Inverse Methods without Optimisation

Peter Challenor
University of Exeter

UNIVERSITY OF EXETER

Overview

- Introduction
- History Matching
- Gaussian Process Emulators
- History Matching Again
- Some Thoughts on Discrepancy
- Conclusions

Inverse Problem

- We have a function y=f(x)
- We collect some data on y and we want to make some inferences about x
- We can set up some loss function or likelihood e.g.

$$\sum (y_i - f(x_i))^2$$

- And optimise it

Bayesian Methods

- Alternatively we can use Bayes theorem to do the inversion

$$p(x|y) = \frac{p(y|x)p(x)}{p(y)}$$

- We can calculate the posterior distribution of x|y

But …

- In general these calculations are difficult
  - They are expensive.
  - Likelihoods and posteriors are often multimodal
  - It doesn't take into account the fact that the model f(.) isn't perfect

An Alternative

- Don’t try to find the ‘best’ set of inputs ($\boldsymbol{x}$)
- Find inputs ($\boldsymbol{x}$) that are *implausible* given the data (y)
- This is a lot easier
- No optimisation
- No sampling posterior

History Matching

- Set up a measure of the distance between the data and the model prediction

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

- If this distance is too far. That value of $\boldsymbol{x}$ is implausible

- We can expand the variance term to give

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

- Where V_y is the variance of y
- and $V_{f(x)}$ is the variance of $f(x)$
- For Imp >3 we say that the inputs (x) are implausible (Pukelsheim (1994))

- but could be expensive to run in which case we can only compute Imp in a small number of places

- Replace $f(x)$ with an approximation $f^*(x)$

- This is known as an emulator (or surrogate or metamodel)

The Gaussian Process Emulator

- We use Bayesian Gaussian Process emulators
- Set up a prior model for the emulator
- Run the model in a designed experiment to span space in a sparse manner
- Calculate the posterior
- Validate the posterior

Gaussian Processes

- A Gaussian Process is a distribution over functions
- A stochastic process where all marginal, joint and conditional distributions are Normal
- It is defined by a mean function and a covariance (or correlation) function

The Mean Function

- Although the theory allows us to have a general mean function we normally use a linear basis function

$$\mu(x) = h(x)^T \beta$$

Often we take the ***h(.)*** functions as monomial terms in a polynomial expansion

The Covariance Function

- Assuming stationarity the covariance function consists of two parts

$$\sigma^2 c(||x_1, x_2||)$$

σ^2 is the variance of the Gaussian Process

$c(.,.)$ is the correlation function that gives the correlation between two points x_1 and x_2 as a function of the distance between them

Some Correlation Functions

Correlation function	Formula	Differentiable?
Exponential power ($\delta_0 < 2$)	$\exp\left(-d^{\delta_0}\right)$	No
Gaussian	$\exp\left(-d^2\right)$	Infinitely
Matérn (general)	$\frac{2^{1-\delta_0}}{\Gamma(\delta_0)}\left(\sqrt{2\delta_0}d\right)^{\delta_0} K_{\delta_0}\left(\sqrt{2\delta_0}d\right)$	$\lfloor \delta_0 \rfloor$ times
Matérn, $\delta_0 = \frac{3}{2}$	$\left(1+\sqrt{3}d\right)\exp\left(-\sqrt{3}d\right)$	Once
Matérn, $\delta_0 = \frac{5}{2}$	$\left(1+\sqrt{5}d+\frac{5}{3}d^2\right)\exp\left(-\sqrt{5}d\right)$	Twice

The Prior

- Set up a prior mean function

$$\mu(x) = h(x)^T \beta$$

- Often a low order polynomial sometimes more complex
- Choose the form of the correlation function
- Usually use vague priors on the GP parameters

Design

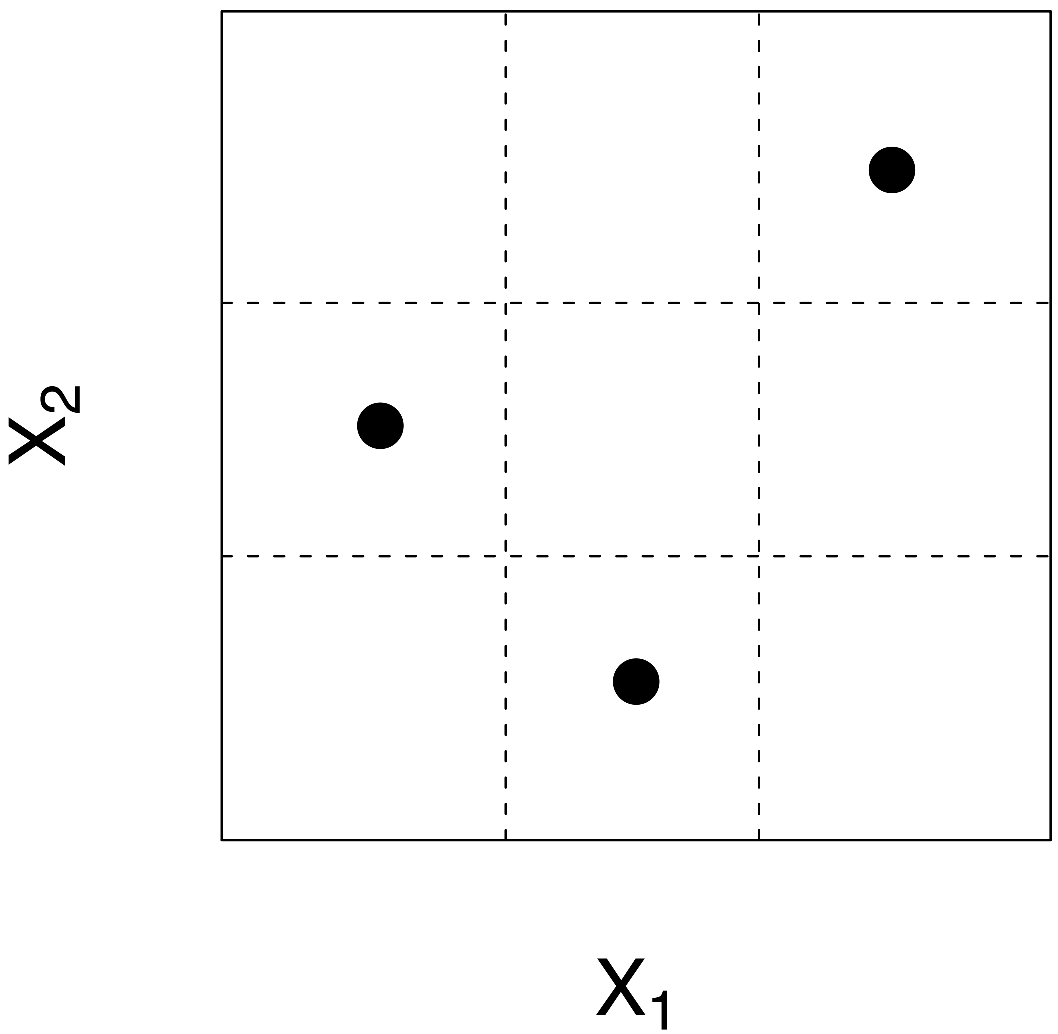
- We want designs that are sparse but also space filling
- Latin hypercubes
- Low discrepancy sequences (Sobol sequence)

