History Matching for Inverse Modelling in Physical and Biological Systems

Peter Challenor
University of Exeter and the Alan Turing Institute

The Alan Turing Institute

UNIVERSITY OF EXETER

The problem

- We are interested in making decisions/inferences about the real world
- We have some numerical solution of mathematical model (simulator) of how the real world works
- And some observations of the real world
- We want to use the observations to improve the simulator to help us make our decisions/inferences

- Denote reality by R
- Measurements of R

  $$d = R + \epsilon_{data}$$
- Simulator

  $$y = f(\theta)$$

  $$y = R + \epsilon_{discrepancy}$$

Model discrepancy

Statisticians (engineers/scientists) are like artists they have an unfortunate tendency to fall in love with their models - George Box

- Input Values
  - We do not know the ‘best’ value of the inputs θ^*
- Discretisation
  - The numerical simulator is an approximation to the actual equations
- The Equations
  - The equations are inevitably only an approximation
- These last two I will call ‘structural error’

The Importance of Structural Error

- There is no reason to believe that the structural error averages out in any sense.
- We cannot write
- $R = f(\theta^*)$
- Even at the ‘best’ value of our inputs our model is not correct
- ‘All models are wrong…’

Slow Models

- The simulators we use are computationally expensive. (Hours to months)
- We can only do a limited number of runs of the simulator
- Build a surrogate model (emulator) and use that for inference

The Emulator

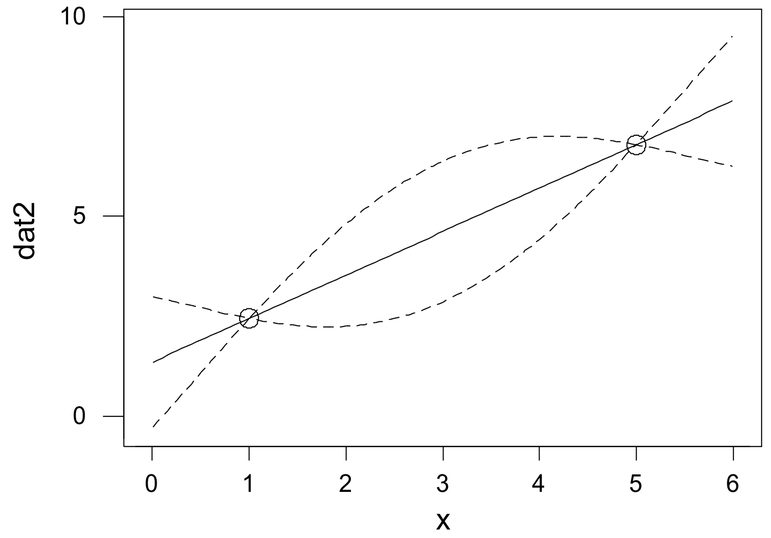
- An emulator is a surrogate model that includes a measure of its own uncertainty.
- We use Gaussian process emulators

Gaussian processes

- Gaussian processes are infinite dimensional stochastic processes all of whose marginal, conditional and joint distributions are Normal
- They are an analog of the Normal distribution for functions
- Defined by a mean function $\mu(x)$ and a covariance function $C(x_1, x_2)$
- Infinitely wide single layer neural net
- Deep Gaussian processes are available

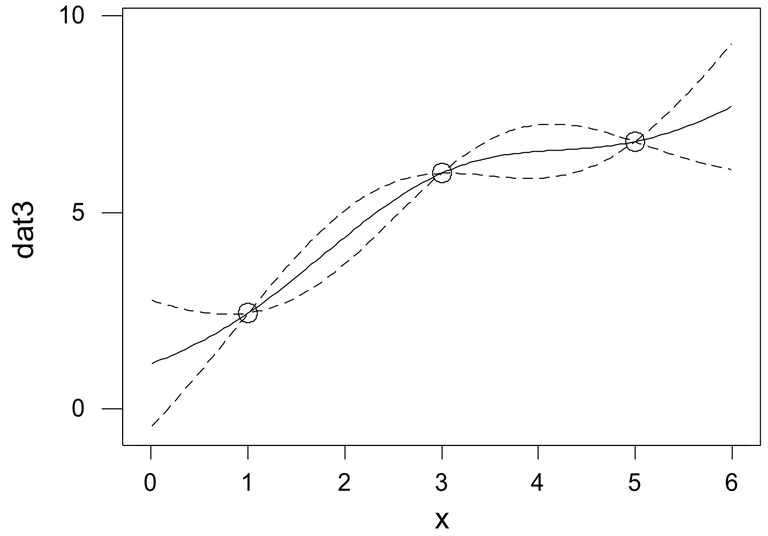
2 code runs

- Consider one input and one output
- Emulator estimate interpolates data
- Emulator uncertainty grows between data points



