk Husmeier's talk with multi-output GPs.

Results: Cross validation

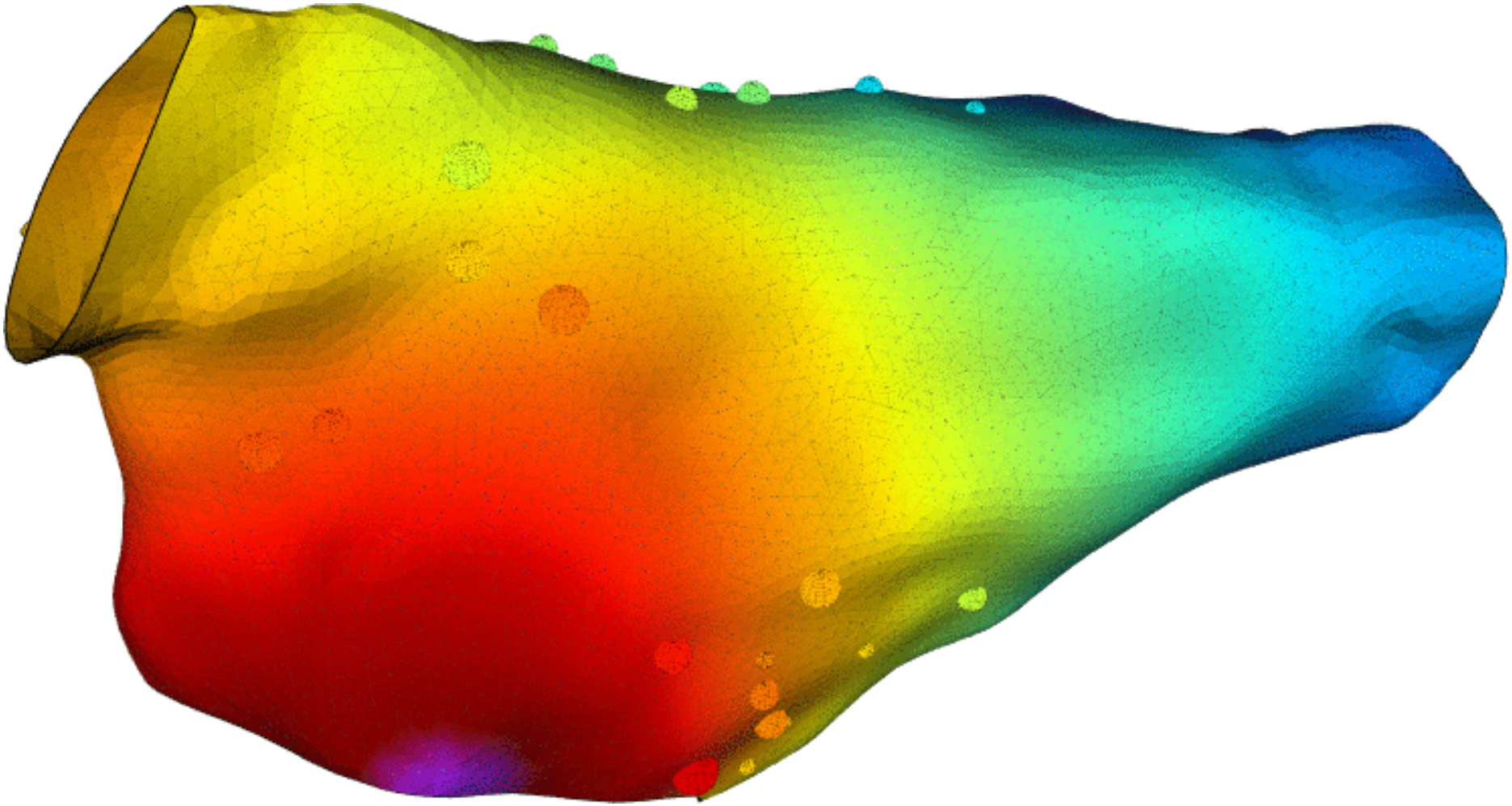


Opens interesting design questions around data collection protocols

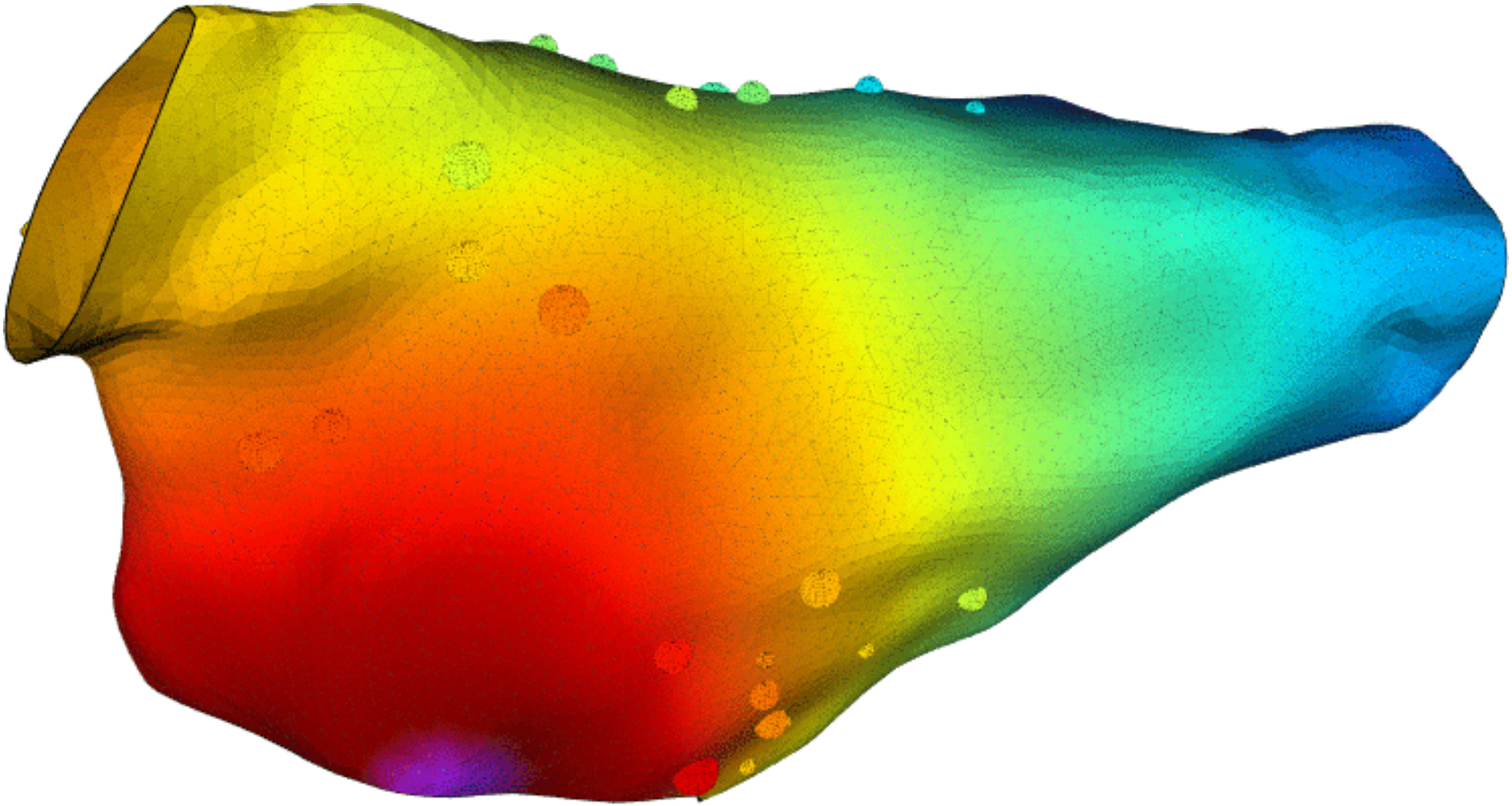
Random samples



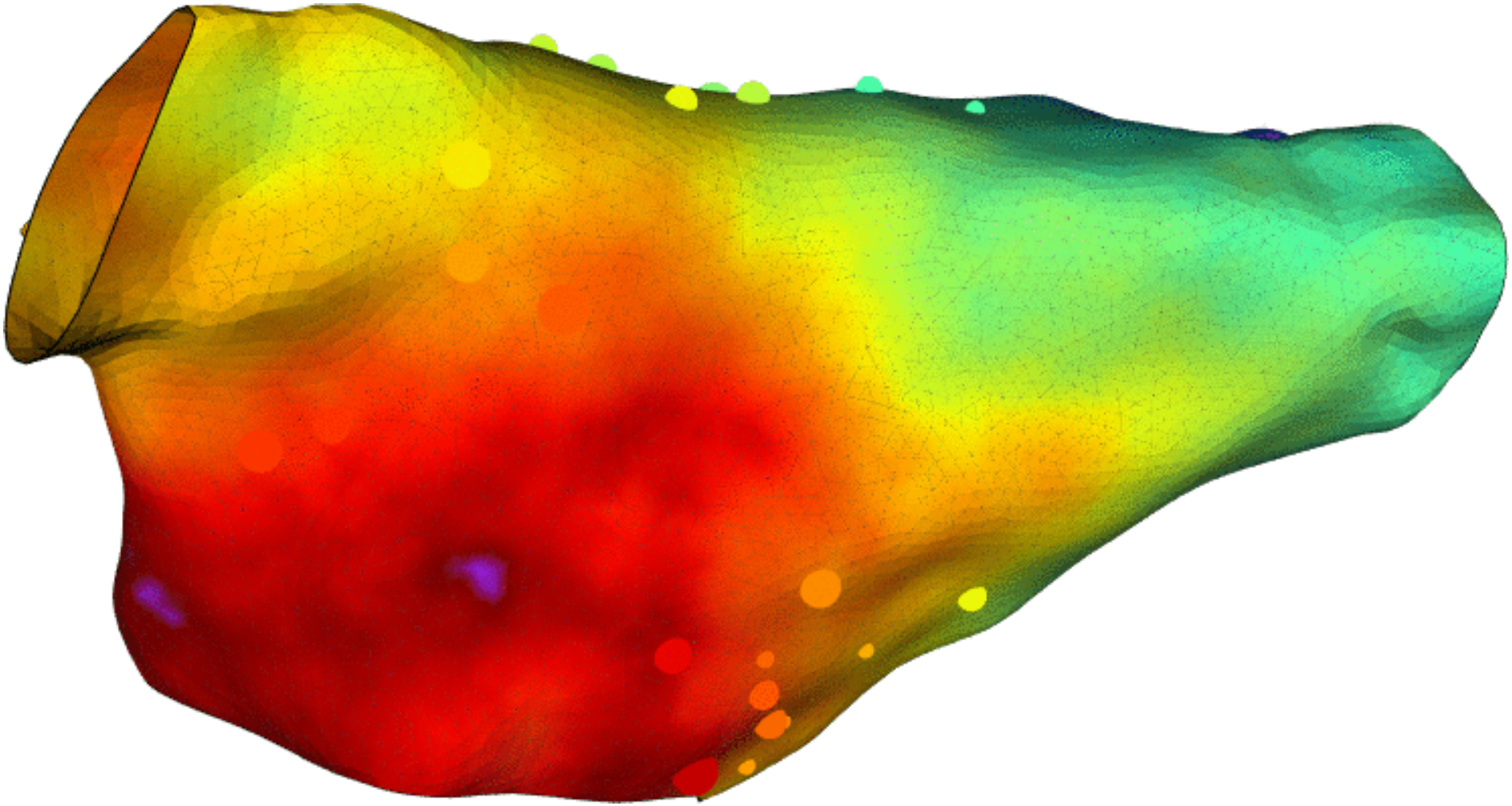
Random samples



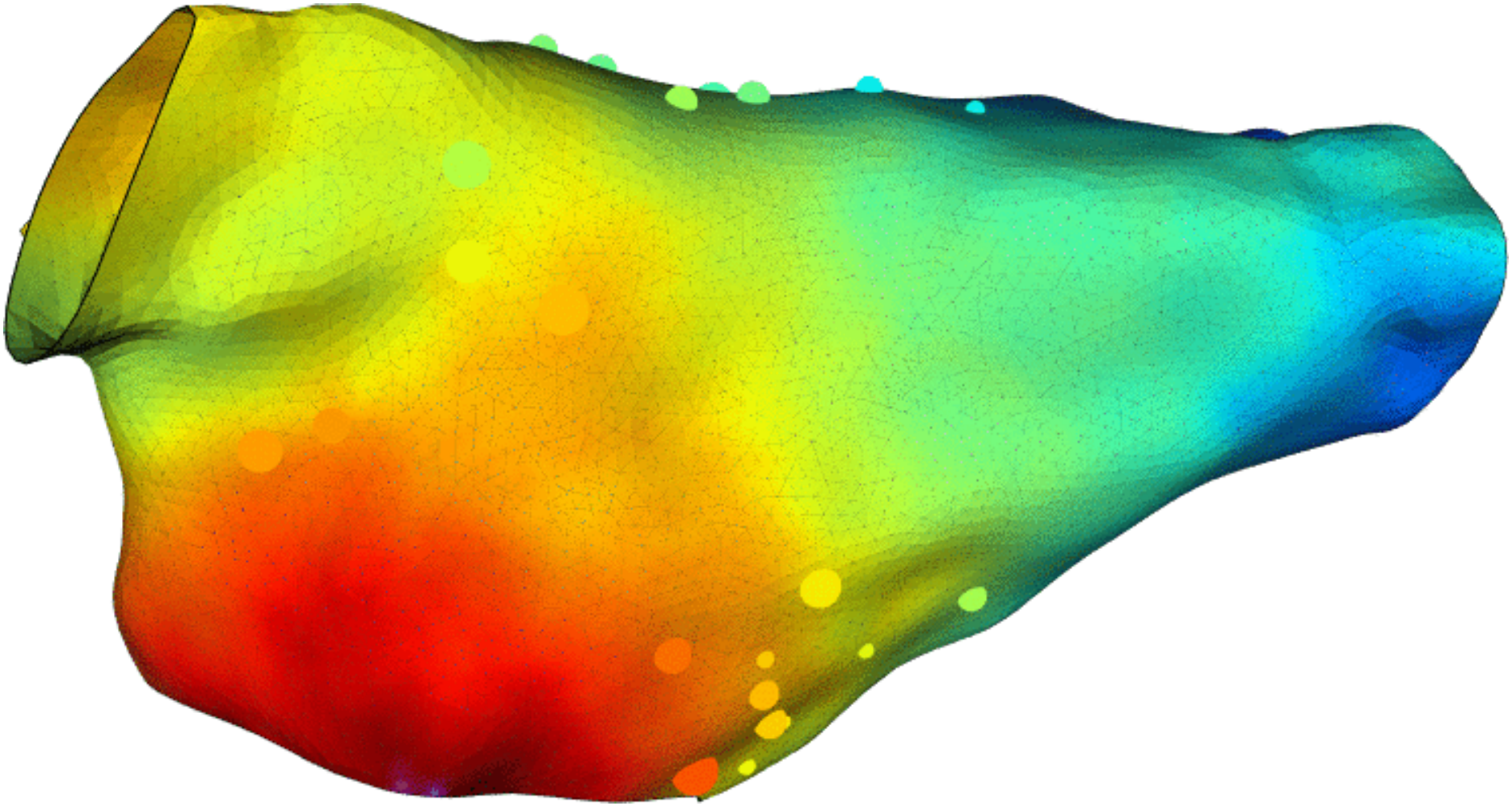
Random samples



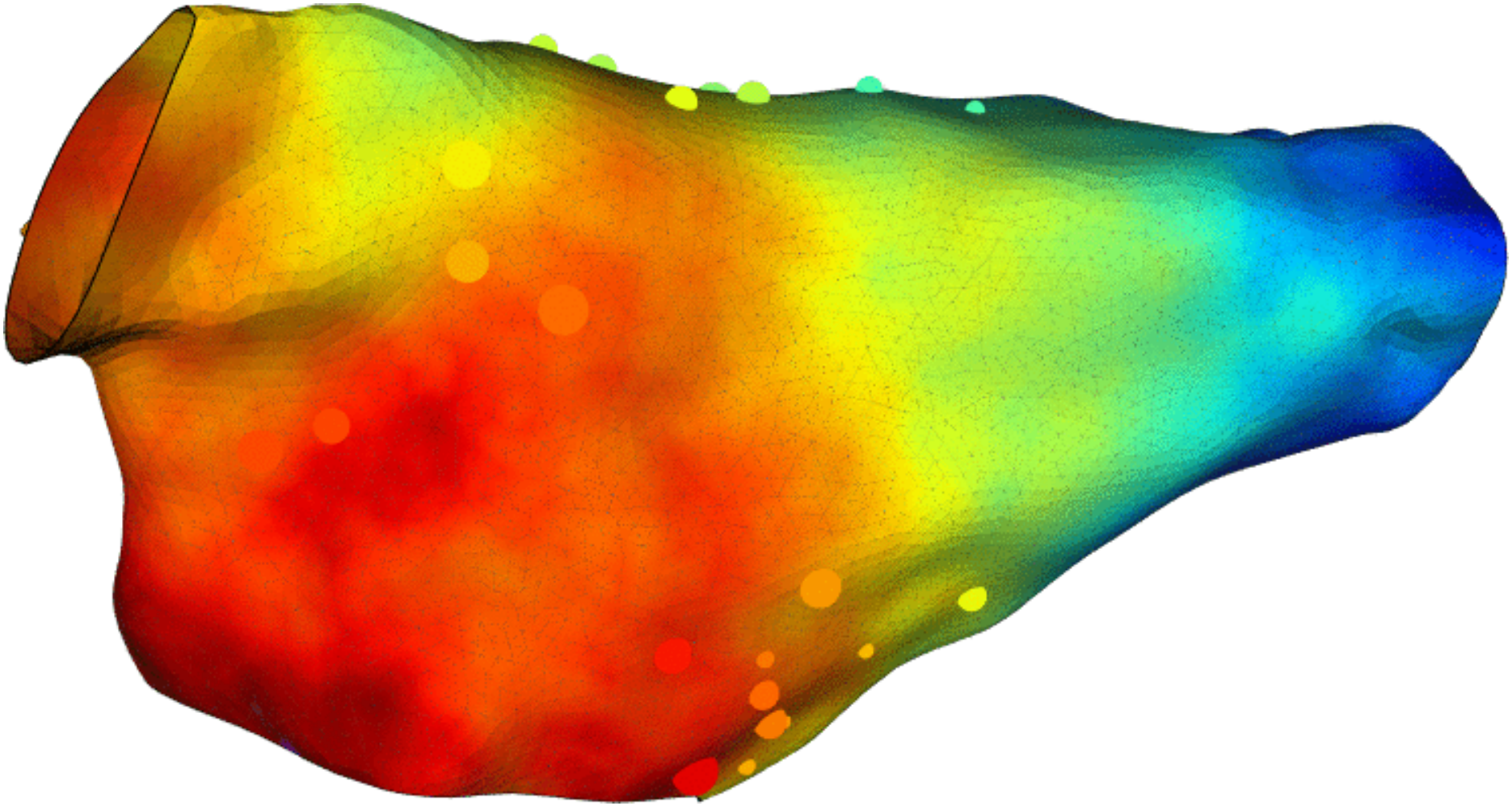
Random samples



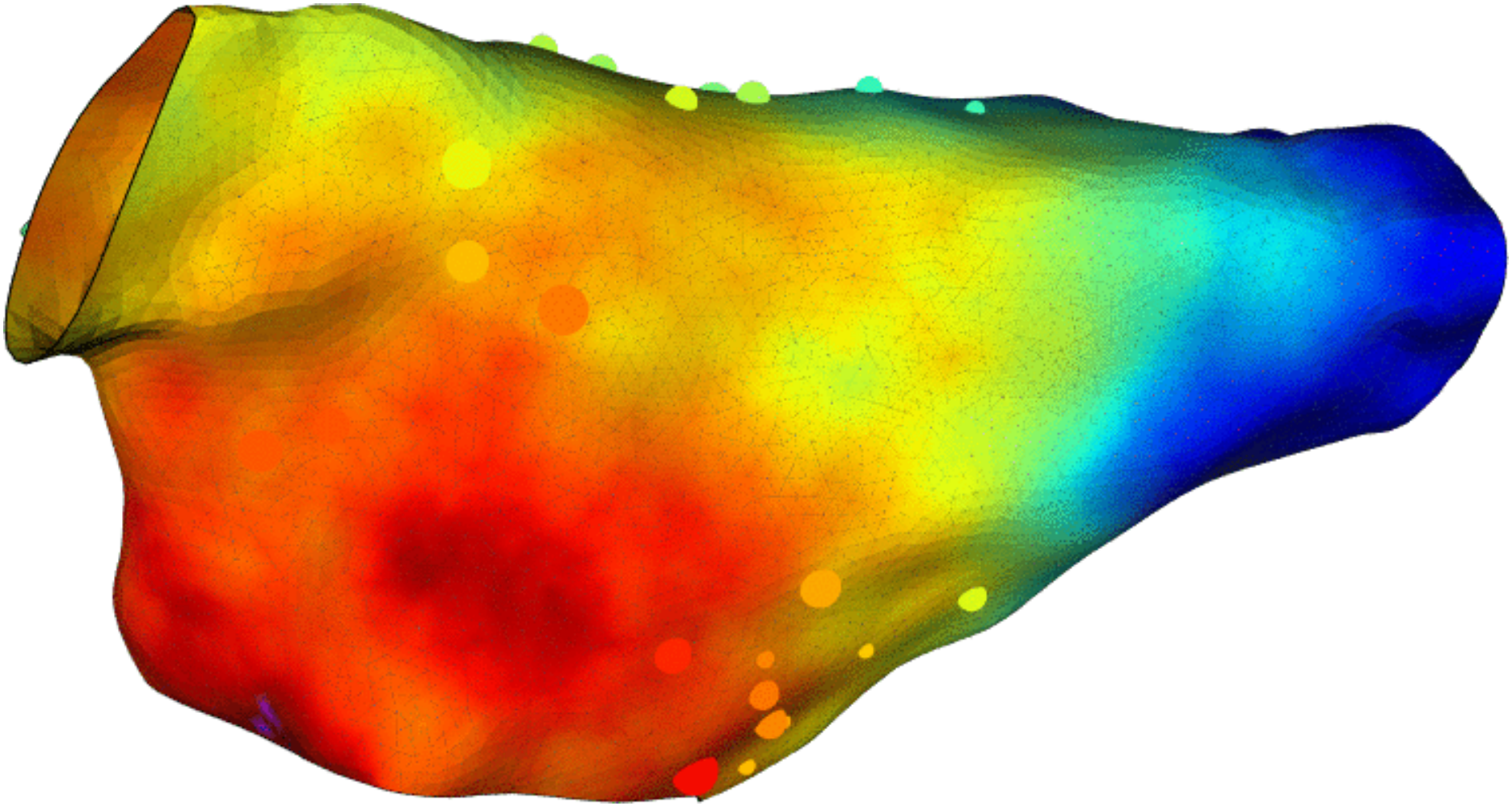
Random samples



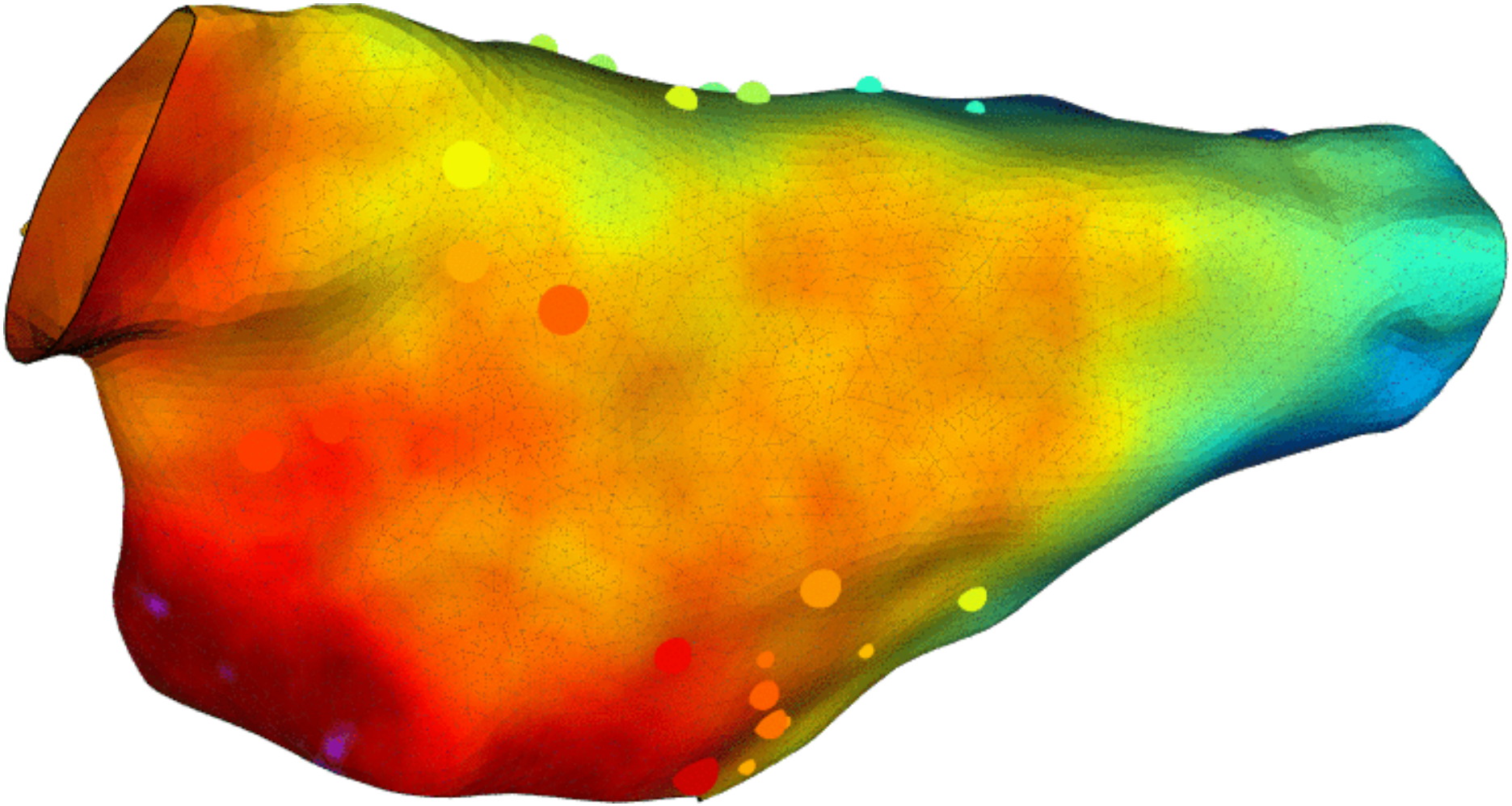
Random samples



Random samples



