

How does/should Uncertainty Quantification apply to tuning climate models?

Danny Williamson, James Salter

University of Exeter

With thanks to: EPSRC, John Scinocca (CCCMA), Slava Kharin (CCCMA), NERC, NOC, Environment Canada, Southampton, Adam Blaker (NOC)

August 24, 2016

- 1 Climate models and the science policy interface
- 2 Climate model tuning
- 3 Uncertainty Quantification
- 4 The Canadian Climate model
- 5 Spatial UQ for climate models

Climate models and the science policy interface

Process based
physical science

Computer science
and numerics

Observation systems
and data products

CMIP/IPCC
protocols/deadlines

Climate modelling
(at global centres)

```
graph TD; A[Process based physical science] --> D[Climate modelling (at global centres)]; B[Computer science and numerics] --> D; C[Observation systems and data products] --> D; E[CMIP/IPCC protocols/deadlines] --> D;
```

Process based
physical science

Computer science
and numerics

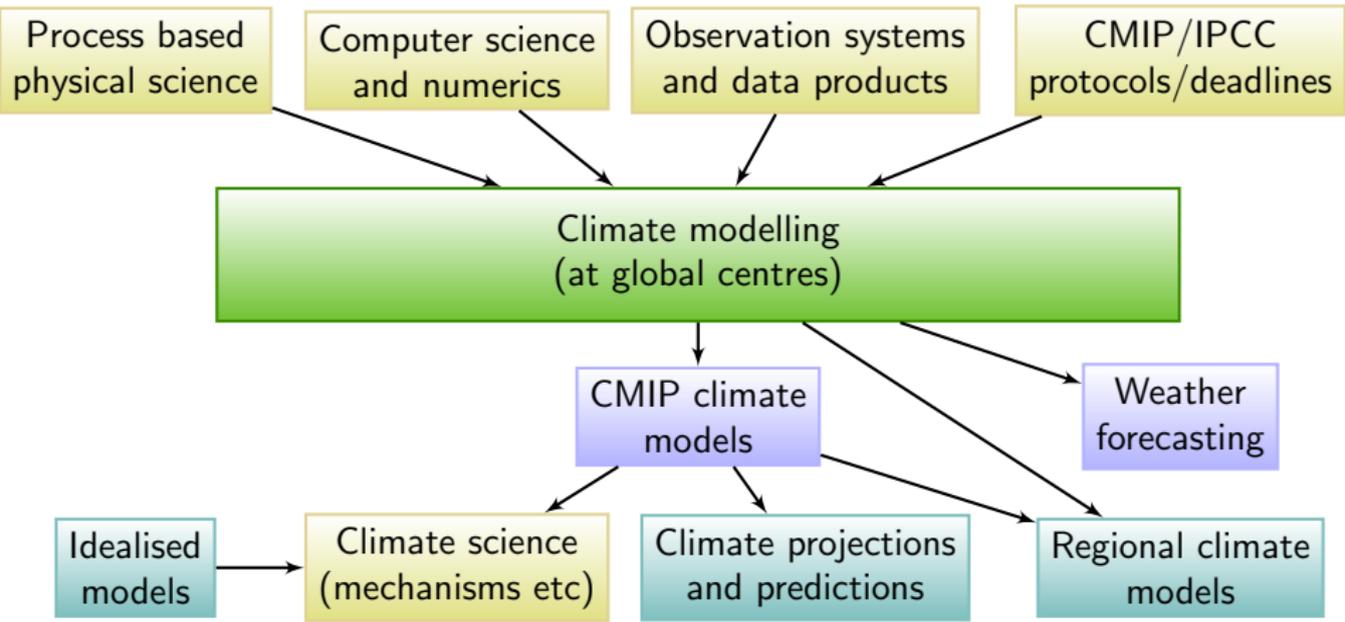
Observation systems
and data products

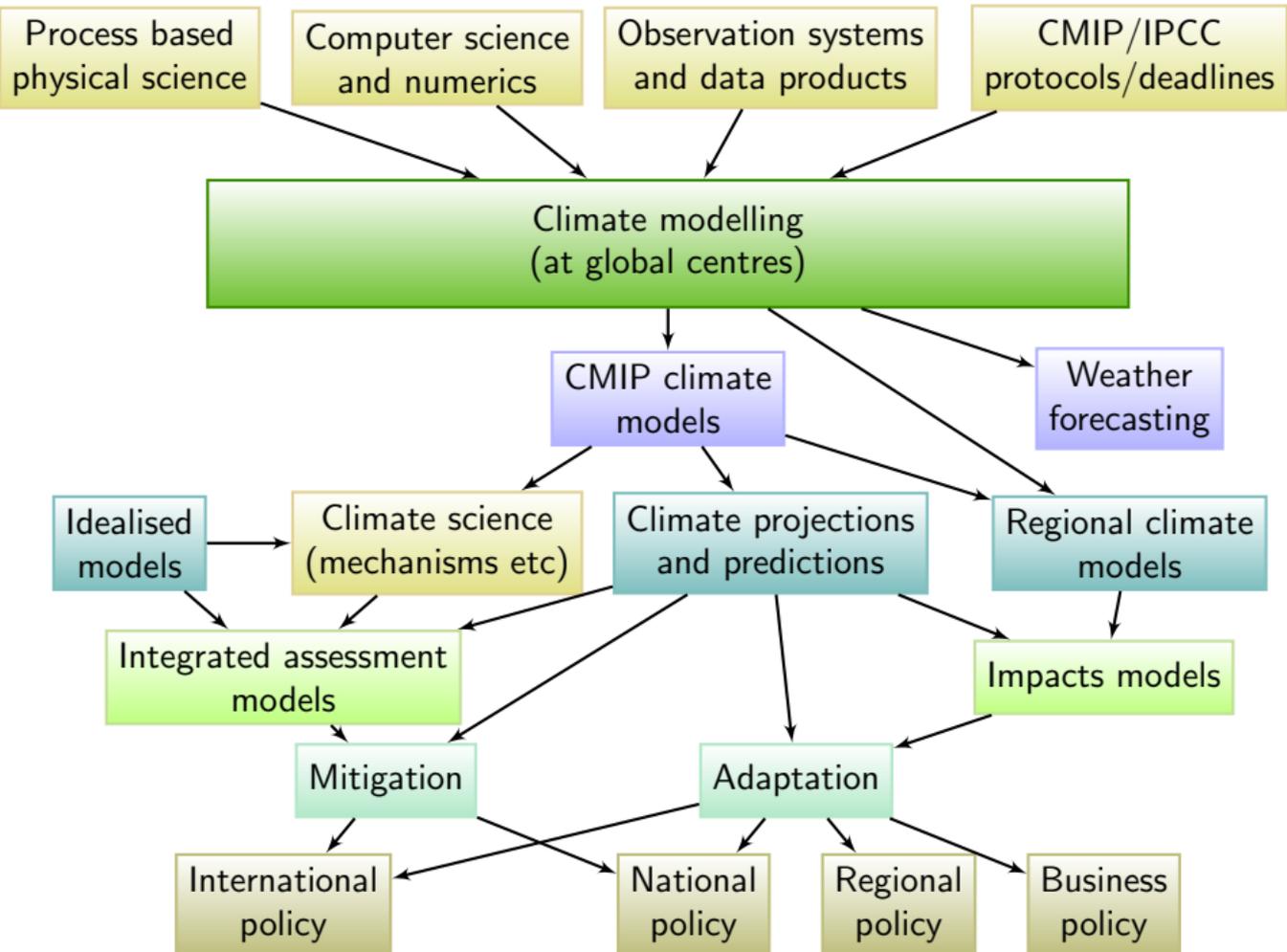
CMIP/IPCC
protocols/deadlines

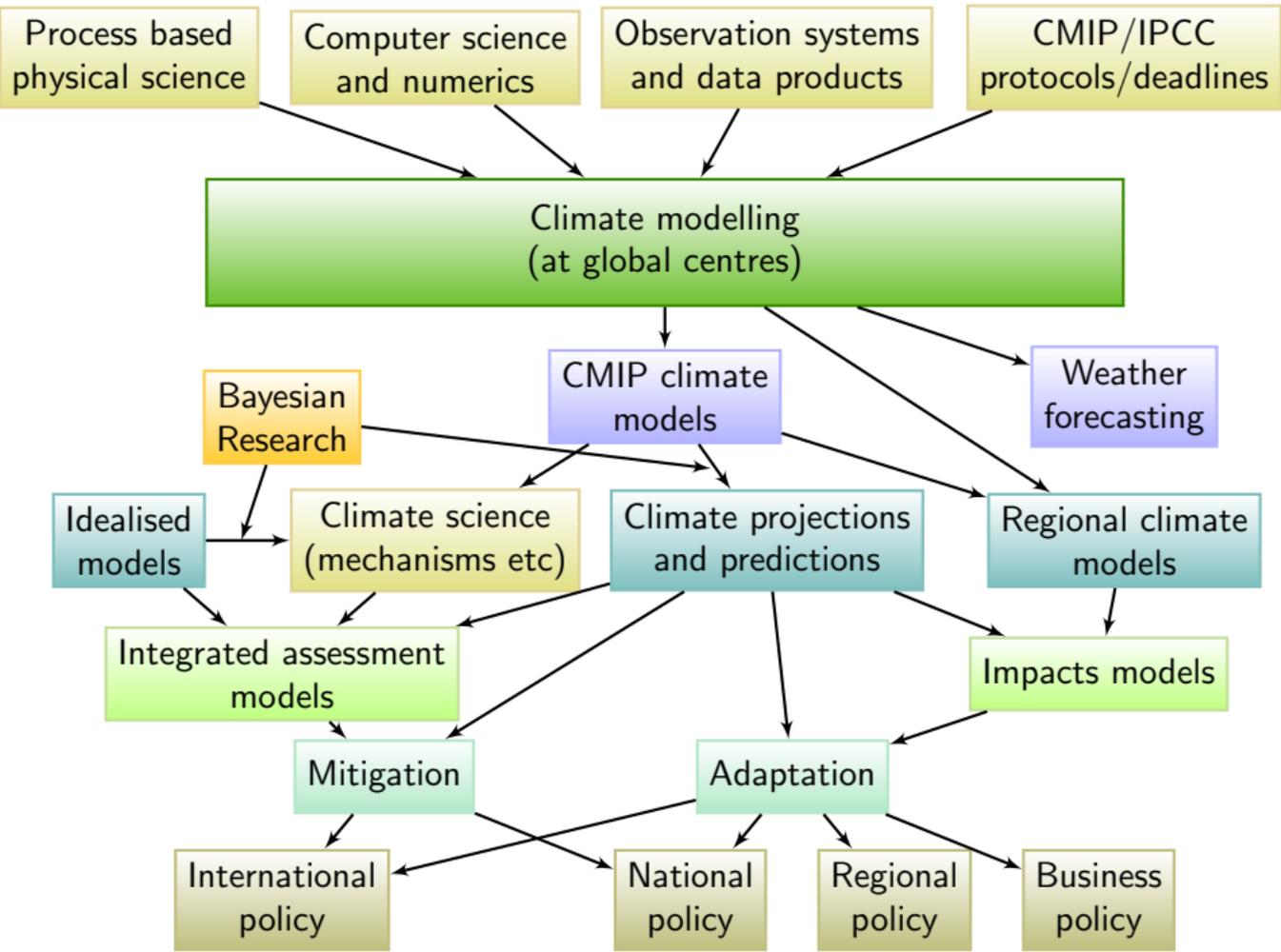
Climate modelling
(at global centres)

