Efficient high-dimensional emulation and calibration

James M Salter*

University of Exeter

April 2021

With thanks to:
Daniel Williamson, Peter Challenor, Tim Dodwell (Exeter)
Lauren Gregoire (University of Leeds)

*j.m.salter@exeter.ac.uk, @jm_salter

Motivation

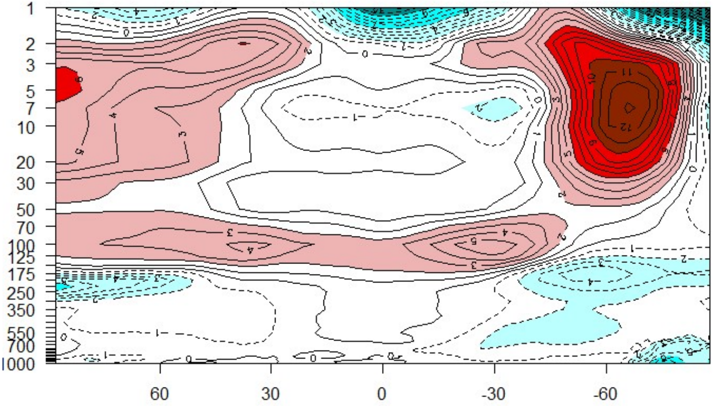
Here, going to deal with emulation/calibration problems where:

1. Output dimension ℓ high
2. True model $f(\mathbf{x})$ expensive, so number of model runs n low ($n << \ell$)
3. Don't see 'good enough' match between real world and model

Aim: use emulator to explore input space, find output as consistent with the real world as possible.

Example - CanAM4 air temperature (TA)

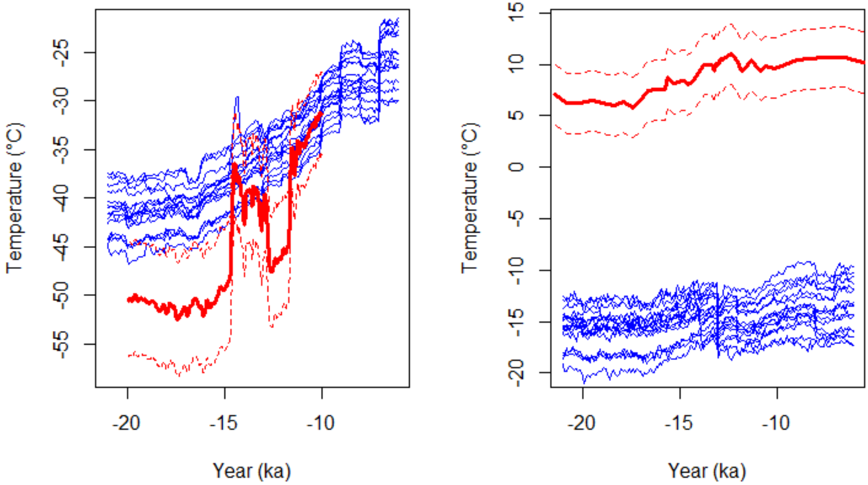
$\ell = 2368, n \sim 60$



Example - FAMOUS/HadCM3

Observations, ensemble

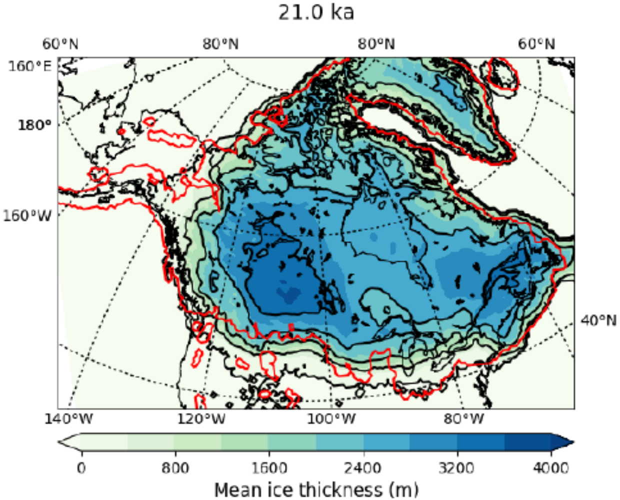
15,000 years, $n = 16$



Example - Glimmer ice sheet model

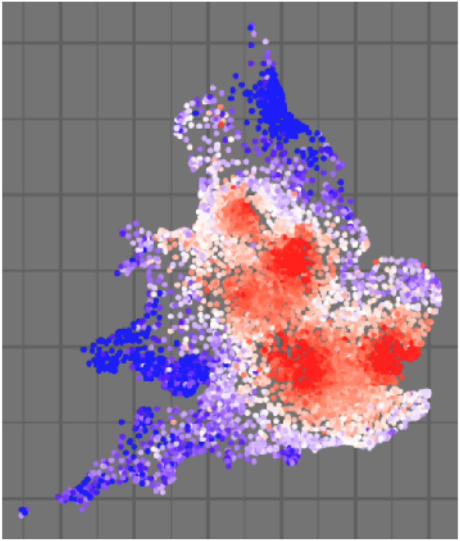
Observations, ensemble

$\ell = 29,100, \; n \sim 500$



Example - Covid modelling

$\ell \sim 8000$



Motivation

Have several options with high dimensional output:

- Emulate summary / several summaries

Motivation

Have several options with high dimensional output:

