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

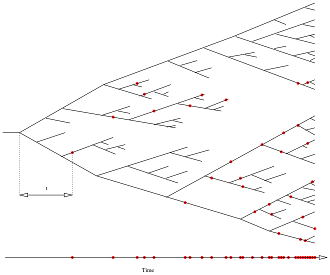
Uncertainty Quantification and inverse problems

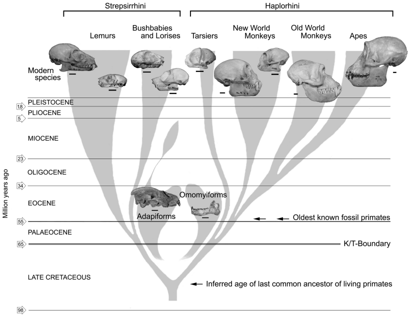
Uncertainty Quantification (UQ) $\equiv$ statistics with complex models

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

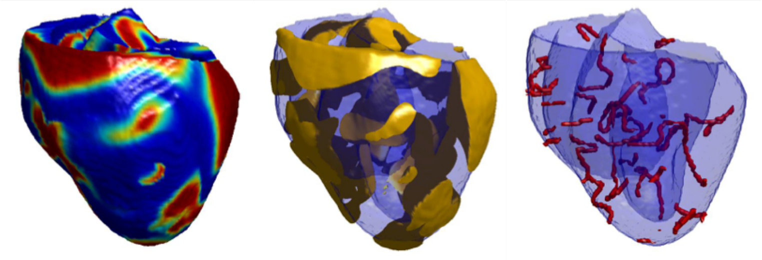
Inverse problems/Calibration/Parameter estimation/...

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





Why do we need UQ?
Atrial fibrillation



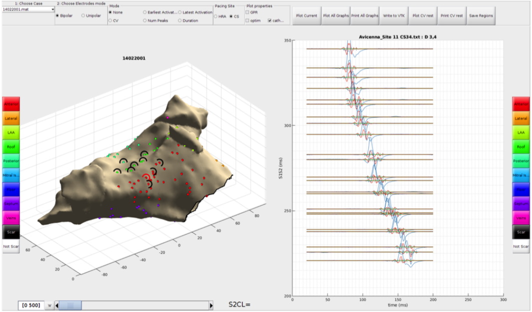
Atrial fibrillation (AF) - rapid and uncoordinated electrical activation (arrhythmia) leading to poor mechanical function.

- 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, Steve Neiderer, Jeremy Oakley

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



Use complex electrophysiology simulation using monodomain eqn on shell anatomy.

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

We need to

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

Tools

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

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

Tools

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

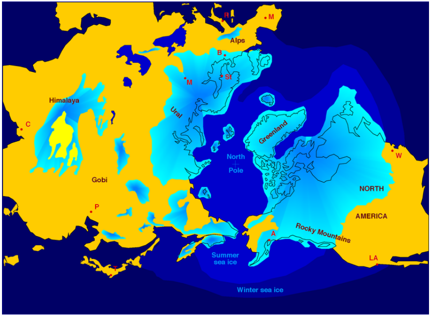
$$\pi(\theta|D) = \frac{\pi(D|\theta)\pi(\theta)}{\pi(D)}$$

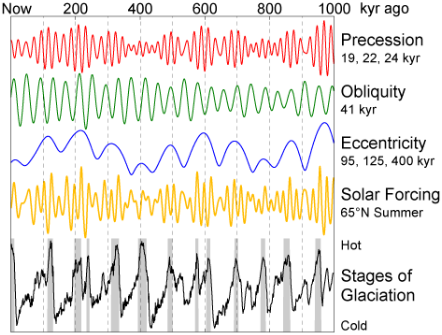
Approaches/tools

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

What drives the glacial-interglacial cycle?

With Jake Carson, Simon Preston, Michel Crucifix





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

- 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)*

Focus is primarily on simple phenomenological models embedded within a statistical state space model

$$\mathrm{d}X_t = g(X_t, \theta)\mathrm{d}t + F(t, \gamma)\mathrm{d}t + \Sigma \mathrm{dW}$$

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

$X_t \in \mathbb{R}^p$ is the state of the climate at time t, and typically $X_{t,1}$ = ice volume, and then $X_{t,2}$ may, eg, represent CO_2, ocean temp, etc, or be undefined.
Typically these models have 10-15 parameters $(\theta, \gamma, s, d, \Sigma, \sigma)$ to be estimated.

