

# Inference for complex models

Richard Wilkinson

School of Maths and Statistics  
University of Sheffield

30 January 2018

# Uncertainty Quantification and inverse problems

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

- Modelling, propagating and updating uncertainties.
- Inter-disciplinary

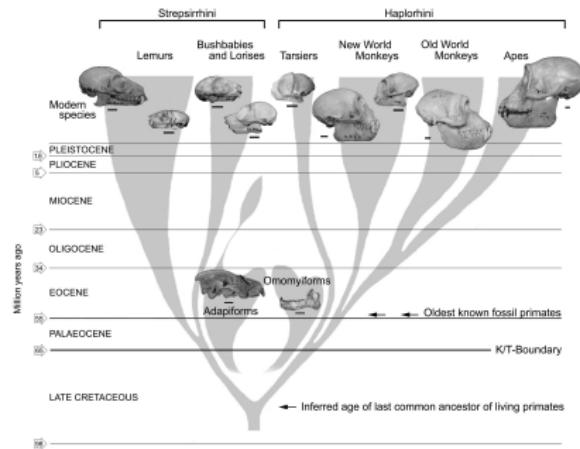
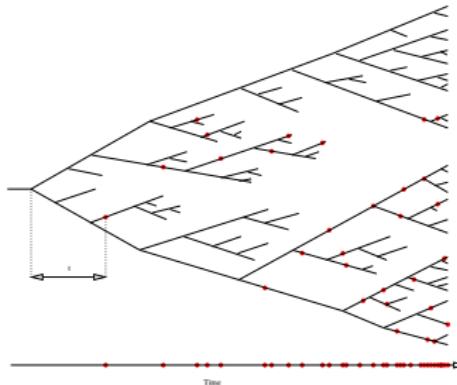
# Uncertainty Quantification and inverse problems

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

- Modelling, propagating and updating uncertainties.
- Inter-disciplinary

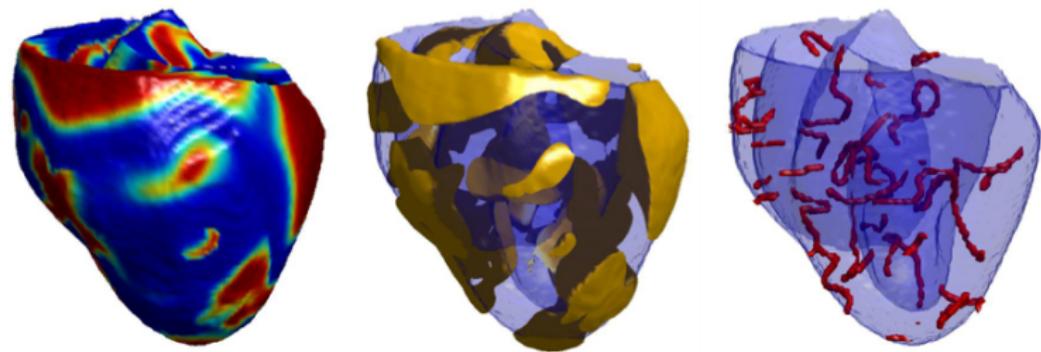
Inverse problems/Calibration/Parameter estimation/...

- For most simulators we specify parameters  $\theta$  and i.c.s and the simulator,  $f(\theta)$ , generates output  $X$ .
- The inverse-problem: observe data  $D$ , estimate parameter values  $\theta$  which explain the data.



# Why do we need UQ?

## Atrial fibrillation



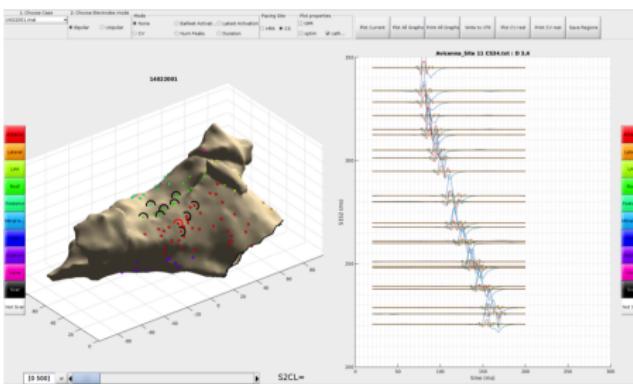
Atrial fibrillation (AF) - rapid and uncoordinated electrical activation (arrhythmia) leading to poor mechanical function.

- 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).

# UQ in Patient Specific Cardiac Models

With Richard Clayton (Sheffield), Steve Neiderer (KCL)

Aim: predict which AF patients will develop AT following ablation, and then treat for both in a single procedure.



Complex electrophysiology simulation using monodomain eqn on shell anatomy with local activation given by the Mitchell-Schaeffer ionic model.

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

- Infer conduction velocity on uncertain atrial anatomy using ECG
- Infer regions of fibrotic material
- Predict AT pathways
- Aid clinical decision making (accounting for uncertainty)

This is hard statistically because the model is slow to evaluate, it has high dimensional inputs and outputs, and we have complex geometry, uncertainty, and data.

## Tools

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

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

## Approaches

- Likelihood-based Monte Carlo
- Approximate Bayesian Computation (ABC) - 'likelihood-free'
- Gaussian process emulation
- Machine learning tools particular for designing scores suitable, and kernel methods
- Applied maths: multi-fidelity methods, reduced order models etc

## Tools

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

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

- $\pi(\theta|D)$  is the posterior distribution – hard to compute.

## Approaches

- Likelihood-based Monte Carlo
- Approximate Bayesian Computation (ABC) - 'likelihood-free'
- Gaussian process emulation
- Machine learning tools particular for designing scores suitable, and kernel methods
- Applied maths: multi-fidelity methods, reduced order models etc

## Tools

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

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

- $\pi(\theta|D)$  is the posterior distribution – hard to compute.
- $\pi(D|\theta)$  is the likelihood function.
  - ▶ For complex models can be slow to compute
  - ▶ Can also be impossible to compute in some cases
  - ▶ Relating simulator to reality can make specifying  $\pi(D|\theta)$  particularly difficult, and the answer sensitive to errors

## Approaches

- Likelihood-based Monte Carlo
- Approximate Bayesian Computation (ABC) - 'likelihood-free'
- Gaussian process emulation
- Machine learning tools particular for designing scores suitable, and kernel methods
- Applied maths: multi-fidelity methods, reduced order models etc

# Tools

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

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

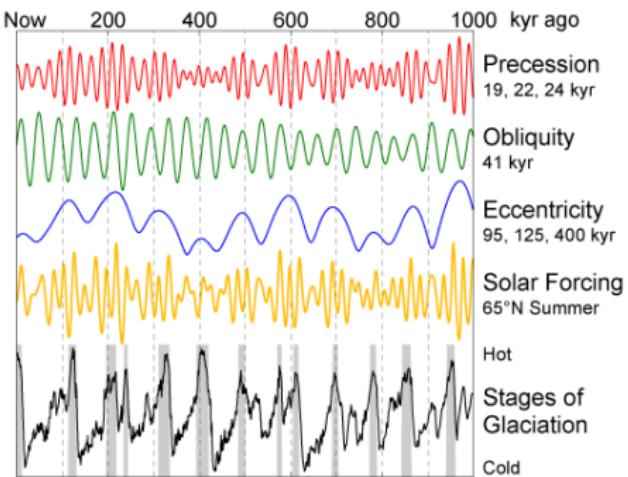
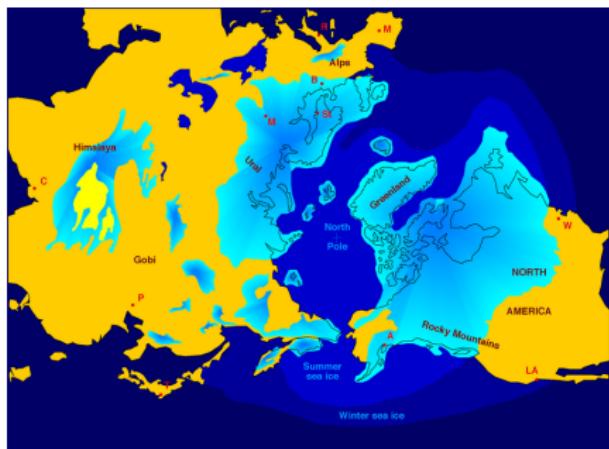
- $\pi(\theta|D)$  is the posterior distribution – hard to compute.
- $\pi(D|\theta)$  is the likelihood function.
  - ▶ For complex models can be slow to compute
  - ▶ Can also be impossible to compute in some cases
  - ▶ Relating simulator to reality can make specifying  $\pi(D|\theta)$  particularly difficult, and the answer sensitive to errors
- $\pi(D)$  is the model evidence or normalising constant.

