ABC for climate

CSML Reading group 4

Yannis Zachos

November 6, 2020

Motivation

- **Task**: Calibration of computer model simulator.
- **Challenge 1**: Finite computational resources.
- **Challenge 2**: Computationally intensive simulation.

- **Example**: Complex climate modelling - Why?
    1. High-dimensional outputs;
    2. Model component coupling;
    3. High spatial resolution.

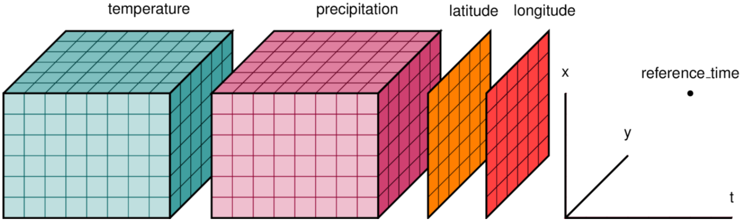

Figure: Data structure `xarray` visualisation for 2 state variables[1].

[1] Image obtained from here.

Motivation II

- We have access to observations C_{obs}.
- Treat climate simulator as a black box mapping $f : \Theta \to \mathcal{C}$ with $f(\theta) = \mathcal{C}_{sim}$.
- Parameter posterior becomes

$$\pi(\theta|\mathcal{C}_{obs}) \propto \underbrace{\pi(\theta)}_{\text{prior}} \int \underbrace{\pi(\mathcal{C}_{obs}|\mathcal{C}_{sim})}_{\text{stats model}} \underbrace{\pi(\mathcal{C}_{sim}|\theta)}_{\text{simulator likelihood}} d\mathcal{C}_{sim}$$

Challenges

1. Can run simulator N times to get $\{\theta^{(i)}, \mathcal{C}_{sim}^{(i)}\}_{i=1}^{N}$ - cannot run MCMC for small N.
2. Simulator-reality discrepancy can be large - prior specification difficulty.

Strategies

1. Make the most out of N simulator runs through careful experimental design:
   1.1 Space-filling domain sampling e.g. *Maximin Latin Hypercube* (MLH).
   1.2 Entropy-minimising domain sampling.
2. Reduce parameter space through *History matching* (HM) or ABC;
3. Approximate simulator with a cheap surrogate model (meta-model) called an *emulator*.

Role of ABC

- Replace statistics model $\pi(\mathcal{C}_{obs}|\mathcal{C}_{sim})$ with acceptance kernel $\mathbb{I}(\rho(\mathcal{C}_{obs}, \mathcal{C}_{sim}) \leq \epsilon)$.
- This induces a uniform distribution on the simulator discrepancy - not good but pragmatic choice?
- (Kennedy and O'Hagan 2001) fitted a stationary Gaussian process to the model discrepancy term.
- Can tune ρ, ϵ to get desired acceptance rate.

History matching

- **Intuition**: Rule out regions of Θ that yield implausible $\mathcal{C}_{sim}$ relative to $\mathcal{C}_{obs}$ after accounting for model discrepancy.
- **Task**: For a set of plausible climate states $\mathcal{P}_{\mathcal{C}}$ find

$$\mathcal{P}_\theta := \{\theta \in \Theta : f(\theta) \in \mathcal{P}_{\mathcal{C}}\}$$

- For deterministic f (as in climate), $\mathcal{P}_\theta = \emptyset$ implies there are no plausible states - poor model.
- For stochastic f, $\mathcal{P}_\theta \neq \emptyset$ as the plausibility-implausibility boundary is soft - same for MCMC.
- Cannot obtain all of $\mathcal{P}_\theta$ - even partial description is useful.

Comparison between History matching and ABC

- $\mathbb{I}(\rho(\mathcal{C}_{obs}, \mathcal{C}_{sim}) \leq \epsilon) = 0$ implies that θ is implausible.
- They both don't use a detailed discrepancy model.
- ABC provides a soft plausibility boundary whereas HM provides a hard one.
- In ABC we select summary statistics on the basis of what is informative of θ.
- In HM outputs that are informative of θ may have large discrepancies, which can lead to type I errors.