Data

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

Variation in the ratio $\delta^{18}O$ in marine sediments and ice cores informs us about historic temperatures.
The raw measurements are of $\delta^{18}O$ as a function of depth in a core: age must be inferred.

Data

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

Variation in the ratio $\delta^{18}O$ in marine sediments and ice cores informs us about historic temperatures.
The raw measurements are of $\delta^{18}O$ 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

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

Where $X_{1:T} = \{X_1, \ldots, X_T\}$.

- 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 for $\pi(X_{1:T}|y_{1:T},\theta)$

- Requires careful proposal design: use Brownian bridge proposals to nudge X_t towards the data.

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

- Requires careful proposal design: use Brownian bridge proposals to nudge X_t 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 (Andrieu & Roberts (2009)) to estimate $\pi(\theta|y_{1:T})$.

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

- Requires careful proposal design: use Brownian bridge proposals to nudge X_t 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 (Andrieu & Roberts (2009)) to estimate $\pi(\theta|y_{1:T})$.

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

- Introduce M parameter particles $\theta_1,\ldots,\theta_M$
- For $t=1,\ldots,T$
  - For each θ_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: ODP677

Carson, Crucifix, Preston, W. 2017

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

- Strongest evidence found for the true model each time

Results: ODP677

Carson, Crucifix, Preston, W. 2017

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

- Strongest evidence found for the true model each time

ODP677 is a marine 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	4.0×10^{24}	1.1×10^{28}
	Unforced	3.5×10^{26}	1.6×10^{18}
T06	Forced	3.3×10^{25}	4.5×10^{29}
	Unforced	1.7×10^{28}	3.3×10^{21}
PP12	Forced	1.5×10^{22}	1.8×10^{34}

Results: ODP677

Carson, Crucifix, Preston, W. 2017

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

- Strongest evidence found for the true model each time

ODP677 is a marine 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	4.0×10^{24}	1.1×10^{28}
	Unforced	3.5×10^{26}	1.6×10^{18}
T06	Forced	3.3×10^{25}	4.5×10^{29}
	Unforced	1.7×10^{28}	3.3×10^{21}
PP12	Forced	1.5×10^{22}	1.8×10^{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?

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

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?

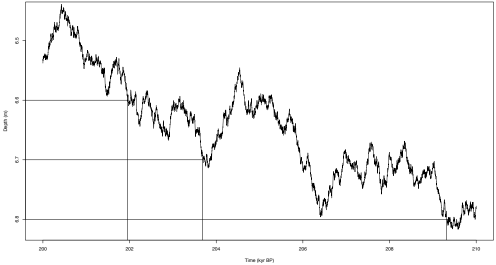
$$\text{Target} \quad \pi(\theta_k, T_{1:N}, X_{1:N}, \mathcal{M}_k | y_{1:N})$$

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 $T_N = 0$

$$\mathrm{d}H = -\mu_s \mathrm{d}T + \sigma \mathrm{d}W$$

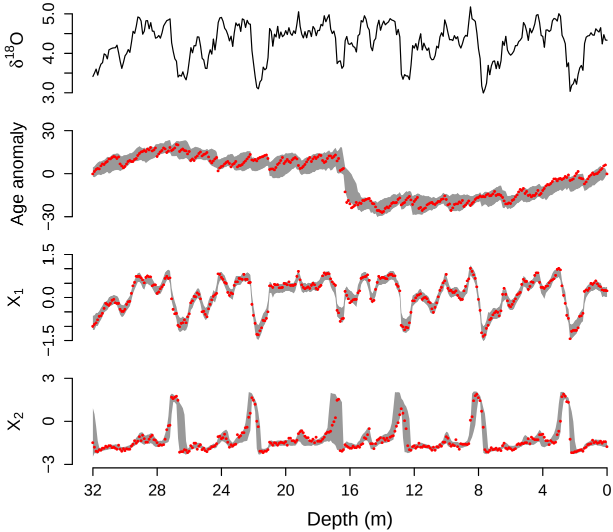
Slices are then taken through the core at specific depths $H_1, \ldots, H_N$.