The Latin Hypercube

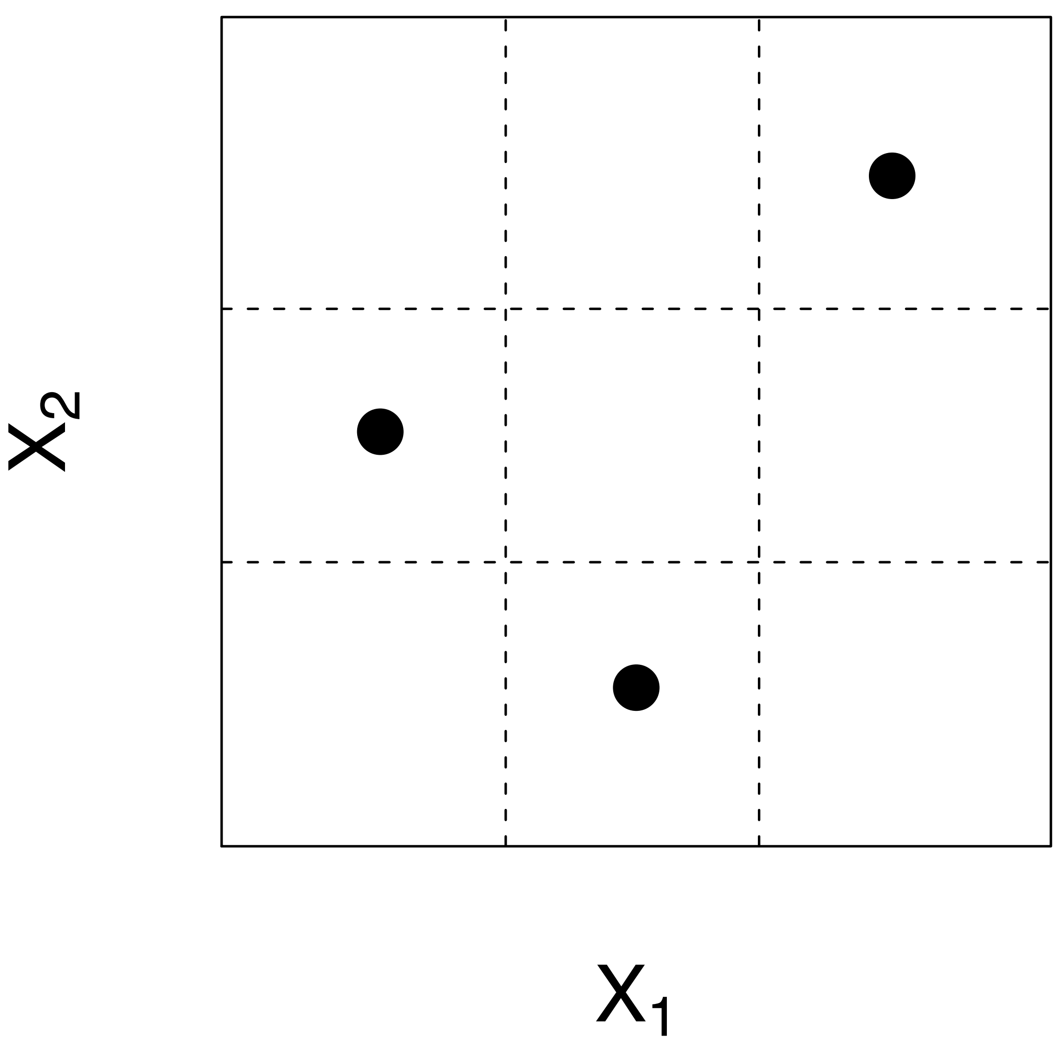
- Latin hypercubes fill space on the margins but not jointly
- What is a good space filling Latin hypercube?
- Maximin
- Orthogonal designs