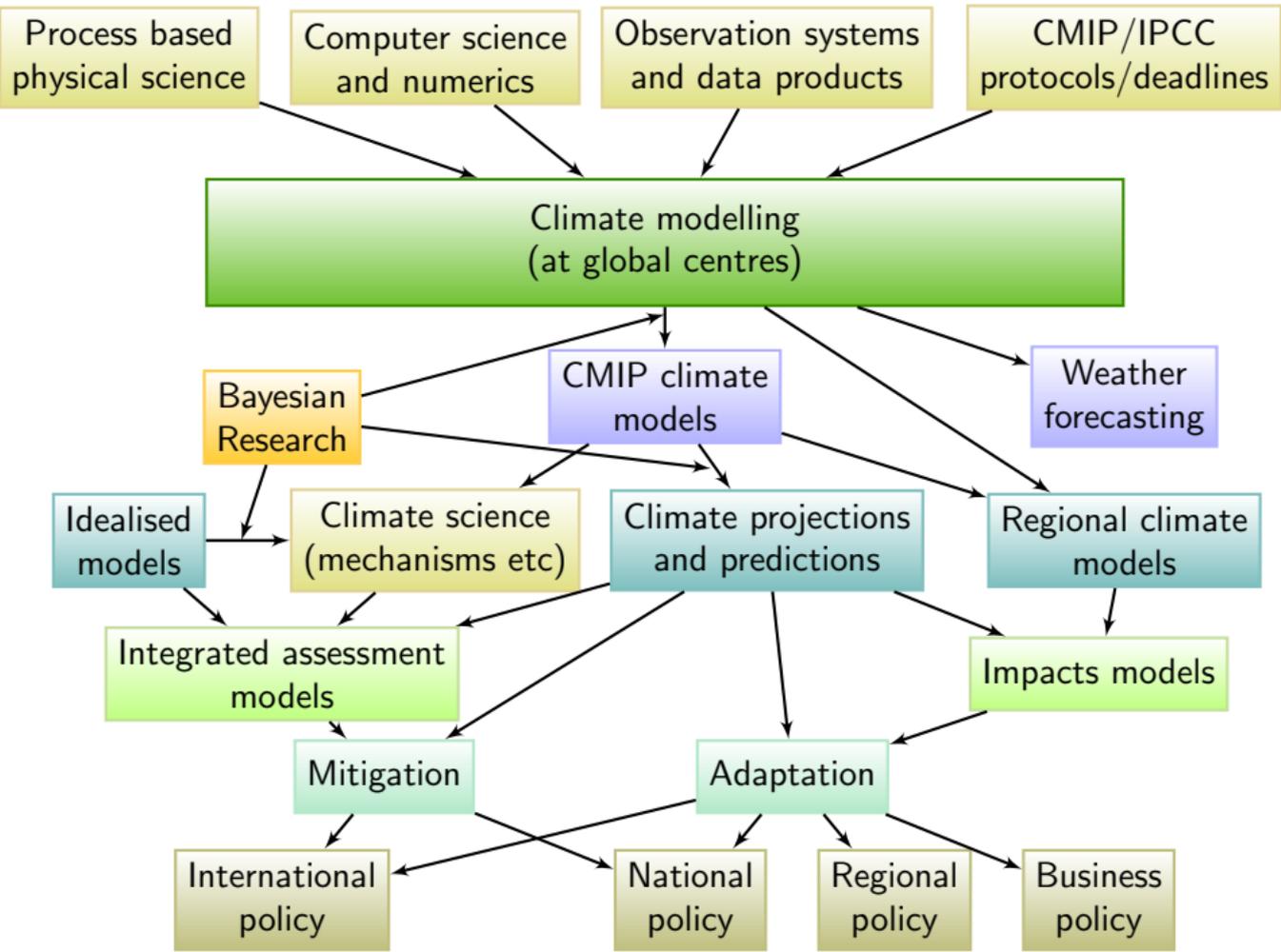
CMIP climate
models

Weather
forecasting









Climate model tuning

How are climate models tuned?

- Behind closed doors
- By committee
- By hand
- By eye

Uncertainty Quantification

Parameter estimation

- “Tuning” to a statistician is “parameter estimation”.
- We usually want to use model $\mathbf{f}(\mathbf{x})$ to learn about the real world \mathbf{y} and we want to choose \mathbf{x}
- If we have observations $\mathbf{z} = \{\mathbf{z}_1, \dots, \mathbf{z}_m\}$ of m parts of \mathbf{y} (e.g. metrics, spatial fields), we can use these to estimate \mathbf{x} .
- So called “automatic” or “objective” tuning methods aim to use \mathbf{z} to find an optimal \mathbf{x} with respect to some cost function

$$C(\mathbf{x}) = \sum_i \|\mathbf{z}_i - \mathbf{f}_i(\mathbf{x})\|_{[i]}$$

- Choice of the $\|\cdot\|_{[i]}$ and methods for estimating $C(\mathbf{x})$ over \mathcal{X} given the expensive $\mathbf{f}(\mathbf{x})$ are crucial here.

Uncertainty

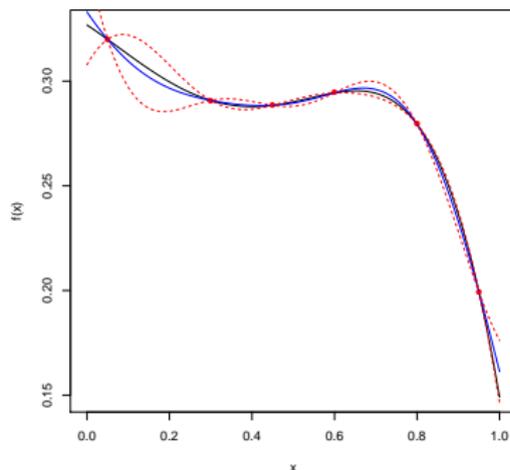
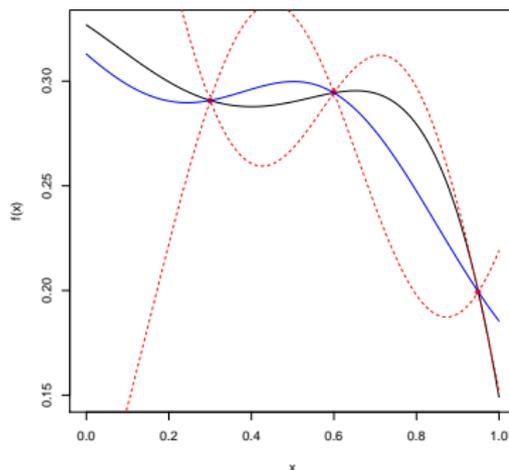
- The $\|z_i - f_i(\mathbf{x})\|_{[i]}$ must account for multiple sources of uncertainty:
- **z**: **Observation uncertainty** (can be large, generally unquantified)
 - Metrics generally not directly observed.
 - Derivation contains many assumptions or involves reanalysis.
- **Internal variability** (generally accounted for)
 - f_i shouldn't match z_i exactly due to sensitive dependence to unknown initial conditions.
- **Structural error/model discrepancy** (almost completely unaddressed)
 - f_i is an imperfect representation of y for any x .

Methods and their application

- Most popular: “Automatic tuning methods” use state of the art optimisation (say with MCMC or genetic algorithms) and embed the actual model inside the calculations.
 - E.g. Zou et al (2014), Zhang et al (2015), Qian et al (2015), Bellprat et al. (2012)
- The “Official statistical approach”: [Bayesian Calibration](#)
 - E.g. Balaji et al. (2016), Sexton et al. (2012), Rougier (2007), LLNL group, ...
- Alternative approaches: [History matching/iterative refocusing](#) (see a few slides time)
 - E.g. Edwards et al (2011), McNeal et al. (2013), Williamson et al. (2013, 2015, 2016),

Emulation with Gaussian processes

- Embedding the full model in an expensive optimisation algorithm won't help the development of state of the art models.
- An emulator is a statistical model that approximates the climate model and gives a fast prediction with uncertainty:



Emulation with Gaussian processes

$$f_i(\mathbf{x}) | \mathbf{F}, \{\beta, \phi_i\} \sim \text{GP}(\mathbf{m}^*(\mathbf{x}), R^*(\cdot, \cdot; \phi))$$

with

$$\mathbf{m}^*(\mathbf{x}) = \sum_j \beta_{ij} \mathbf{g}_j(\mathbf{x}) + \mathbf{K}(\mathbf{x}) V^{-1} \left(\mathbf{F} - \sum_j \beta_{ij} \mathbf{g}_j(\mathbf{X}) \right)$$

$$R^*(\mathbf{x}, \mathbf{x}'; \phi) = R(\mathbf{x}, \mathbf{x}'; \phi) - \mathbf{K}(\mathbf{x}) V^{-1} \mathbf{K}(\mathbf{x}')^T$$

and $K(\mathbf{x}) = R(\mathbf{x}, \mathbf{X}, \phi)$

$$R(\mathbf{x}, \mathbf{x}'; \phi) = \sigma^2 \left(\nu + (1 - \nu) \prod_{k=1}^d \exp \left\{ -\theta_k |x_k - x'_k|^{\kappa_k} \right\} \right), \quad \phi = \{\sigma, \nu, \theta, \kappa\}.$$

Bayesian Calibration

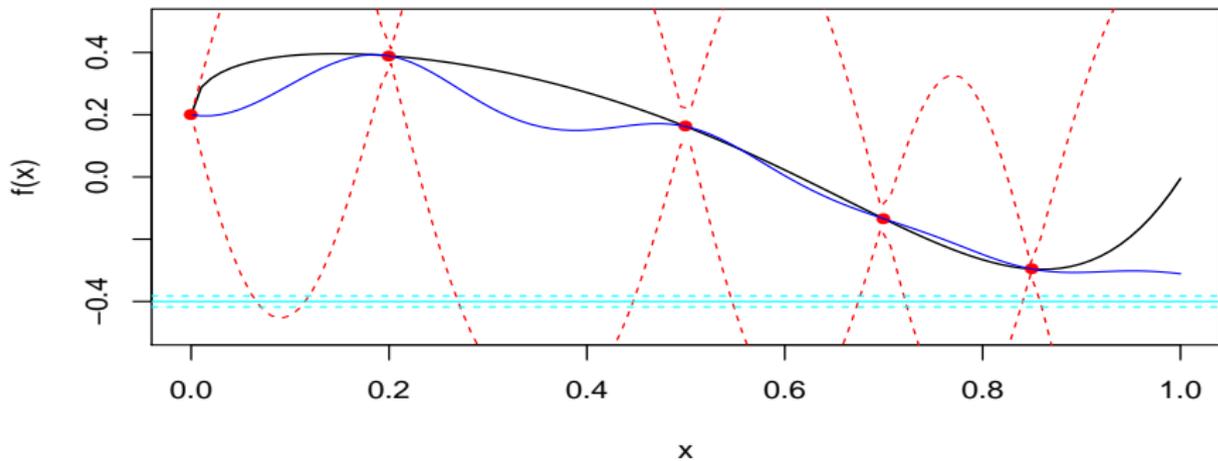
- The Kennedy O'Hagan framework for Bayesian calibration of computer models has over 1200 citations and is now ubiquitous.

$$\begin{aligned}\mathbf{y} &= \mathbf{f}(\mathbf{x}^*) + \boldsymbol{\eta}; & \boldsymbol{\eta} &\sim GP(\mathbf{m}(\cdot), c_{\eta}(\cdot, \cdot)) \\ \mathbf{z} &= \mathbf{y} + \mathbf{e}; & \mathbf{e} &\sim N(0, \Sigma_e)\end{aligned}$$

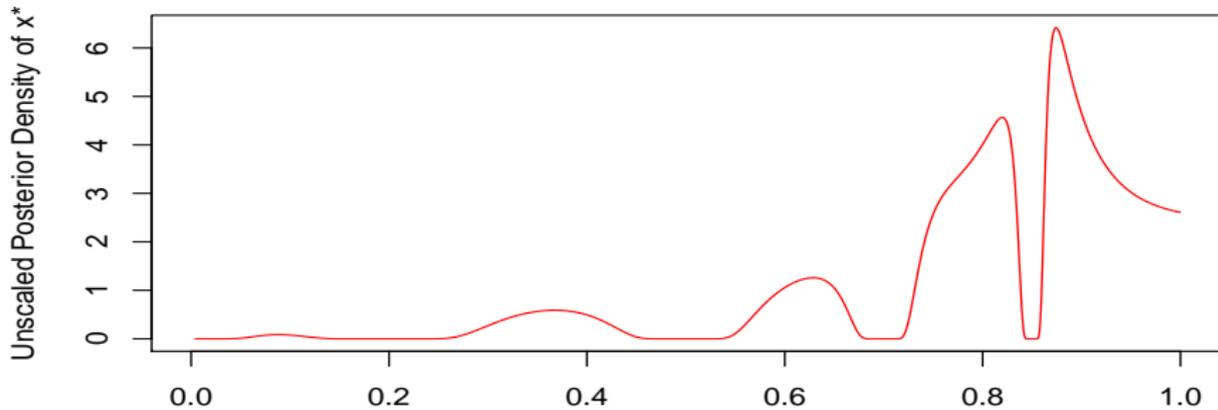
- We observe n training runs of the model, $\mathbf{F} = \mathbf{f}(\mathbf{x}_1), \dots, \mathbf{f}(\mathbf{x}_n)$
- Bayesian calibration exploits the above structure and uses Gaussian processes to allow us to sample from $\pi(\mathbf{x}^* | \mathbf{F}, \mathbf{z})$.