Random samples



Unfortunately random samples produce unphysical (non-monotonic) patterns...
Not a surprise - the GP isn't a scientific model - it doesn't 'know' it is modelling a wave.....

Problem 3: Learning tissue parameters from complex simulators - ongoing

Incorporating physics

Model cellular electrophysiology using the Mitchell-Schaeffer (MS) ionic model that captures conduction velocity and refractory restitution properties.

- 5 parameters

Problem 3: Learning tissue parameters from complex simulators - ongoing

Incorporating physics

Model cellular electrophysiology using the Mitchell-Schaeffer (MS) ionic model that captures conduction velocity and refractory restitution properties.

- 5 parameters

The electrophysiology of the left atrium is simulated $S(\cdot)$ using a monodomain equation in a shell anatomy with local activation given by the MS model, isotropic tissue conductivity, and infarcted, dense fibrotic and ablation regions modelled as non-conducting tissue

- ie 5 parameters at every location $\theta(x)$
- Think of the simulator as a black box $S(\theta)$ where $\theta \in \mathbb{R}^{5N_{cell}}$

Emulation in high dimension

S is expensive to evaluate so we wish to resort to emulation.

- given a training set $\{\theta_i, y_i = S(\theta_i)\}_{i=1}^n$ we want to learn a statistical representation of the mapping $S : \theta \to y$.

Emulation in high dimension

S is expensive to evaluate so we wish to resort to emulation.

- given a training set $\{\theta_i, y_i = S(\theta_i)\}_{i=1}^n$ we want to learn a statistical representation of the mapping $S : \theta \to y$.

θ is high dimensional, we must reduce dimension to predict well