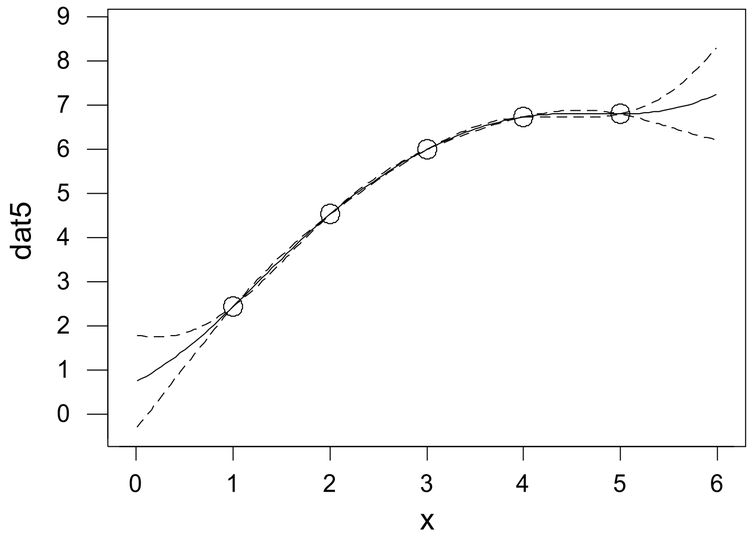
3 code runs

- Adding another point changes estimate and reduces uncertainty



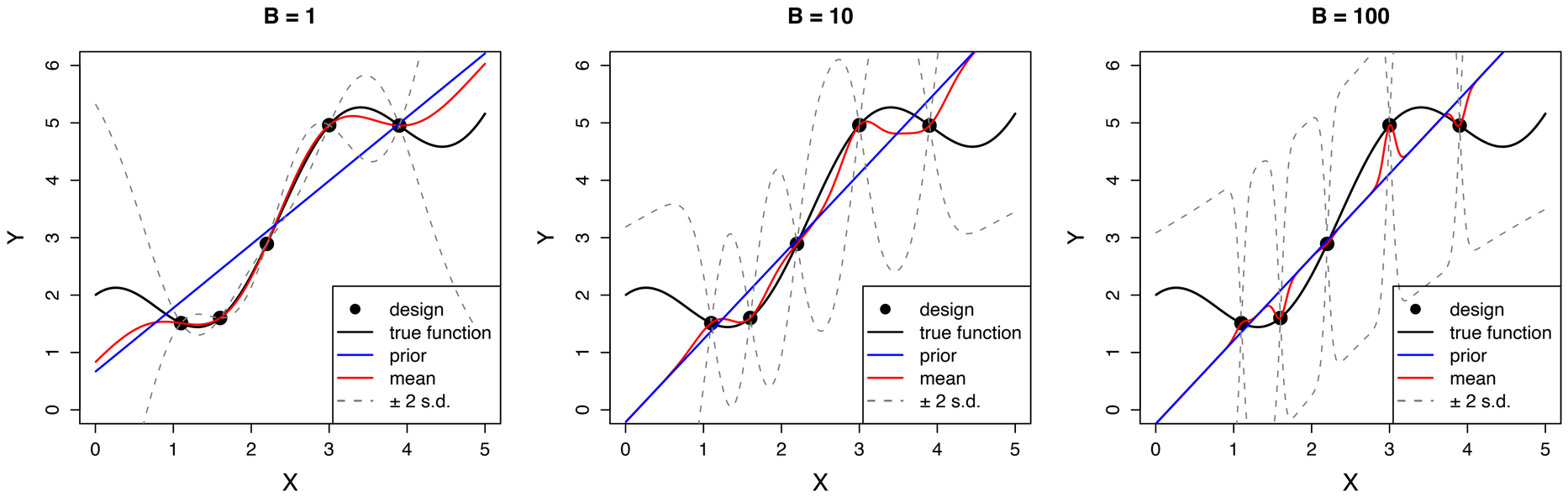
5 code runs

- And so on



Fitting the Gaussian Process Emulator

- Set up priors for the mean function and the parameters of the GP
- Run the simulator in a carefully designed experiment
- Find the posteriors for the GP parameters
- Validate the emulator (Leave one out, Bastos and O'Hagan, 2009, Technometrics)


B = 1
B = 10
B = 100
Y
X
0
1
2
3
4
5
6
design
true function
prior
mean
± 2 s.d.

Using the GP Emulator

- Prediction
- Sensitivity Analysis
- Uncertainty Analysis
- Inverse Modelling (calibration, tuning)

Inverse Modelling

- Have some observations of the real world
- And a numerical simulator
- Use the observations to make inferences about the simulator, in particular about its inputs

The Classical Methods

- Minimise a loss function (usually the squared difference) to get point estimates
- Use Bayesian Calibration to get posteriors on the inputs
- BUT because of the structural error term neither of these is appropriate
- Serious risk of over-fitting

Kennedy and O'Hagan

- Kennedy and O'Hagan (2001, JRSSB) simultaneously fit Gaussian process emulators to both the simulator and the discrepancy term.
- Works well for prediction but there are identifiability problems.
- Strong priors can get around these Brynjarsdottir and O'Hagan (2014 Inverse Problems)
- Or we could limit the form of the discrepancy term

History Matching

- An alternative is known as history matching
- Instead of trying to find the ‘best’ set of simulator inputs (θ^*) find all those sets of inputs that give implausible model outputs.
- Remove these and what is left must contain the best set
- Optimisation is hard