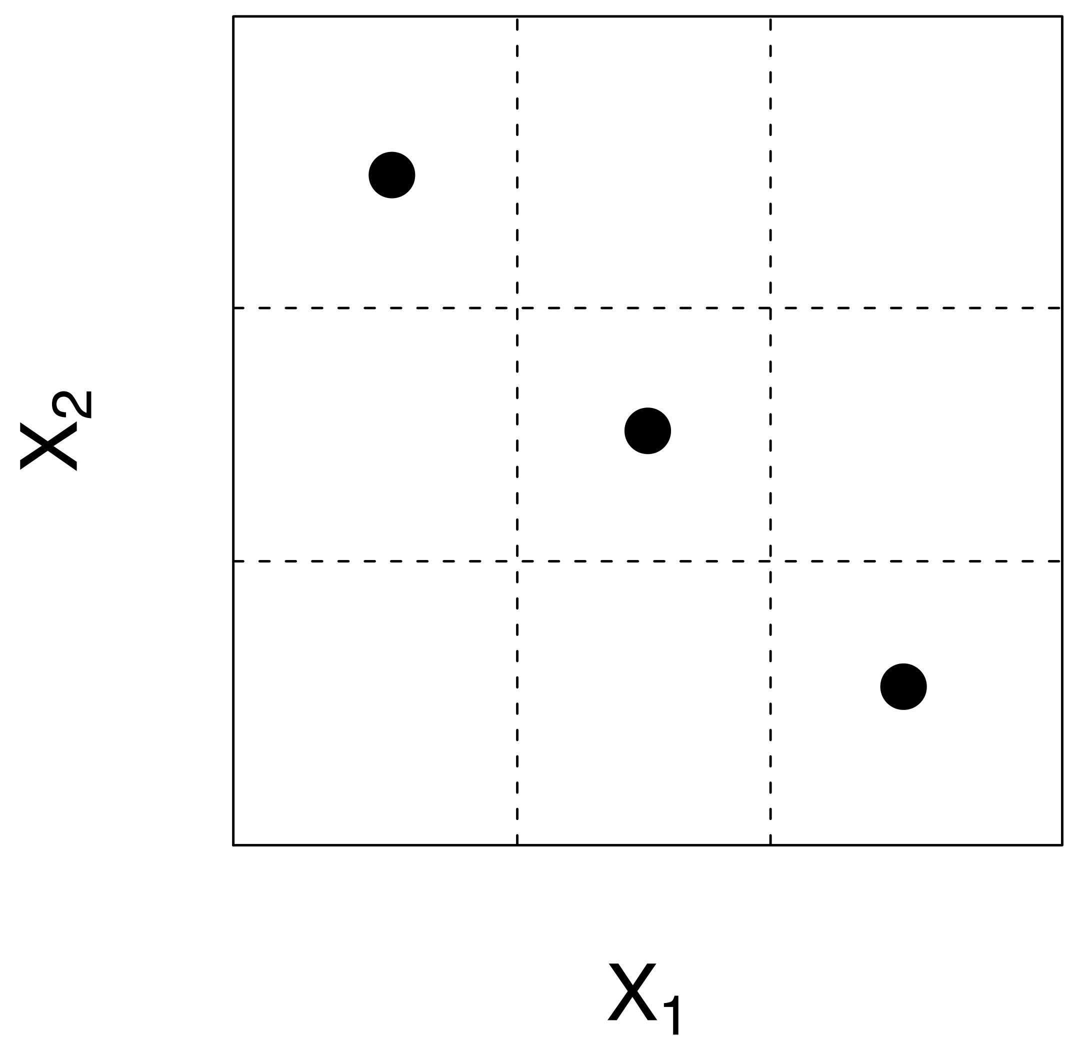
The Latin Hypercube

The Latin Hypercube

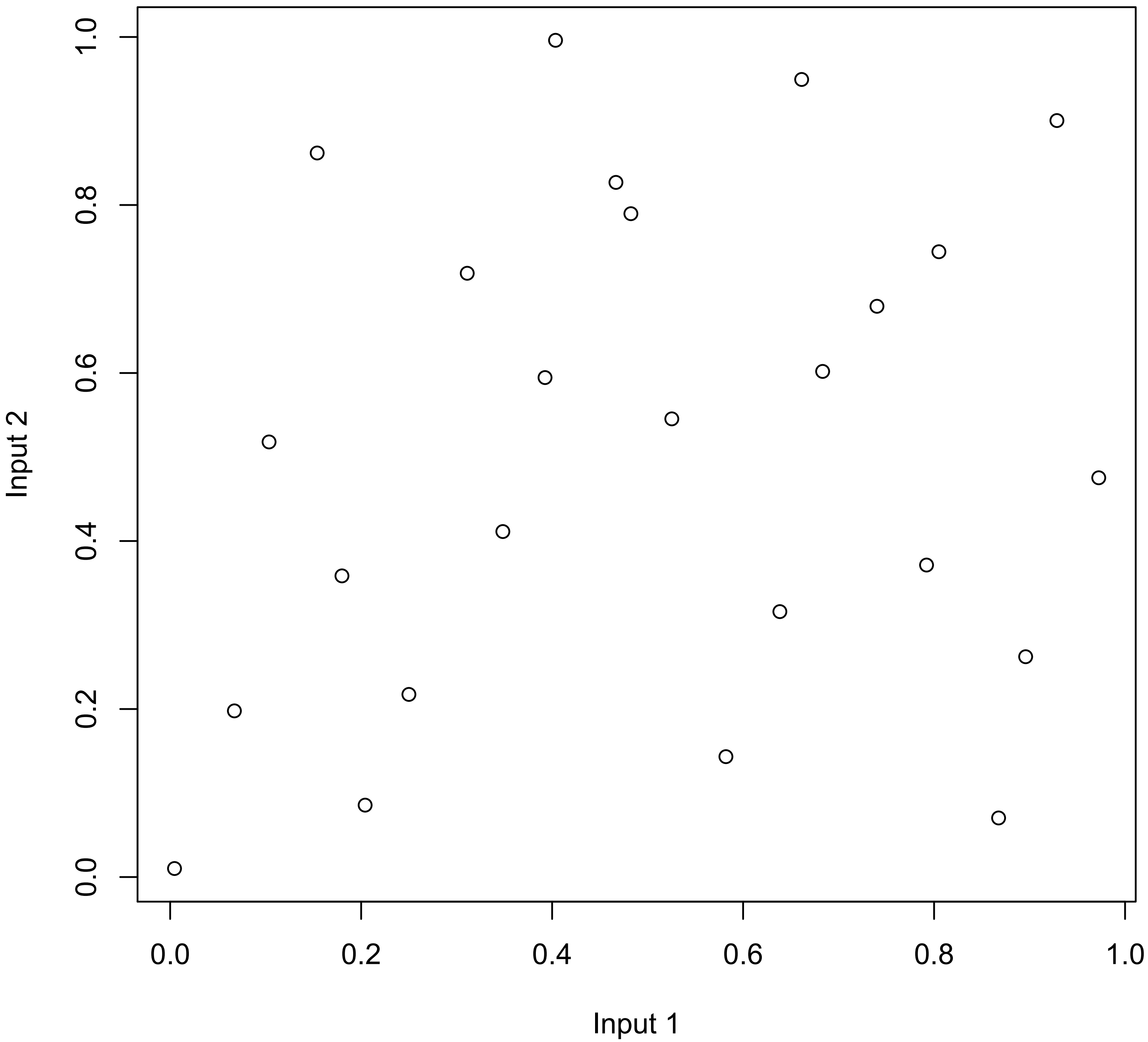


The Latin Hypercube

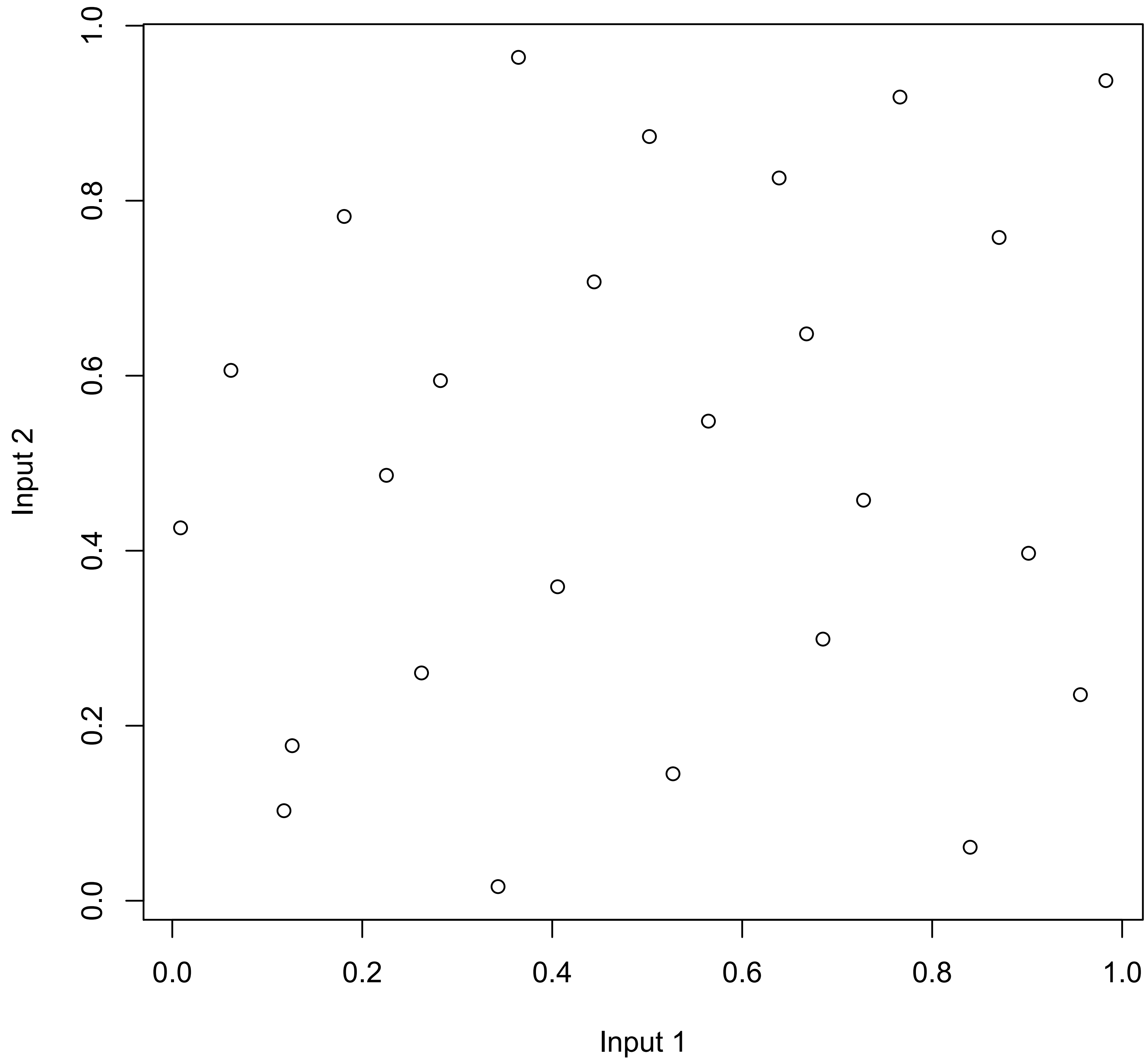




A Latin Hypercube



A maximin LHC



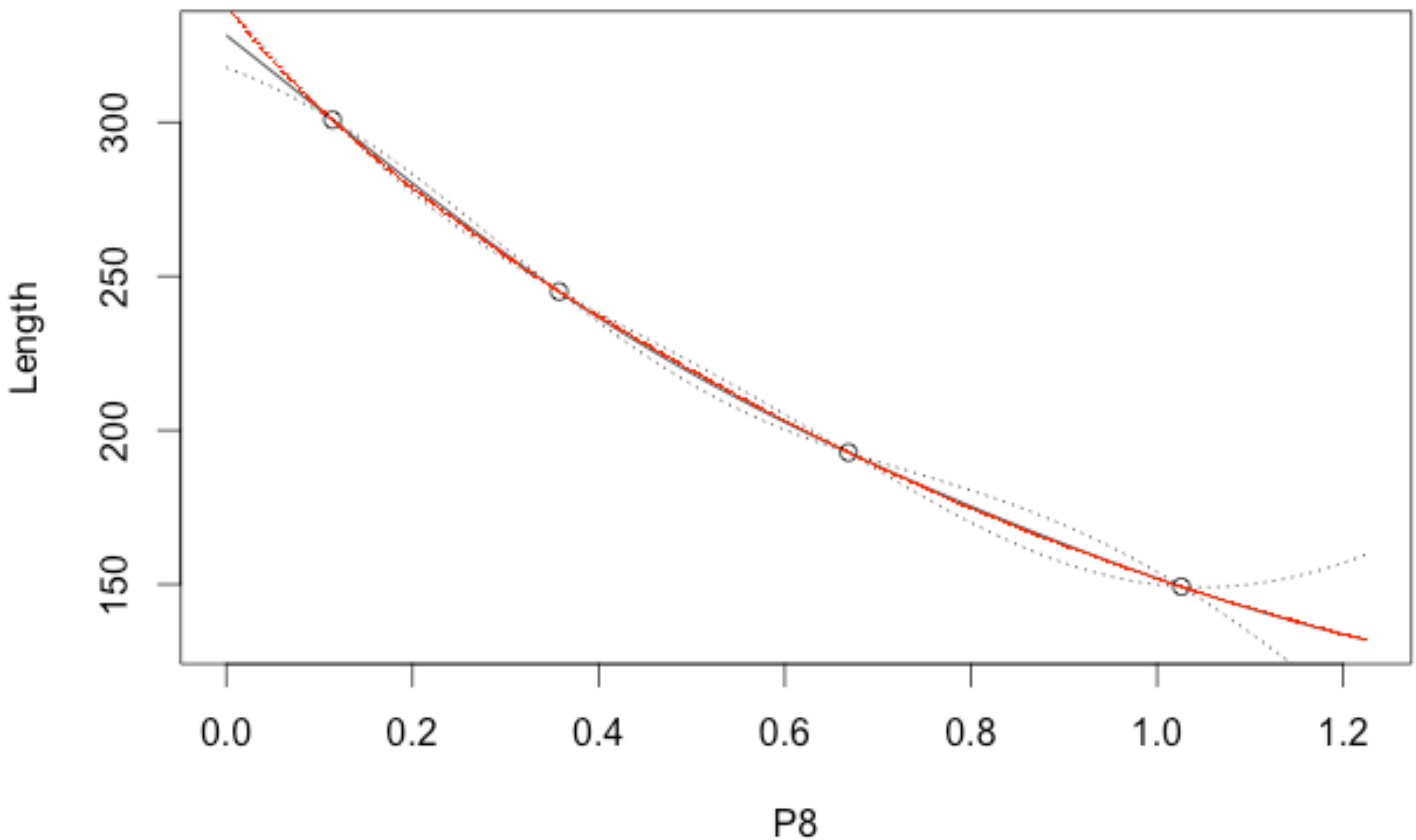
The Posterior

- Once we have run the simulator/model we can calculate the posterior GP - the emulator

