Gaussian process emulation to rigorously explore uncertainty in complex models of the atmosphere and climate

Jill S Johnson

Collaborators: Leighton Regayre, Rachel Sansom and Ken Carslaw (University of Leeds); Lauren Marshall (University of Durham).

Gaussian processes, surrogates and digital twins workshop

14th September 2023

Presentation Outline

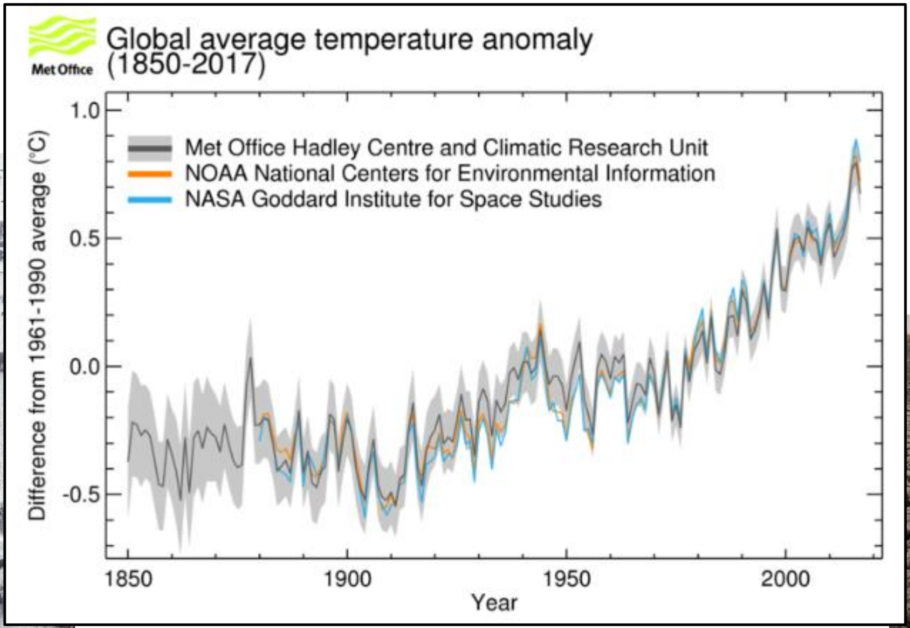
- **Context**
  - Aerosols, and their effects on the climate...
  - What do we mean by ‘uncertainty’?
  - Models of complex systems and uncertainty…
- **Methods - A statistical framework for UQ in complex models**
  - Gaussian Process ‘emulation’ to densely sample the model
- **Applications in Climate Science**
  - Volcanic eruptions…
  - Uncertainty constraint of a global aerosol-climate model…
- **Challenges…**
- **Summary**

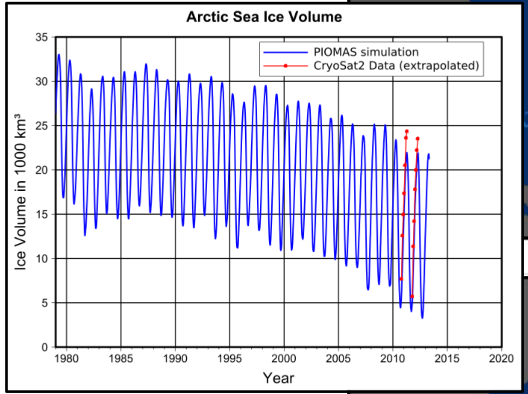
Motivation: The changing climate…

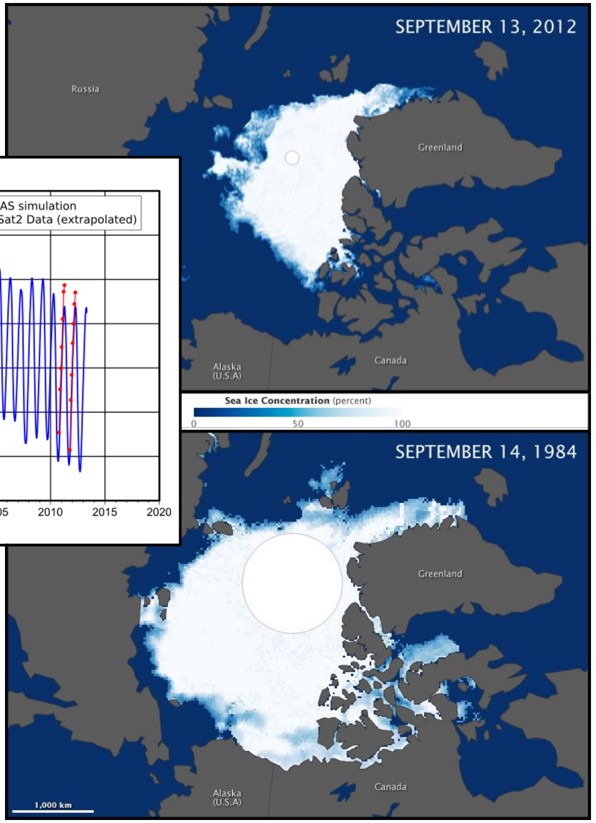
In recent years, the Earth's climate has been changing (e.g. the average global temperature has been increasing)