- Emulate summary / several summaries
- Emulate all outputs

Motivation

Have several options with high dimensional output:

- Emulate summary / several summaries
- Emulate all outputs
- Emulate projection onto low-dimensional basis

Outline

1. Emulation, calibration
2. Dimension reduction
3. Pros, cons, examples
4. Summary

Gaussian process emulation

- Have design $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n)$ in parameter space $\mathcal{X} \subset \mathbb{R}^p$.
- Run model $f(\cdot)$ at $\mathbf{X}$, gives $\ell \times n$ matrix of model output:

$$\mathbf{F} = (f(\mathbf{x}_1), \dots, f(\mathbf{x}_n))$$

- Fit Gaussian process to output:

$$f_i(\mathbf{x})|\mathbf{F}, \boldsymbol{\beta}, \phi \sim \mathrm{GP}(m_i(\mathbf{x}), K_i(\mathbf{x}, \mathbf{x})), \quad i = 1, \dots, \ell,$$

  for mean function $m_i(\mathbf{x})$, covariance function $K_i(\mathbf{x}, \mathbf{x})$.
- Possibly includes nugget so doesn't interpolate data exactly.
- At any point $\mathbf{x}'$, can evaluate expectation, variance.

Calibration

$$\mathbf{z} = f(\mathbf{x}^*) + \mathbf{e} + \boldsymbol{\eta}$$

- $f(\cdot)$ - computer model representing real-world system, e.g. climate, ice sheet evolution, heartbeats, spread of infectious diseases, ...
- $\mathbf{z}$ - observations of the real-world system
- $\mathbf{e}$ - observation error (imperfect observations)
- $\boldsymbol{\eta}$ - model discrepancy/inadequacy
- $\mathbf{x}^*$ - the 'best' setting of the input parameters, $\mathbf{x} \in \mathcal{X} \subset \mathbb{R}^p$
- Generally replace $f(\cdot)$ with our GP emulator for speed.

Calibration

$$\mathbf{z} = f(\mathbf{x}^*) + \mathbf{e} + \boldsymbol{\eta}$$

- $f(\cdot)$ - computer model representing real-world system, e.g. climate, ice sheet evolution, heartbeats, spread of infectious diseases, ...
- $\mathbf{z}$ - observations of the real-world system
- $\mathbf{e}$ - observation error (imperfect observations)
- $\boldsymbol{\eta}$ - model discrepancy/inadequacy
- $\mathbf{x}^*$ - the 'best' setting of the input parameters, $\mathbf{x} \in \mathcal{X} \subset \mathbb{R}^p$
- Generally replace $f(\cdot)$ with our GP emulator for speed.
- Identifiability issue between $\mathbf{x}^*$ and $\boldsymbol{\eta}$: when see a difference between $\mathbf{z}$, $\mathbf{F}$, don't know whether problem is due to choice of $\mathbf{x}$, or discrepancy - need to explore $\mathcal{X}$

Calibration/history matching

- **Bayesian calibration** (Kennedy & O'Hagan 2001): constructs posterior distribution for best input, $\mathbf{x}^*$: $\pi(\mathbf{x}^*|\mathbf{F}, \mathbf{z})$.
- **History matching** (Craig et al. 1996): rules out regions of parameter space that are not consistent with observations using implausibility:

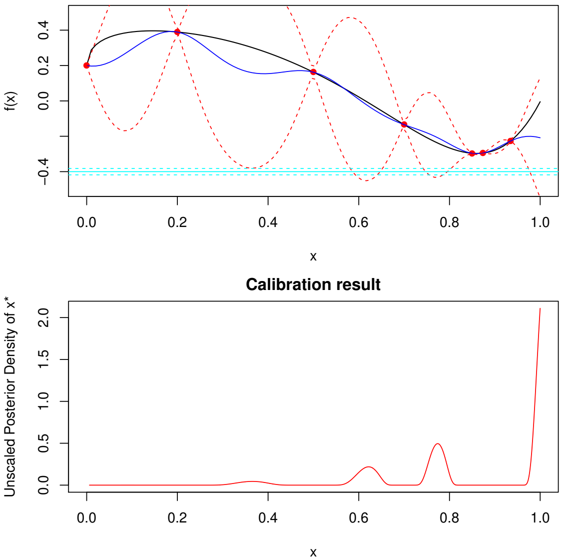
$$\mathcal{I}(\mathbf{x}) = (\mathbf{z} - \mathrm{E}[f(\mathbf{x})])^T (\mathrm{Var}[f(\mathbf{x})] + \mathbf{\Sigma_e} + \mathbf{\Sigma}_{\boldsymbol{\eta}})^{-1} (\mathbf{z} - \mathrm{E}[f(\mathbf{x})])$$

- Points that are not ruled out are said to be in '**Not Ruled Out Yet**' (**NROY**) space, the space of not implausible points:

$$\mathcal{X}_{NROY} = \{\mathbf{x} \in \mathcal{X} | \mathcal{I}(\mathbf{x}) < T\},$$

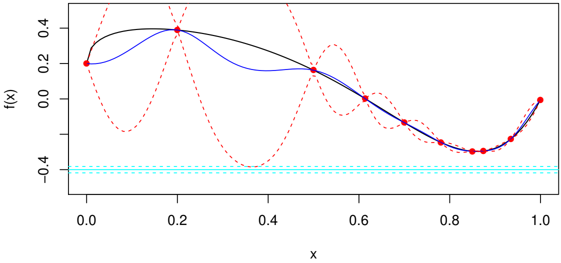
  where T often 3^2 in 1D, $\chi^2_{\ell,0.995}$ in ℓD.
- Can perform multiple waves, iteratively refocussing in the current NROY space.