- Note the posterior is a stochastic process

- Often we just show the mean and variance

Emulator for a single input in a cardiac cell model



Validation

- Building emulators isn't hard
- Building good emulators can be
- Important to validate any emulator
    - Leave one out
    - Residual Analysis - Bastos and O'Hagan (2009)
    - Separate validation experiment

The Forward Problem

- We can use a validated emulator for:-
  - prediction
  - sensitivity analysis
  - uncertainty analysis

Calibration

- Kennedy and O'Hagan (2001) calibrate a model using two GPs fitted simultaneously. One as an emulator and one for the model discrepancy
- In practice hard to do without prior information (see Brynjarsdóttir and O'Hagan. 2014)

History matching revisited

- Returning to history matching
- The implausibility equation is

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

- Expanding the variance as before gives

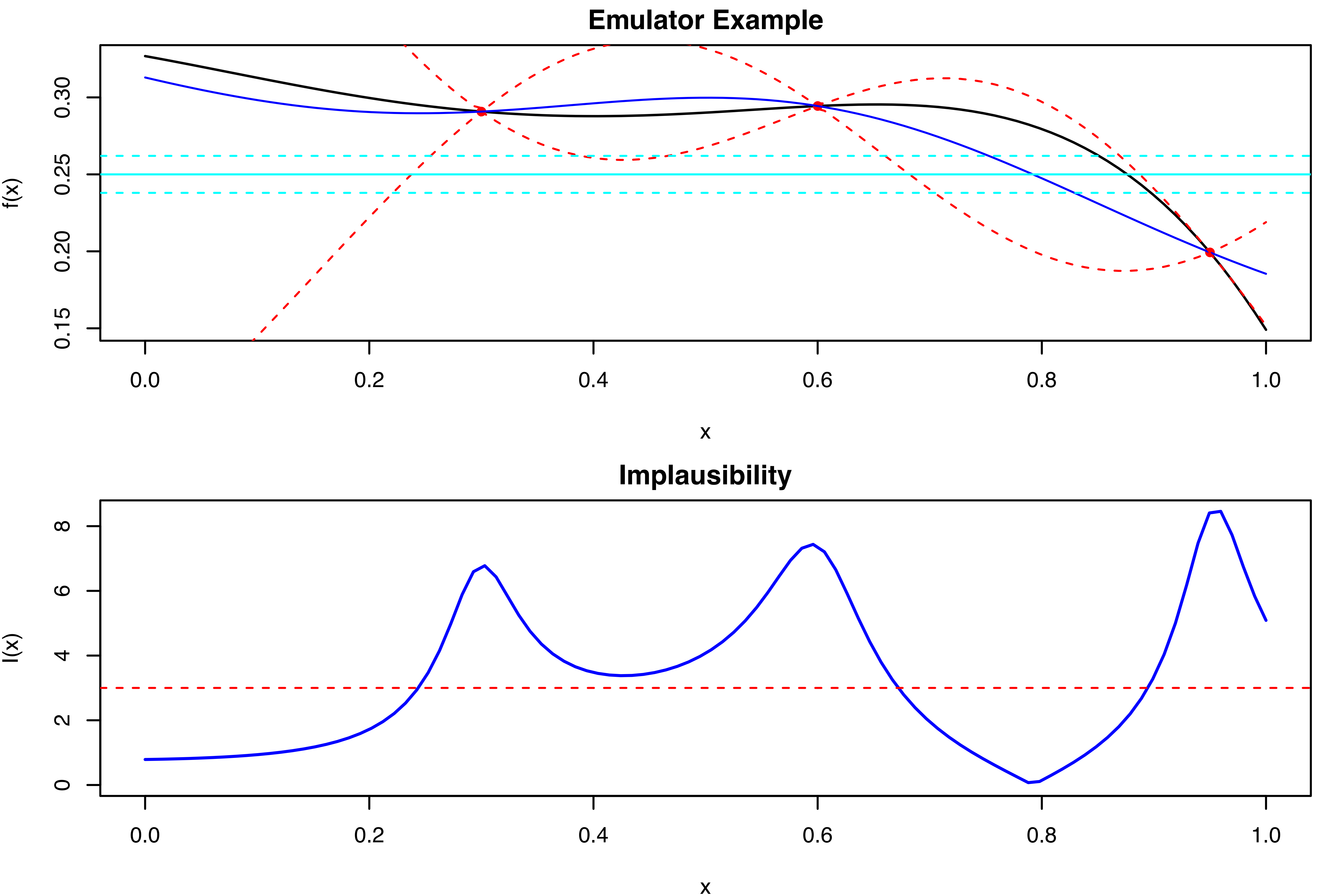
$$Imp = \sqrt{\frac{y - E(f(x))^2}{V_y + V_{emul} + V_{disc}}}$$

- V_y is the variance of the data y
- V_{emul} is the emulator variance
- V_{disc} is the model discrepancy