## Approaches

- Likelihood-based Monte Carlo
- Approximate Bayesian Computation (ABC) - 'likelihood-free'
- Gaussian process emulation
- Machine learning tools particular for designing scores suitable, and kernel methods
- Applied maths: multi-fidelity methods, reduced order models etc

# Climate science

What drives the glacial-interglacial cycle?



**Eccentricity:** orbital departure from a circle, controls duration of the seasons

**Obliquity:** axial tilt, controls amplitude of seasonal cycle

**Precession:** variation in Earth's axis of rotation, affects difference between seasons

## Model selection

What drives the glacial-interglacial cycle?

- Which aspect of the astronomical forcing is of primary importance?
- Which models best represent the cycle?

*Most models of the [...] glacial cycles have at least four degrees of freedom [parameters], and some have as many as twelve. Unsurprisingly [this is...] insufficient to distinguish between the skill of the various models (Roe and Allen 1999)*

## Model selection

What drives the glacial-interglacial cycle?

- Which aspect of the astronomical forcing is of primary importance?
- Which models best represent the cycle?

*Most models of the [...] glacial cycles have at least four degrees of freedom [parameters], and some have as many as twelve. Unsurprisingly [this is...] insufficient to distinguish between the skill of the various models (Roe and Allen 1999)*

Simple phenomenological models based on hypothesised relationships that capture some aspect of the climate system.

Let  $X_t \in \mathbb{R}^P$  be the state of the climate at time  $t$ . Typically  $X_{1,t}$  = ice volume, and other components many represent  $\text{CO}_2$ , ocean temp, etc, or be undefined.

## Model selection

What drives the glacial-interglacial cycle?

- Which aspect of the astronomical forcing is of primary importance?
- Which models best represent the cycle?

*Most models of the [...] glacial cycles have at least four degrees of freedom [parameters], and some have as many as twelve. Unsurprisingly [this is...] insufficient to distinguish between the skill of the various models (Roe and Allen 1999)*

Simple phenomenological models based on hypothesised relationships that capture some aspect of the climate system.

Let  $X_t \in \mathbb{R}^p$  be the state of the climate at time  $t$ . Typically  $X_{1,t}$  = ice volume, and other components many represent  $\text{CO}_2$ , ocean temp, etc, or be undefined.

Embed these models within a statistical state space model

$$dX_t = g(X_t, \theta)dt + F(t, \gamma)dt + \Sigma dW$$

$$\text{Observe } Y_t = d + sX_{1,t} + \epsilon_t$$

Typically these models have 10-15 parameters  $(\theta, \gamma, s, d, \Sigma, \sigma)$  to be estimated.

# Data



$^{18}\text{O}$  is heavier than  $^{16}\text{O}$ , and so its circulation behaviour varies with temperature.

Variation in the ratio  $\delta^{18}\text{O}$  in marine sediments and ice cores informs us about historic temperatures.

The raw measurements are of  $\delta^{18}\text{O}$  as a function of depth in a core: age must be inferred.

- Climate reconstruction

$$\pi(X_{1:T}|y_{1:T}, \theta_m, \mathcal{M}_m)$$

- Model calibration

$$\pi(\theta_m|y_{1:T}, \mathcal{M}_m)$$

- Model selection (model evidence/Bayes factors)

$$\pi(y_{1:T}|\mathcal{M}_m)$$

Progressively more difficult to calculate as  $\pi(X_{t+1}|X_t, \theta_m, \mathcal{M}_m)$  is unknown.

**Filtering:** Sequential Monte Carlo (SMC) methods are the natural approach for finding the filtering distributions  $\pi(X_{1:T}|y_{1:T}, \theta)$

- Requires careful proposal design: use Golightly and Wilkinson (2006) Brownian bridge proposals to nudge parameters towards the data.

**Filtering:** Sequential Monte Carlo (SMC) methods are the natural approach for finding the filtering distributions  $\pi(X_{1:T}|y_{1:T}, \theta)$

- Requires careful proposal design: use Golightly and Wilkinson (2006) Brownian bridge proposals to nudge parameters towards the data.

**Parameter estimation:** SMC provides an unbiased estimate of the marginal likelihood

$$\pi(y_{1:T}|\theta) = \pi(y_1|\theta) \prod_{t=2}^T \pi(y_t|y_{1:t-1}, \theta)$$

Use these estimates in a pseudo marginal scheme to estimate  $\pi(\theta|y_{1:T})$

**Filtering:** Sequential Monte Carlo (SMC) methods are the natural approach for finding the filtering distributions  $\pi(X_{1:T}|y_{1:T}, \theta)$

- Requires careful proposal design: use Golightly and Wilkinson (2006) Brownian bridge proposals to nudge parameters towards the data.

**Parameter estimation:** SMC provides an unbiased estimate of the marginal likelihood

$$\pi(y_{1:T}|\theta) = \pi(y_1|\theta) \prod_{t=2}^T \pi(y_t|y_{1:t-1}, \theta)$$

Use these estimates in a pseudo marginal scheme to estimate  $\pi(\theta|y_{1:T})$

**Model selection:** SMC<sup>2</sup> (Chopin *et al.* 2011) allows us to also estimate  $\pi(y_{1:T})$ . Basic idea:

- Introduce  $M$  parameter particles  $\theta_1, \dots, \theta_M$
- For  $t = 1, \dots, T$ 
  - ▶ For each  $\theta_i$  run a particle filter targeting  $\pi(X_{1:t}|y_{1:t}, \theta_i)$
  - ▶ Recalculate all the importance weights and resample if necessary

This takes 3-4 days on a standard server, or 4-6 hours on a GPU.

# Results

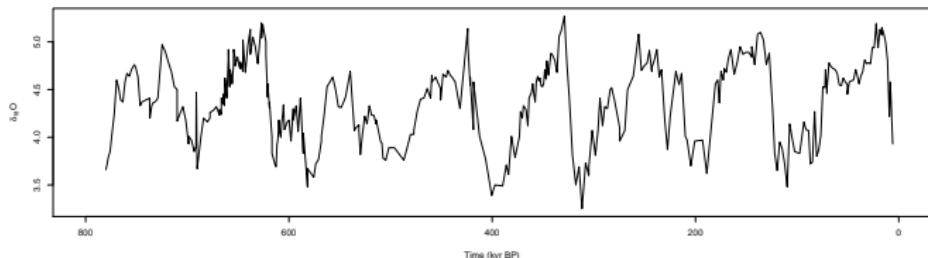
Carson, Crucifix, Preston, W. 2017

Simulation studies show we can accurately estimate state trajectory, parameters (including the correct forcing), and choose between competing models.

Model		Evidence $\pi(y_{1:N} \mathcal{M}_m)$	
		SM91-unforced	SM91-forced
SM91	Forced	$5.6 \times 10^{28}$	$1.4 \times 10^{41}$
	Unforced	$1.1 \times 10^{30}$	$2.4 \times 10^{18}$
T06	Forced	$3.6 \times 10^{20}$	$2.6 \times 10^{30}$
	Unforced	$1.1 \times 10^{22}$	$2.9 \times 10^{14}$
PP12	Forced	$2.8 \times 10^8$	$2.1 \times 10^{18}$

- Strongest evidence for the true model found each time
- For the data generated from the forced model, the forced version of the wrong model is preferred.
- MC error is large: typically an order of magnitude, so confident conclusions only if Bayes factor (= ratio of evidences) > 1000.