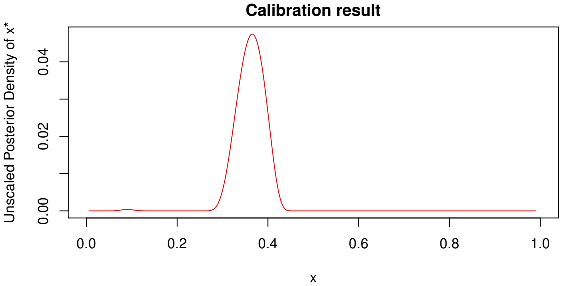
Example*



*'whackamole', h/t Danny

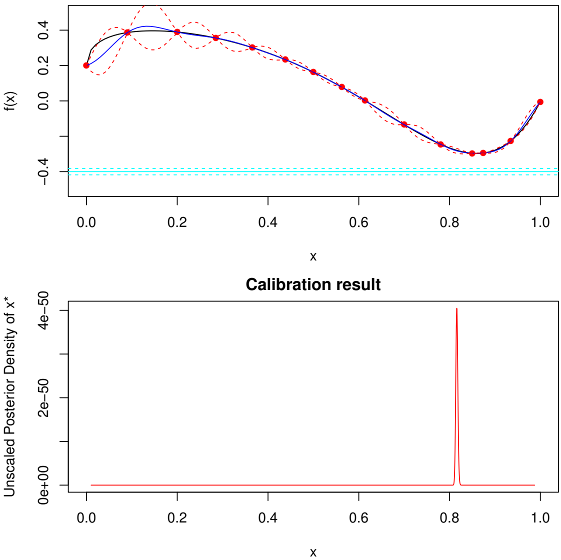
Example*





*'whackamole', h/t Danny

Example*



* 'whackamole', h/t Danny

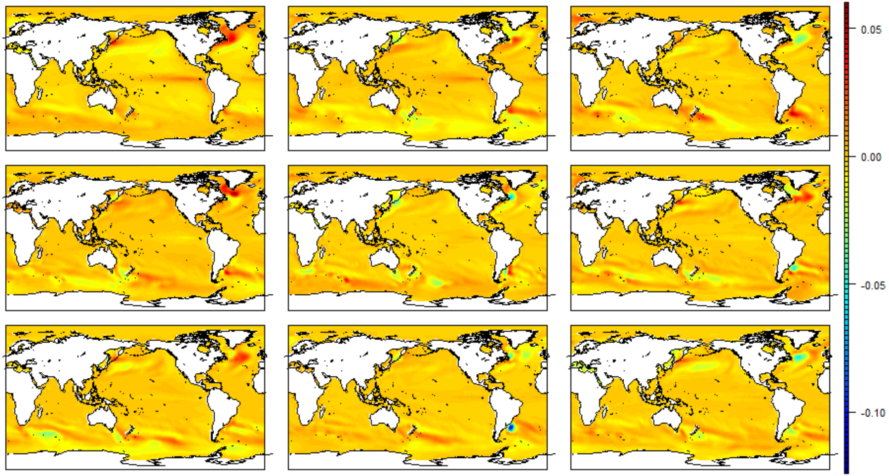
Outline

1. Emulation, calibration
2. Dimension reduction
3. Pros, cons, examples
4. Summary

Basis approach for large ℓ

- Calculate SVD/PCA/EOF across the (centred) ensemble $\mathbf{F}_{\boldsymbol{\mu}} \longmapsto \mathbf{\Gamma}$.

Basis approach for large ℓ



Basis approach for large ℓ

- Calculate SVD/PCA/EOF across the (centred) ensemble $\mathbf{F}_{\boldsymbol{\mu}} \longmapsto \mathbf{\Gamma}$.
- Project output fields onto the leading q directions of $\mathbf{\Gamma}$.
  At $\mathbf{x}$: ℓ outputs $\rightarrow q$ coefficients via projection*:

$$\mathbf{c}(\mathbf{x}) = \mathbf{P_W}(f(\mathbf{x}) - \boldsymbol{\mu}), \quad \mathbf{P_W} = (\mathbf{\Gamma}_q^T \mathbf{W}^{-1} \mathbf{\Gamma}_q)^{-1} \mathbf{\Gamma}_q^T \mathbf{W}^{-1}.$$

- Build Gaussian process emulators for each set of coefficients :

$$c_i(\mathbf{x}) \sim \mathrm{GP}(m_i(\mathbf{x}), K_i(\mathbf{x}, \mathbf{x})), \quad i = 1, \ldots, q.$$

- Can then map back to original field:

$$\mathrm{E}[f(\mathbf{x})] = \mathbf{\Gamma}_q \mathrm{E}[\mathbf{c}(\mathbf{x})], \quad \mathrm{Var}[f(\mathbf{x})] = \mathbf{\Gamma}_q \mathrm{Var}[\mathbf{c}(\mathbf{x})] \mathbf{\Gamma}_q^T.$$

*wrt some positive definite $\mathbf{W}$

Choosing basis

Basis choice needs to satisfy a few rules:

- $q << \ell$
- Possible to build emulators - so need some element of explaining variability in data
- Ability to represent $\mathbf{z}$ in the subspace (e.g. check $\mathcal{R}_{\mathbf{W}}(\mathbf{\Gamma}_q, \mathbf{z})$)

Choosing basis

Basis choice needs to satisfy a few rules:

- $q << \ell$
- Possible to build emulators - so need some element of explaining variability in data
- Ability to represent $\mathbf{z}$ in the subspace (e.g. check $\mathcal{R}_{\mathbf{W}}(\mathbf{\Gamma}_q, \mathbf{z})$)

Reconstruction error:

For basis $\mathbf{B}$, positive definite matrix $\mathbf{W}$, define[†]:

$$\mathcal{R}_{\mathbf{W}}(\mathbf{B}, \mathbf{z}) = \|\mathbf{z} - \mathbf{r}(\mathbf{z})\|_{\mathbf{W}} = (\mathbf{z} - \mathbf{r}(\mathbf{B}, \mathbf{z}))^T \mathbf{W}^{-1} (\mathbf{z} - \mathbf{r}(\mathbf{B}, \mathbf{z})),$$
$$\mathbf{r}(\mathbf{B}, \mathbf{z}) = \mathbf{B}(\mathbf{B}^T \mathbf{W}^{-1} \mathbf{B})^{-1} \mathbf{B}^T \mathbf{W}^{-1} \mathbf{z}.$$

Note that in perfect emulation case (zero emulator variance) and with $\mathbf{W} = \mathbf{\Sigma}_{\mathbf{e}} + \mathbf{\Sigma}_{\eta}$, this is the same distance metric as in $\mathcal{I}(\mathbf{x})$.

[†] JMS, Williamson et al. 2019

Choosing basis

Basis choice needs to satisfy a few rules:

- $q << \ell$
- Possible to build emulators - so need some element of explaining variability in data
- Ability to represent $\mathbf{z}$ in the subspace (e.g. check $\mathcal{R}_{\mathbf{W}}(\mathbf{\Gamma}_q, \mathbf{z})$)

The terminal case:

$$\mathcal{R}_{\mathbf{W}}(\mathbf{\Gamma}_q, \mathbf{z}) > T.$$

Natural consequence of this: choice of basis may guarantee that we rule out $f(\mathbf{x}) = \mathbf{z}$, even if we had a perfect emulator: the '**terminal case**'.

Choosing basis

SVD is variance-maximising, doesn't care about $\mathbf{z}$. May not be a good choice for calibration...

Choose basis representation that allows to properly explore $\mathcal{X}$, find whether can tune inputs, avoid assigning everything to discrepancy

Can often fix by rearranging information in $\mathbf{F}$ (rotation) such that important, low eigenvalue directions are 'blended' with variance maximising directions†

† see JMS et al. 2019, JASA for how to choose $\mathbf{\Gamma}_q$, proofs, tricks for finding...

Fast history matching in ℓ dimensions

Want:

$$\mathcal{I}(\mathbf{x}) = (\mathbf{z} - \mathsf{E}[f(\mathbf{x})])^T (\mathsf{Var}[f(\mathbf{x})] + \boldsymbol{\Sigma}_{\mathbf{e}} + \boldsymbol{\Sigma}_{\eta})^{-1} (\mathbf{z} - \mathsf{E}[f(\mathbf{x})]).$$

☹ Expensive, ℓ-dimensional matrix inversion varies with $\mathbf{x}$.

Fast history matching in ℓ dimensions

Thanks to emulator basis structure, we can rewrite this as[‡]:

$$\hat{\mathcal{I}}(\mathbf{x}) = (\mathbf{P}_{\mathbf{W}}\mathbf{z} - \mathrm{E}[\mathbf{c}(\mathbf{x})])^T(\mathrm{Var}[\mathbf{c}(\mathbf{x})] + \mathbf{P}_{\mathbf{W}}(\boldsymbol{\Sigma}_{\mathbf{e}} + \boldsymbol{\Sigma}_{\eta}))^{-1}(\mathbf{P}_{\mathbf{W}}\mathbf{z} - \mathrm{E}[\mathbf{c}(\mathbf{x})]),$$
$$\mathbf{W} = \boldsymbol{\Sigma}_{\mathbf{e}} + \boldsymbol{\Sigma}_{\eta},$$
$$\mathcal{I}(\mathbf{x}) = \mathcal{R}_{\mathbf{W}}(\boldsymbol{\Gamma}_q, \mathbf{z}) + \hat{\mathcal{I}}(\mathbf{x}).$$

🙂 Fast, one-off inversion of **W**, then everything works in q dimensions, no loss of information.

Hence NROY space becomes:

$$\mathcal{X}_{NROY} = \{\mathbf{x} \in \mathcal{X} | \hat{\mathcal{I}}(\mathbf{x}) < T - \mathcal{R}_{\mathbf{W}}(\boldsymbol{\Gamma}_q, \mathbf{z})\}.$$

‡ JMS and Williamson 2020, arXiV

Outline

1. Emulation, calibration
2. Dimension reduction
3. Pros, cons, examples
4. Summary

Why use a basis method?