- hard/impossible to learn a non-linear function in high dim space (all points are far apart)

We need to find a representation λ of θ that allows us to build emulator $\tilde{S} : \lambda \to y$ st

$$\tilde{S}(\lambda) \approx S(\theta)$$

eg with $\tilde{S}$ ∼GP.

Emulation in high dimension

S is expensive to evaluate so we wish to resort to emulation.

- given a training set $\{\theta_i, y_i = S(\theta_i)\}_{i=1}^n$ we want to learn a statistical representation of the mapping $S : \theta \to y$.

θ is high dimensional, we must reduce dimension to predict well

- hard/impossible to learn a non-linear function in high dim space (all points are far apart)

We need to find a representation λ of θ that allows us to build emulator $\tilde{S} : \lambda \to y$ st

$$\tilde{S}(\lambda) \approx S(\theta)$$

eg with $\tilde{S}$ ∼GP.

Conflicting pressures... The representation needs to be

- sufficiently detailed to allow us to answer the question about reentry waves (so can't be too low dim)
- low enough dimension to allow GP emulation to be done.

Options

- Given a prior distribution $p(\theta)$ we could repeat the PCA trick and use a truncated Karhunen-Loeve expansion

$$\theta(x) \approx \sum_{j=1}^{d} \lambda_j \phi_j(x)$$

where ϕ_j are the eigenfunctions of the linear operator associated with the covariance function of $p(x)$.

Options

- Given a prior distribution $p(\theta)$ we could repeat the PCA trick and use a truncated Karhunen-Loeve expansion

$$\theta(x) \approx \sum_{j=1}^{d} \lambda_j \phi_j(x)$$

where ϕ_j are the eigenfunctions of the linear operator associated with the covariance function of $p(x)$.