- Why is this happening?
- What are the potential impacts?

http://www.metoffice.gov.uk/climate/uk/interesting/2014-janwind







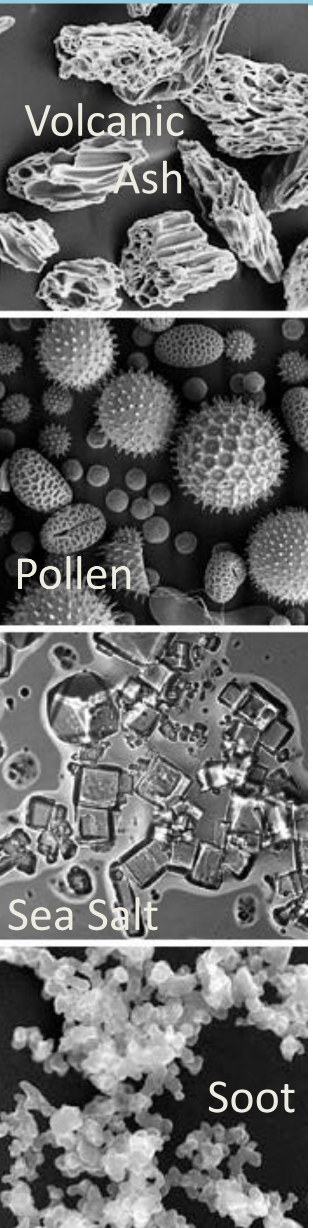
Context: Aerosols, and their effects on the climate

Aerosols are particles suspended in the atmosphere

- They come from **natural** and **anthropogenic** sources

Source: http://earthobservatory.nasa.gov/Features/Aerosols/

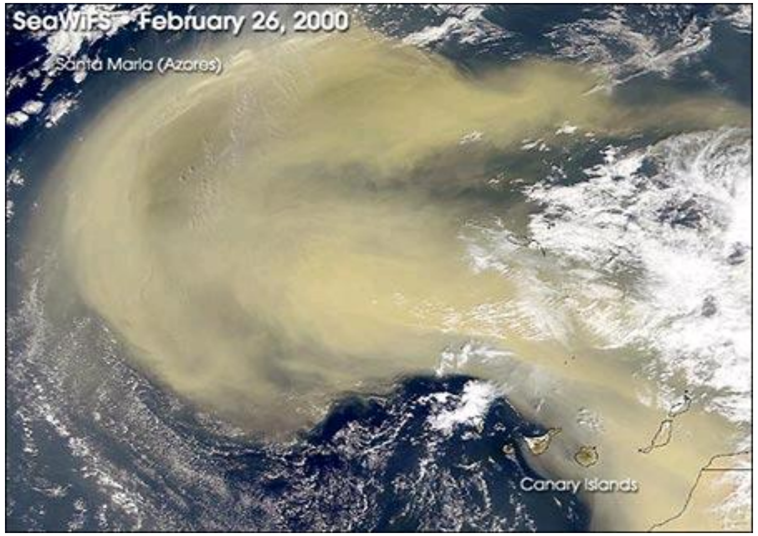
Context: Aerosols, and their effects on the climate



Aerosols are particles suspended in the atmosphere

- They come from **natural** and **anthropogenic** sources
- They **affect the radiative balance** of the earth-atmosphere system in two ways**:**

1. Aerosols scatter and absorb energy directly: The Aerosol Radiation Interaction (ARI) effect



Source: http://earthobservatory.nasa.gov/Features/Aerosols/, NASA image by Robert Simmon.