Simulator emulation

- Derive cheap approximation $\tilde{f}(\theta)$ to $f(\theta)$.
- Emulator is built based on $\mathcal{D} := \{\theta_i, f(\theta_i)\}_{i=1}^{N}$.
- Gaussian processes are great emulators! Why?
- Non-stationary GP kernels (treed GPs) can provide flexible emulators for all of Θ, multi-task GPs ...
- It is critical to discriminate between $\theta \in \mathcal{P}_\theta$ and $\theta \notin \mathcal{P}_\theta$ even if f is a poor model.

Sequential history matching

- How do we learn complex $\mathcal{P}_\theta$?
  1. *A priori* start with $\mathcal{P}_\theta^{(0)} = \Theta$.
  2. Choose design $\mathcal{D}_\theta^{(1)} = \{\theta_i \in \Theta : i = 1, \ldots, n_1\}$ and run the simulator to get ensemble $\mathcal{D}^{(1)} = \{(\theta_i, C_i) = f(\theta_i) : \theta_i \in \mathcal{D}_\theta^{(1)}\}$.
  3. Build emulator $\tilde{f}_{(1)}$ and use it to predict $\tilde{\mathcal{P}}_\theta^{(1)}$.
  4. DITTO for $\mathcal{D}_\theta^{(2)}$, $\mathcal{D}^{(2)}$, $\tilde{f}^{(2)}$, $\tilde{\mathcal{P}}_\theta^{(2)}$ and so on...
- $\tilde{f}^{(i)}$ can be built for $\theta \in \mathcal{P}_\theta^{(i-1)}$ instead of $\theta \in \Theta$ to reduce f-variability.
- *Greedy* approach - no way to backtrack.
- Trade-off between optimatility and efficiency. When to explore/exploit?

Experimental design

Approach 1

- For exploration of entire Θ or $\mathcal{P}_\theta^{(i)}$ we can use space filling designs e.g. Maximin Latin Hypercube (MLH).
- MLH generates near-random samples from a multidimensional distribution of θ.

Approach 2

- Select a new design point that minimises the entropy of parameter regions we are uncertain about.
- This is equivalent to maximising new information or maximally reducing emulator uncertainty.
- Formally add $\theta = \arg\min_{\theta_i} \mathbb{E}\left[\bar{H}|\mathcal{D}^{(i-1)} \cup \{\theta_i\}\right]$ to the design, where
  $\bar{H}$: average entropy of the emulator prediction of the plausibility space.

Plausibility criteria

- **Example 1**: (Williamson et al. 2013) used $\mathcal{P}_\theta = \{\theta \in \Theta : |\mathcal{I}(\theta)| \leq \alpha\}$ with

$$\mathcal{I}(\theta) = \frac{|C_{obs} - \mathbb{E}[f(\theta)]|}{\sqrt{\mathrm{Var}\left(C_{obs} - \mathbb{E}[f(\theta)]\right)}}.$$

- Criteria can become progressively more stringent, slowly approaching the final desired criterion $\mathcal{P}_{\mathcal{C}}^{(W)}$.
- Keep track of $|\mathcal{P}_\theta^i - \mathcal{P}_\theta^{i+1}|$ to avoid type-I errors and relax criteria if necessary.
- What about deciding plausibility of regions where emulator is uncertain?
- **Example 2**: Plausibility criterion is $D_- \leq f(\theta) \leq D_+$.
- For GP emulators, $\tilde{f} \sim \mathcal{N}\left(\mu_\theta, \Sigma_\theta\right)$ and therefore $\mathbb{P}_{\tilde{f}}(\theta \in \mathcal{P}_\theta)$ is available in closed form.