Implausibility

- The idea of history matching is based on the idea of implausibility

$$I_{mp}(\theta) = \sqrt{\frac{E[y - f(\theta)]^2}{V(y - f(\theta))}}$$

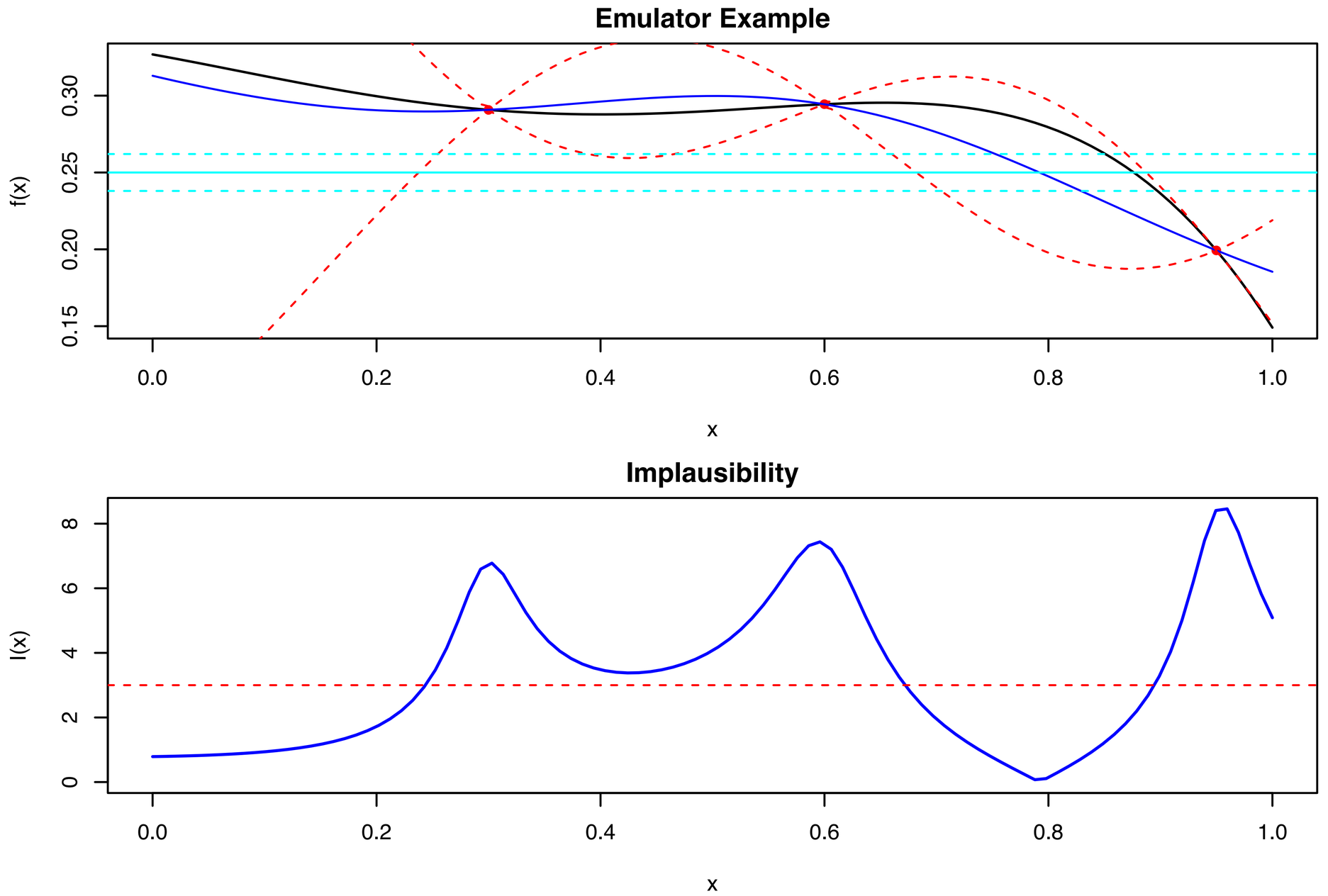
Expanding

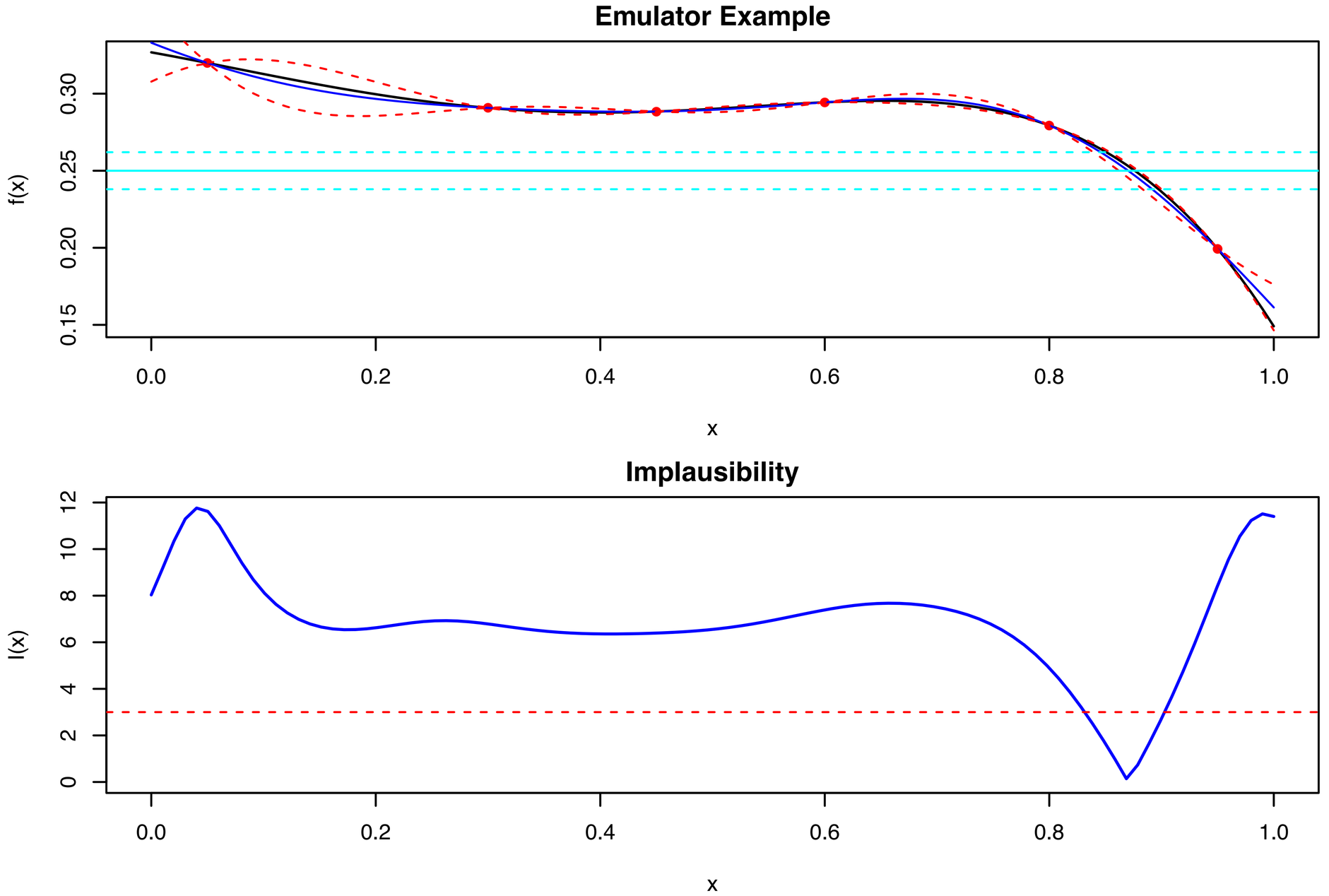
$$I_{mp}(\theta) = \sqrt{\frac{(y_{obs} - E[\tilde{f}(\theta)])^2}{\sigma^2_{emul}(\theta) + \sigma^2_{obs} + \sigma^2_{discrep}}}$$