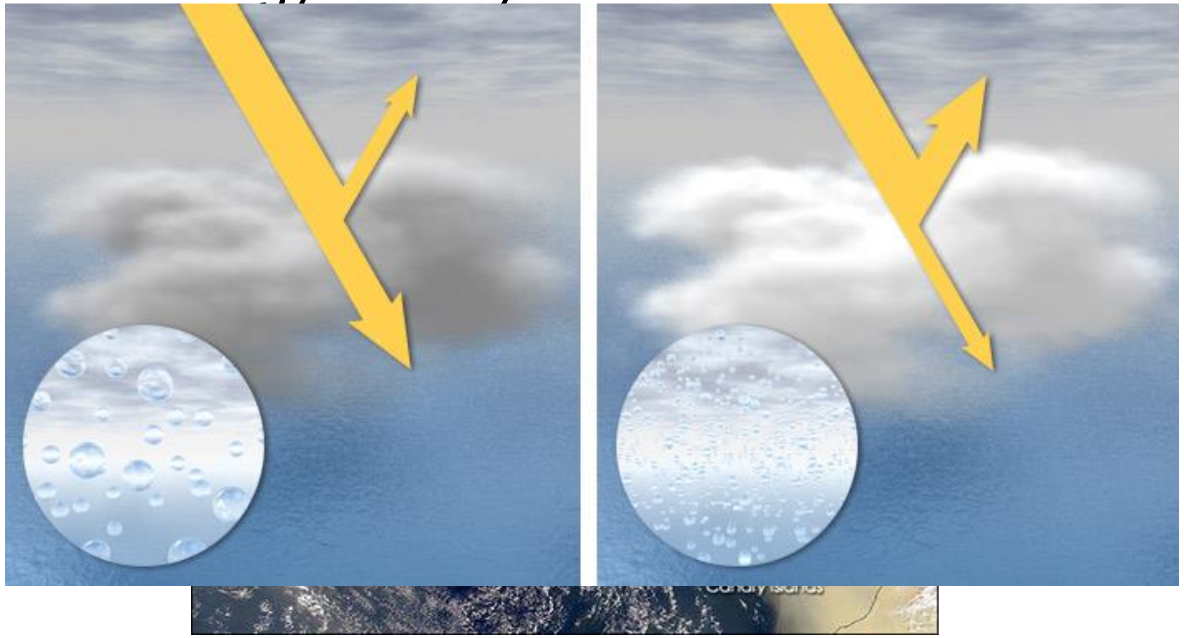
Source: http://earthobservatory.nasa.gov/Features/Aerosols/

Context: Aerosols, and their effects on the climate

Aerosols are particles suspended in the atmosphere

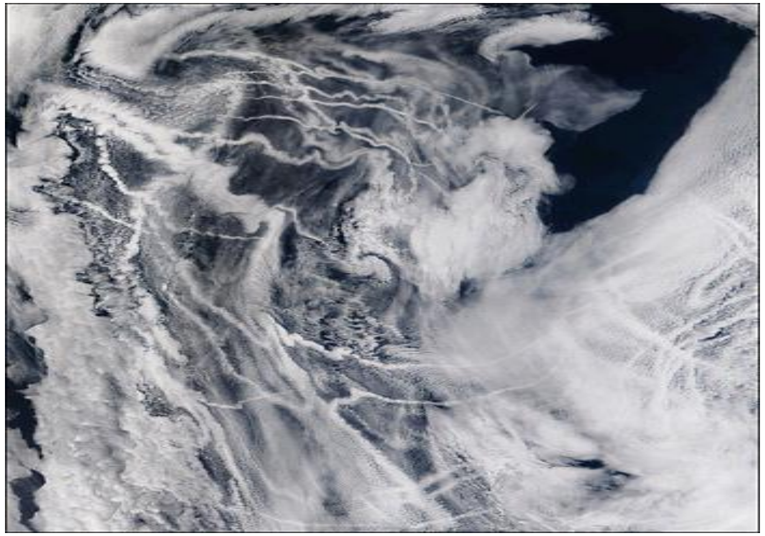
- They come from **natural** and **anthropogenic** sources
- They **affect the radiative balance** of the earth-atmosphere system in two ways**:**

1. Aerosols scatter and absorb energy directly: The Aerosol



Source: http://earthobservatory.nasa.gov/Features/Aerosols/, NASA image by Robert Simmon.