## Results: ODP677

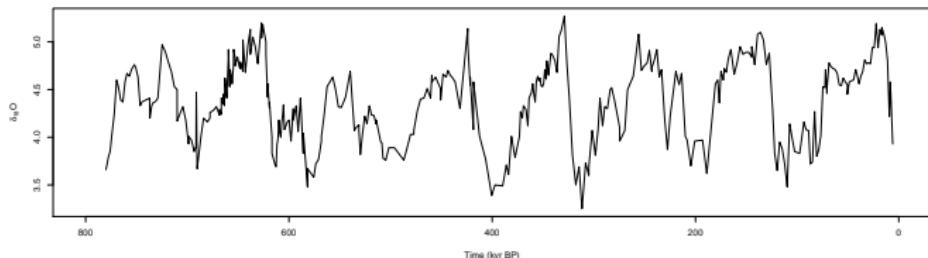


ODP677 stack which has been dated by

- Lisiecki and Raymo (2005) used orbital tuning
- Huybers 2007 used a depth-derived age model (no orbital tuning)

Model		Evidence	
		ODP677: H07(unforced)	ODP677: LR04(forced)
SM91	Forced	$4.0 \times 10^{24}$	$1.1 \times 10^{28}$
	Unforced	$3.5 \times 10^{26}$	$1.6 \times 10^{18}$
T06	Forced	$3.3 \times 10^{25}$	$4.5 \times 10^{29}$
	Unforced	$1.7 \times 10^{28}$	$3.3 \times 10^{21}$
PP12	Forced	$1.5 \times 10^{22}$	$1.8 \times 10^{34}$

## Results: ODP677



ODP677 stack which has been dated by

- Lisiecki and Raymo (2005) used orbital tuning
- Huybers 2007 used a depth-derived age model (no orbital tuning)

Model		Evidence	
		ODP677: H07(unforced)	ODP677: LR04(forced)
SM91	Forced	$4.0 \times 10^{24}$	$1.1 \times 10^{28}$
	Unforced	$3.5 \times 10^{26}$	$1.6 \times 10^{18}$
T06	Forced	$3.3 \times 10^{25}$	$4.5 \times 10^{29}$
	Unforced	$1.7 \times 10^{28}$	$3.3 \times 10^{21}$
PP12	Forced	$1.5 \times 10^{22}$	$1.8 \times 10^{34}$

The dating method applied changes the answer - theory laden data!

# Age model

Carson, Crucifix, Preston, W. (in submission)

Can we also quantify chronological uncertainty?

# Age model

Carson, Crucifix, Preston, W. (in submission)

Can we also quantify chronological uncertainty?

$$\text{Target} \quad \pi(\theta, T_{1:N}, X_{1:N}, \mathcal{M}_k | y_{1:N})$$

where  $T_{1:N}$  are the times of the observation  $y_{1:N}$  (previously assumed).

# Age model

Carson, Crucifix, Preston, W. (in submission)

Can we also quantify chronological uncertainty?

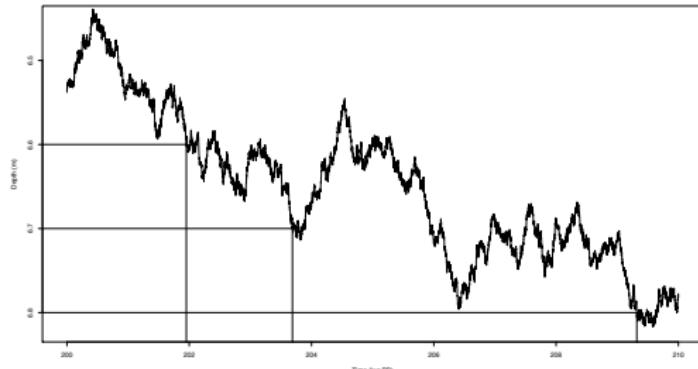
$$\text{Target} \quad \pi(\theta, T_{1:N}, X_{1:N}, \mathcal{M}_k | y_{1:N})$$

where  $T_{1:N}$  are the times of the observation  $y_{1:N}$  (previously assumed).

Propose a simple age model for sediment accumulation. Let  $H$  be the depth in the core, with  $H_N = 0$  at  $T_N = 0$

$$dH = -\mu_s dT + \sigma dW$$

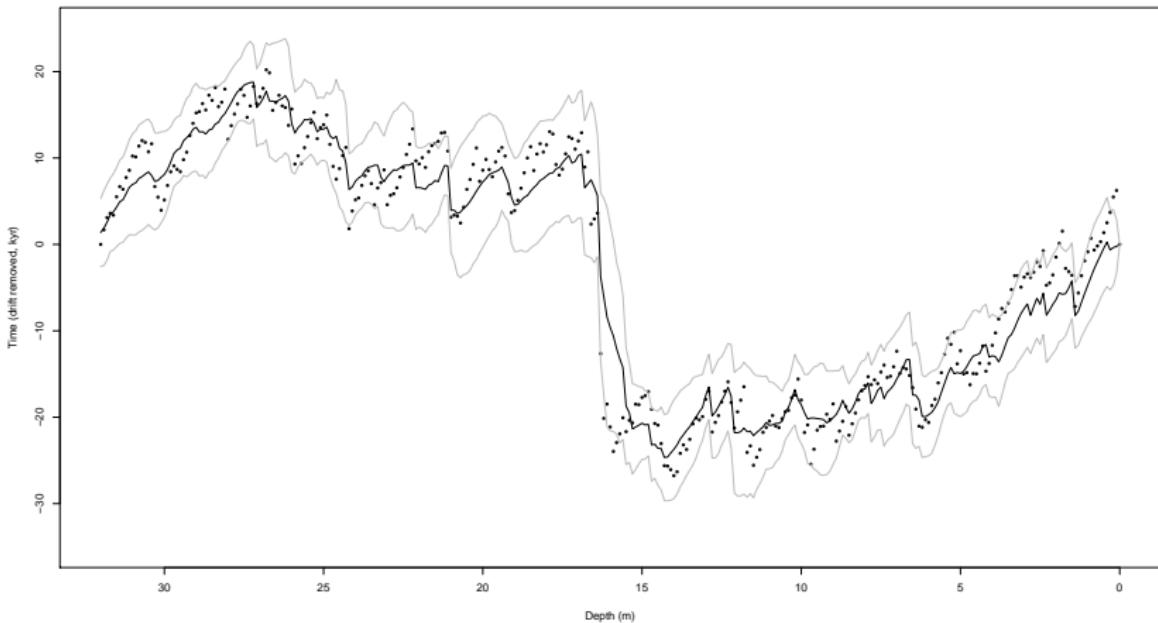
Slices are then taken through the core at specific depths  $H_1, \dots, H_N$ .



The age of a slice is the last time that depth was reached, which we can convert to a first passage problem. We also add a compaction model.

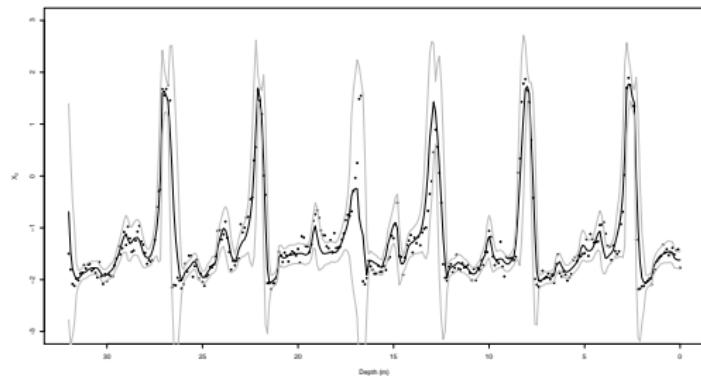
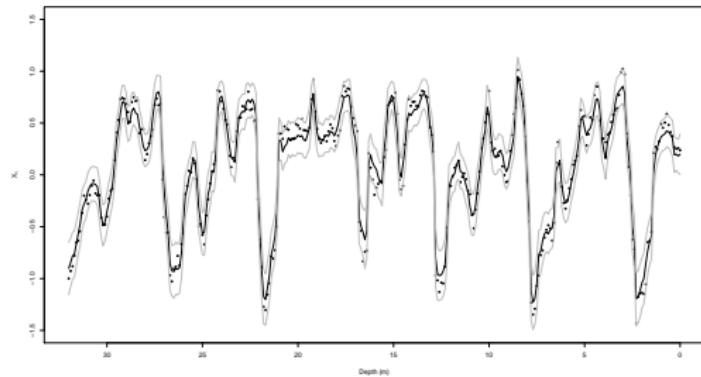
# Simulation study results - age vs depth (trend removed)