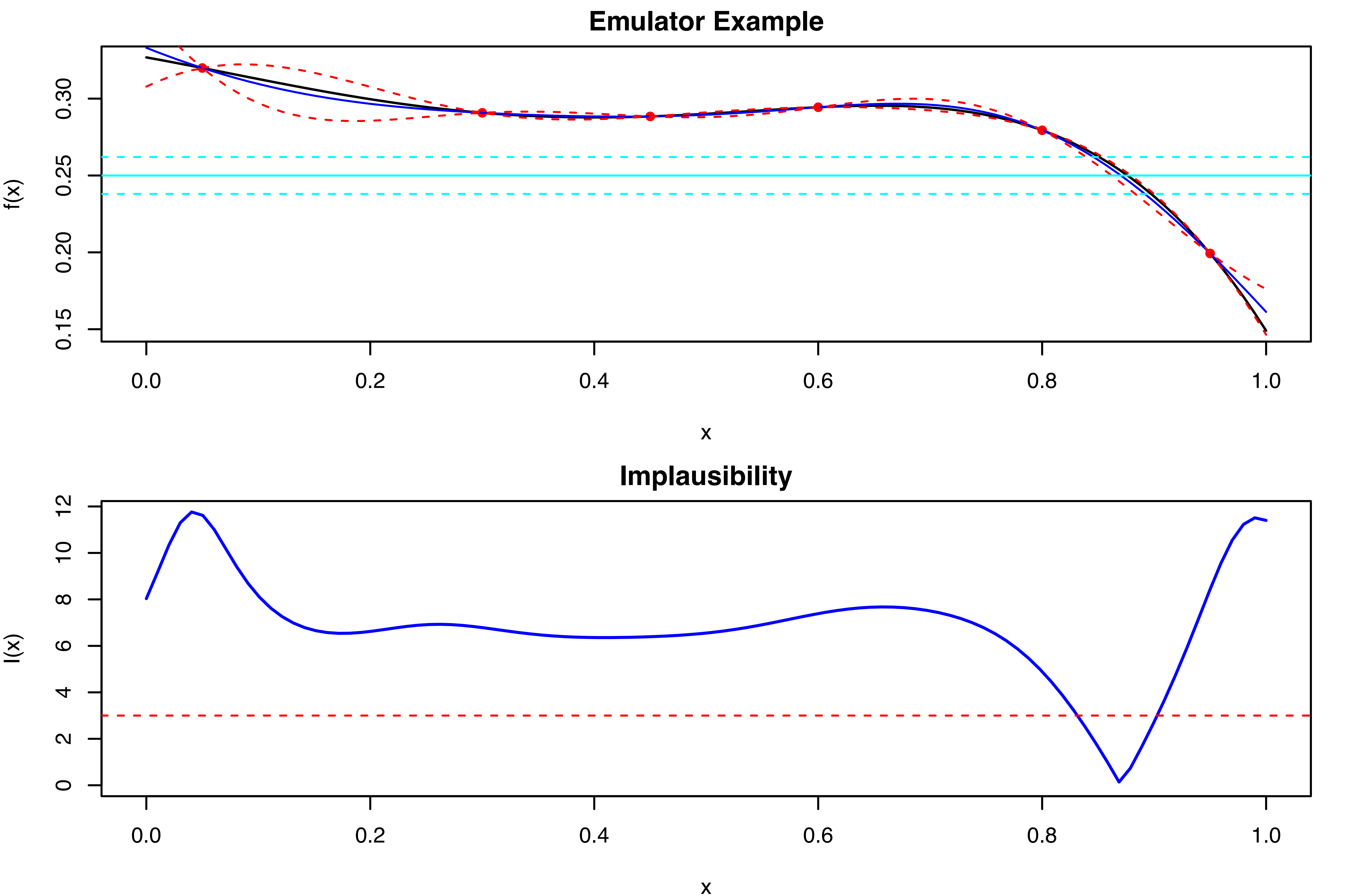
Procedure

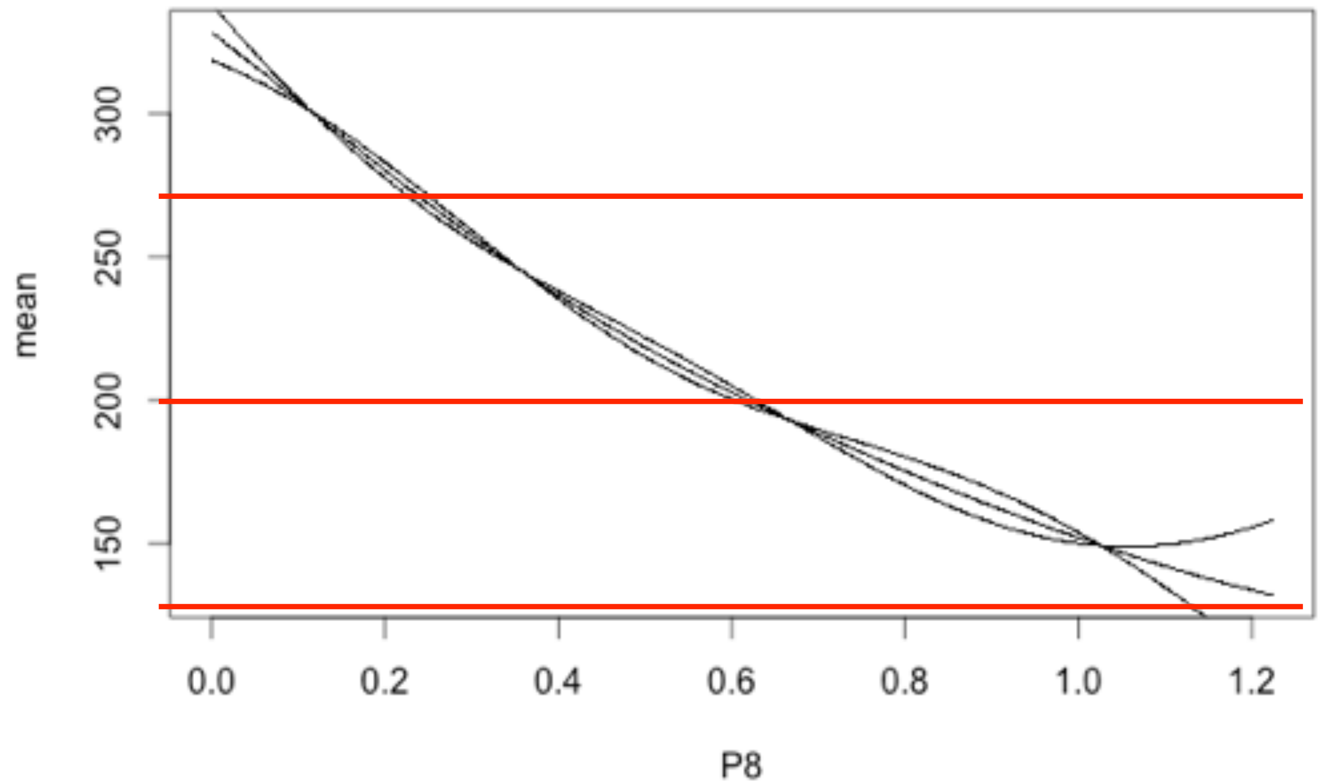
- Collect data
- Run designed experiment
- Build emulator
- Perform history matching
- All points with `Imp` <3 deemed *not implausible*
- If we have many metrics take `max(Imp)`
- These constitute the *Not Ruled Out Yet* (NROY) space

- Design additional experiment within NROY space (wave 2)
- Rebuild emulator
- History match
- Repeat until NROY is either small enough or does not shrink
- At which point we may need more data

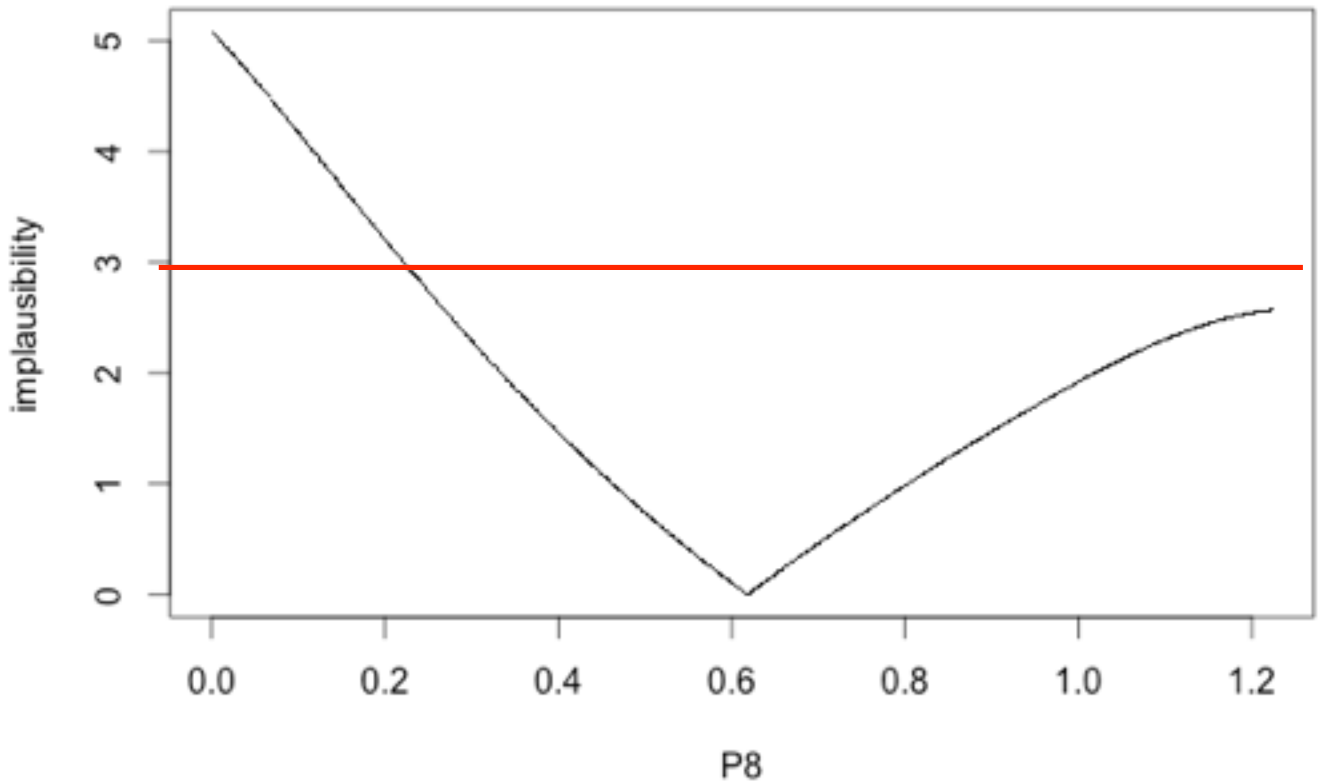


Emulator Example
f(x)
0.15
0.20
0.25
0.30
0.0
0.2
0.4
0.6
0.8
1.0
x
Implausibility
I(x)
0
2
4
6
8
0.0
0.2
0.4
0.6
0.8
1.0
x