With increasing computer power, greater parallelisation, can always fit GPs to each output individually:

$$f_i(\mathbf{x})|\mathbf{F} \sim \mathrm{GP}(m_i(\mathbf{x}), K_i(\mathbf{x}, \mathbf{x})), \quad i = 1, \ldots, \ell.$$

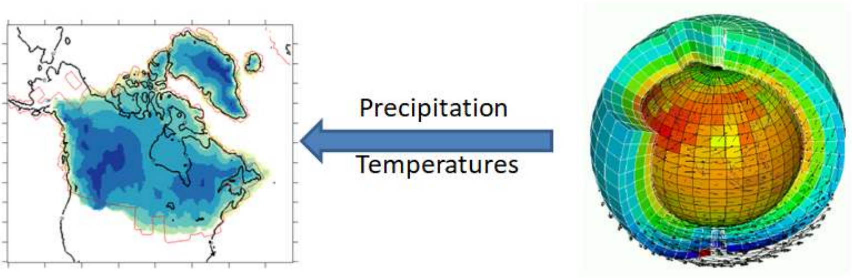
What benefits do we get if we ***don't*** do this? (and equally, what bonus do we get above emulating summaries?)

1. Interpretation/prior assessment
2. Efficiency (emulation, validation, calibration)
3. Coherence

Simulating ice evolution

Ice sheet model, Glimmer-CISM

General Circulation Models



7 input parameters (e.g. heat flux, basal sliding, lapse rate)

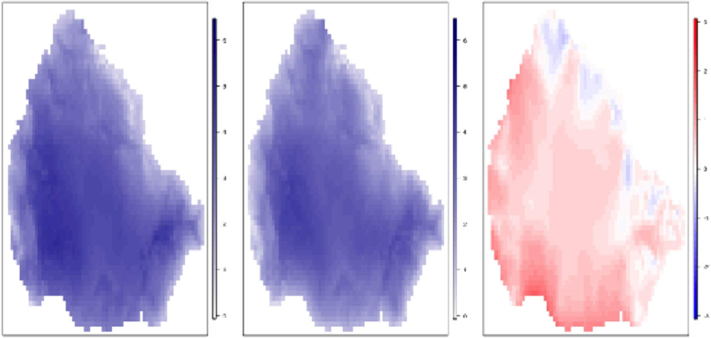
Spatio-temporal fields required for temperature and precipitation (from low-resolution GCM, e.g. FAMOUS, 48×37 spatial field for 15,000 years)

$\ell = 194 \times 150 = 29,100$ field output every 100 years

Example - ice sheet reconstructions

North American ice sheet at 21 ka.
Restricted to $\ell = 8922$, $n = 100$.

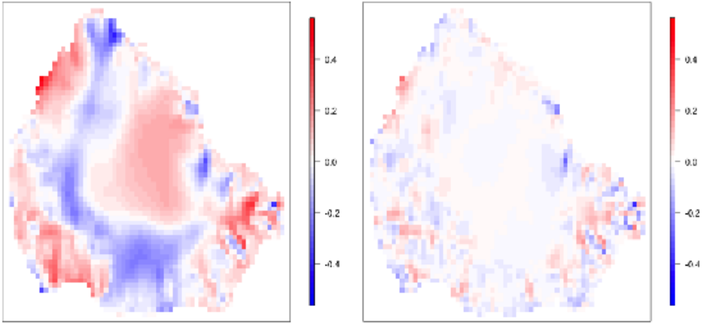
Proxy observations | Training data mean | Difference



Example - ice sheet reconstructions

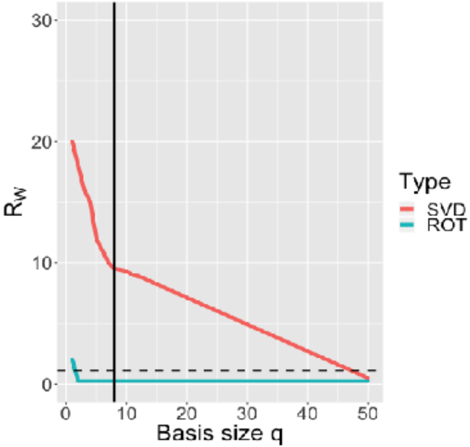
Difference between proxy observations and their basis reconstruction

Truncated SVD　　　　Truncated ROT



(so able to attribute more to model rather than discrepancy, hopefully search in better directions with emulator)

Example - ice sheet reconstructions



- - - History matching cutoff, T

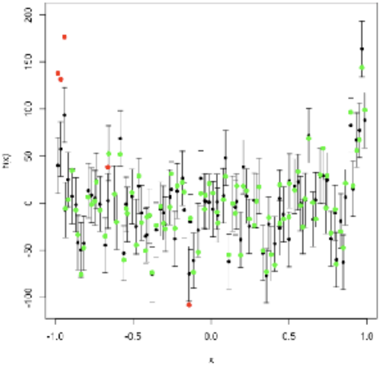
| Basis truncation

Why use a basis method: **Emulator efficiency**

- Higher initial cost (finding basis, inverting $\mathbf{W}$), but one-off
- Thousands of emulators (ℓ) vs a few (q)
- Cost of basis approach doesn't increase much as evaluate emulators at more $\mathbf{x} \in \mathcal{X}$ (and will generally do millions when HM) - useful when have multiple waves and/or multiple spatial field outputs and/or small, hard-to-sample-from NROY space

Why use a basis method: **Validation efficiency**

Might look good overall (e.g. LOOs, prediction over validation set with 5% outside 2SDs), so passes checks.
Could go wrong in particular parts of space, demonstrate non-stationarity.
Easier to rigorously validate, re-fit handful of emulators.
Fewer failure points.