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

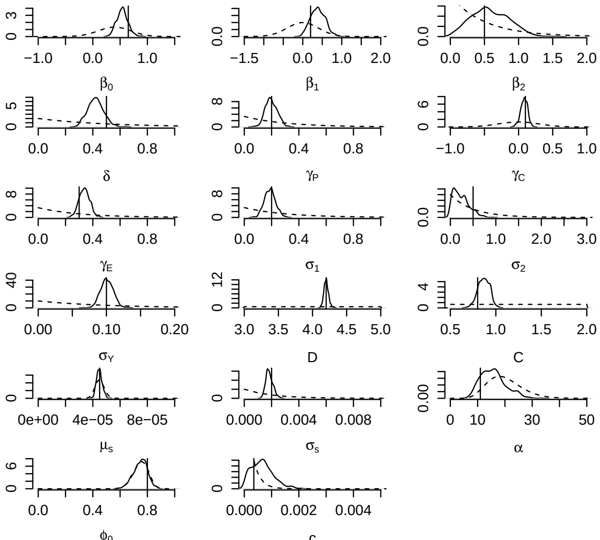
Data (black line) and estimated 95% HDR intervals (grey). True values in red.



The trend has been removed from the age-depth plot.

Simulation study results - parameter estimation

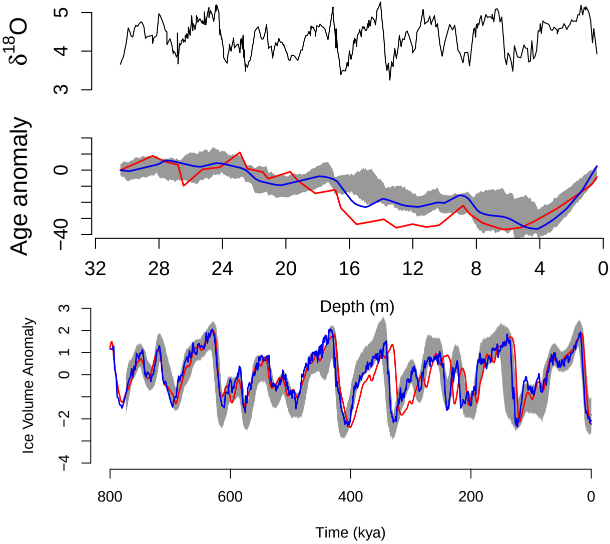
Posterior (solid lines), prior (dashed lines), and true values (vertical lines)



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)

Blue: LR04 reconstruction, Red: H07 reconstruction, Grey: 95% HDR

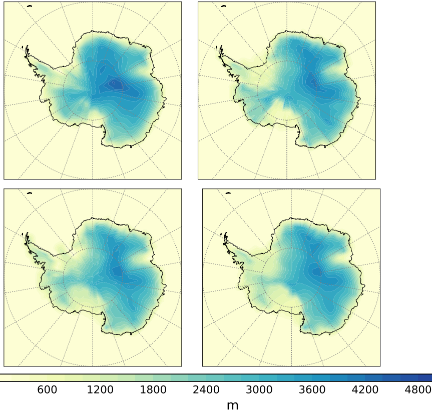


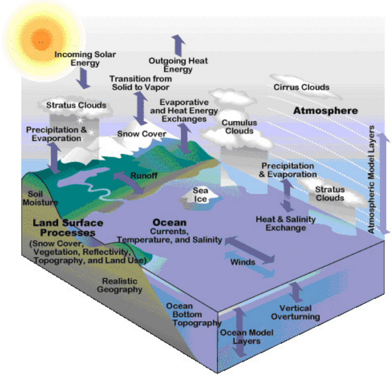
Advantages: full UQ, model selection, simultaneous parameter estimation and climate reconstruction

Expensive and high dimensional simulators

Louise Sime (BAS)