Dots = truth, black line = estimate, grey = 95% CI

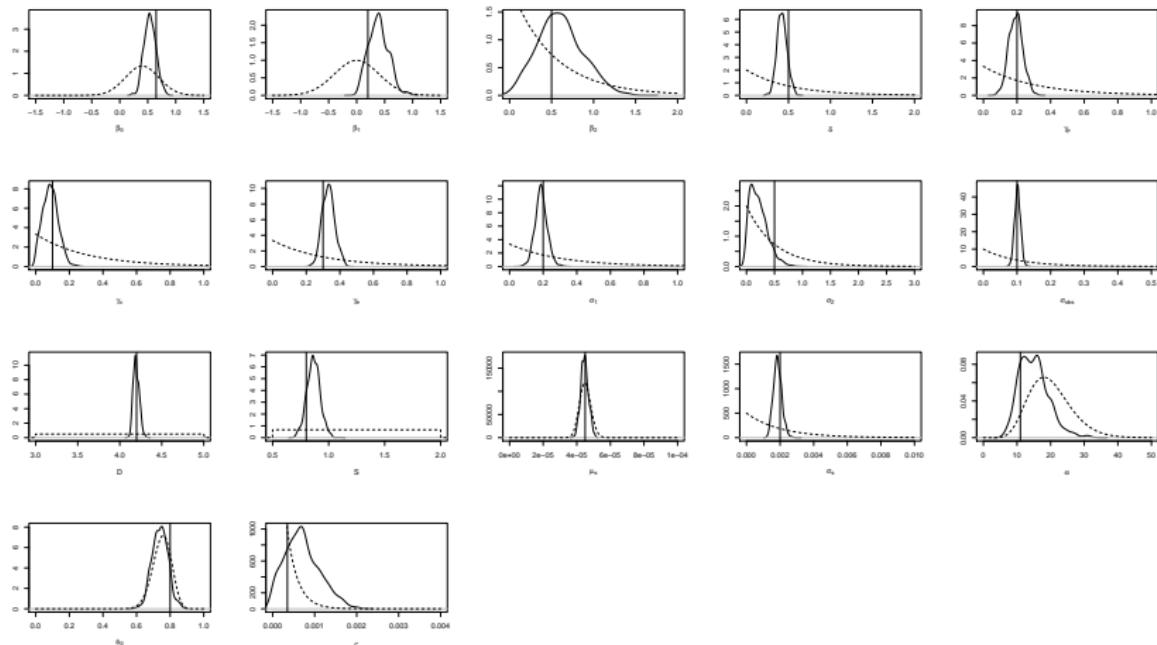


# Simulation study results - climate reconstruction

Dots = truth, black line = estimate, grey = 95% CI



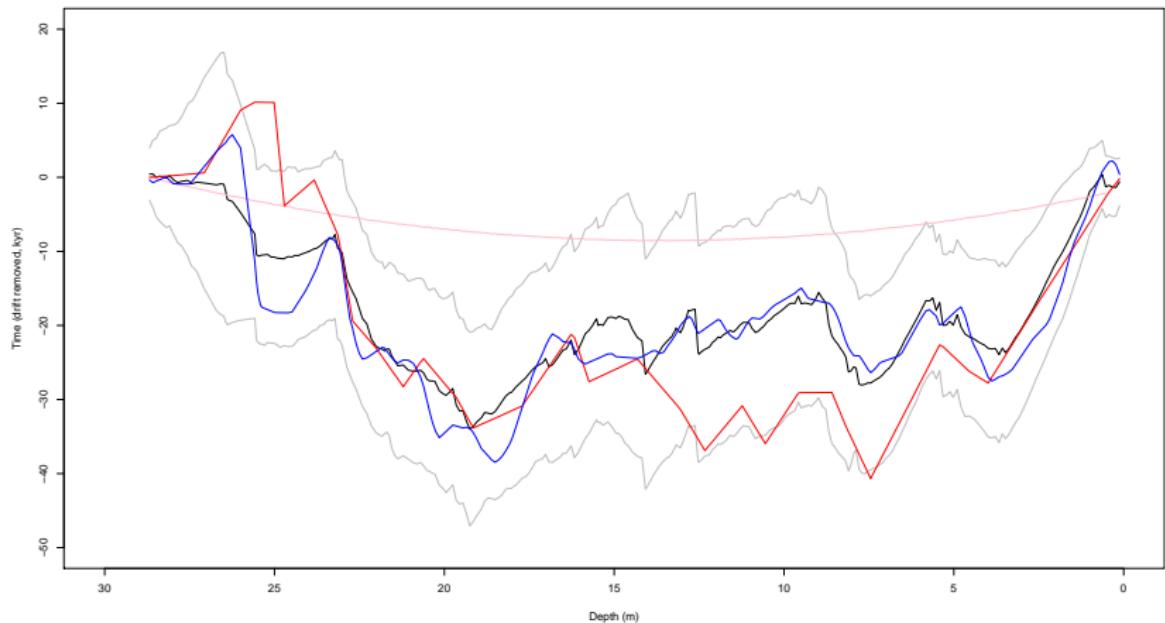
# Simulation study results - parameter estimation



Simultaneous inference of the choice between 5 models, 17 parameters, 800 ages, 1600 climate variables, using just 800 observations.

# Results for ODP846 - age vs depth (trend removed)

Black = posterior mean, grey = 95%CI, red = Huybers 2007, blue = Lisieki and Raymo 2004

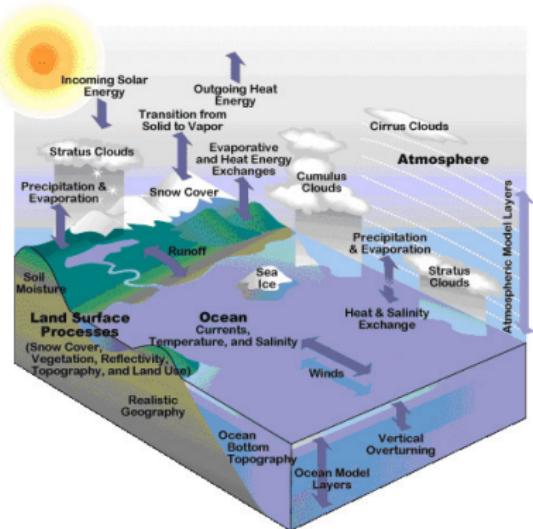
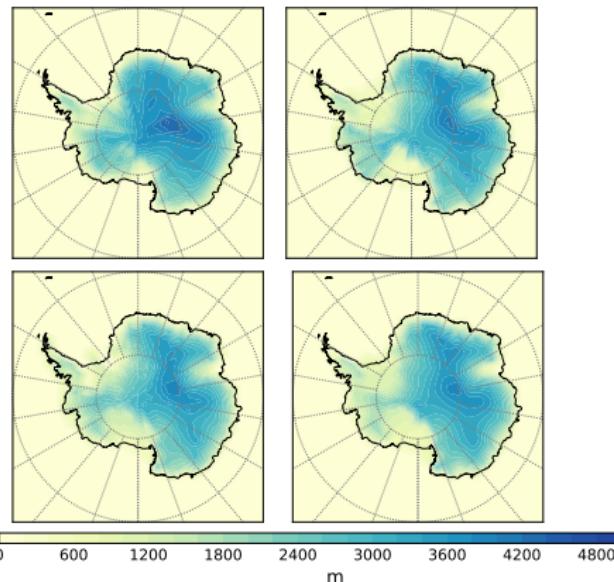


Advantages: full UQ, model selection, simultaneous parameter estimation and climate reconstruction

# Expensive and high dimensional simulators

Louise Sime (BAS)

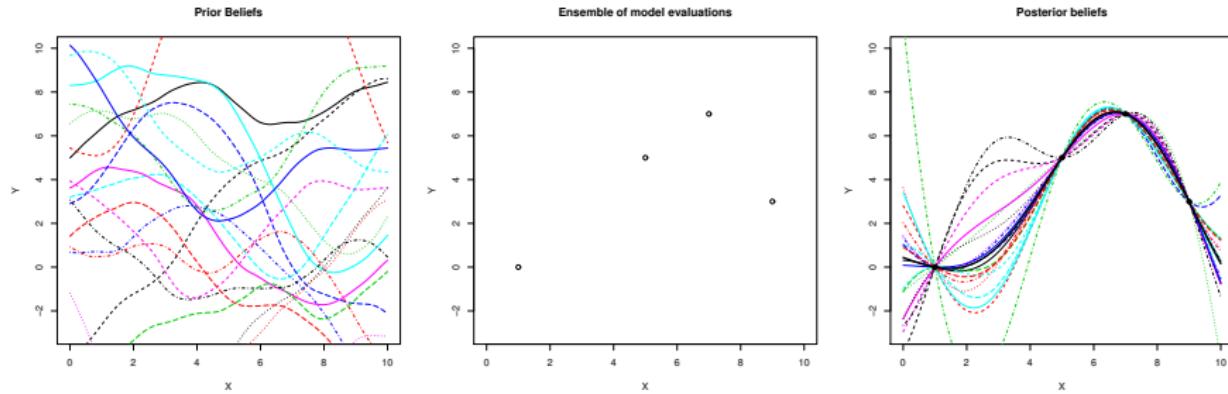
Climate models (GCMs) are expensive, complex, and take use dimensional inputs and give high dimensional outputs



How can we infer the shape of the ice-sheet at the last glacial maximum?