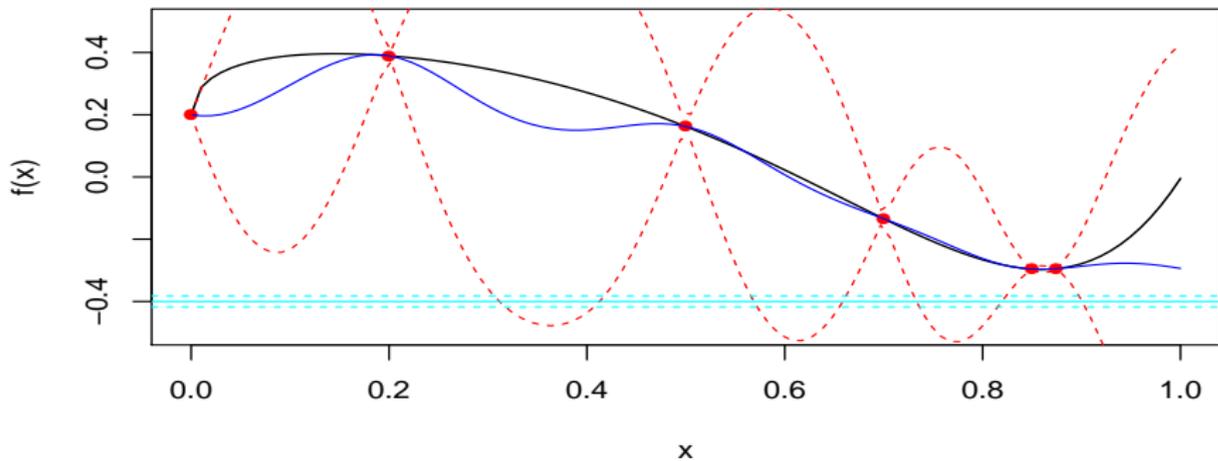
Main issues with calibration for climate models

1. Over tuning/Over fitting: **Optimising the partial state vector biases other parts of it that the modellers really care about**
2. Calibration without informative prior discrepancy is not a good idea (Brynjarsdottir and O'Hagan, 2014)
3. The model is unfit for purpose cannot be an answer: “whack-a-mole”
4. For high dimensional output, principal component methods are likely to lead to whack-a-mole (see later).