- Instead, we will use

$$\theta(x) = \lambda_{LAT} LAT(x) + \lambda_{Fib} Fib(x) + \sum_{j=1}^{d'} \lambda_j \phi'_j(x)$$

where ϕ'_j are a Karhunen-Loeve basis orthogonal to the handpicked basis vectors $LAT(x)$ and $Fib(x)$. Hope that the spatial heterogeneity in θ will be similar to the spatial heterogeneity in LAT and Fibrosis estimates.

Options

- Given a prior distribution $p(\theta)$ we could repeat the PCA trick and use a truncated Karhunen-Loeve expansion

$$\theta(x) \approx \sum_{j=1}^{d} \lambda_j \phi_j(x)$$

  where ϕ_j are the eigenfunctions of the linear operator associated with the covariance function of $p(x)$.

- Instead, we will use

$$\theta(x) = \lambda_{LAT} LAT(x) + \lambda_{Fib} Fib(x) + \sum_{j=1}^{d'} \lambda_j \phi'_j(x)$$

  where ϕ'_j are a Karhunen-Loeve basis orthogonal to the handpicked basis vectors $LAT(x)$ and $Fib(x)$. Hope that the spatial heterogeneity in θ will be similar to the spatial heterogeneity in LAT and Fibrosis estimates.

We then seek to solve the inverse problem for λ and thus θ.

Technical aside: PCA with two datasets

with Howard Elman

We want to solve the following optimization problem:

$$\max_x x^\top Ax \text{ subject to } x^\top x = 1 \text{ and } B^\top x = 0$$

where A is an $n \times n$ positive definite matrix. B here is a set of orthonormal basis vectors (an incomplete basis) for $\mathbb{R}^n$, i.e. B is $n \times p$. This is a non-convex problem with quadratic constraint.

Technical aside: PCA with two datasets

with Howard Elman

We want to solve the following optimization problem:

$$\max_x x^\top A x \text{ subject to } x^\top x = 1 \text{ and } B^\top x = 0$$

where A is an $n \times n$ positive definite matrix. B here is a set of orthonormal basis vectors (an incomplete basis) for $\mathbb{R}^n$, i.e. B is $n \times p$. This is a non-convex problem with quadratic constraint.

Solution:

The condition $B^\top x = 0$ is equivalent to $x \in \text{null}(B^\top) := Z$ i.e., x lies in the null space of B. We then have that

$$\mathbb{R}^n = Z \oplus Z^\perp = Z \oplus B$$

Thus $x = Zw$ for some w, and so the problem becomes

$$\max_w w^\top Z^\top A Z w \text{ subject to } w^\top Z^\top Z w = w^\top w = 1$$

This is the original eigenvalue problem $\rightarrow$ solve with the SVD!

I.e., if $A = X^\top X$, then we do the SVD of XZ.

Supervised dimension reduction

it does seem optimisitic to think that the marginal of θ would necessarily be structured so that the leading principal components contain the essential information about the response (Dennis Cook - Fisher Lecture 2007)

Supervised dimension reduction

> *it does seem optimisitic to think that the marginal of θ would necessarily be structured so that the leading principal components contain the essential information about the response (Dennis Cook - Fisher Lecture 2007)*

More generally, we would like a supervised dimension reduction approach to find a reduced space that targets both

- Regression problem
- Inverse problem

Supervised dimension reduction

> *it does seem optimisitic to think that the marginal of θ would necessarily be structured so that the leading principal components contain the essential information about the response (Dennis Cook - Fisher Lecture 2007)*

More generally, we would like a supervised dimension reduction approach to find a reduced space that targets both

- Regression problem
- Inverse problem

Currently looking into sufficient dimension reduction: find B so that

$$y \perp\!\!\!\perp \theta | B\theta$$

$\exists$ kernel sufficient dimension reduction, kernel CCA, kernel PCA, eg instead of $\theta = \sum \langle \theta, u_i \rangle u_i$ use

$$\theta = \sum \langle \phi(\theta), v_i \rangle_F v_i$$

Conclusions

It is inappropriate to be concerned about mice when there are tigers abroad. George Box

Conclusions

It is inappropriate to be concerned about mice when there are tigers abroad. George Box

Where are our tigers?

- 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
  - Can some uncertainties be ignored?
  - Do some uncertainties need more degrees of freedom to be described than others?
- Simulator discrepancy - can we model our way out of trouble?

Conclusions

It is inappropriate to be concerned about mice when there are tigers abroad. George Box

Where are our tigers?

- 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
  - Can some uncertainties be ignored?
  - Do some uncertainties need more degrees of freedom to be described than others?
- Simulator discrepancy - can we model our way out of trouble?

Thank you for listening!