# Computationally expensive: Gaussian process emulators

If the model is expensive, we can build a surrogate/emulator of it.



How do we find emulators that obey physical constraints ( fugacity, symmetry etc) and have the correct asymptotic behaviour?

- Uteva, Graham, W., & Wheatley (2017), Cresswell, Wheatley, Wilkinson, & Graham (2016)

How do we deal with high dimensional inputs and outputs?

- Holden, Edwards, Ridgwell, Wilkinson, ... (under revision) show that limiting warming to  $< 1.5^{\circ}\text{C}$  target of the 2015 Paris Agreement is still achievable with 50% probability under rapid decarbonisation.

# Storage Panacea

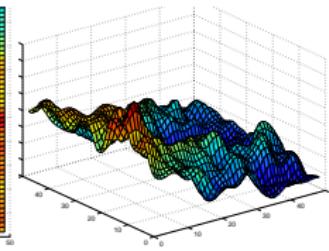
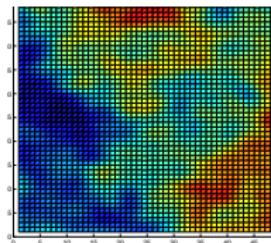
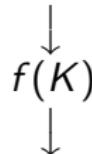
With Andrew Cliffe, Henry Power

Knowledge of the physical problem is encoded in a simulator  $f$

$$\nabla \cdot \mathbf{u} = 0, \quad \mathbf{u} = -\frac{K}{\mu}(\nabla P + \rho g \mathbf{e}_z), \quad \phi \frac{\partial C}{\partial t} + \mathbf{u} \cdot \nabla C = \phi \nabla \cdot (\mathbf{D} \nabla C) - \gamma_c C$$

Inputs:

Permeability field,  $K$   
(2d field)

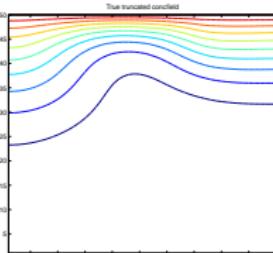
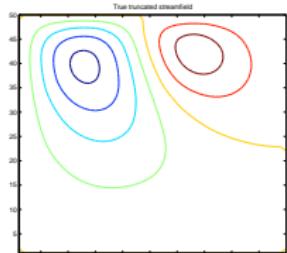


$\downarrow f(K)$

Outputs:

Stream func. (2d field),  
concentration (2d field),  
surface flux (1d scalar),

⋮



## Emulating simulators with high dimensional input

Crevilln-Garca, W., Shah, & Power (2017), Tian, W., Yang, Power, Fagerlund, & Niemi(2017)

For the CCS simulator, the input is a permeability field which only needs to be known at a finite but large number of locations,

- e.g. if we use a  $100 \times 100$  grid in the solver,  $K$  contains  $10^4$  entries
- Impossible to directly model  $f : \mathbb{R}^{10,000} \rightarrow \mathbb{R}$

# Emulating simulators with high dimensional input

Crevilln-Garca, W., Shah, & Power (2017), Tian, W., Yang, Power, Fagerlund, & Niemi(2017)

For the CCS simulator, the input is a permeability field which only needs to be known at a finite but large number of locations,

- e.g. if we use a  $100 \times 100$  grid in the solver,  $K$  contains  $10^4$  entries
- Impossible to directly model  $f : \mathbb{R}^{10,000} \rightarrow \mathbb{R}$

Instead, use a Karhunen-Loève (KL) expansion of  $K$  to reduce dimension:

- $K = \exp(Z)$  where  $Z \sim GP(m, C)$
- $Z$  can be represented as

$$Z(\cdot) = \sum_{i=1}^{\infty} \lambda_i \xi_i \phi_i(\cdot)$$

where  $\lambda_i$  and  $\phi_i(\cdot)$  are eigen-pairs of the Hilbert-Schmidt integral operator of the covariance function, and  $\xi_i \sim N(0, 1)$ .

By truncating

$$K(x) \approx \exp \left( \sum_{i=1}^n \lambda_i \xi_i \phi_i(x) \right)$$

we reduce the modelling problem to one of modelling  $f : \mathbb{R}^n \rightarrow \mathbb{R}$

# Emulating from fields to fields

W. (2011), Holden, Edwards, Garthwaite & W (2015), Bounceur, Crucifix, & W. (2015)

Now consider emulating the stream function and concentration fields ( $100 \times 100$  matrices).

We can use a similar trick, and use the singular value decomposition (PCA) to reduce the dimension.

Form the SVD of  $Y = LDR^T$

- Form a reduced rank approximation to  $Y$  by ignoring all but the first  $k$  eigenvectors:

$$Y \approx L_* D_* R_*^T$$

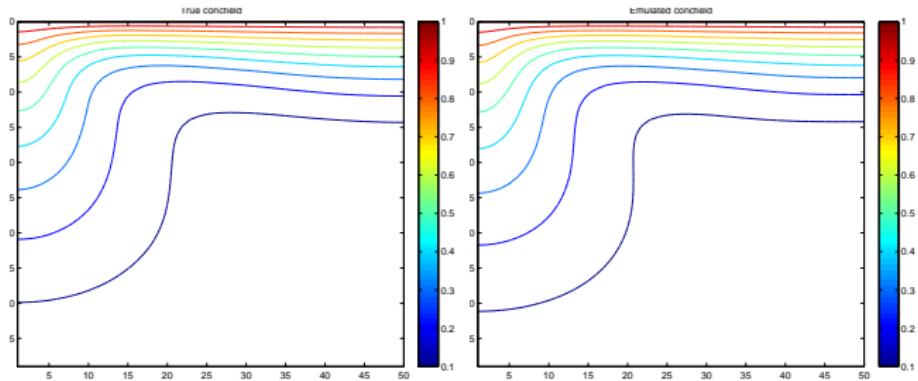
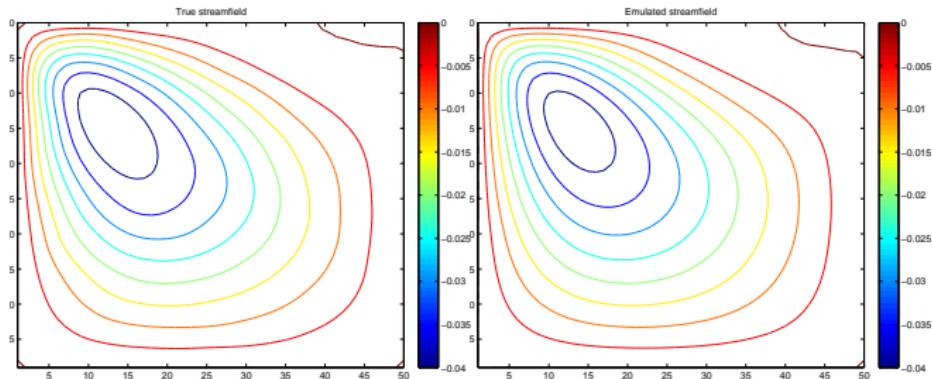
- If  $R_*^T = (t_1, \dots, t_N)$ , where each  $t_i$  is a vector of length  $k$ , then

$$L_* D_* t_1 \approx \mathbf{y}_1$$

To build an emulator from  $\mathbf{x}$  to  $\mathbf{y}$  we can build  $k$  separate emulators from  $\mathbf{x}$  to each element in the vector  $t$ .