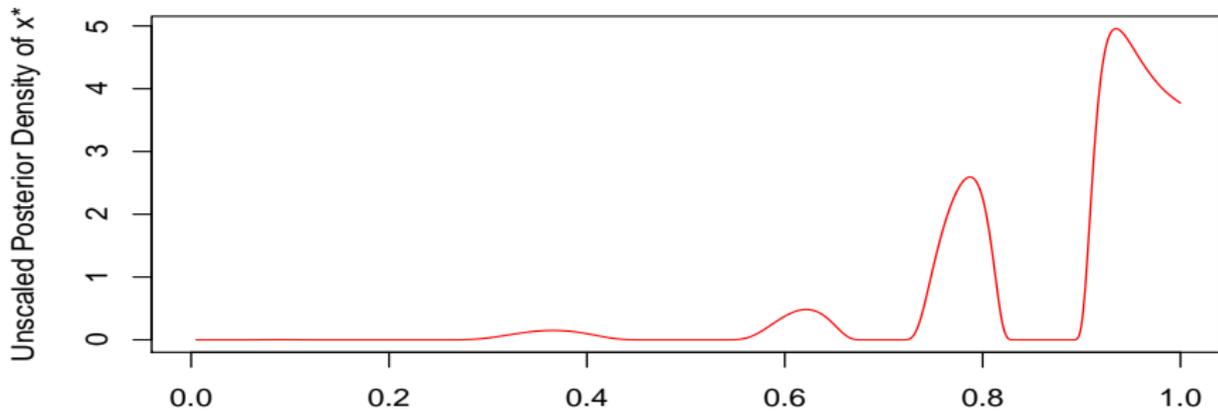


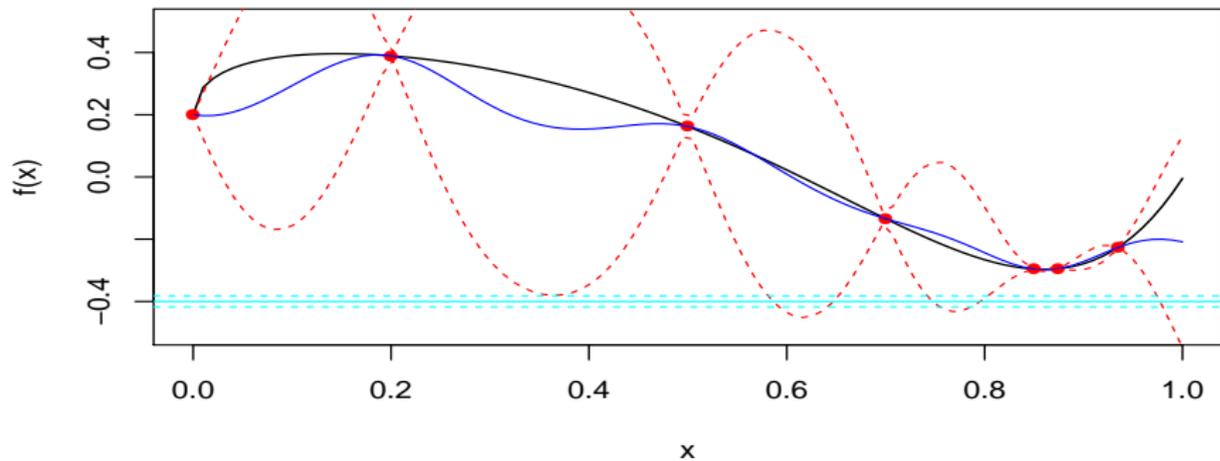
Calibration result



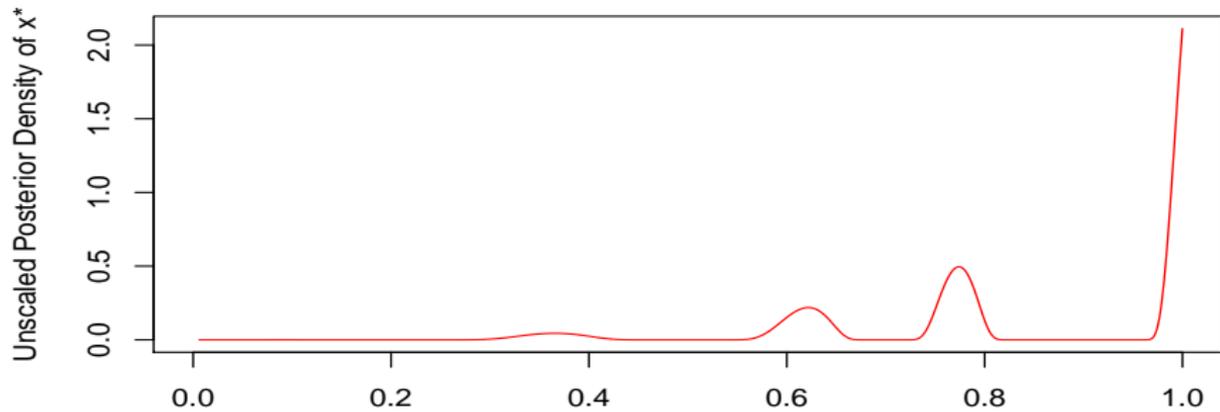


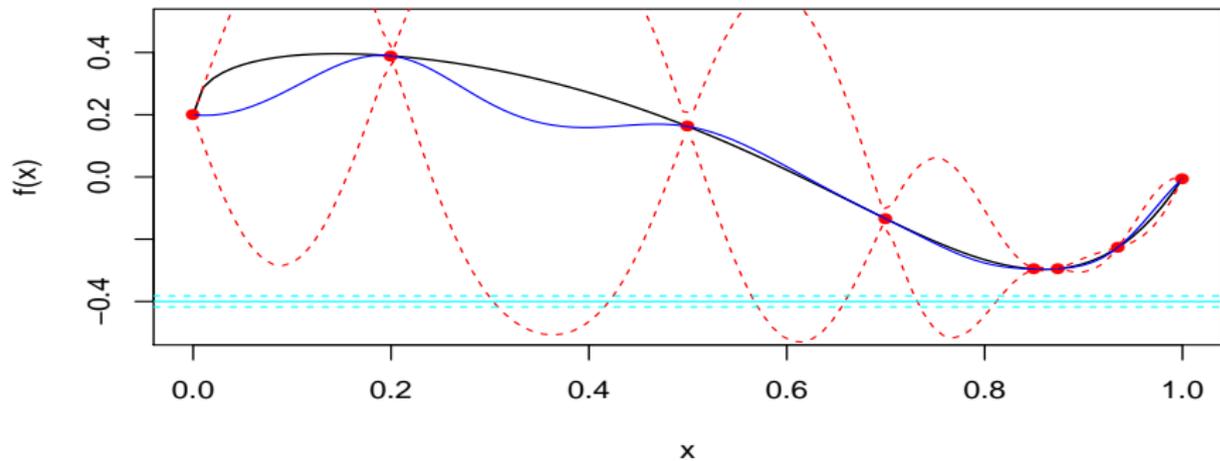
Calibration result



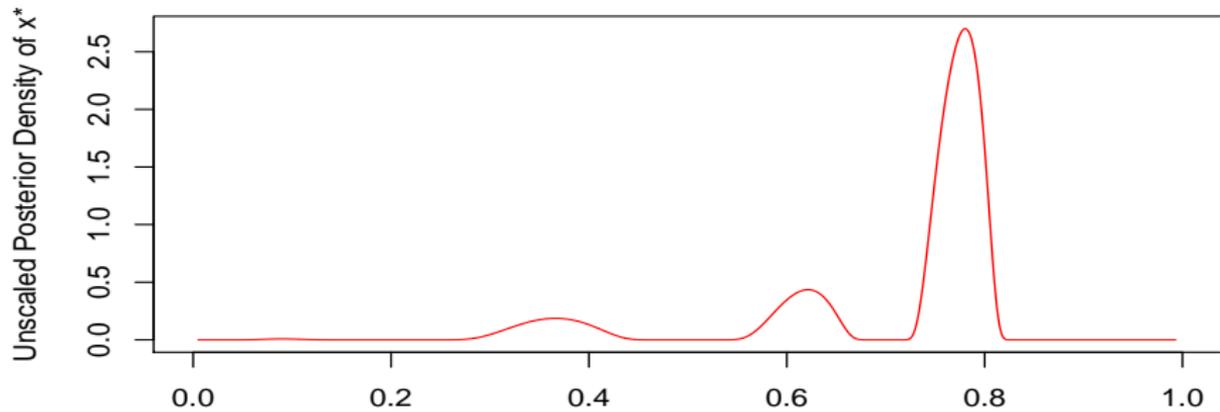


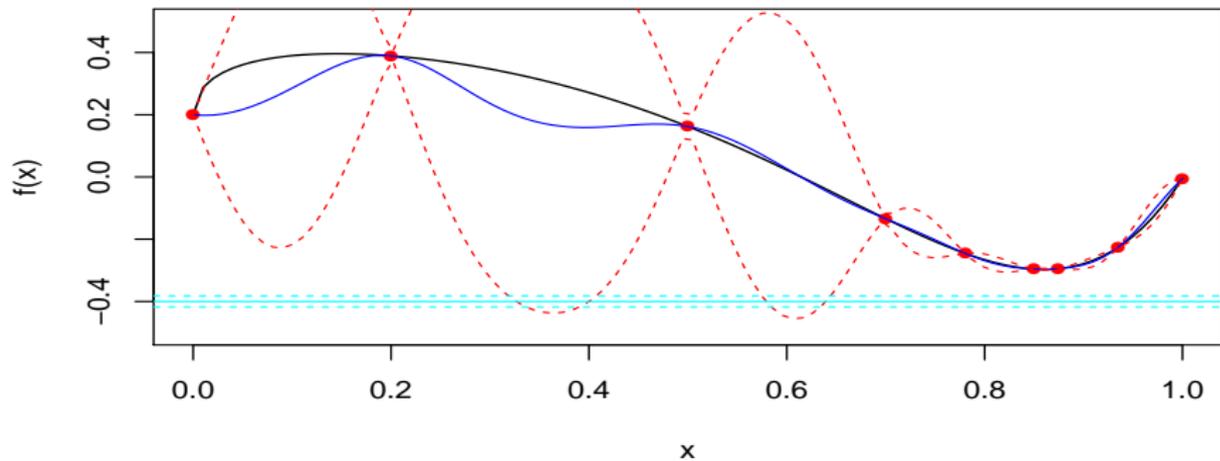
Calibration result



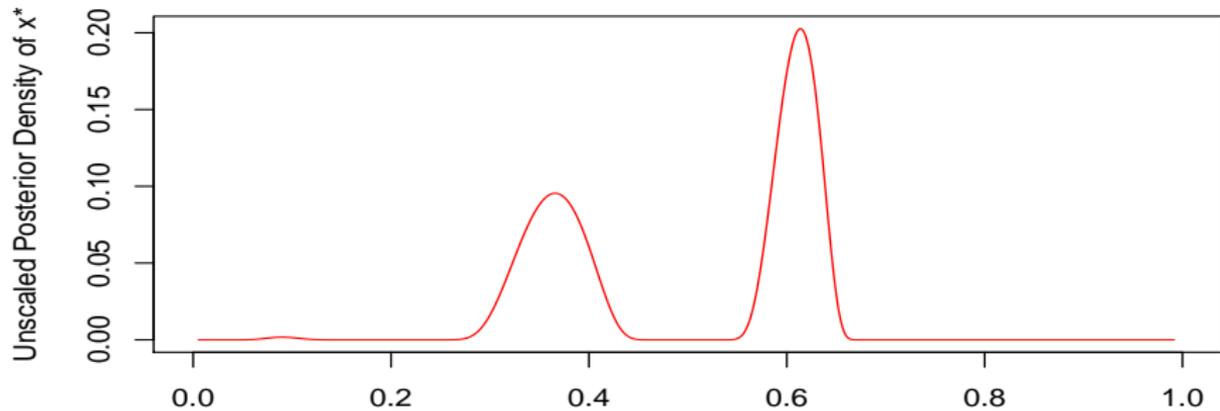


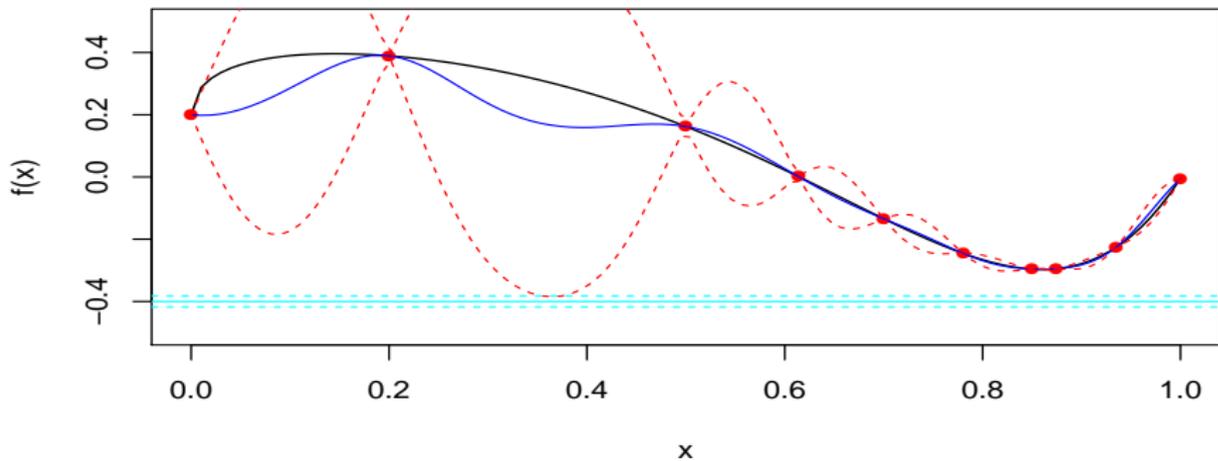
Calibration result



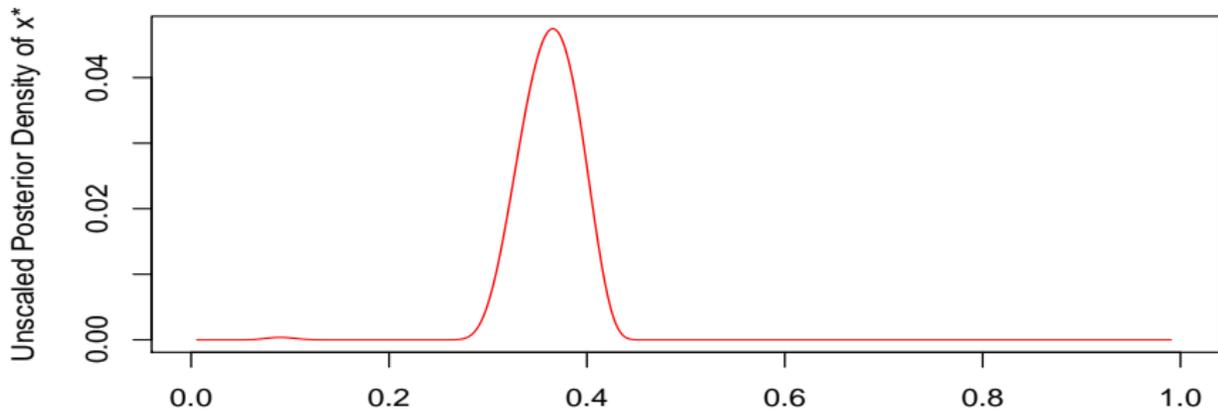


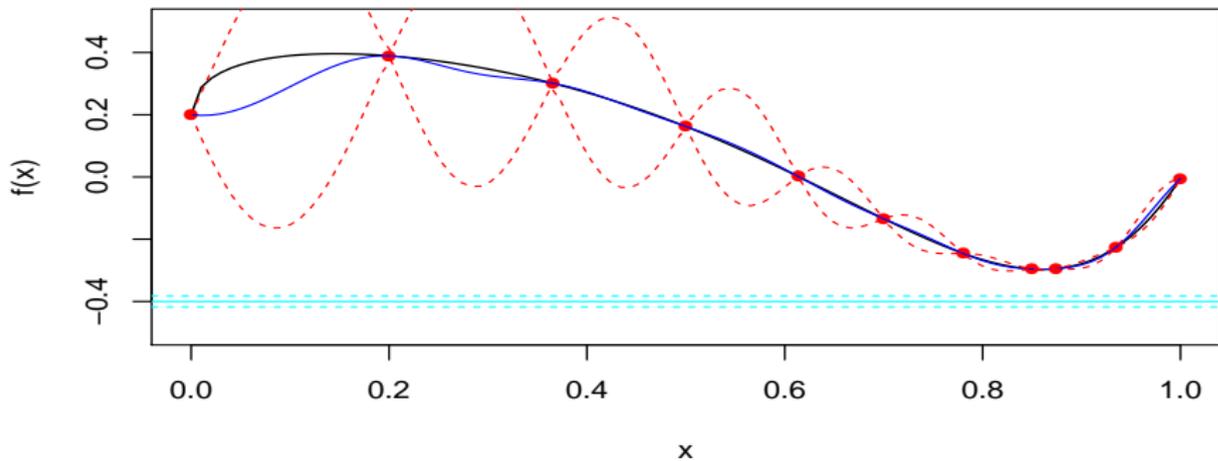
Calibration result



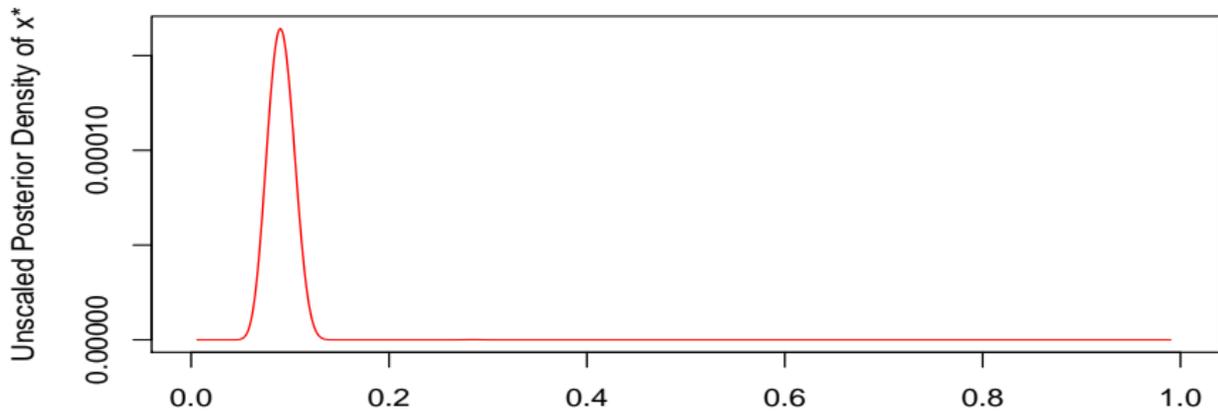


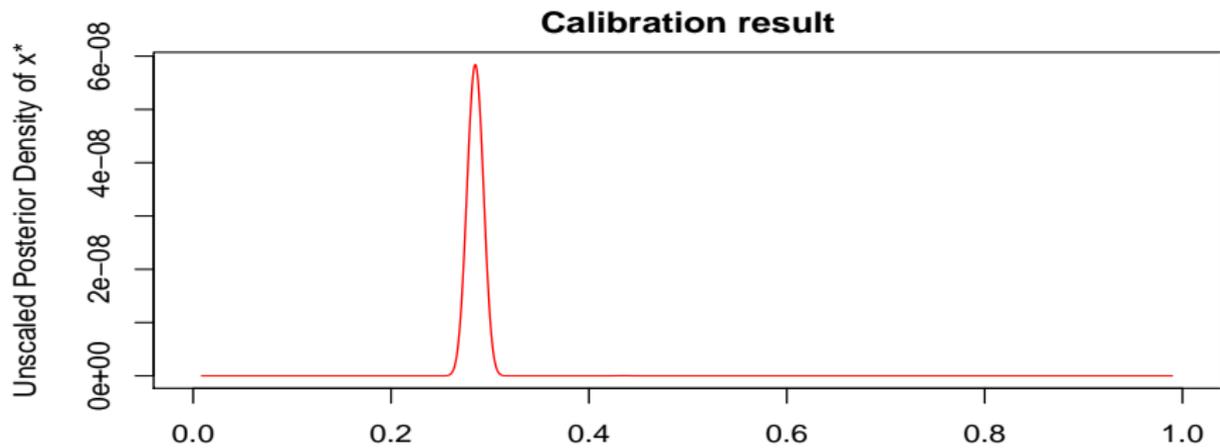
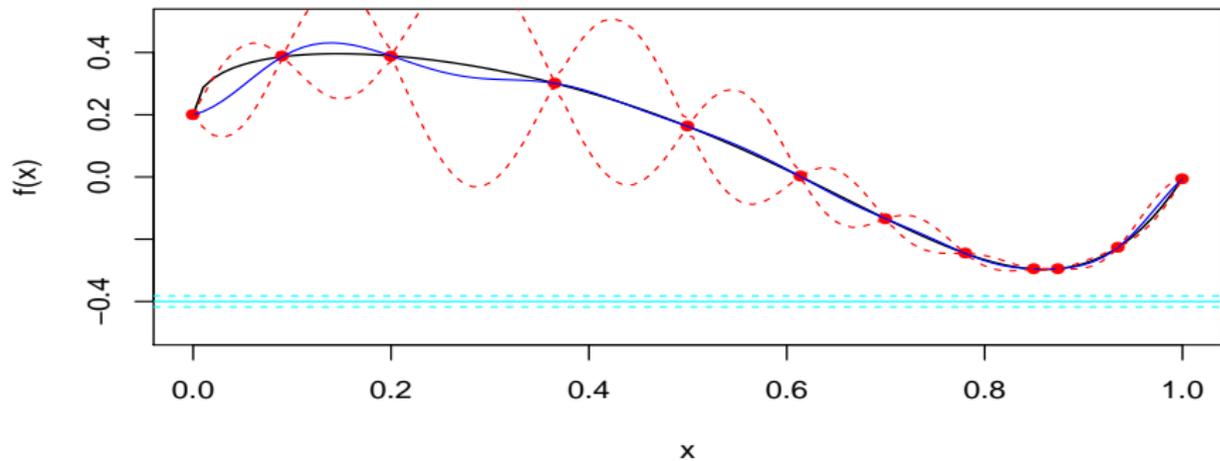
Calibration result