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

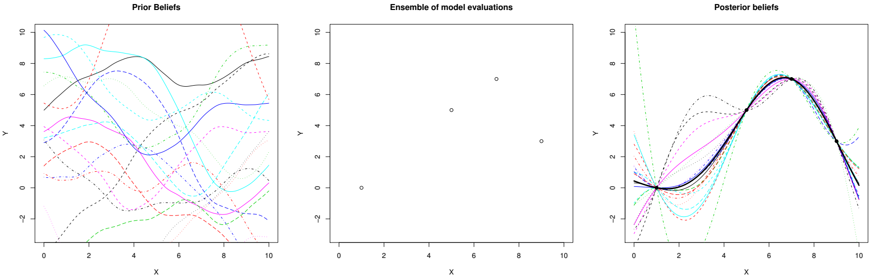




How can we infer the shape of the ice-sheet at the last glacial maximum?

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, W., & Graham (2016)

How do we deal with high dimensional inputs and outputs?

- Holden, Edwards, Ridgwell, Wilkinson, ... (under revision) show that limiting warming to $< 1.6°C$ is still achievable with 50% probability under rapid decarbonisation.

Carbon capture and storage (CCS) Panacea

With Andrew Cliffe, Henry Power

Knowledge of the physical problem is encoded in a simulator f

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

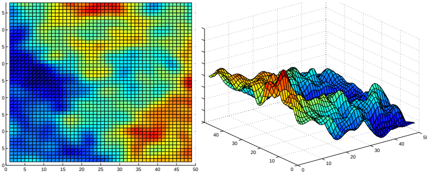
Inputs:

Permeability field, K
(2d field)

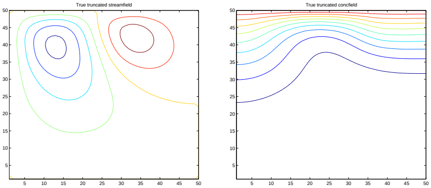
$$\downarrow$$
$$f(K)$$
$$\downarrow$$

Outputs:

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



$\downarrow f(K)$



Emulating simulators with high dimensional input

Crevillén-Garca, W., Shah, & Power (2017), Tian, W., Yang, Power, Fargerlund, & Niemi(2017)

The permeability field needs to be known at a large (but finite) number of locations, eg, if solver grid is 100×100, $\dim K = 10^4$

- Impossible to directly model $f : \mathbb{R}^{10,000} \to \mathbb{R}$

Emulating simulators with high dimensional input

Crevillén-Garca, W., Shah, & Power (2017), Tian, W., Yang, Power, Fargerlund, & Niemi(2017)

The permeability field needs to be known at a large (but finite) number of locations, eg, if solver grid is 100×100, dim $K = 10^4$

- Impossible to directly model $f : \mathbb{R}^{10{,}000} \to \mathbb{R}$

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

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

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

  where λ_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

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

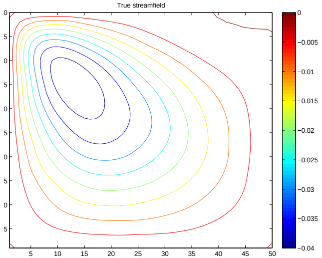
we reduce the modelling problem to one of modelling $f : \mathbb{R}^n \to \mathbb{R}$

Emulating from fields to fields

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

Now consider emulating the stream function and concentration field outputs (also 100×100 matrices).

We can use the singular value decomposition (SVD) to reduce the output dimension.



Form the SVD of $Y = LDR^T$

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

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

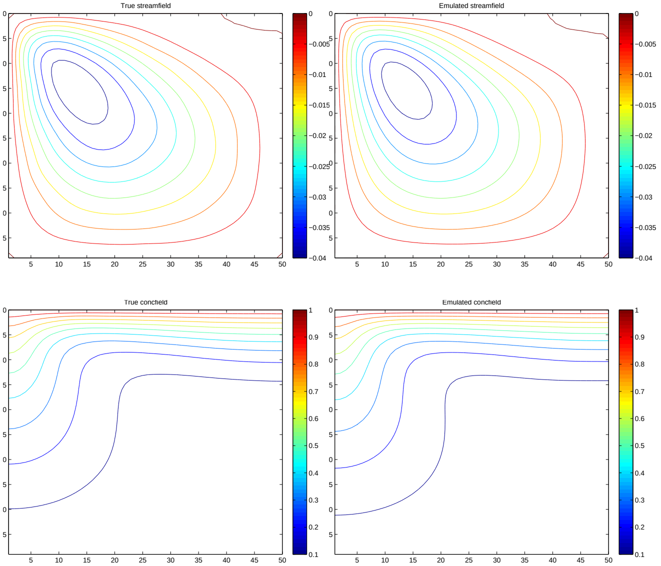
- If $R_*^T = (t_1, \ldots, t_N)$, where each t_i is a vector of length k, then