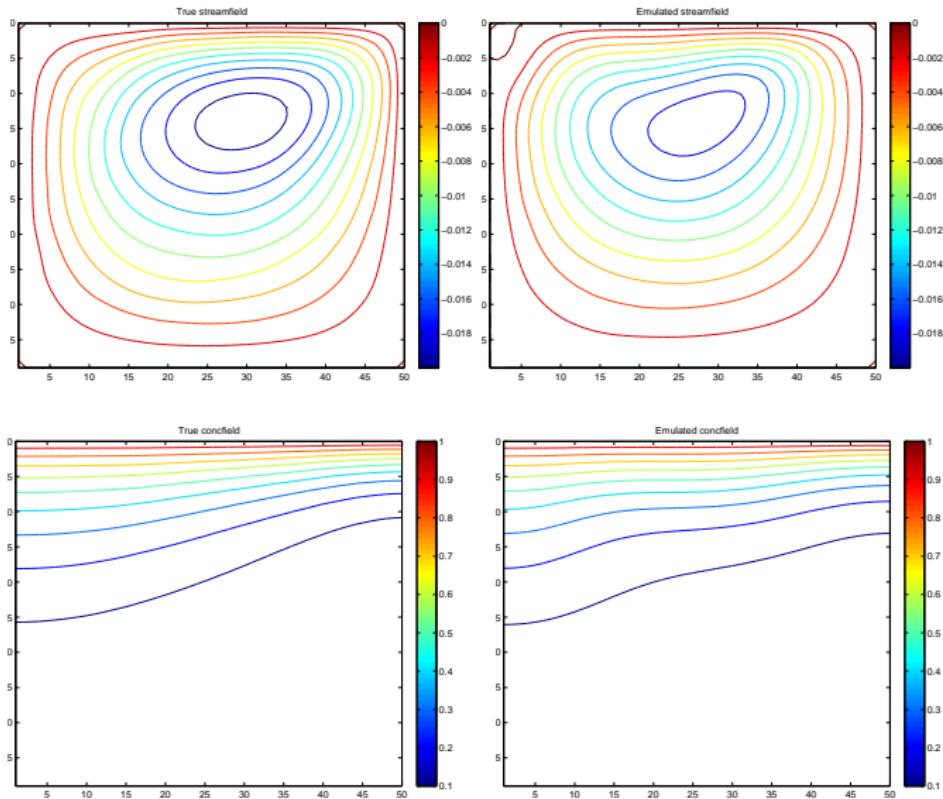
# Emulating the stream function and concentration fields

Left=true, right = emulated, 118 training runs, held out test set.



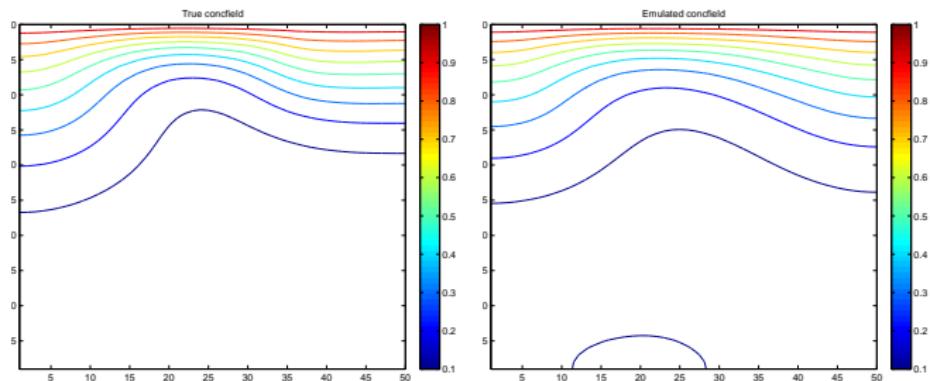
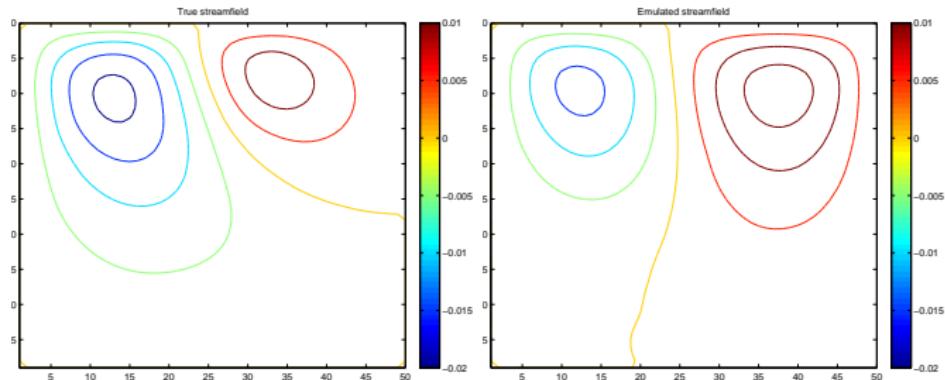
# Emulating the stream function and concentration fields

Left=true, right = emulated, 118 training runs, held out test set.

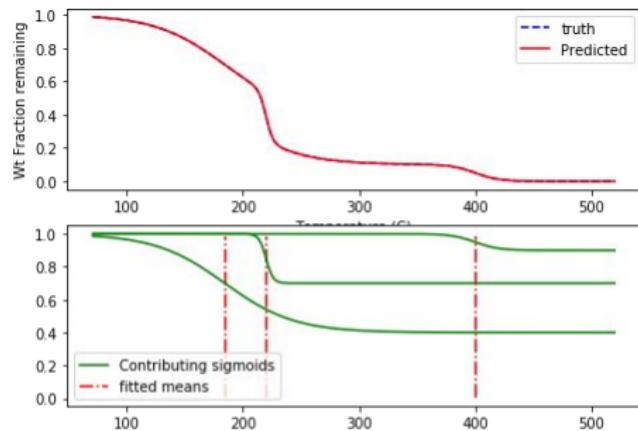


# Emulating the stream function and concentration fields

Left=true, right = emulated, 118 training runs, held out test set.



# Other decompositions



**Physical dimension reduction** With Tony Ryan, industrial support from SC Johnson, Thermogravimetric analysis

**Calibration focused dimension reduction** Emulation typically finds a global approximation. PCA/KL dimension reduction is unsupervised. Instead, can we do optimal model/dimension reduction when focussed on the calibration problem?

- Solved for linear Gaussian systems. Use RKHS methods to solve for non-linear problems.

PhD studentship provided by 

# Stochastic intractable models

## Approximate Bayesian Computation (ABC)

ABC algorithms are a collection of Monte Carlo methods used for calibrating simulators

- they do not require explicit knowledge of the likelihood function
- inference is done using simulation from the model (they are ‘likelihood-free’).

ABC methods are popular in biological disciplines, particularly genetics.  
They are

- Simple to implement
- Intuitive
- Embarrassingly parallelizable
- Can usually be applied

# Rejection ABC

## Uniform Rejection Algorithm

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

# Rejection ABC

## 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.

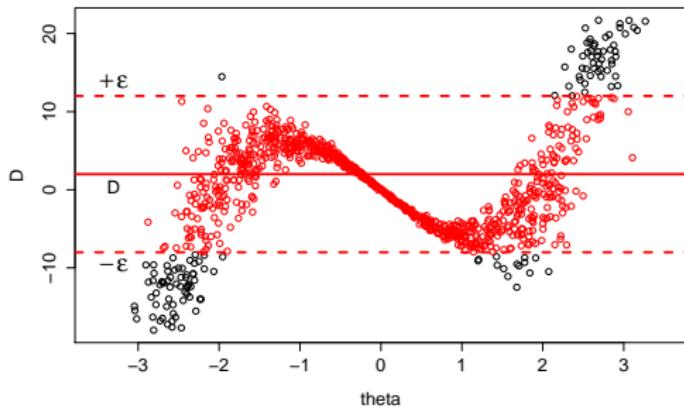
- As  $\epsilon \rightarrow \infty$ , we get observations from the prior,  $\pi(\theta)$ .
- If  $\epsilon = 0$ , we generate observations from  $\pi(\theta | D)$ .

Rejection sampling is inefficient, but we can adapt other MC samplers such as MCMC and SMC.