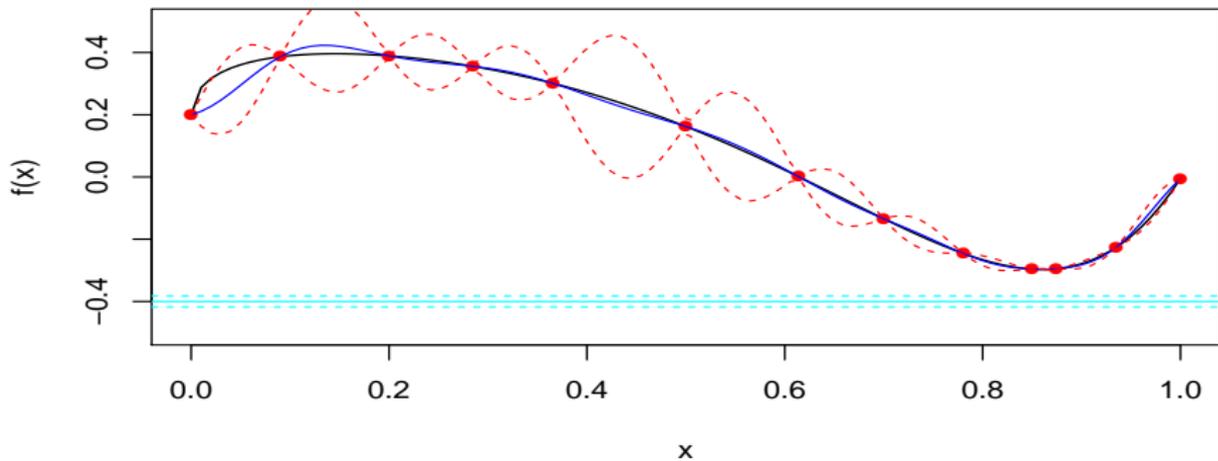




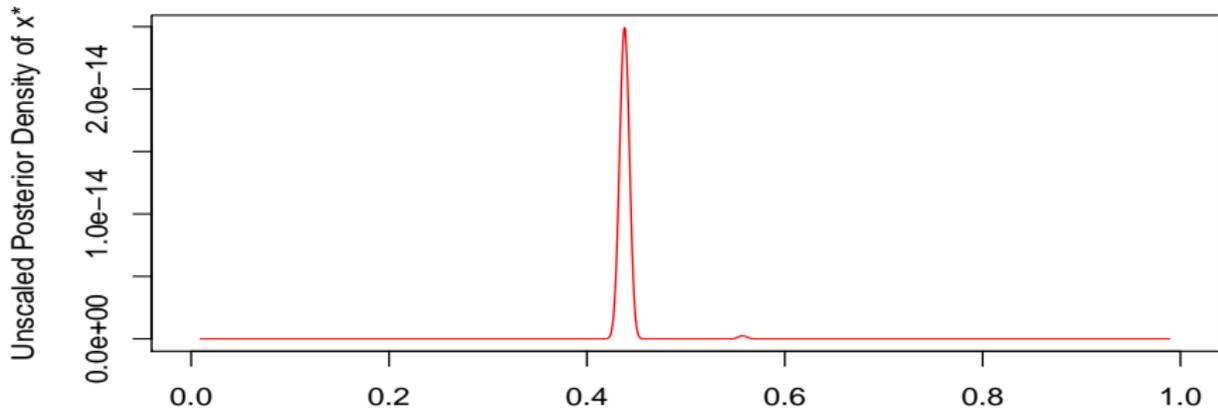
Calibration result

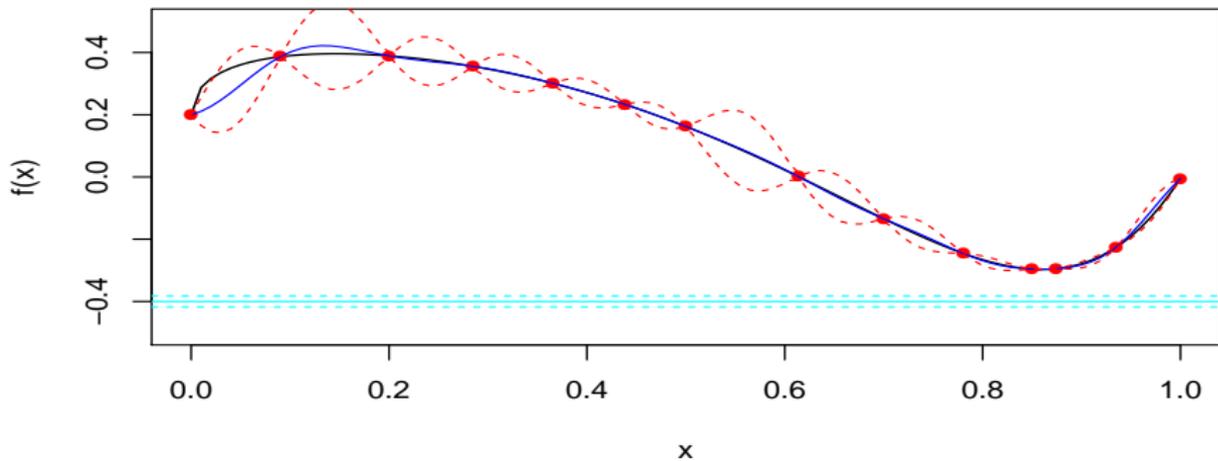




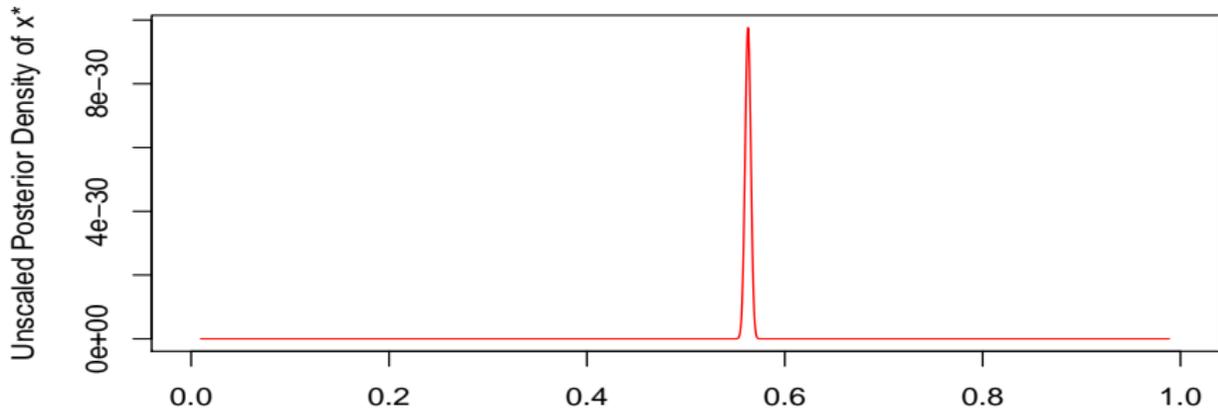


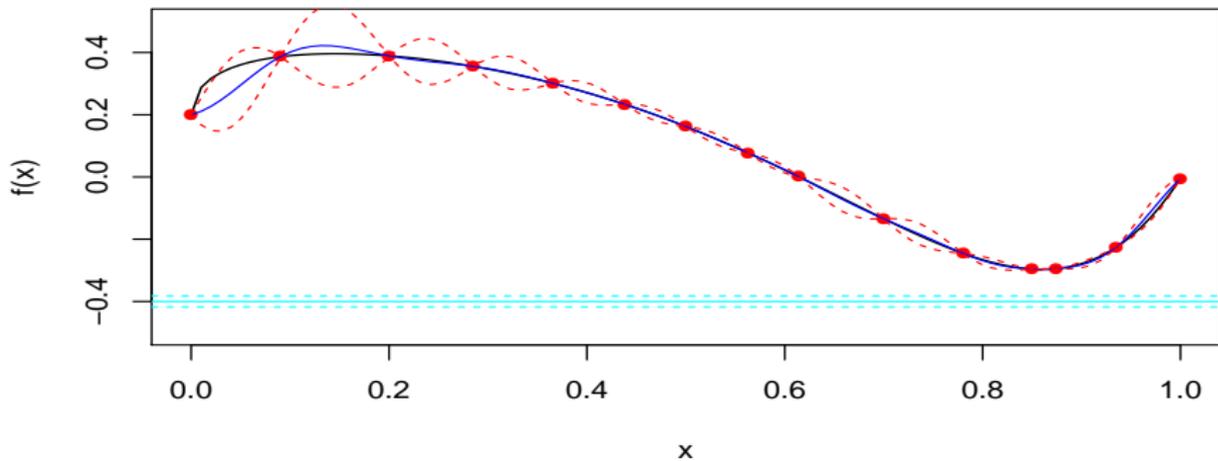
Calibration result



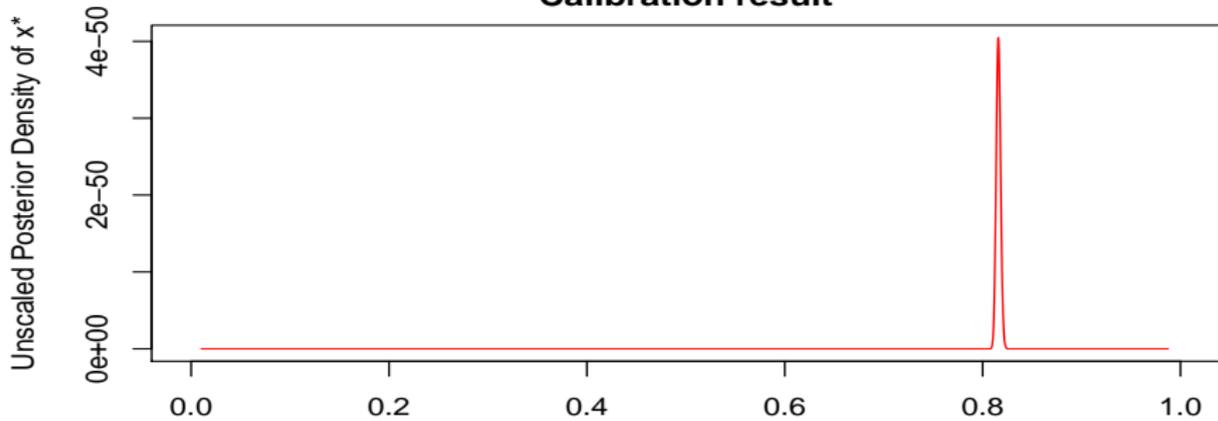


Calibration result





Calibration result

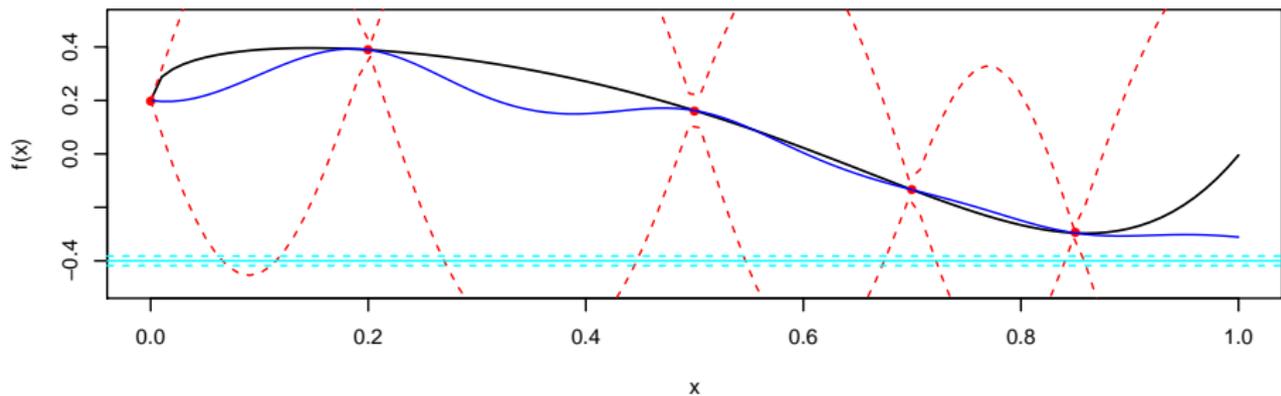


History matching/Iterative Refocussing

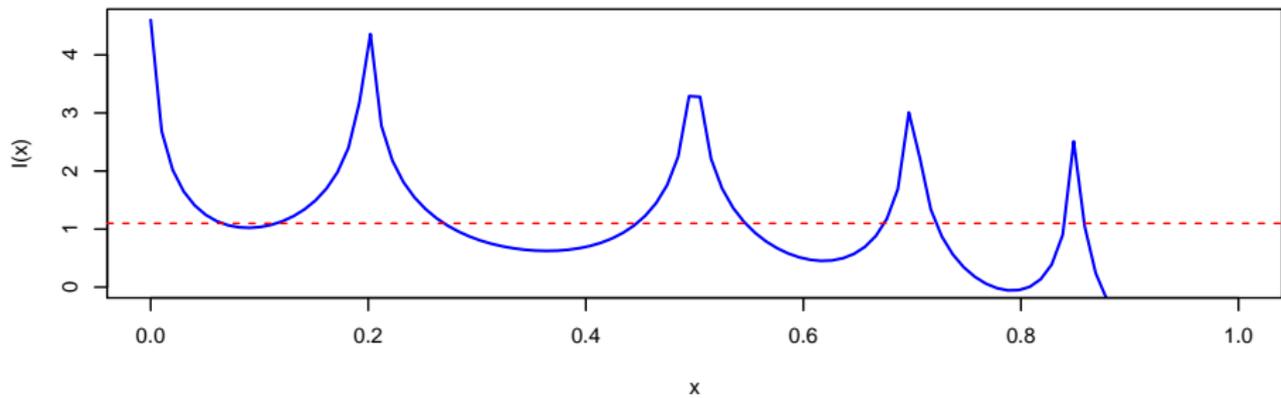
- Looking for \mathbf{x}^* using partial observations leads to **overtuning** climate models.
- The method of **history matching** looks to see if observations are able to rule any regions of parameter space out.