History Matching in practice

1. Run our simulator in a designed experiment
2. Build and validate a GP emulator
3. Calculate the implausibility
4. All points with implausibility > 3 are ruled implausible (Pukelsheim (1994))
5. What remains is termed Not Ruled Out Yet (NROY) space

- Repeat steps 1-5 in waves until we reach a stopping rule


Emulator Example
f(x)
0.15
0.20
0.25
0.30
0.0
0.2
0.4
0.6
0.8
1.0
x
Implausibility
I(x)
0
2
4
6
8
0.0
0.2
0.4
0.6
0.8
1.0
x


Emulator Example
f(x)
0.15
0.20
0.25
0.30
0.0
0.2
0.4
0.6
0.8
1.0
x
Implausibility
I(x)
0
2
4
6
8
10
12
0.0
0.2
0.4
0.6
0.8
1.0
x

Stopping Rules

- NROY shrinks to some prespecified value and we do a K&OH calibration in this reduced space
- NROY becomes so small we can effectively use it as a point estimate
- NROY disappears completely. The simulator and the data are not compatible

NROY Disappearing

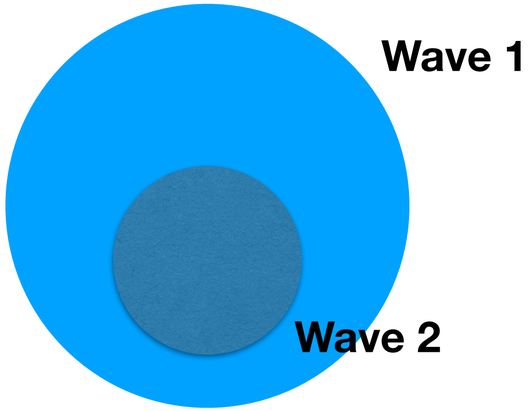
- If the simulator and the data are incompatible NROY will go to zero (all points are implausible)
- If you do classical calibration this will not be apparent. You will get the set of inputs closest to the data (even if they are a long way away) and this estimator will appear to get less and less uncertain even though the simulator and data are incompatible
- The discrepancy between the simulator and the reality, $\sigma^2_{discrep}$, is too small. By increasing this term we can make NROY finite again.
- This gives us a ‘tolerance to error’ to discuss with the modeller/decision maker.

A Non-Trivial Example

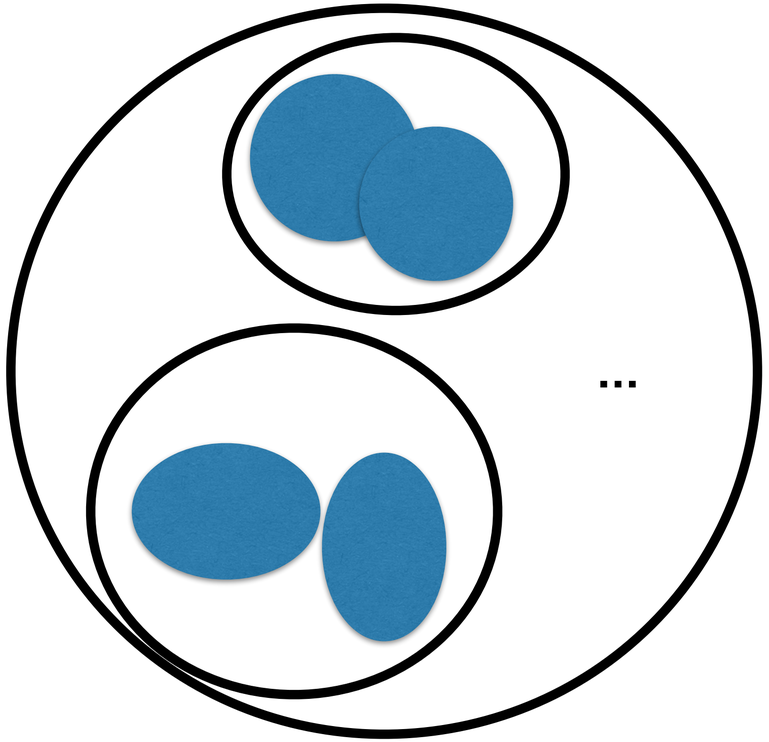
Diastolic Heart Disease

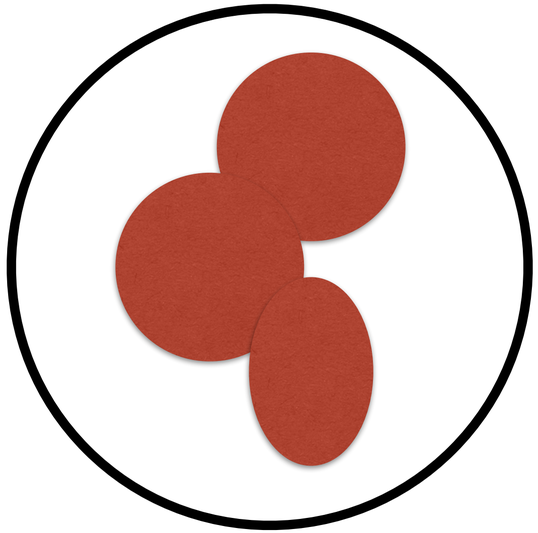
- Diastolic heart failure is an untreatable cardiac condition.
- Affects about 450,000 people in the UK
- The heart becomes stiff and cannot behave normally.
- 9 unsuccessful drug trials.
- Could be more than one condition
- Can a numerical cardiac model help with diagnosis?
- As a case study we compared a single healthy patient with a single unhealthy one.