Why use a basis method: **Coherency**

Why might it be useful to be able to sample from the emulator posterior in a physical coherent way?

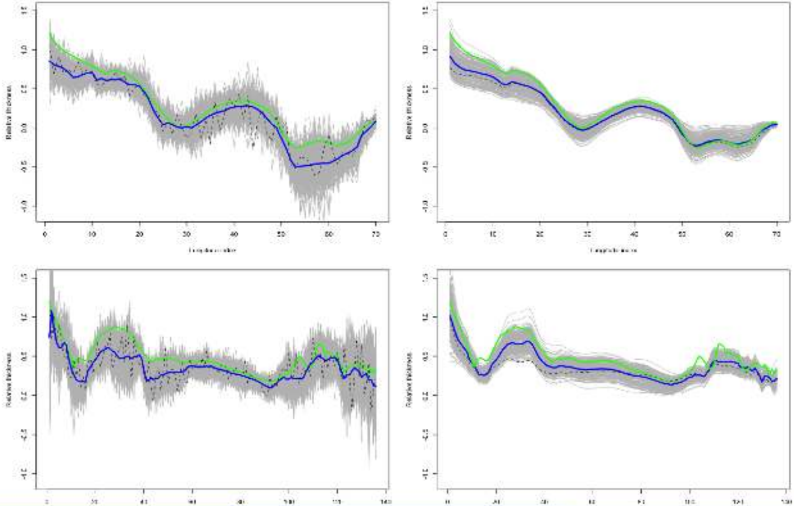
- If using as input (e.g. boundary condition) to another model.
- For exceedance probabilities (several co-located grid cells contributing to a risk of e.g. inundation, important to account for correlated outputs).

Why use a basis method: **Coherency**

Sampled ice sheet profile at fixed latitude

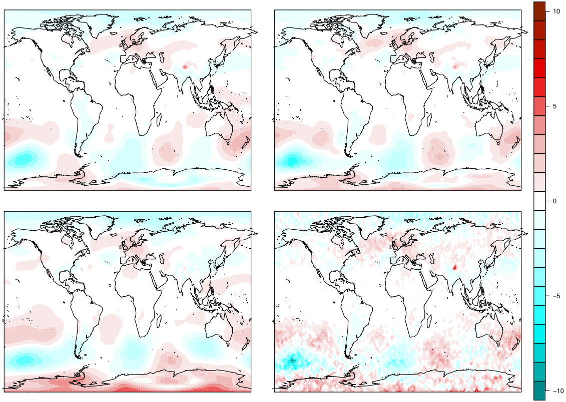
Univariate Basis

Observations Emulator mean



Why use a basis method: **Coherency**

Expectations look similar, but individual samples look like model output for basis approach - not true for independently-sampled GPs unless posterior variance $\to 0$.

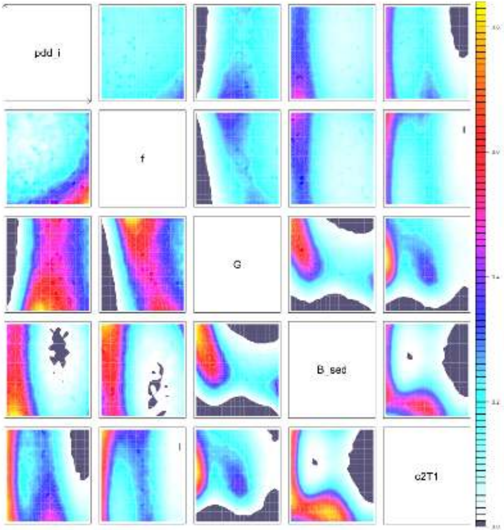


Expectation

Sample

Why use a basis method: **Calibration efficiency**

Extremely fast to run for millions of points



Why use a basis method: **Summary**

Good:

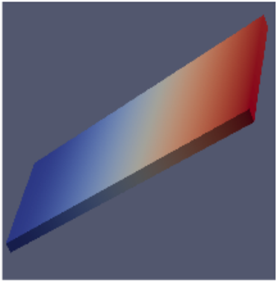
- Fewer emulators to build, validate (fewer failure points, easier diagnosis of problems)
- Hence much faster predictions, particularly useful when multiple waves
- Similar accuracy in experience
- Basis structure gives fast calibration, no loss of information
- Captures patterns/correlations/physicality from model, interpretability
- Emulator posterior behaves like model output
- Better exploitation of information in ensemble vs emulating summaries - if expensive to run, want to extract as much signal as possible.

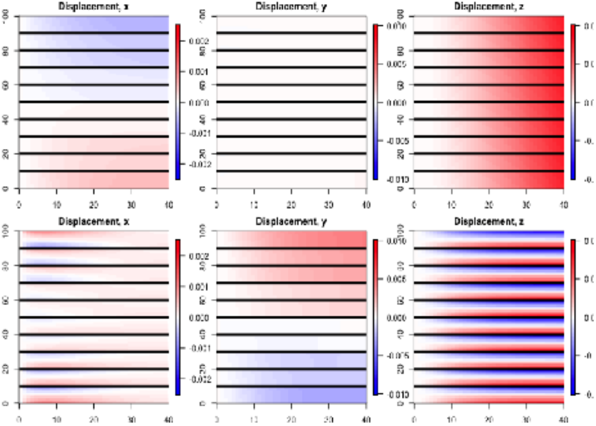
Why use a basis method: **Summary**

Bad:

- Initial expense of inverting **W**, calculating SVD increases with ℓ
- Observations may lie outside subspace defined by ensemble (but either fixable, or a problem in all cases)
- Basis selection/emulation can be challenging
- Maybe not enough degrees of freedom?
- Patterns don't necessarily align ($\Longrightarrow$ kernel PCA)
- Harder to specify $\ell \times \ell$ variance matrices for observation error, discrepancy vs if these are scalars.