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

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.

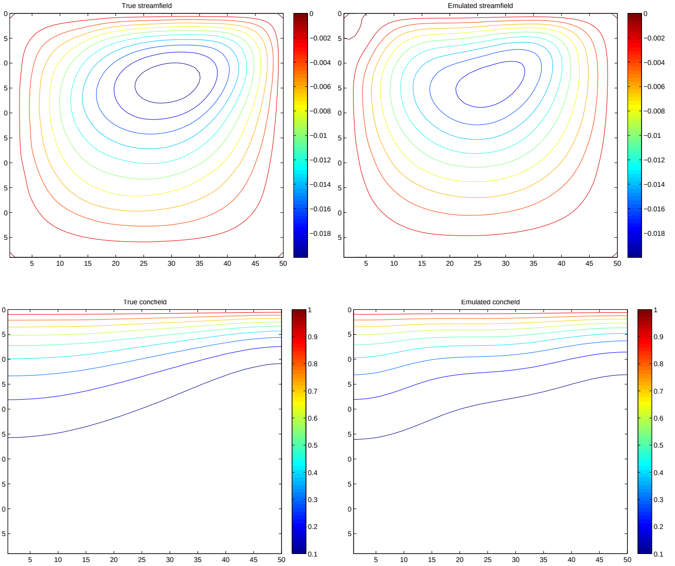
Emulating the stream function and concentration fields

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



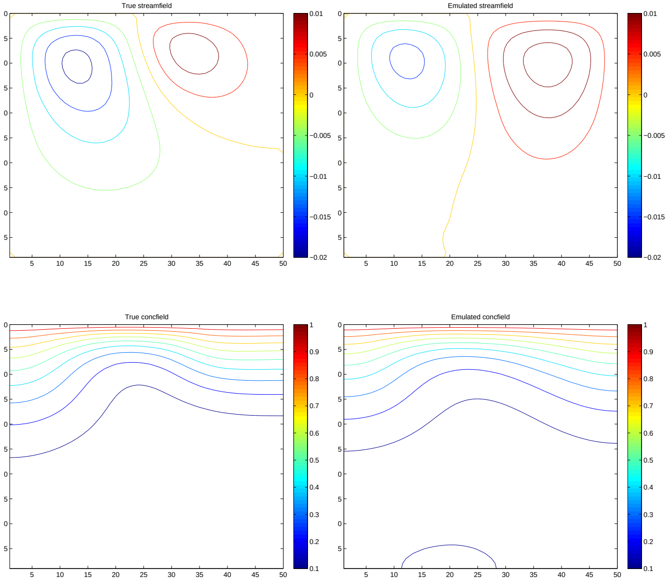
Emulating the stream function and concentration fields

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

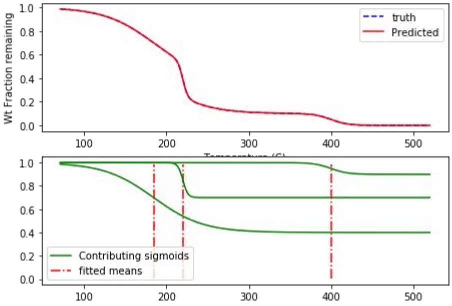


Emulating the stream function and concentration fields

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

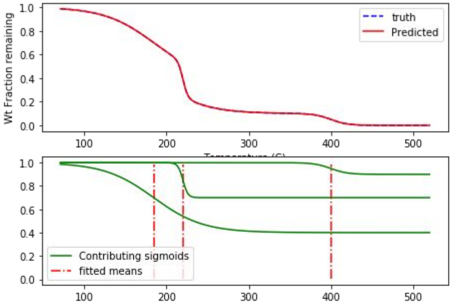


Other decompositions



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

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. SVD/KL dimension reduction is unsupervised. Instead, can we do optimal model/dimension reduction when focussed on the calibration problem?