NROY for patient A



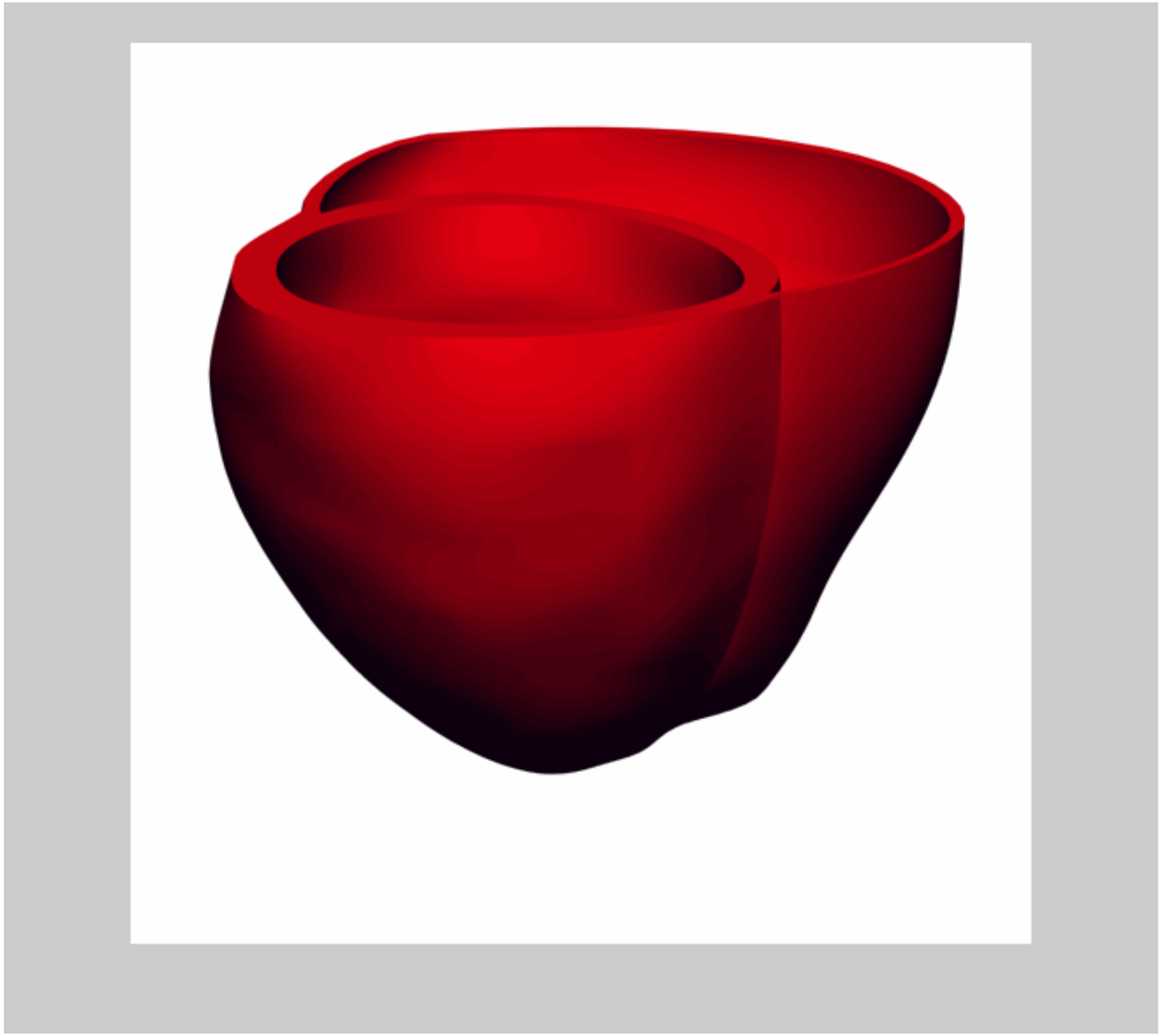
NROY for patients with condition





NROY for patients without condition

A Cardiac Model



Thanks to Steve Neiderer, KCL/St Thomas

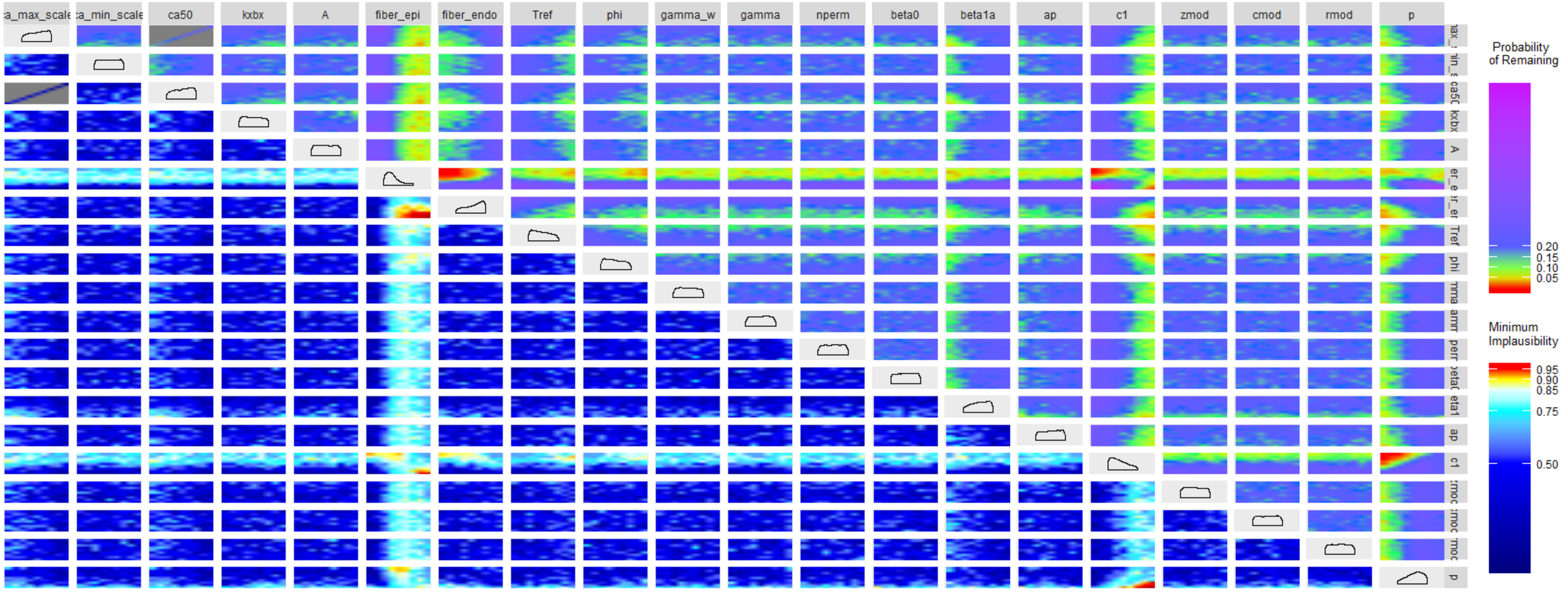
Preprocessing the data

- We treat all the simulator output (in space and time) as a single vector.
- We reduce the dimensionality by using a modified version of PCA

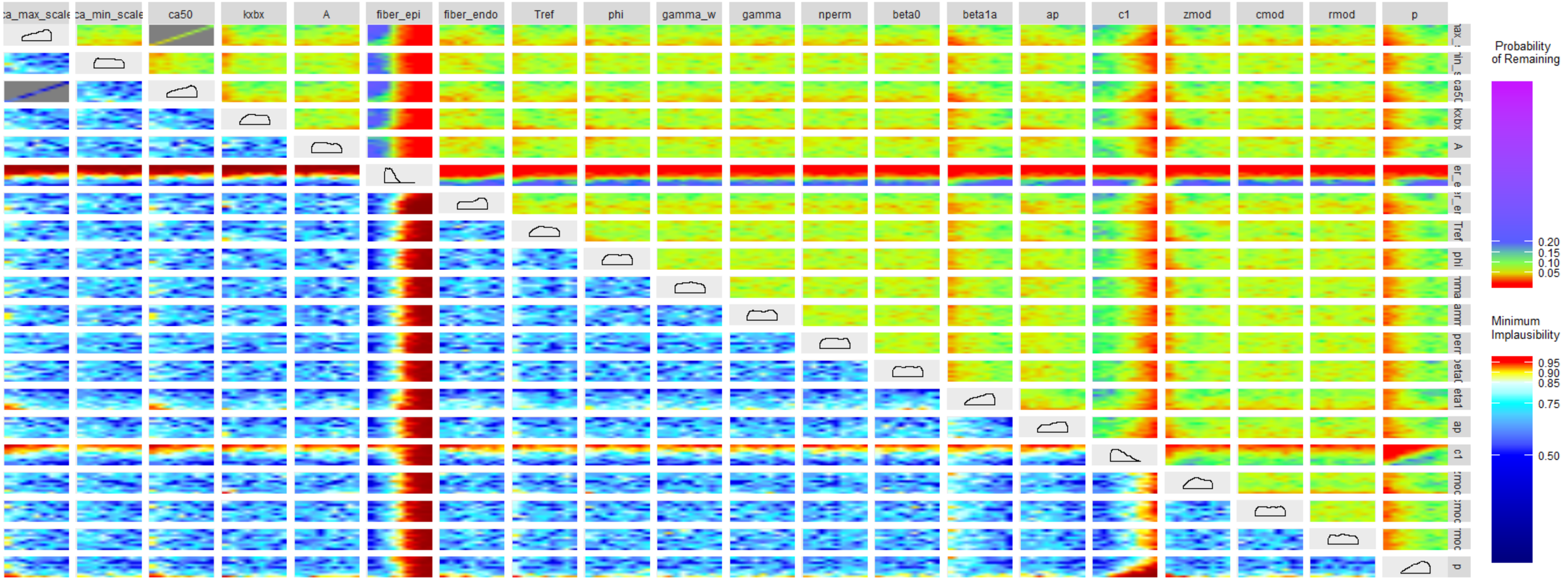
Salter et al (2019)Uncertainty Quantification for Computer Models With Spatial Output Using Calibration-Optimal Bases. JASA. http://doi.org/10.1080/01621459.2018.1514306

- The results are shown for the first principal component; the second is similar
- Initial analysis - elicited no discrepancy. NROY goes to zero.
- Elicit more reasonable discrepancy term
- Compare to an MRI scan for a healthy patient