Composite/PDE model

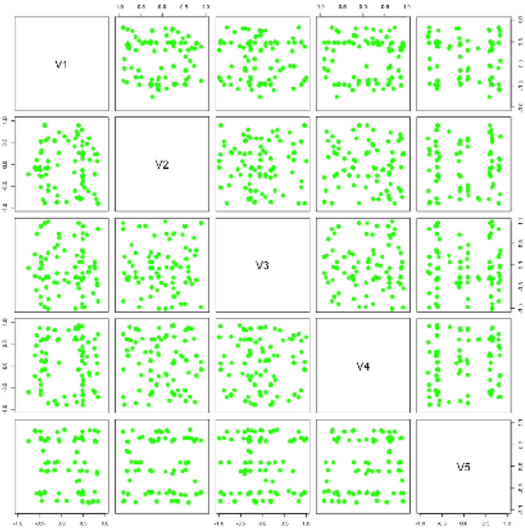




$q = 2$ basis elements do the job thanks to high correlations! (Unsurprising - material can only deform in certain physically-coherent ways)

Composite/PDE model

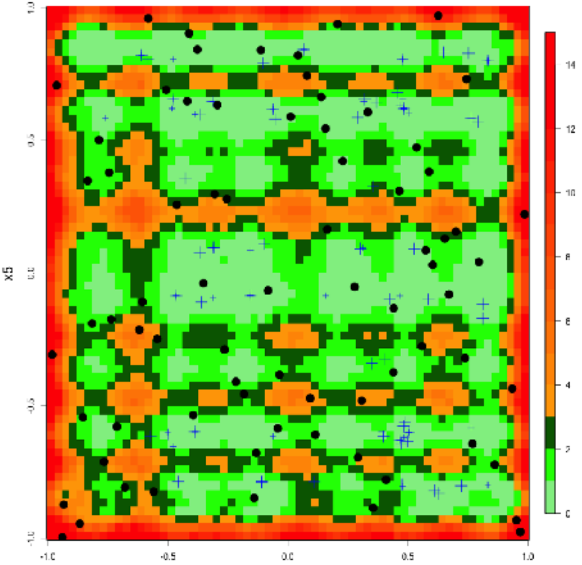
Sketch of true* NROY space



*zero emulator variance

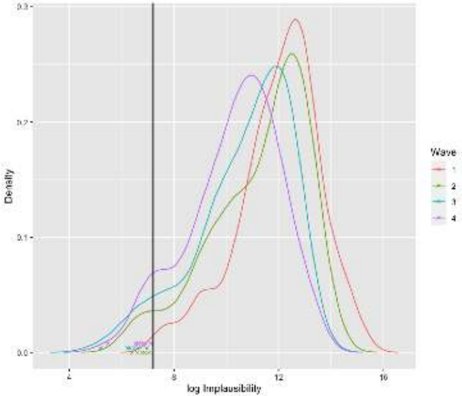
Composite/PDE model

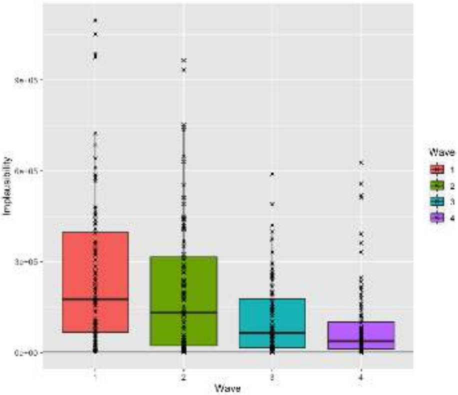
Sketch of true* NROY space



Does it work?

83.6% → 45.2% → 30.1% → 22.8% of $\mathcal{X}$

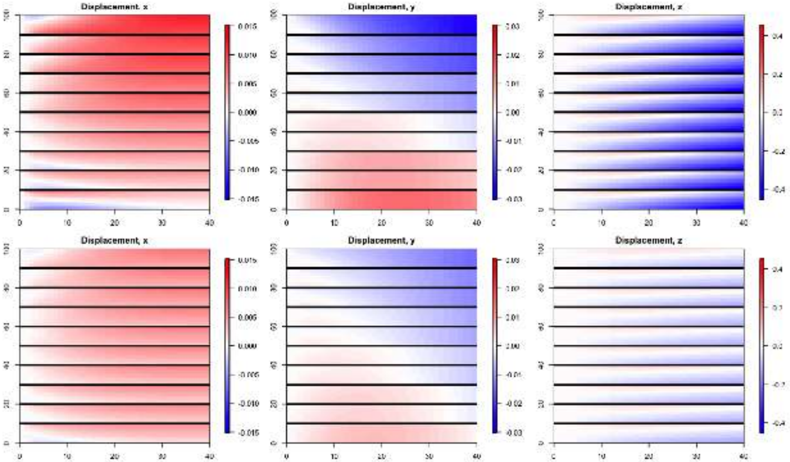




We get a lot of the way, but as zoom in identify non-stationarities etc. so need to be smarter