Application Ia

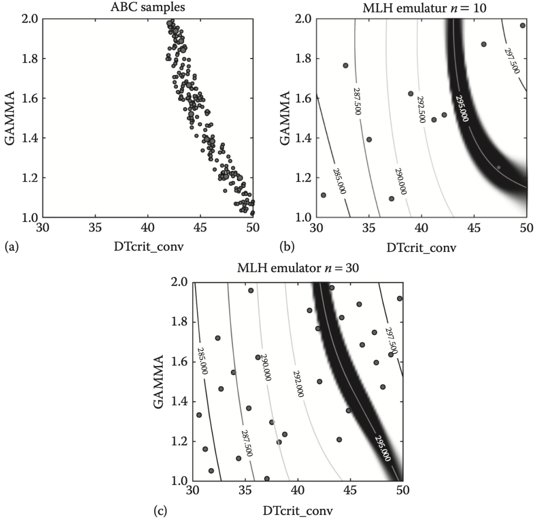

FIGURE 19.1
(a) Accepted samples from the rejection ABC algorithm after 100 (light grey) and 1,000 (dark grey) simulator evaluations. (b and c) The estimated plausible region using an emulator trained with a maximin Latin hypercube design (points shown in grey) with 10 (middle) and 30 (right) simulator evaluations.

Application Ib

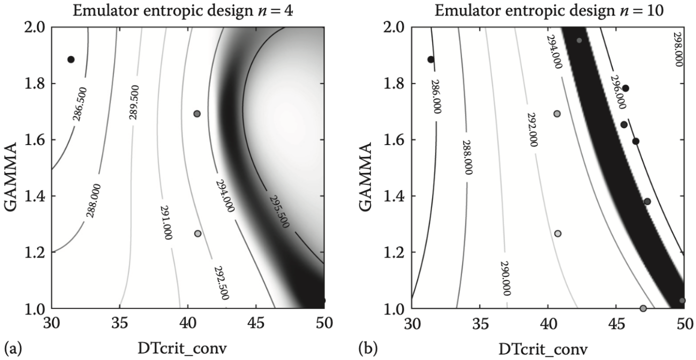

FIGURE 19.2
Results from using an entropy based sequential design. The left-hand column shows the estimated response surface (contours) and $\mathbb{P}(\theta \in \mathcal{P}_\theta)$ (shading), with the design points overlaid. The large dark grey point is the most recently added point. The right-hand column shows the entropy surface. The top row uses four simulator evaluations, and the bottom row uses ten simulator evaluations, all added according to the entropy criterion.

Application II

- $\dim \theta = 24$, 8 plausibility metrics/model outputs.
- $n_{MLH} = 500$ (4 plausible), $n_{EFPS} = 885$ (471 plausible).

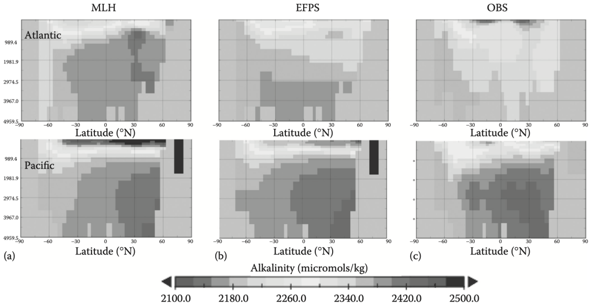

FIGURE 19.3
Cross-sections of ocean alkalinity through the Atlantic (25°W) and Pacific (155°W) Oceans. The figure compares the mean of the training MLH ensemble (a) and the plausibility filtered EFPC ensemble (b) with observations (c).

Discussion

- Is ABC really necessary in this type of modelling? Can we use a better model for $\pi(\mathcal{C}_{obs}|\mathcal{C}_{sim})$?
- Is there a way to efficiently backtrack in SHM?
- Are there more robust plausibility criteria?

Recommended reading I

Kennedy, Marc C and Anthony O'Hagan (2001). "Bayesian calibration of computer models". In: 63, pp. 425–464 (cit. on p. 5).

Williamson, Daniel et al. (Oct. 2013). "History matching for exploring and reducing climate model parameter space using observations and a large perturbed physics ensemble". In: *Climate Dynamics* 41.7-8, pp. 1703–1729. ISSN: 0930-7575. DOI: 10.1007/s00382-013-1896-4. URL: http://link.springer.com/10.1007/s00382-013-1896-4 (cit. on p. 11).