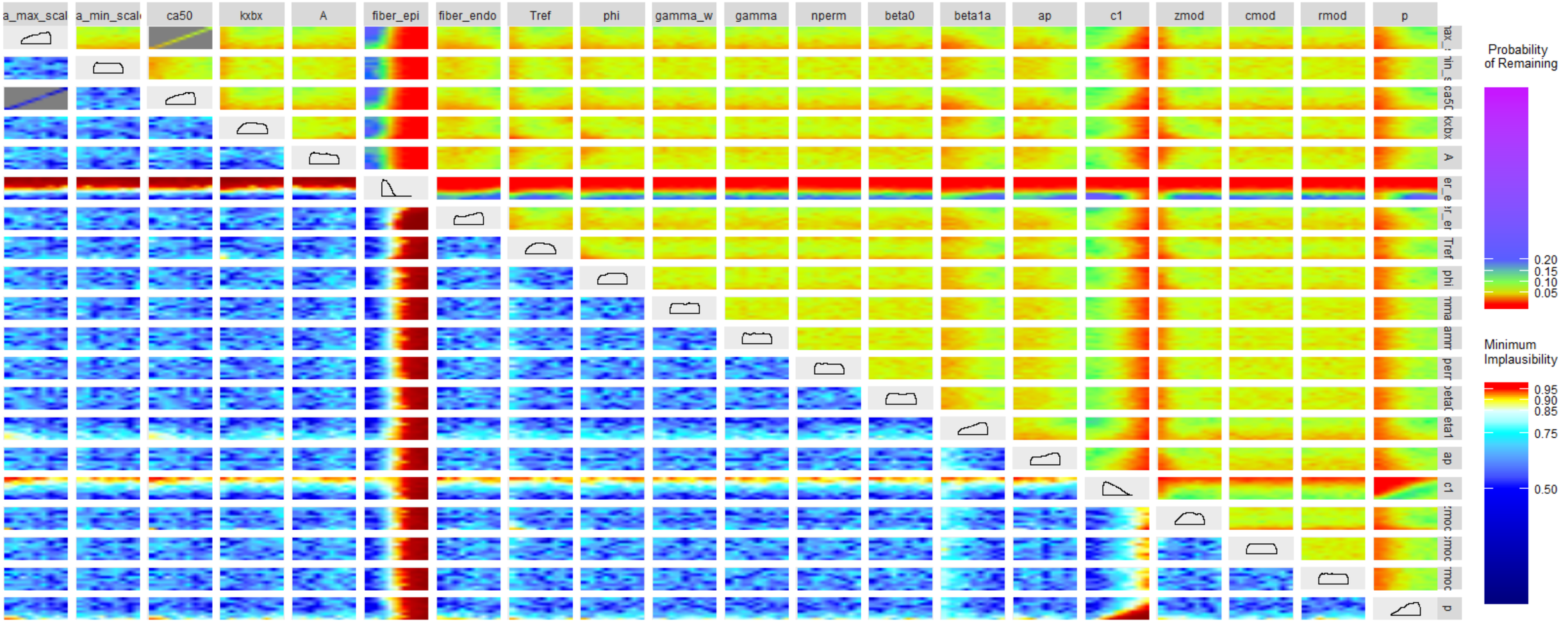
Wave 1: 25% of the parameter space remains



Wave 2: 6% of the parameter space remains



Wave 3: 5% of the parameter space remains



Results

- History matching for the unhealthy patient reduces to a few percent
- The final NROYs do not overlap
- Need more patients, more MRI scans

Advantages and Disadvantages of History Matching

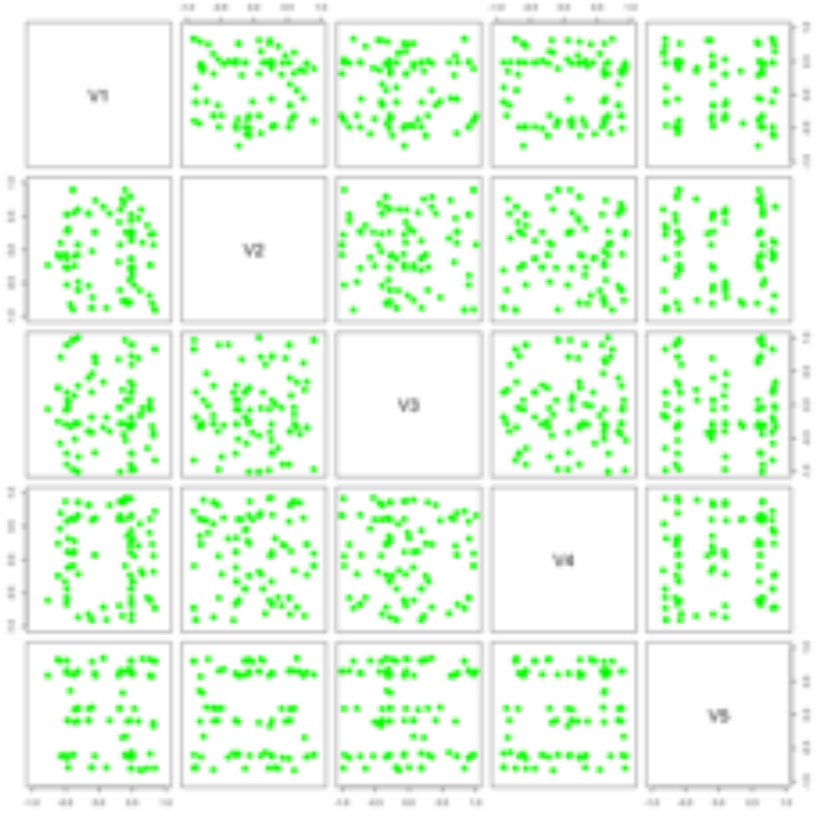
- Gives a range not a point value or posterior
- Not probabilistic - geometric
- NROY can become empty
- Bayesian calibration finds the closest point to the data