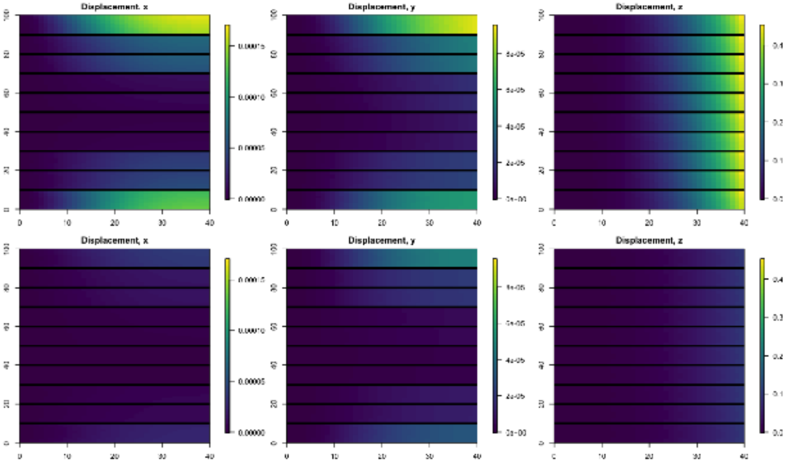
Does it work?

$$f(\mathbf{x}^*) - E_1[f(\mathbf{x}^*)], f(\mathbf{x}^*) - E_4[f(\mathbf{x}^*)]$$

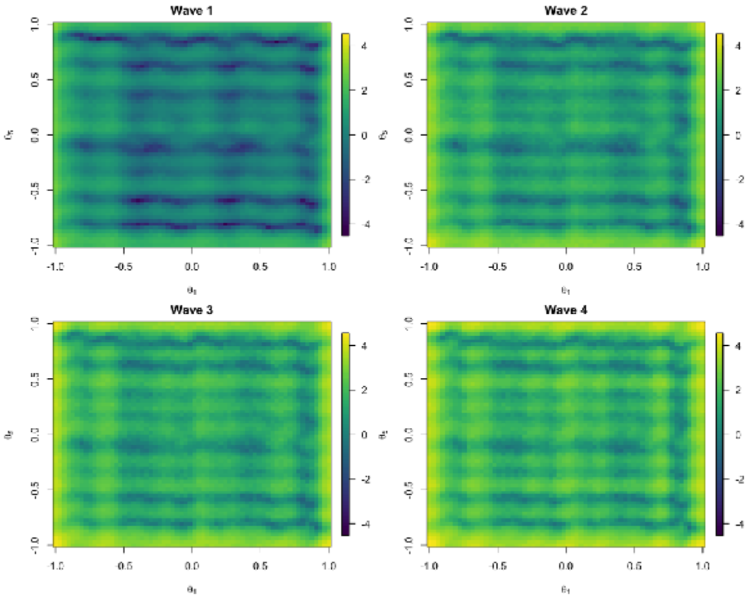


Does it work?

$Var_1[f(\mathbf{x}^*)]$, $Var_4[f(\mathbf{x}^*)]$



NROY space



Outline

Summary

- Given a good choice of basis, can emulate, predict, and calculate implausibility metric efficiently, for high ℓ.
- Useful tool for many problems - won't always care about what the full output looks like, but get a lot of nice properties almost for free.
- Type of basis not important - as long as satisfies rules! Doesn't need to be SVD-based (but often a good starting point).
- Discrepancy clearly important, potential for identifying within this framework.

Some references

- JMS & Williamson D. B. (2020). "Efficient calibration for high-dimensional computer model output using basis methods". arXiv preprint arXiv:1906.05758
- JMS, Dodwell T.J., et al. (2021) "A History Matching Approach to Building Full-Field Emulators in Composite Analysis". In preparation.
- JMS, Williamson D. B., Scinocca J., and Kharin V. (2019). "Uncertainty quantification for computer models with spatial output using calibration-optimal bases". Journal of the American Statistical Association, 114.528, 1800-1814.

Bayesian calibration:

- Kennedy, M. C. and O'Hagan, A. (2001). "Bayesian calibration of computer models." Journal of the Royal Statistical Society: Series B (Statistical Methodology), 63(3):425–464.
- Higdon, Dave, et al. "Computer model calibration using high-dimensional output." Journal of the American Statistical Association 103.482 (2008): 570-583.
- Chang, Won, et al. "Probabilistic calibration of a Greenland Ice Sheet model using spatially resolved synthetic observations: toward projections of ice mass loss with uncertainties." Geoscientific Model Development 7.5 (2014): 1933-1943.

History matching:

- Craig, P. S., Goldstein, M., Seheult, A., and Smith, J. (1996). "Bayes linear strategies for matching hydrocarbon reservoir history." Bayesian statistics, 5:69–95.
- Vernon, I., Goldstein, M., and Bower, R. G. (2010). "Galaxy formation: a Bayesian uncertainty analysis." Bayesian Analysis, 5(4):619–669.
- Williamson, Daniel, et al. "Identifying and removing structural biases in climate models with history matching." Climate Dynamics 45.5-6 (2015): 1299-1324.