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

PhD studentship provided by TOTAL .

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

Rejection ABC

Uniform Rejection Algorithm

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

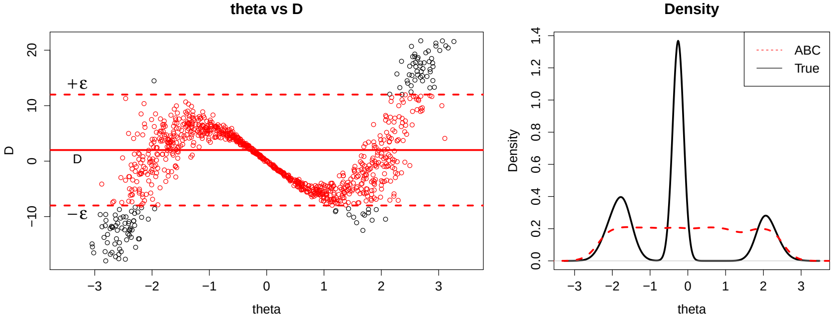
ϵ reflects the tension between computability and accuracy.

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

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

Simple $\rightarrow$ Popular with non-statisticians

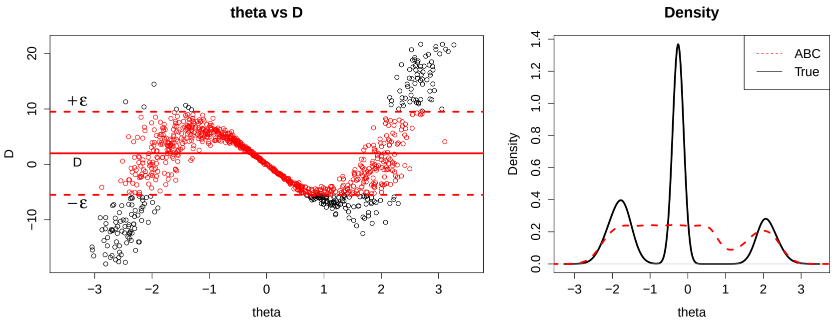
$\epsilon = 10$



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

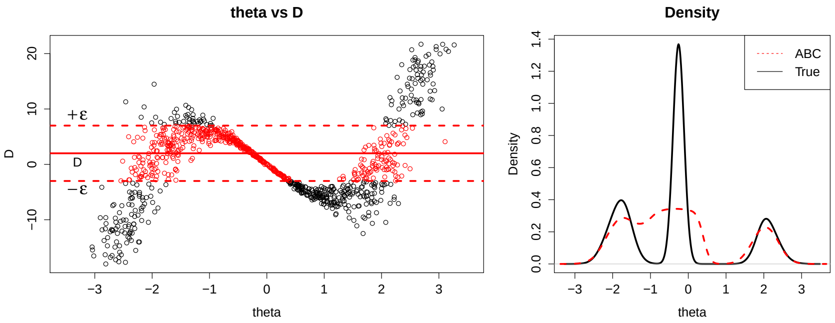
$$\rho(D, X) = |D - X|, \qquad D = 2$$

$\epsilon = 7.5$



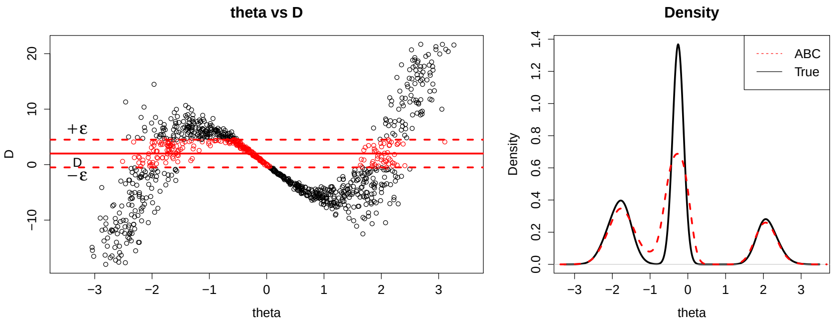
theta vs D
20
10
0
−10
D
+ε
D
−ε
−3
−2
−1
0
1
2
3
theta
Density
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
Density
ABC
True
−3
−2
−1
0
1
2
3
theta