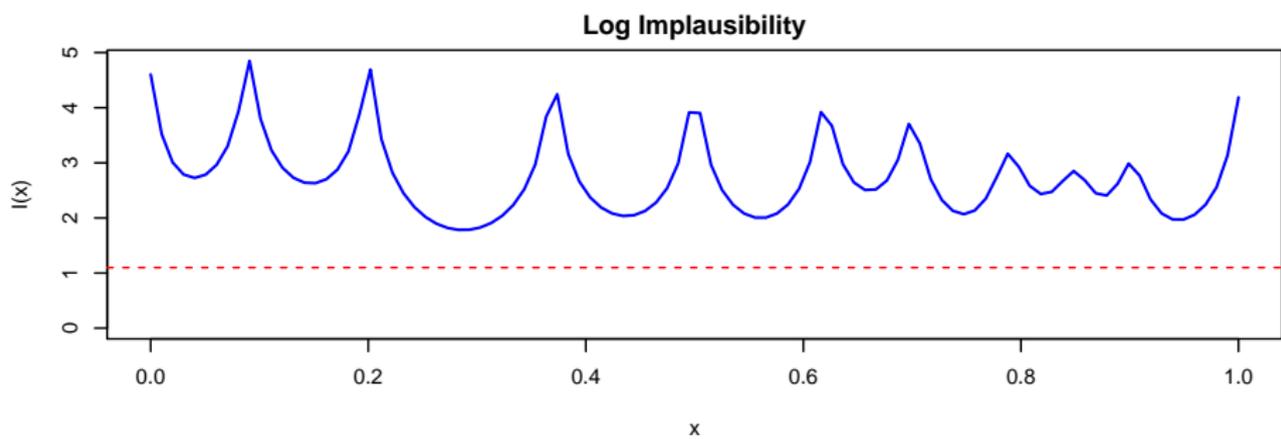
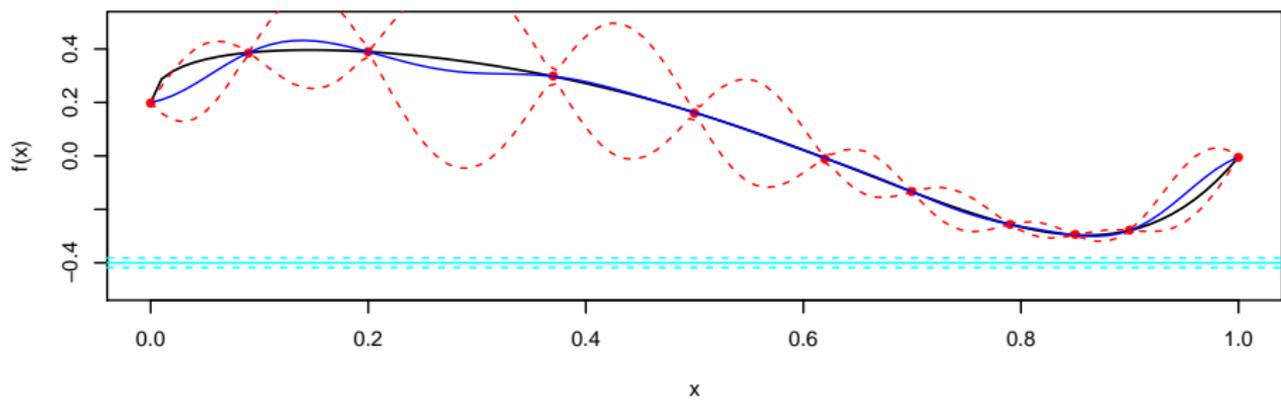
$$\|\mathbf{z} - \mathbf{f}(\mathbf{x})\|_h = (\mathbf{z} - \mathbb{E}[\mathbf{f}(\mathbf{x})])^T \text{Var}[\mathbf{z} - \mathbb{E}[\mathbf{f}(\mathbf{x})]]^{-1} (\mathbf{z} - \mathbb{E}[\mathbf{f}(\mathbf{x})])$$

- A point \mathbf{x}_0 is *implausible* if $\|\mathbf{z} - \mathbf{f}(\mathbf{x}_0)\|_h > a$ for some threshold a .
- We call the points we cannot rule out "Not Ruled Out Yet" (NROY).
- We design NROY ensembles and iteratively reduce NROY space.



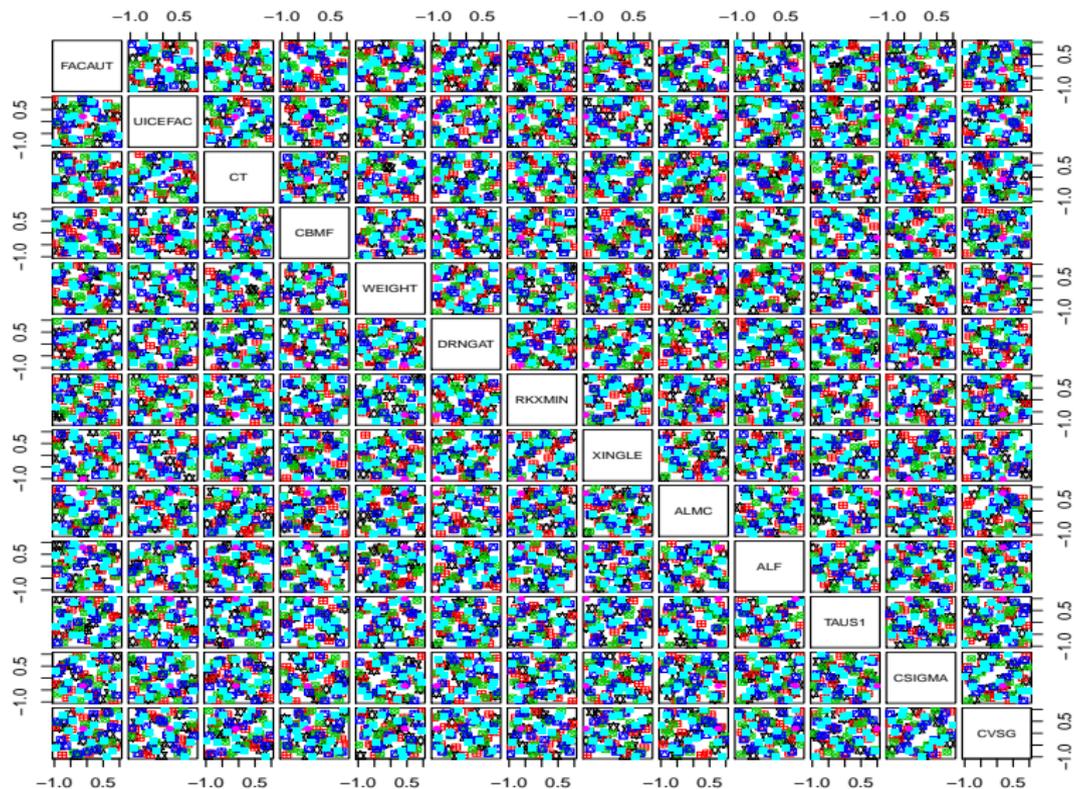
Log Implausibility



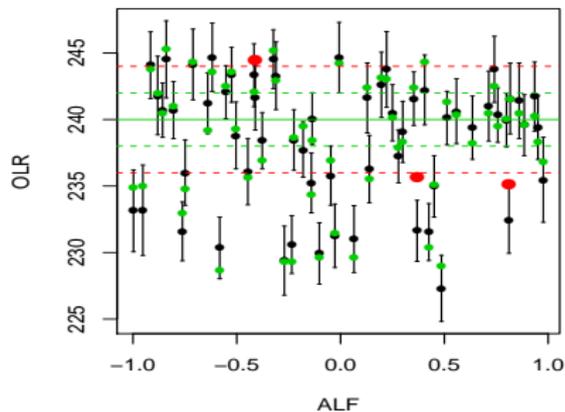
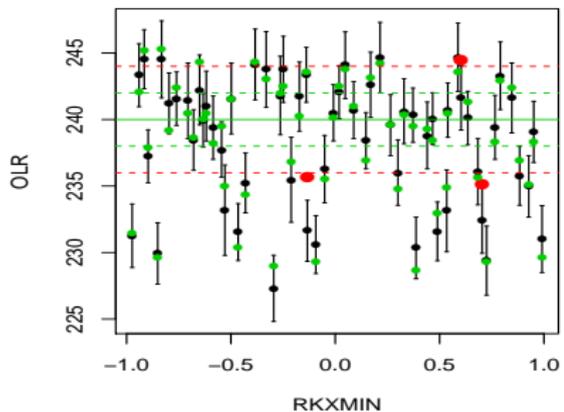
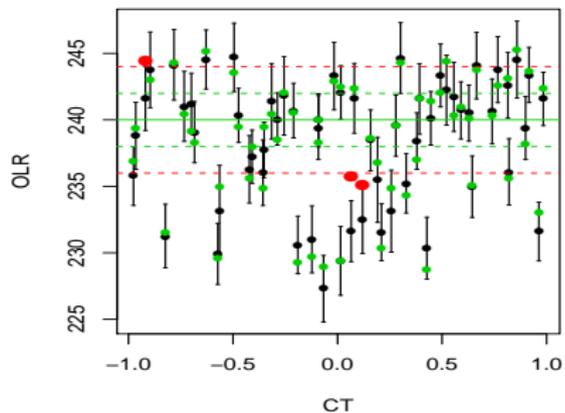
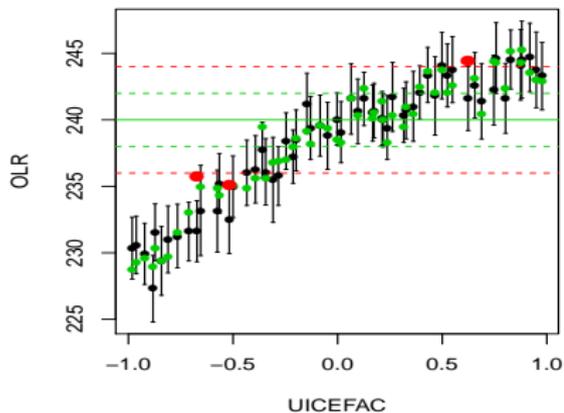


The Canadian Climate model

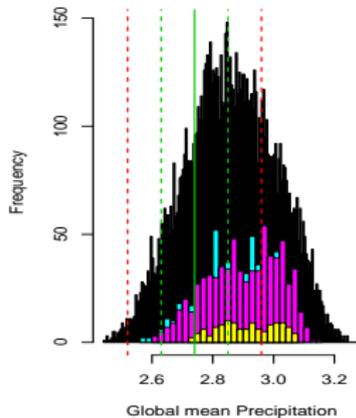
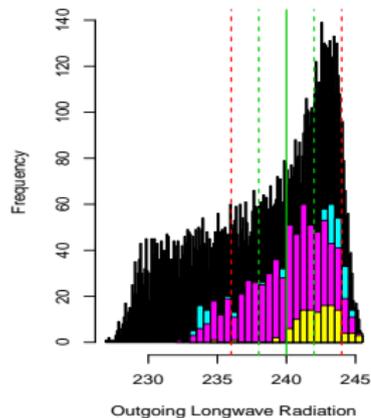
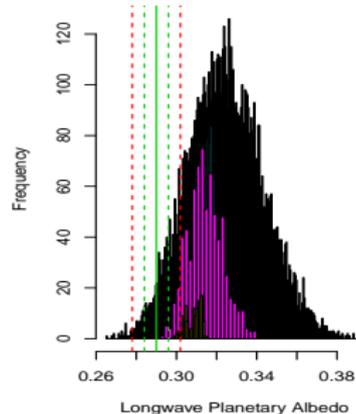
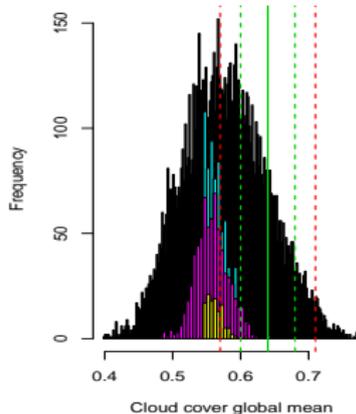
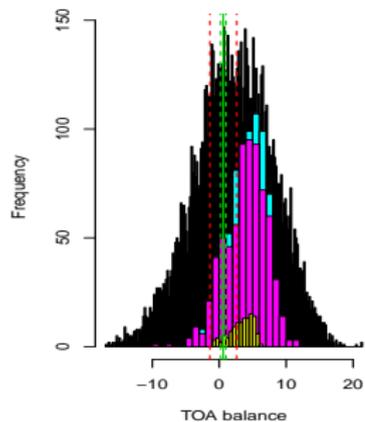
Wave 1 design for CanAGCM4



Emulating Simple Metrics (e.g. Global OLR)



NROY space



Stalled

- Tuning experience is not really with globally averaged quantities.
- Uncertainties didn't exist for these.
- We still wanted to explore the model spatially.
- “Good” parameter choices (passing our tests), did not pass the “Anomaly deck test”

Spatial UQ for climate models

Emulation of spatial fields

- The SVD/EOF basis is commonly used for emulating spatial output.
- The model runs in \mathbf{F} ($l \times n$) are projected onto the basis given by the first q right singular vectors of the ensemble.
- Think “EOFs in parameter space”.