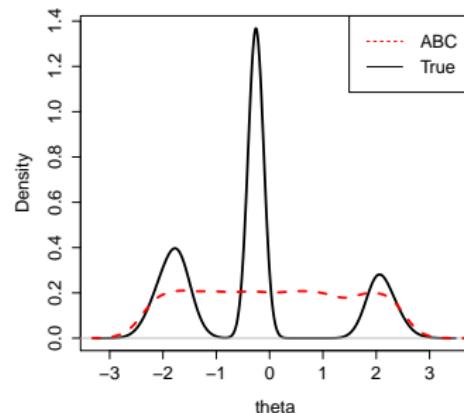
Simple → Popular with non-statisticians

$$\epsilon = 10$$

theta vs D



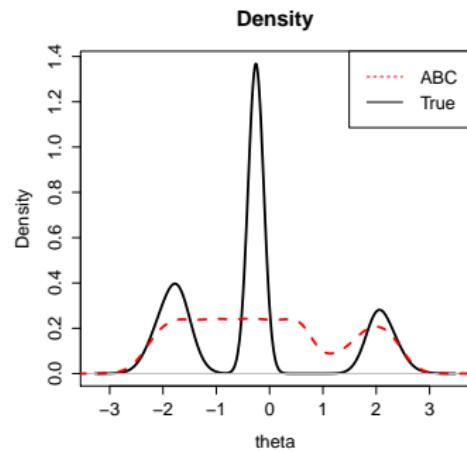
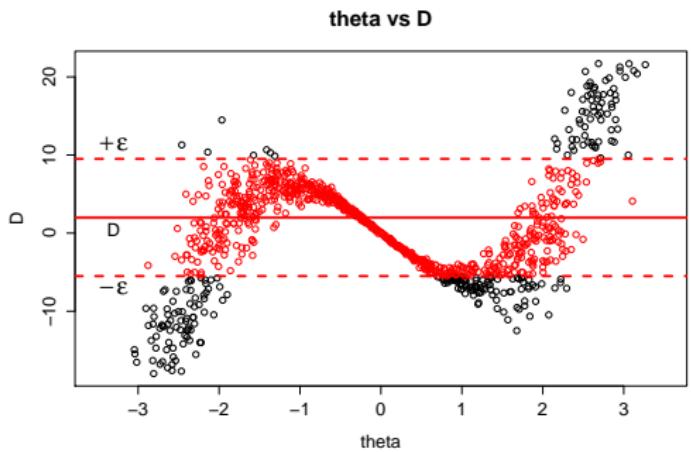
Density



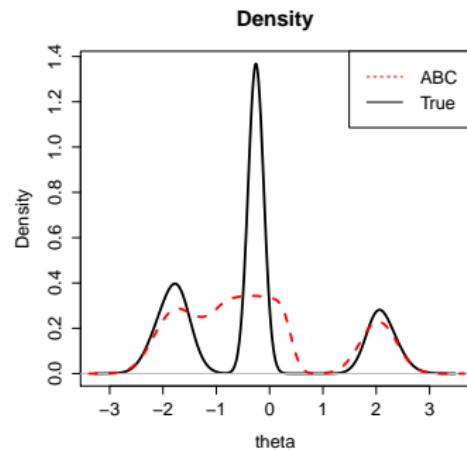
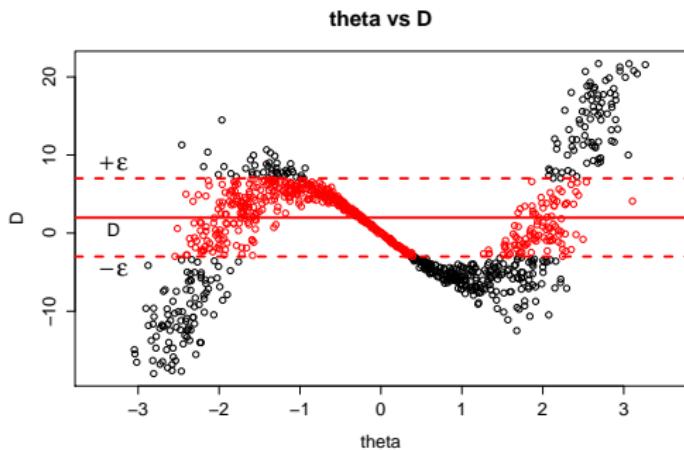
$$\theta \sim U[-10, 10], \quad X \sim N(2(\theta + 2)\theta(\theta - 2), 0.1 + \theta^2)$$

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

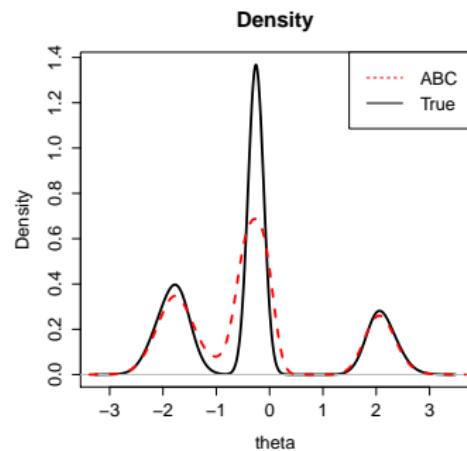
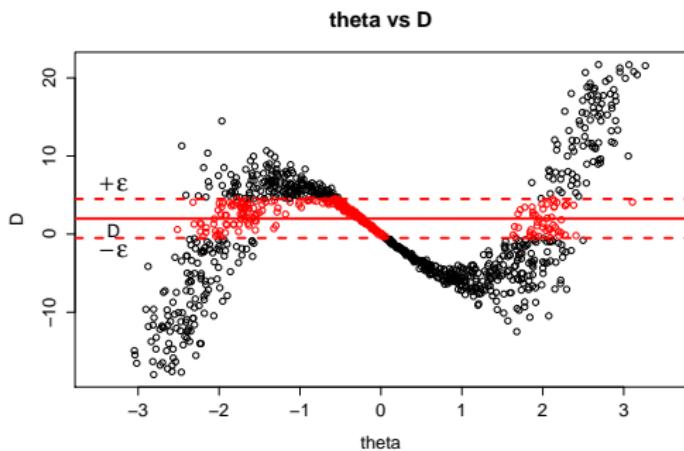
$$\epsilon = 7.5$$



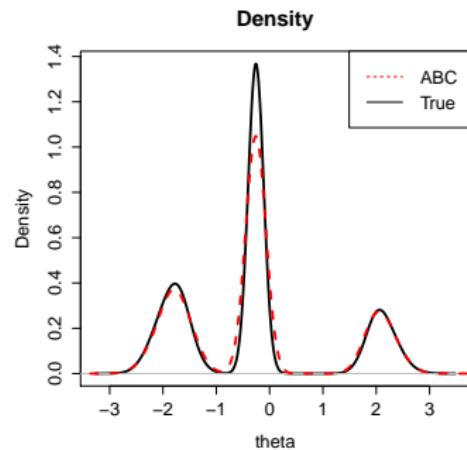
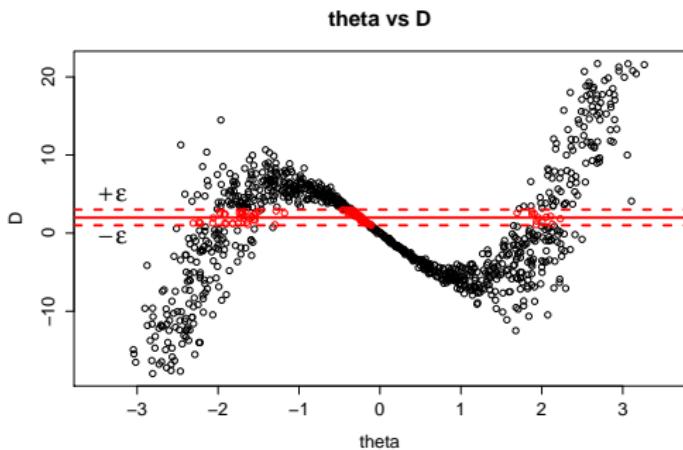
$$\epsilon = 5$$



$$\epsilon = 2.5$$



$$\epsilon = 1$$



# ABC as a probability model

W. (2013)

We wanted to solve the inverse problem

$$D = f(\theta)$$

but instead ABC solves

$$D = f(\theta) + e.$$

# ABC as a probability model

W. (2013)

We wanted to solve the inverse problem

$$D = f(\theta)$$

but instead ABC solves

$$D = f(\theta) + e.$$

ABC gives ‘exact’ inference under a different model!

We can show that

## Proposition

If  $\rho(D, X) = |D - X|$ , then ABC samples from the posterior distribution of  $\theta$  given  $D$  where we assume  $D = f(\theta) + e$  and that

$$e \sim U[-\epsilon, \epsilon]$$

## Accelerating ABC

Monte Carlo methods are generally guaranteed to succeed if we run them for long enough.

This guarantee is costly and can require more simulation than is possible.

## Accelerating ABC

Monte Carlo methods are generally guaranteed to succeed if we run them for long enough.

This guarantee is costly and can require more simulation than is possible.

However,

- Most methods sample naively - they don't learn from previous simulations.
- They don't exploit known properties of the likelihood function, such as continuity
- They sample randomly, rather than using careful design.

We can use methods that don't suffer in this way, but at the cost of losing the guarantee of success.