Issues

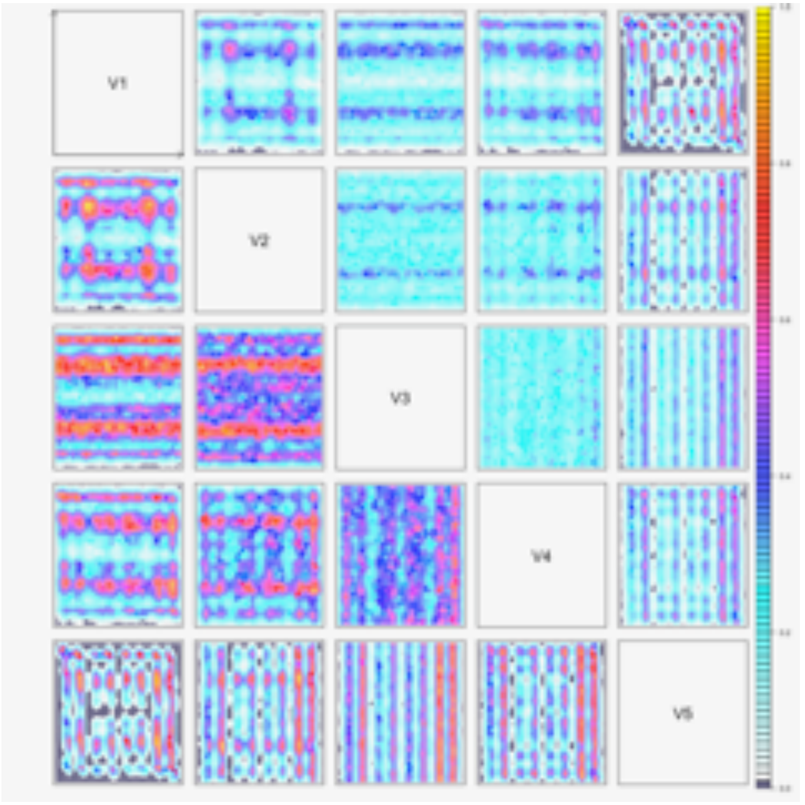
- Design
- Multiple metrics
- Perfect models
- Relationship to ABC
- Discrepancy
- Physical and biological systems

Design

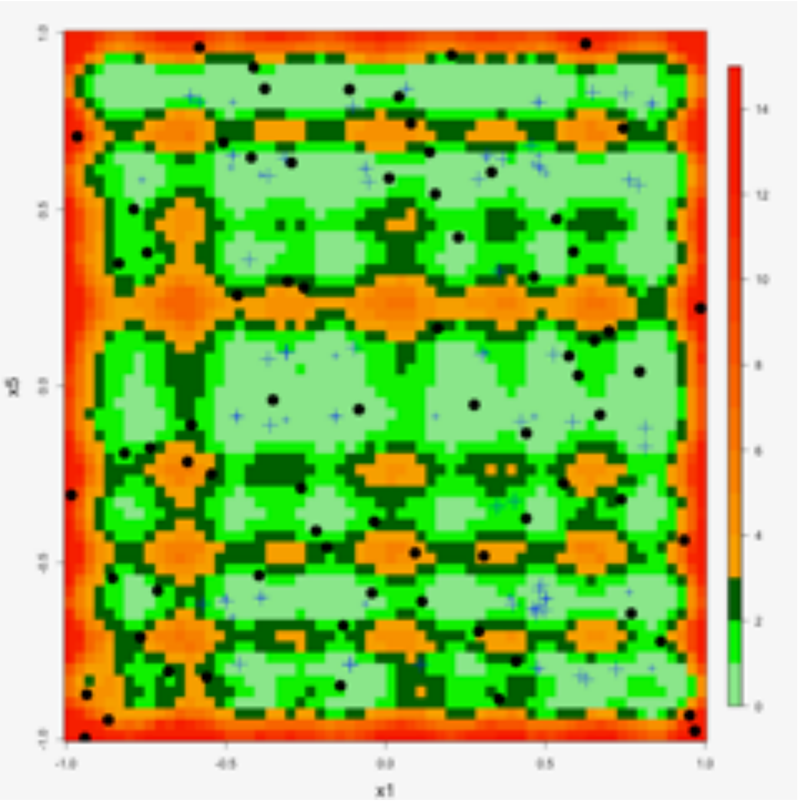
- Design for Wave 1 is standard
- For later waves there are issues
  - Put all new points in NROY?
- Optimal Design (Volodina Thesis)



Green dots are good points found by evaluating the true model



Depth plot of NROY space at wave 4



After 1 wave, just looking at the 2 most active parameters
(blue +s true good points, black dots wave 1 design,
green = NROY, orange/red = not NROY)

James Salter, Tim Dodwell

Multiple Metrics

- Combining metrics
  - *Max(Imp)* (Vernon et al 2010)
  - Second, third largest
- Multivariate methods

$$I_{mp}(\theta)^2 = (y - E(f(\theta)))^T Var(y - E(f(\theta)))^{-1}(y - E(f(\theta)))$$

‘Perfect’ models

- In a ‘perfect’ model V_{disc} = 0
- Add ‘perfect’ data -> V_y=0

$$Imp = \frac{(y - E(f(x))^2}{V_{emul}}$$

- Both of these go to zero as we increase the number of model runs (under our assumptions)
- But which goes fastest?

Physical vs Biological Systems

- One of the components of the implausibility measure is σ^2_{data}
- For physical systems it is reasonable to think of this as a number
- The data error is the ‘measurement error’
- All jet engines are the same; all rabbits are different
- Variation between and within populations

Relationship to ABC

- Approximate Bayesian Computation (ABC) rejects models that are far from the data
- It is similar to history matching (Wilkinson et al)
- The calculations are the same but the motivation is different

Discrepancy

- Hard to specify
- ‘Unknown unknowns’
- Unmodelled processes
- Assumptions in the model
- Discretisations
- How could the model be improved?

Physical vs Biological Systems

- One of the components of the implausibility measure is σ^2_{data}
- For physical systems it is reasonable to think of this as a number
- The data error is the ‘measurement error’
- All jet engines are the same; all rabbits are different
- Variation between and within populations

Uncertainty in Biological Systems

- Calibrating the model on the population (large variance) is not very precise
- Sub-populations have less variability - more precise calibration (need a better model)
- Need to decide why you need a calibrated model and for what purpose
- Personalised medicine?

Conclusions

- History matching is an alternative solution to inverse models
- Related to ABC
- No optimisation required

Thanks

- ICERM for organising this

- You for listening

- James Salter, Hossein Mohammedi, Danny Williamson, Victoria Volodina, Tim Dodwell at Exeter

- Michael Goldstein, Ian Vernon at Durham, Richard Wilkinson at Sheffield, Steve Neiderer at KCL