$\epsilon = 5$



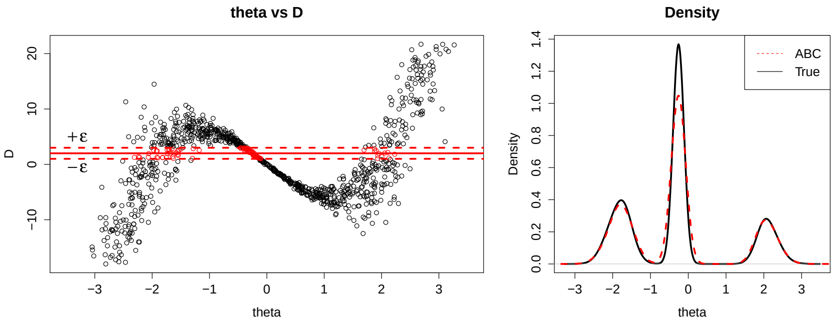
theta vs D
+ε
D
−ε
20
10
0
−10
−3
−2
−1
0
1
2
3
theta
Density
ABC
True
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
−3
−2
−1
0
1
2
3
theta

$\epsilon = 2.5$



theta vs D
D
theta
+ε
D
−ε
20
10
0
−10
−3
−2
−1
0
1
2
3
Density
Density
theta
ABC
True
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4

$\epsilon = 1$


theta vs D
D
theta
+ε
−ε
20
10
0
−10
−3
−2
−1
0
1
2
3
Density
Density
ABC
True
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
theta

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 θ 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
- $\vdots$

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

Can lead to orders of magnitude speed up in computation.

Surrogate ABC

- Wilkinson 2014
- Meeds and Welling 2014
- Gutmann and Corander 2015
- Strathmann, Sejdinovic, Livingstone, Szabo, Gretton 2015
- $\vdots$

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

Can lead to orders of magnitude speed up in computation.

Constituent elements:

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

Building surrogates through history-matching waves

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

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

Building surrogates through history-matching waves

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

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

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

- 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 θ is implausible if

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

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

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

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

Active learning for history-matching/GP-ABC

With James Hensman

Sequential design is the key to further reducing computational burden.

- Given our current knowledge, where should we next run the simulator to most improve our knowledge?

One option is to minimise the expected average entropy of the history match

Active learning for history-matching/GP-ABC

With James Hensman

Sequential design is the key to further reducing computational burden.

- Given our current knowledge, where should we next run the simulator to most improve our knowledge?

One option is to minimise the expected average entropy of the history match

- Any GP emulator allows us to calculate a probabilistic classification

$$p(\theta) = \mathbb{P}(\theta \text{ implausible})$$

- The entropy of our belief at θ is

$$E(\theta) = -p\log p - (1-p)\log(1-p)$$

  and the average entropy is

$$\mathcal{E} = \int E(\theta)\mathrm{d}\theta$$

- Choose the next design point to minimise the expected value of $\mathcal{E}$.

Challenges: Inference under discrepancy

How should we do inference if the model is imperfect?

[1] 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

$$y \sim G$$

[1] 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[1].

[1] 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[1].

How should we proceed if

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

[1] 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[1].

How should we proceed if

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

Interest lies in inference of θ not calibrated prediction.

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

[1] 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

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

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

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

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 ϵ.

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

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

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 ϵ.

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 θ?

Conclusions

- 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 is a blend of statistics (and increasingly machine learning), applied maths, and application knowledge.
- One of the key challenges is understanding the errors, deciding which are important, and thus where to spend our time and energy.

Conclusions

- 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 is a blend of statistics (and increasingly machine learning), applied maths, and application knowledge.
- One of the key challenges is understanding the errors, deciding which are important, and thus where to spend our time and energy.

Thank you for listening!