Let

$$\mathbf{\Gamma} = (\mathbf{\Gamma}_q, \mathbf{\Gamma}_\epsilon) \quad \mathbf{\Gamma}_q = (\gamma_1, \dots, \gamma_q) \quad (1)$$

then

$$f(\mathbf{x}) = \mathbf{\Gamma}_q \mathbf{c}(\mathbf{x}) + \epsilon$$

giving q coefficients associated with each parameter choice \mathbf{x}_i :

$$\mathbf{c}(\mathbf{x}_i) = (\mathbf{\Gamma}_q^T \mathbf{\Gamma}_q)^{-1} \mathbf{\Gamma}_q^T f(\mathbf{x}_i) \quad (2)$$

- Emulators are fitted for each of the elements, $\mathbf{c}_k(\mathbf{x})$ $k = 1, \dots, q$.

The 'whack-a-mole' problem

- A principal feature of the problem is $l \gg n > q$.
- Observations \mathbf{z} may not have signal in the directions specified by $\mathbf{\Gamma}_q$.
- Projecting \mathbf{z} onto $\mathbf{\Gamma}_q$ and going back may lead to a spuriously bad looking field.

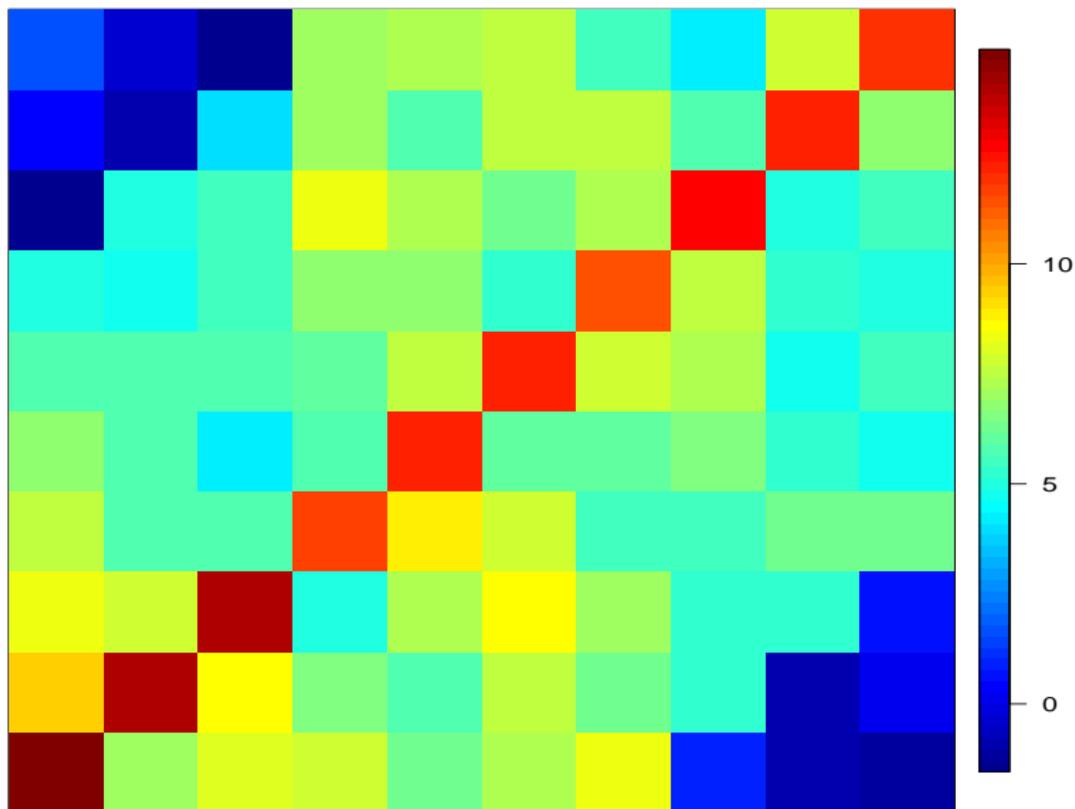
The 'whack-a-mole' problem

- A principal feature of the problem is $l \gg n > q$.
- Observations \mathbf{z} may not have signal in the directions specified by $\mathbf{\Gamma}_q$.
- Projecting \mathbf{z} onto $\mathbf{\Gamma}_q$ and going back may lead to a spuriously bad looking field.
- E.g. suppose $\exists \mathbf{x}^*$ such that $f(\mathbf{x}^*) = \mathbf{z}$.
- Then nothing guarantees $\|\mathbf{z} - \mathbf{\Gamma}_q(\mathbf{\Gamma}_q^T \mathbf{\Gamma}_q)^{-1} \mathbf{\Gamma}_q^T f(\mathbf{x}^*)\|_h < a$.
- I.e. Even a perfect emulator for the coefficients may mean $\|\mathbf{z} - \mathbf{\Gamma}_q c(\mathbf{x}^*)\|_h > a$.

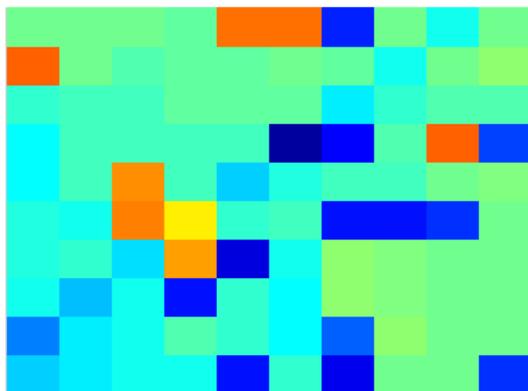
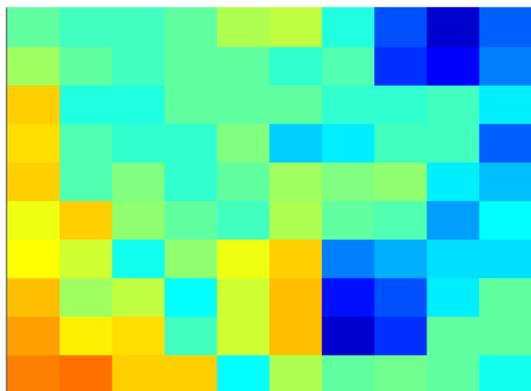
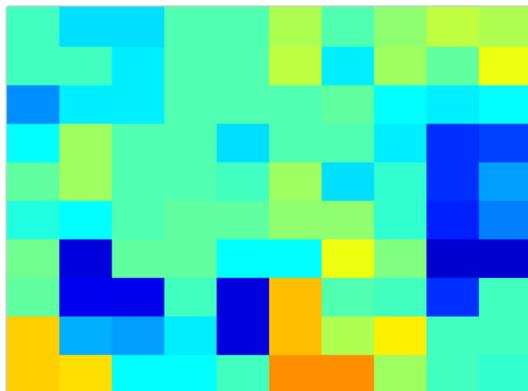
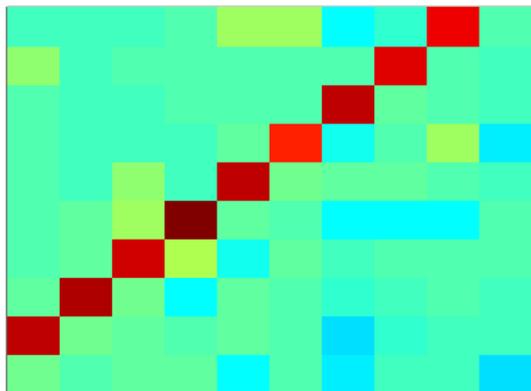
The 'whack-a-mole' problem

- A principal feature of the problem is $l \gg n > q$.
- Observations \mathbf{z} may not have signal in the directions specified by $\mathbf{\Gamma}_q$.
- Projecting \mathbf{z} onto $\mathbf{\Gamma}_q$ and going back may lead to a spuriously bad looking field.
- E.g. suppose $\exists \mathbf{x}^*$ such that $f(\mathbf{x}^*) = \mathbf{z}$.
- Then nothing guarantees $\|\mathbf{z} - \mathbf{\Gamma}_q(\mathbf{\Gamma}_q^T \mathbf{\Gamma}_q)^{-1} \mathbf{\Gamma}_q^T f(\mathbf{x}^*)\|_h < a$.
- I.e. Even a perfect emulator for the coefficients may mean $\|\mathbf{z} - \mathbf{\Gamma}_q c(\mathbf{x}^*)\|_h > a$.
- In this situation, we will never find existing models of interest.
- We will either cut out all/none of the parameter space for the wrong reasons or, if using Bayesian calibration, be stuck playing 'whack-a-mole'.

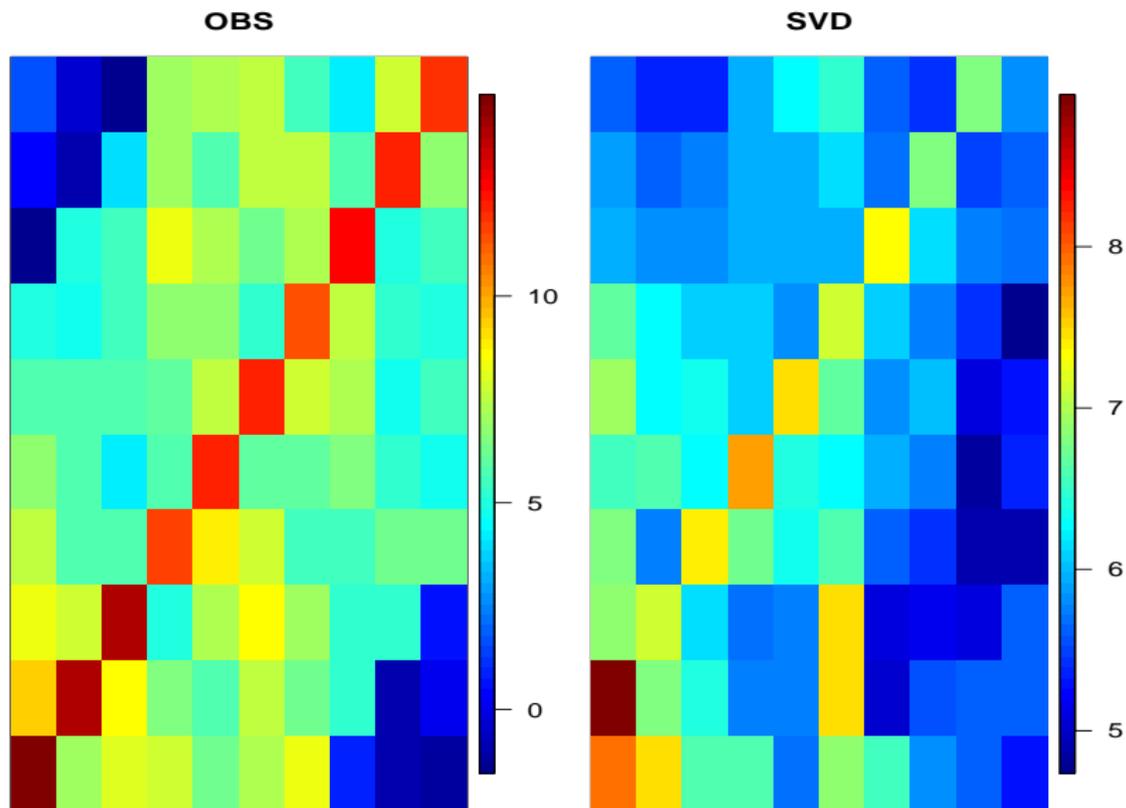
Toy Example: Obs Pattern



Toy Example: Ensemble Variability (SVD)



Toy Example: Failed Reconstruction



Ideal Alternative

- We can project \mathbf{F} onto any orthogonal basis \mathbf{B} (nothing is special about $\mathbf{\Gamma}$).
- Each basis vector, \mathbf{b}_i is a spatial pattern.
- If we chose p important spatial patterns $\mathbf{b}_1, \dots, \mathbf{b}_p$ (e.g. with low structural error) we could engineer $\|\mathbf{z} - \mathbf{B}_p(\mathbf{B}_p^T \mathbf{B}_p)^{-1} \mathbf{B}_p^T \mathbf{z}\|_h < a$.
- E.g. (cheating) $B = (\mathbf{b}_1, \mathbf{B}_\epsilon)$, $\mathbf{b}_1 = \mathbf{z}$, \mathbf{B}_ϵ formed from the right singular vectors of $(\mathbf{F} - \mathbf{b}_1(\mathbf{b}_1^T \mathbf{b}_1)^{-1} \mathbf{b}_1^T \mathbf{F})$.