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



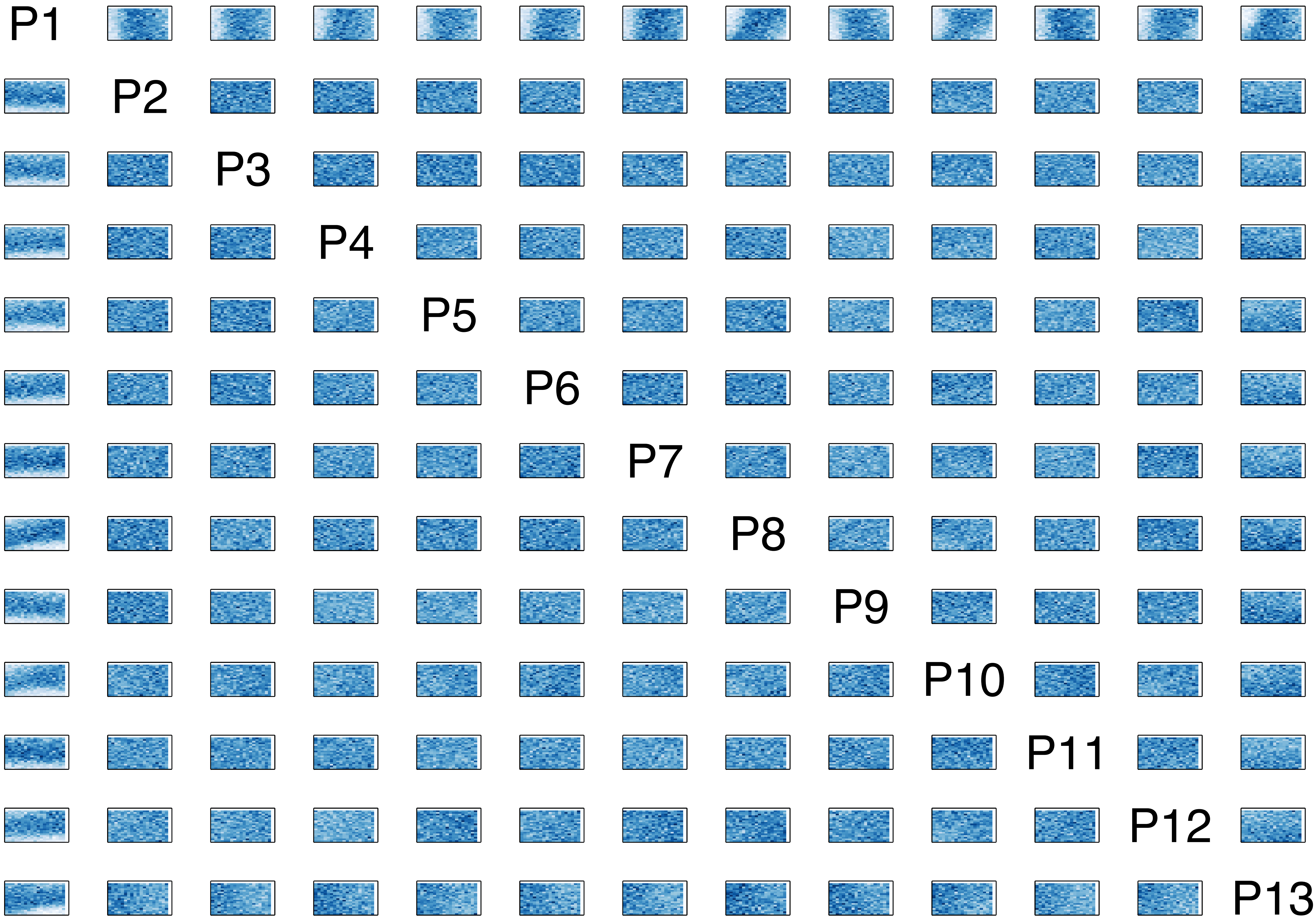
mean
300
250
200
150
0.0
0.2
0.4
0.6
0.8
1.0
1.2
P8



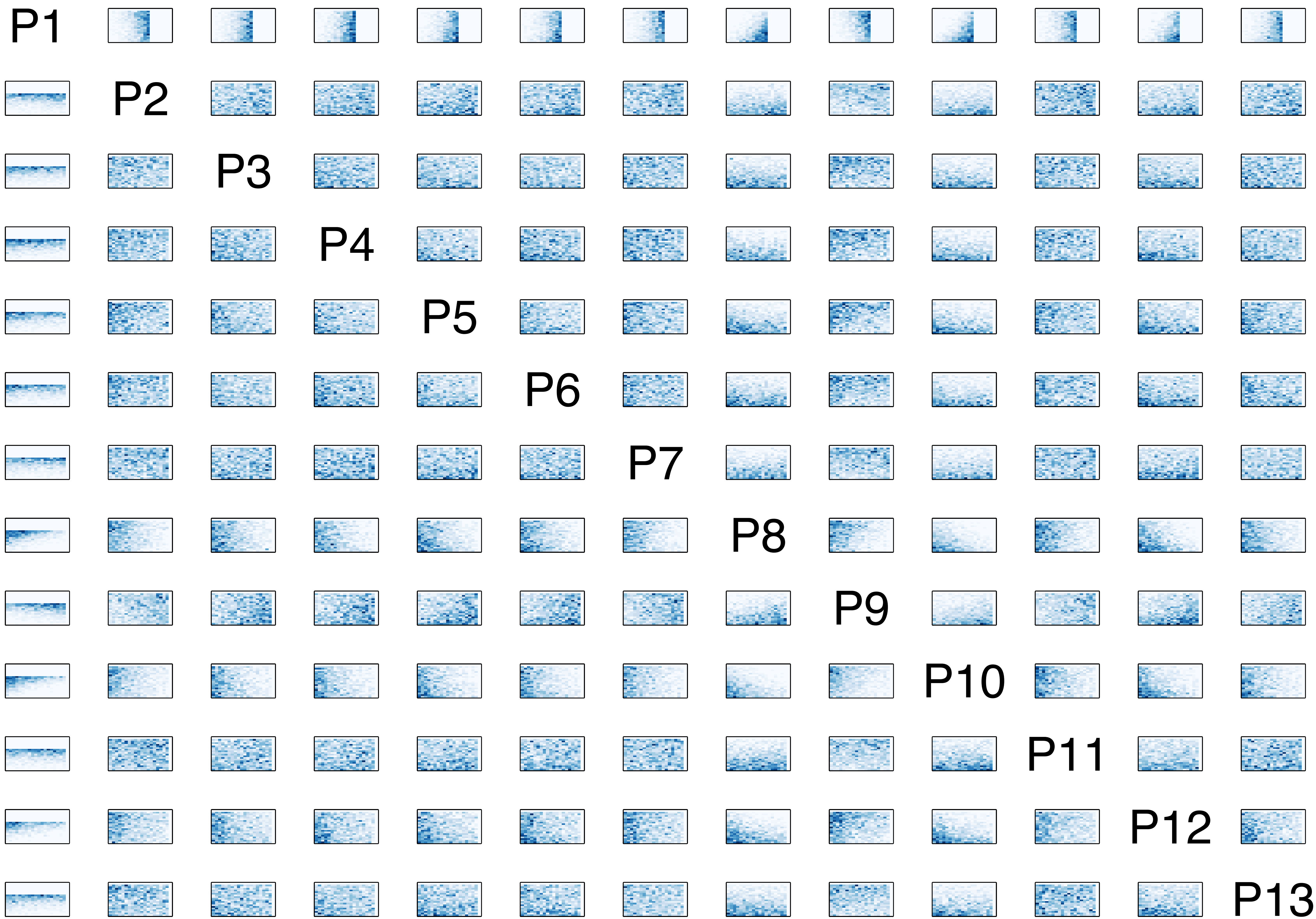
implausibility
5
4
3
2
1
0
0.0
0.2
0.4
0.6
0.8
1.0
1.2
P8

History matching on cardiac cell model with 13 uncertain parameters

- A single observation (mean =200, sd = 15) (made up)
- No discrepancy
- 130 model runs in maximinLHC
- Rules out 24% of 13-d space