# Surrogate ABC

- Wilkinson 2014
- Meeds and Welling 2014
- Gutmann and Corander 2015
- Strathmann, Sejdinovic, Livingstone, Szabo, Gretton 2015
- :

With obvious influence from emulator community (e.g. Sacks, Welch, Mitchell, and Wynn 1989, Craig *et al.* 2001, Kennedy and O'Hagan 2001)

Constituent elements:

- Target of approximation
- Aim of inference and inference scheme
- Choice of surrogate/emulator
- Training/acquisition rule

⊣ a relationship to probabilistic numerics

## Building surrogates through history-matching waves

Craig *et al.* (1997), W. (2014)

The ABC log-likelihood  $I(\theta) = \log L(\theta)$  typical ranges across a wide range of values. Consequently, it is hard to approximate  $I(\theta)$  accurately for all  $\theta$ .

# Building surrogates through history-matching waves

Craig *et al.* (1997), W. (2014)

The ABC log-likelihood  $I(\theta) = \log L(\theta)$  typical ranges across a wide range of values. Consequently, it is hard to approximate  $I(\theta)$  accurately for all  $\theta$ .

- But we only need to make good predictions near  $\hat{\theta}$
- Introduce a sequence of surrogates, cf. waves of **history matching**.
- In each wave, build a GP model that can rule out regions of space as **implausible** according to some heuristic.

# Building surrogates through history-matching waves

Craig *et al.* (1997), W. (2014)

The ABC log-likelihood  $I(\theta) = \log L(\theta)$  typical ranges across a wide range of values. Consequently, it is hard to approximate  $I(\theta)$  accurately for all  $\theta$ .

- But we only need to make good predictions near  $\hat{\theta}$
- Introduce a sequence of surrogates, cf. waves of **history matching**.
- In each wave, build a GP model that can rule out regions of space as **implausible** according to some heuristic.

For example, decide that  $\theta$  is implausible if

$$\mathbb{P}(\tilde{I}(\theta) > \max_{\theta_i} I(\theta_i) - T) \leq 0.001$$

where  $\tilde{I}(\theta)$  is the GP model of  $I(\theta)$

Choose  $T$  so that if  $I(\hat{\theta}) - I(\theta) > T$  then  $\pi(\theta|y) \approx 0$ .

- Ruling  $\theta$  to be implausible is to set  $\pi(\theta|y) = 0$

# Entropic designs

## Challenges: Inference under discrepancy

How should we do inference if the model is imperfect?

---

<sup>1</sup>Even if we can't agree about it!

## Challenges: Inference under discrepancy

How should we do inference if the model is imperfect?

Data generating process

$$y \sim G$$

---

<sup>1</sup>Even if we can't agree about it!

## Challenges: Inference under discrepancy

How should we do inference if the model is imperfect?

Data generating process

$$y \sim G$$

Model (complex simulator, finite dimensional parameter)

$$\mathcal{F} = \{F_\theta : \theta \in \Theta\}$$

If  $G = F_{\theta_0} \in \mathcal{F}$  then we know what to do<sup>1</sup>.

---

<sup>1</sup>Even if we can't agree about it!

## Challenges: Inference under discrepancy

How should we do inference if the model is imperfect?

Data generating process

$$y \sim G$$

Model (complex simulator, finite dimensional parameter)

$$\mathcal{F} = \{F_\theta : \theta \in \Theta\}$$

If  $G = F_{\theta_0} \in \mathcal{F}$  then we know what to do<sup>1</sup>.

How should we proceed if

$$G \notin \mathcal{F}$$

---

<sup>1</sup>Even if we can't agree about it!

## Challenges: Inference under discrepancy

How should we do inference if the model is imperfect?

Data generating process

$$y \sim G$$

Model (complex simulator, finite dimensional parameter)

$$\mathcal{F} = \{F_\theta : \theta \in \Theta\}$$

If  $G = F_{\theta_0} \in \mathcal{F}$  then we know what to do<sup>1</sup>.

How should we proceed if

$$G \notin \mathcal{F}$$

Interest lies in inference of  $\theta$  not calibrated prediction.

Modelling our way out of trouble has proven to be unsuccessful.

---

<sup>1</sup>Even if we can't agree about it!

## History matching

ABC was proposed as a method of last resort, but there is evidence it works particularly well for mis-specified models.

## History matching

ABC was proposed as a method of last resort, but there is evidence it works particularly well for mis-specified models.

**History matching** was designed for inference in mis-specified models. It seeks to find a NROY set

$$\mathcal{P}_\theta = \{\theta : S_{HM}(\hat{F}_\theta, y) \leq 3\}$$

where

$$S_{HM}(F_\theta, y) = \frac{|\mathbb{E}_{F_\theta}(Y) - y|}{\sqrt{\text{Var}_{F_\theta}(Y)}}$$

## History matching

ABC was proposed as a method of last resort, but there is evidence it works particularly well for mis-specified models.

**History matching** was designed for inference in mis-specified models. It seeks to find a NROY set

$$\mathcal{P}_\theta = \{\theta : S_{HM}(\hat{F}_\theta, y) \leq 3\}$$

where

$$S_{HM}(F_\theta, y) = \frac{|\mathbb{E}_{F_\theta}(Y) - y|}{\sqrt{\text{Var}_{F_\theta}(Y)}}$$

ABC approximates the posterior as

$$\pi_\epsilon(\theta) \propto \pi(\theta) \mathbb{E}(\mathbb{I}_{S(\hat{F}_\theta, y) \leq \epsilon})$$

for some choice of  $S$  (typically  $S(\hat{F}_\theta, y) = \rho(\eta(y), \eta(y'))$  where  $y' \sim F_\theta$ ) and  $\epsilon$ .

## History matching

ABC was proposed as a method of last resort, but there is evidence it works particularly well for mis-specified models.

**History matching** was designed for inference in mis-specified models. It seeks to find a NROY set

$$\mathcal{P}_\theta = \{\theta : S_{HM}(\hat{F}_\theta, y) \leq 3\}$$

where

$$S_{HM}(F_\theta, y) = \frac{|\mathbb{E}_{F_\theta}(Y) - y|}{\sqrt{\text{Var}_{F_\theta}(Y)}}$$

ABC approximates the posterior as

$$\pi_\epsilon(\theta) \propto \pi(\theta) \mathbb{E}(\mathbb{I}_{S(\hat{F}_\theta, y) \leq \epsilon})$$

for some choice of  $S$  (typically  $S(\hat{F}_\theta, y) = \rho(\eta(y), \eta(y'))$  where  $y' \sim F_\theta$ ) and  $\epsilon$ .

They have thresholding of a score in common and are algorithmically comparable (thresholding).

## History matching and ABC

These methods (anecdotally) seem to work better in mis-specified situations.

Why?

## History matching and ABC

These methods (anecdotally) seem to work better in mis-specified situations.

Why?

They differ from likelihood based approaches in that

- They only use some aspect of the simulator output
  - ▶ Typically we hand pick which simulator outputs to compare, and weight them on a case by case basis.
- Potentially use generalised scores/loss-functions
- The thresholding type nature potentially makes them somewhat conservative
  - ▶ Bayes/Max-likelihood estimates usually concentrate asymptotically. If  $G \notin \mathcal{F}$  can we hope to learn precisely about  $\theta$ ?

## History matching and ABC

These methods (anecdotally) seem to work better in mis-specified situations.

Why?

They differ from likelihood based approaches in that

- They only use some aspect of the simulator output
  - ▶ Typically we hand pick which simulator outputs to compare, and weight them on a case by case basis.
- Potentially use generalised scores/loss-functions
- The thresholding type nature potentially makes them somewhat conservative
  - ▶ Bayes/Max-likelihood estimates usually concentrate asymptotically. If  $G \notin \mathcal{F}$  can we hope to learn precisely about  $\theta$ ?

# Conclusions

## Uncertainty quantification

- The challenge for a statistician is to be involved in several fields of application and to use that to motivate theoretical contributions.
- Uncertainty quantification (UQ) has grown in importance as a field, and has penetrated the scientific and industrial consciousness.
- UQ (as a field) is a blend of statistics (and increasingly machine learning), applied maths, and application knowledge.
- One of the key challenges for the future is understanding the errors, deciding which are important, and where to our spend time and energy.

# Conclusions

## Uncertainty quantification

- The challenge for a statistician is to be involved in several fields of application and to use that to motivate theoretical contributions.
- Uncertainty quantification (UQ) has grown in importance as a field, and has penetrated the scientific and industrial consciousness.
- UQ (as a field) is a blend of statistics (and increasingly machine learning), applied maths, and application knowledge.
- One of the key challenges for the future is understanding the errors, deciding which are important, and where to our spend time and energy.

Thank you for listening!