Ideal Alternative

- We can project \mathbf{F} onto any orthogonal basis \mathbf{B} (nothing is special about $\mathbf{\Gamma}$).
- Each basis vector, \mathbf{b}_i is a spatial pattern.
- If we chose p important spatial patterns $\mathbf{b}_1, \dots, \mathbf{b}_p$ (e.g. with low structural error) we could engineer $\|\mathbf{z} - \mathbf{B}_p(\mathbf{B}_p^T \mathbf{B}_p)^{-1} \mathbf{B}_p^T \mathbf{z}\|_h < a$.
- E.g. (cheating) $B = (\mathbf{b}_1, \mathbf{B}_\epsilon)$, $\mathbf{b}_1 = \mathbf{z}$, \mathbf{B}_ϵ formed from the right singular vectors of $(\mathbf{F} - \mathbf{b}_1(\mathbf{b}_1^T \mathbf{b}_1)^{-1} \mathbf{b}_1^T \mathbf{F})$.
- This is a challenging(!!!) elicitation problem and 3 years of discussion with collaborators have gotten nowhere.

Rotating the SVD basis

- We can rotate Γ by multiplying it on the right by $n \times n$ rotation matrix Λ .
- A rotation matrix is formed by selecting $\binom{n}{2}$ angles θ that define the rotation by angle θ about each pair of basis vectors.
- The goal is to structure Λ so that

$$\|\mathbf{z} - \Gamma\Lambda_q(\Lambda_q^T\Gamma^T\Gamma\Lambda_q)^{-1}\Lambda_q^T\Gamma^T\mathbf{z}\|_h < \xi_{tol}$$

Rotating the SVD basis

- We can rotate Γ by multiplying it on the right by $n \times n$ rotation matrix Λ .
- A rotation matrix is formed by selecting $\binom{n}{2}$ angles θ that define the rotation by angle θ about each pair of basis vectors.
- The goal is to structure Λ so that

$$\|\mathbf{z} - \Gamma\Lambda_q(\Lambda_q^T\Gamma^T\Gamma\Lambda_q)^{-1}\Lambda_q^T\Gamma^T\mathbf{z}\|_h < \xi_{tol}$$

- Blindly choosing Λ to minimise the above is highly likely to lead to having q coefficients with no signal we can emulate.

Rotating the SVD basis

- Our method is to choose Λ to minimise

$$\|\mathbf{z} - \Gamma \Lambda_q (\Lambda_q^T \Gamma^T \Gamma \Lambda_q)^{-1} \Lambda_q^T \Gamma^T \mathbf{z}\|_h$$

subject to projection onto each of the first q basis vectors:

$(\Gamma \Lambda)_{.k}$, $k = 1, \dots, q$ explaining a given percentage of the ensemble variability in \mathbf{F} .

Rotating the SVD basis

- Our method is to choose Λ to minimise

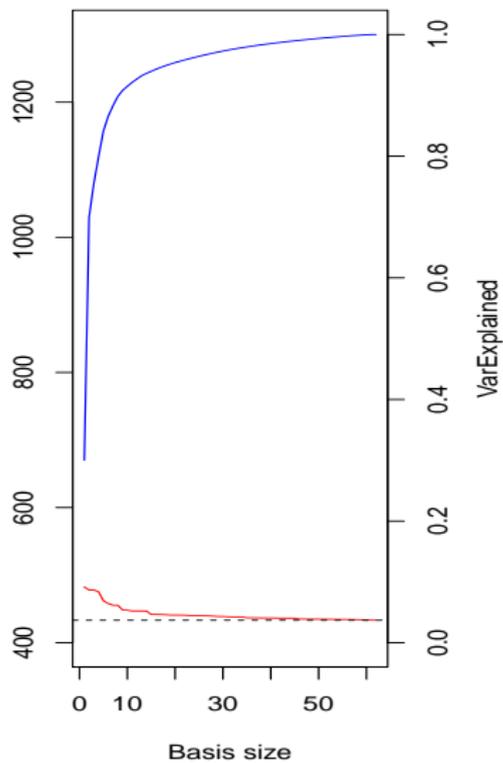
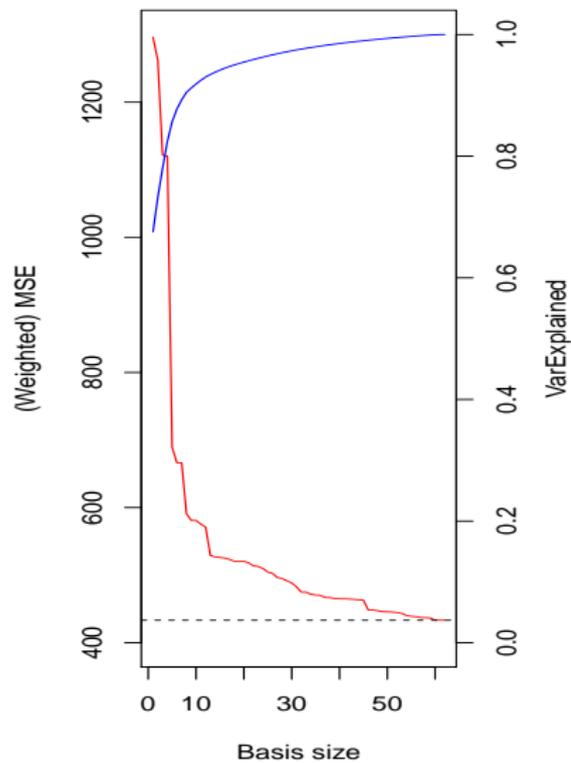
$$\| \mathbf{z} - \Gamma \Lambda_q (\Lambda_q^T \Gamma^T \Gamma \Lambda_q)^{-1} \Lambda_q^T \Gamma^T \mathbf{z} \|_h$$

subject to projection onto each of the first q basis vectors:

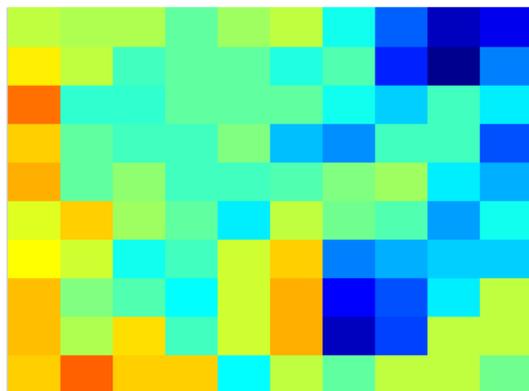
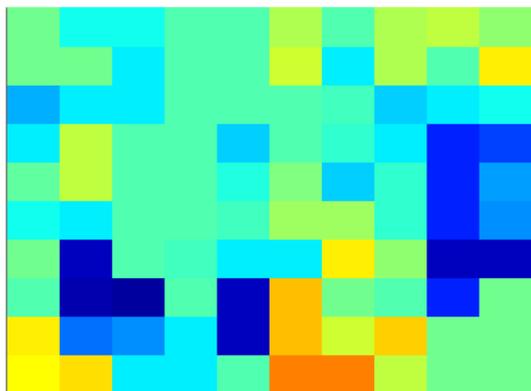
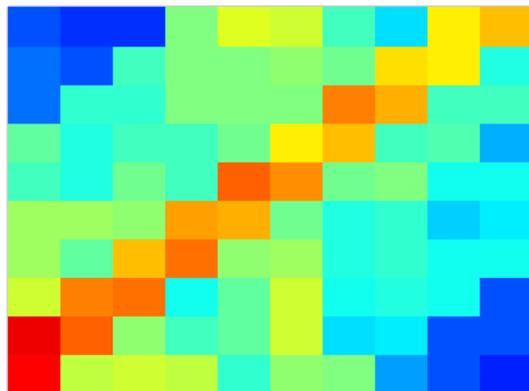
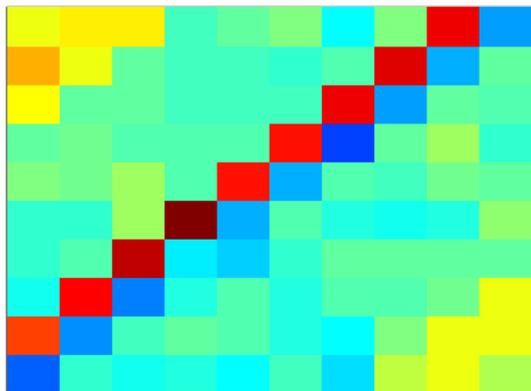
$(\Gamma \Lambda)_{.k}$, $k = 1, \dots, q$ explaining a given percentage of the ensemble variability in \mathbf{F} .

- To do this, we exploit the fact that each rotated basis vector is a re-weighting of the original basis vectors, with the columns of Λ , the weights.
- We can therefore obtain each rotated basis vector sequentially, optimising just n parameters each time rather than $\binom{n}{2}$.
- We usually do this 1 – 2 times until we are close enough to the minimum, we fill the basis by taking the first few right singular vectors of the residual ensemble following projection onto the newly chosen rotated basis vectors.

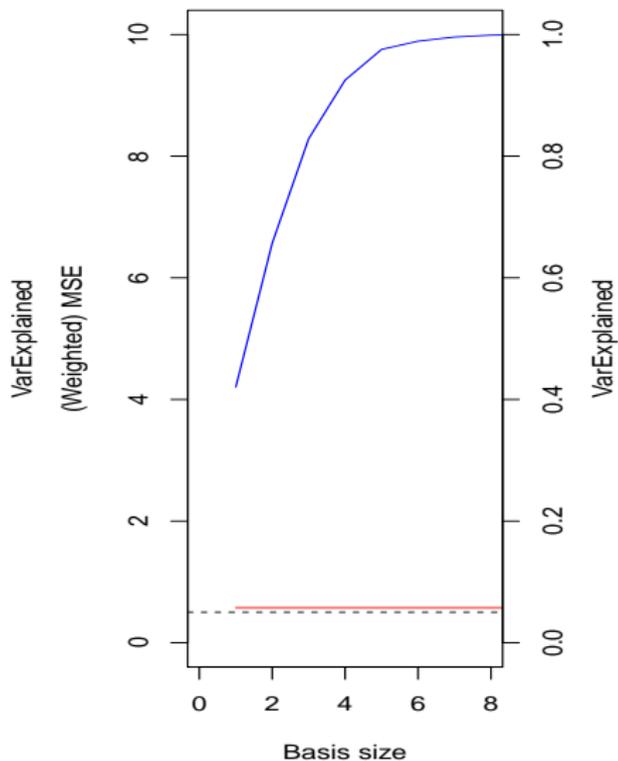
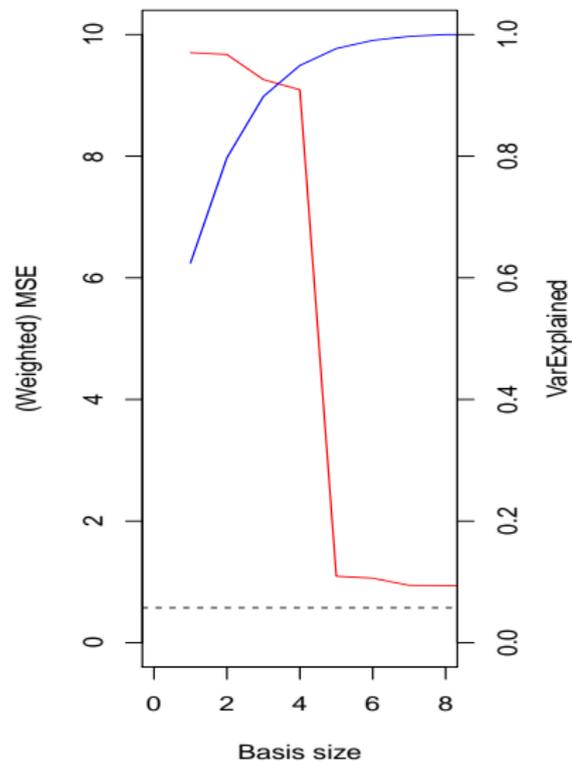
Optimal Rotation



Toy Example: Optimal Rotation



Toy Example: Optimal Rotation



Toy Example: 'Good' model reconstruction