P1
P2
P3
P4
P5
P6
P7
P8
P9
P10
P11
P12
P13

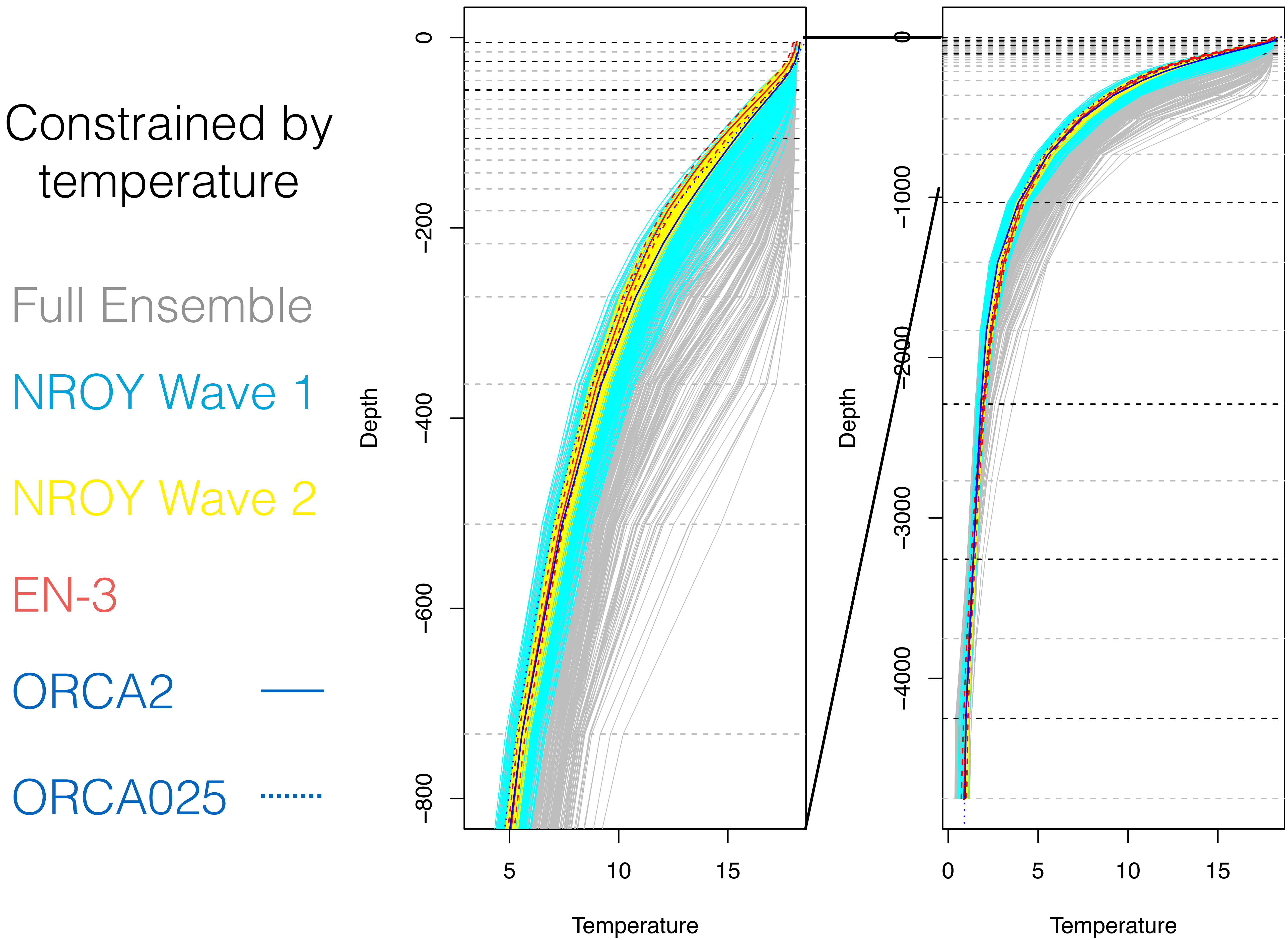
- A more discriminating example
- mean = 450, sd =5
- Only 13% space left

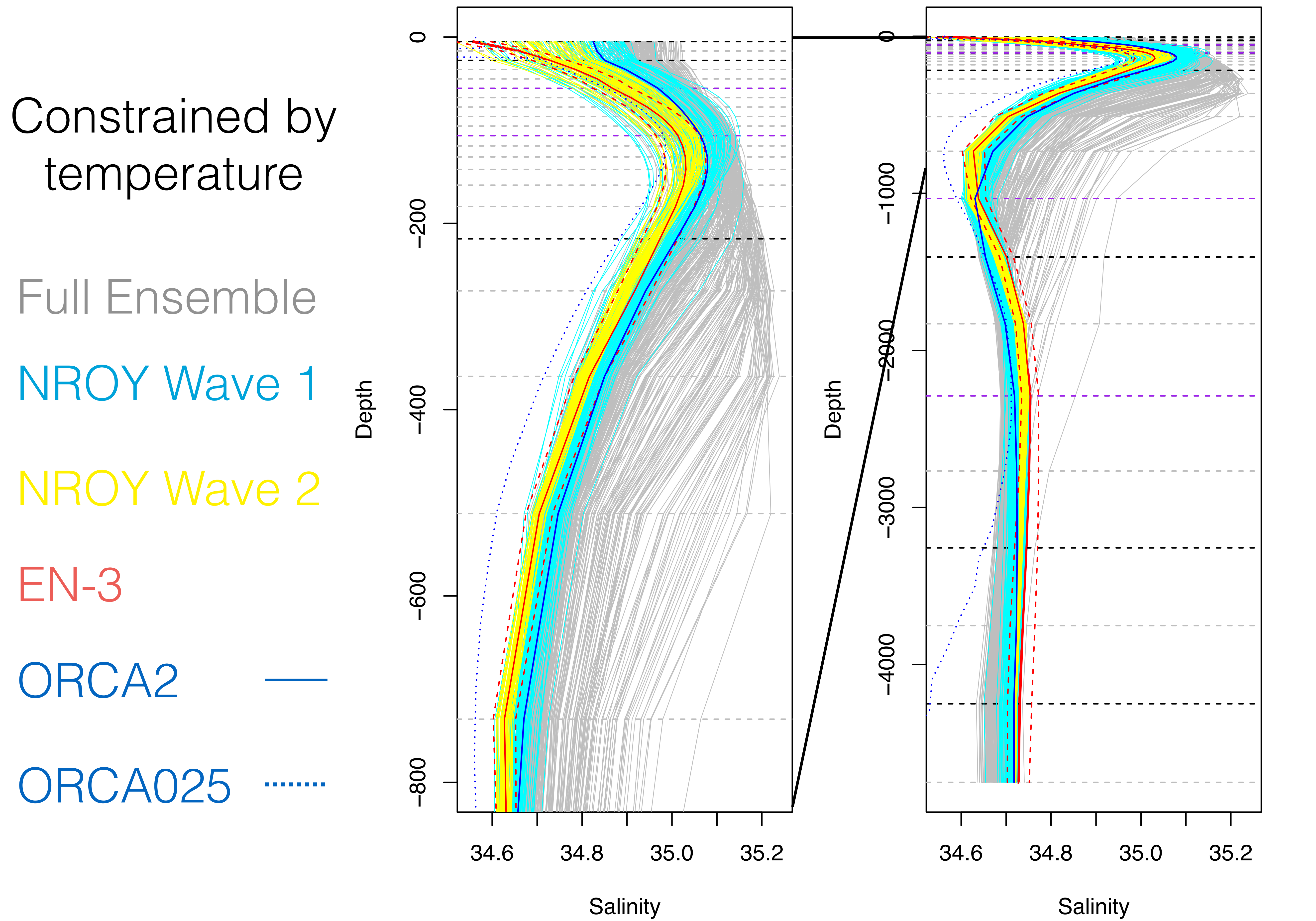


P1
P2
P3
P4
P5
P6
P7
P8
P9
P10
P11
P12
P13

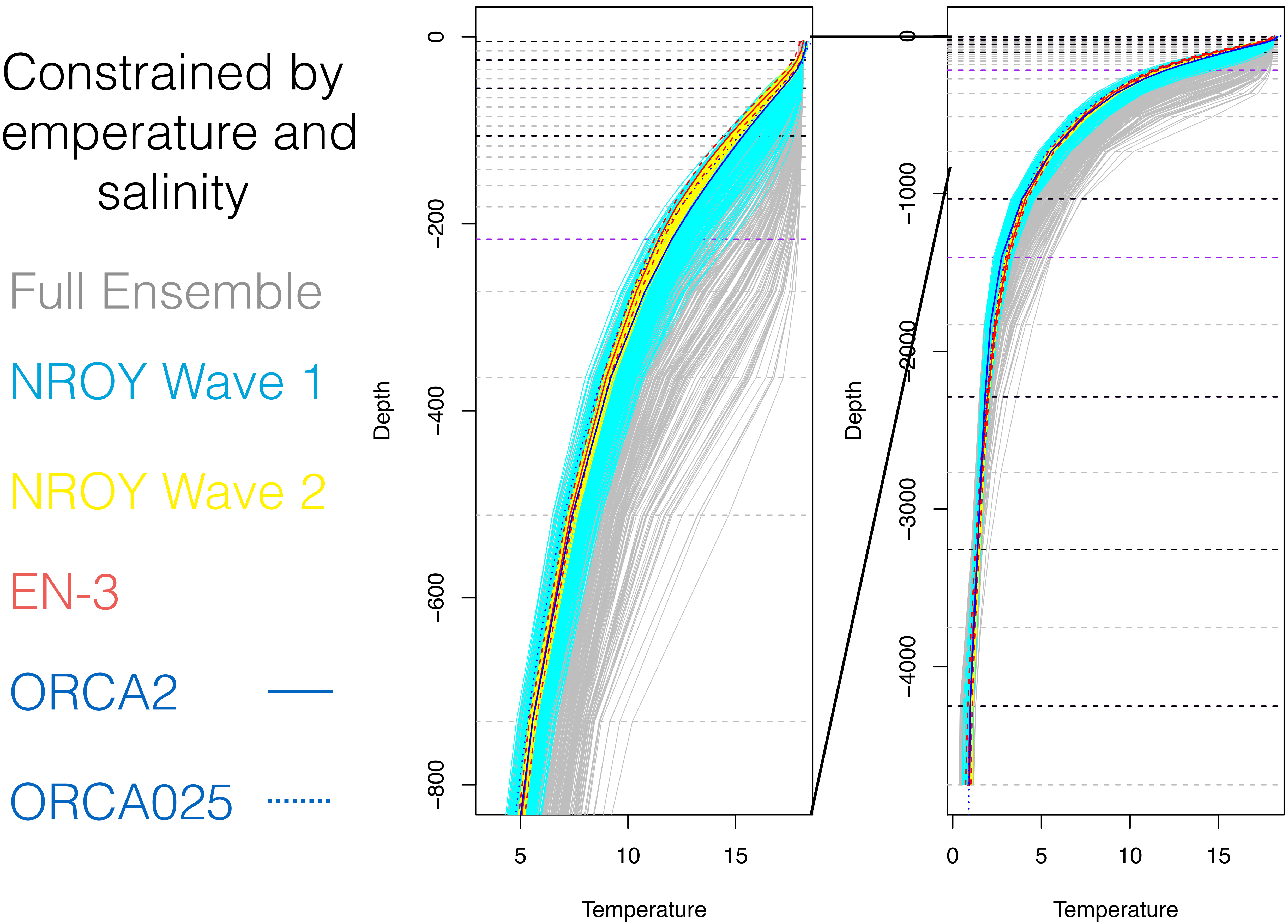
ORCA2

- State of the art ocean model 2° resolution
- ‘Climatological’ forcing (Normal years)
- Matching temperature at 8 depths with EN3 climatology
- Removes 95% of parameter space. (Wave 1)
- Adding salinity
- Thanks to Danny Williamson and Adam Blaker


Constrained by temperature
Full Ensemble
NROY Wave 1
NROY Wave 2
EN-3
ORCA2
ORCA025
Depth
Temperature
0
−200
−400
−600
−800
5
10
15
−1000
−2000
−3000
−4000
0


Constrained by temperature
Full Ensemble
NROY Wave 1
NROY Wave 2
EN-3
ORCA2
ORCA025
Depth
0
−200
−400
−600
−800
Salinity
34.6
34.8
35.0
35.2
Depth
0
−1000
−2000
−3000
−4000
Salinity
34.6
34.8
35.0
35.2

Constrained by temperature and salinity



Constrained by temperature and salinity

Full Ensemble

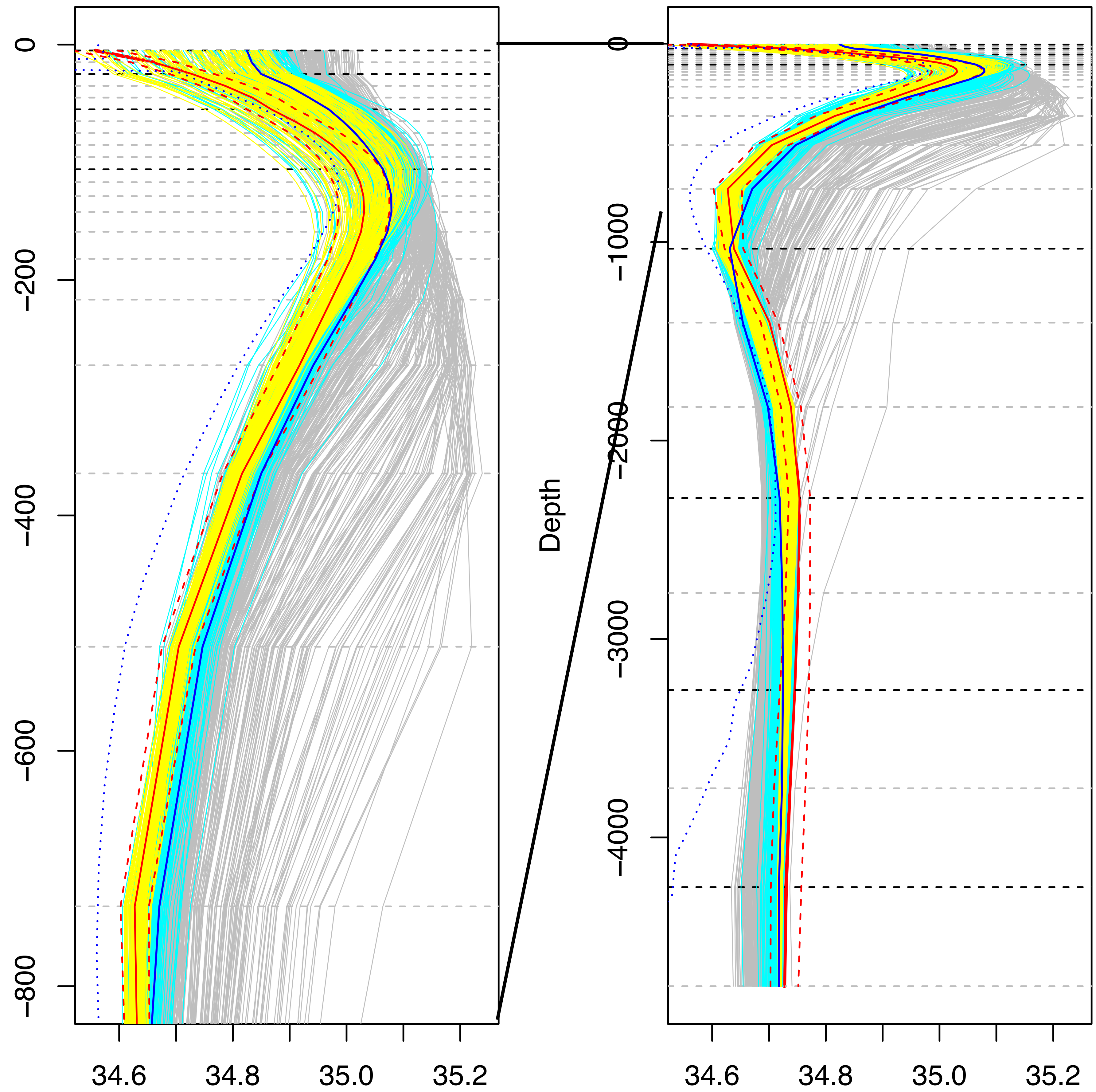
NROY Wave 1

NROY Wave 2

EN-3

ORCA2 ———

ORCA025 ········



Discrepancy

- None of our models is perfect
- Kennedy and O'Hagan estimate discrepancy
- In history matching it is an input
- We need to elicit it

‘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?

Stochastic Models

- So far all the models have been deterministic
- But we can generalise to stochastic models
- Emulate mean and variance of the model
- Split the discrepancy into a stochastic part (model variance) + the discrepancy
- Andrianakis et al (2015)

Random Effects

- Some models should be fitted to individuals (e.g. cardiac models)

- But we aggregate data across groups

- This adds an additional uncertainty

- Data variance or additional discrepancy?

Tolerance to error

- Often our NROY space will go to zero as we add more waves of model runs
- This implies the discrepancy variance is too small
- An alternative interpretation is that the discrepancy is our *tolerance to error*
- How bad are we prepared to let our models be to fit the data?

Research Areas

- Which metric to match?
- Combining metrics
  - *Max(Imp)* (Vernon et al 2010)
  - Second, third largest
- Multivariate methods

$$Imp^2 = (y - E(f(x)))^T Var(y - E(f(x)))^{-1}(y - E(f(x)))$$

- Spatial methods and dimension reduction
- Relating different models to each other
- For an interesting application in ABC (approximate Bayesian computation); see Richard Wilkinson’s 2014 ArXiv paper

Conclusions

- History matching (and GP emulators) allows us to do inverse problems without optimisation (or estimating posteriors)
- Even if we want to do conventional methods in the final NROY space, for example we may need a posterior, because of the limited region we expect the function to be much better behaved.