2. Aerosols can affect and change the properties of clouds: The Aerosol Cloud Interaction (ACI) effect



Source: http://earthobservatory.nasa.gov/Features/Aerosols/

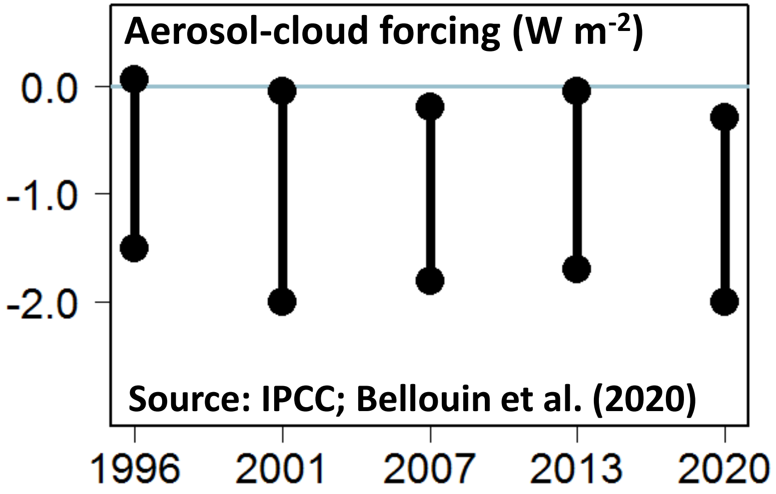
Context: Aerosols, and their effects on the climate

Aerosols are particles suspended in the atmosphere

- They come from **natural** and **anthropogenic** sources
- They **affect the radiative balance** of the earth-atmosphere system in two ways**:**

1. Aerosols scatter and absorb energy directly: The Aerosol Radiation Interaction (ARI) effect
2. Aerosols can affect and change the properties of clouds: The Aerosol Cloud Interaction (ACI) effect

Their effects on the **change in Earth's radiative balance** over the industrial period – **Aerosol Radiative Forcing** – is the **most uncertain factor** in international assessments of climate change…



Source: http://earthobservatory.nasa.gov/Features/Aerosols/

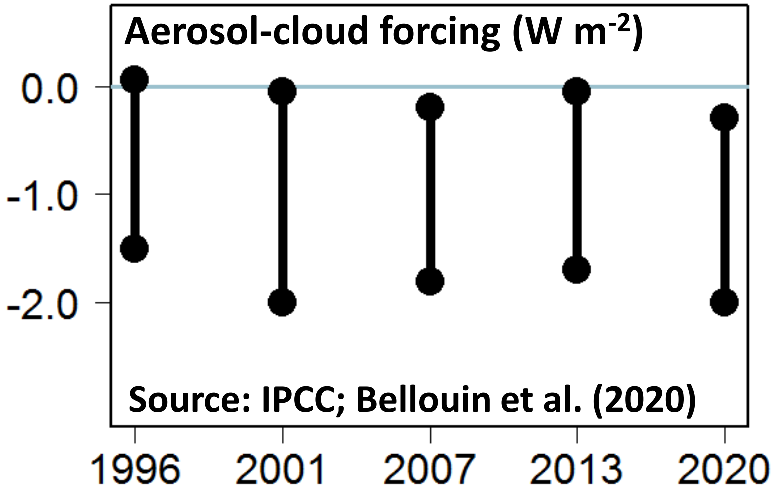
Context: Aerosols, and their effects on the climate

Aerosols are particles suspended in the atmosphere

- They come from **natural** and **anthropogenic** sources
- They **affect the radiative balance** of the earth-atmosphere system in two ways**:**

1. Aerosols scatter and absorb energy directly: The Aerosol Radiation Interaction (ARI) effect
2. Aerosols can affect and change the properties of clouds: The Aerosol Cloud Interaction (ACI) effect

Their effects on the **change in Earth's radiative balance** over the industrial period – **Aerosol Radiative Forcing** – is the **most uncertain factor** in international assessments of climate change…



➡ We need to **explore the effects of uncertainty** in cloud / aerosol / climate models in order to understand it and ultimately constrain it….

Source: http://earthobservatory.nasa.gov/Features/Aerosols/

Context: What is uncertainty?

– **Aleatory** uncertainty, due to **randomness**:

- Natural variability in climate processes.
- Stochastic variables: rainfall, wind speed…

Context: What is uncertainty?

- **Aleatory** uncertainty, due to **randomness**:
  - Natural variability in climate processes.
  - Stochastic variables: rainfall, wind speed…
- **Epistemic** uncertainty, due to **lack of knowledge**:
  - Emissions levels.
  - The characterisation of the variability.
  - Process interactions that we don’t yet know about…

Context: Descriptions of uncertainty

Our descriptions of uncertainty can **vary** greatly…

– **Qualitative**

- Statements such as “likely” and “unlikely”…
- Low, medium or high confidence (IPCC).

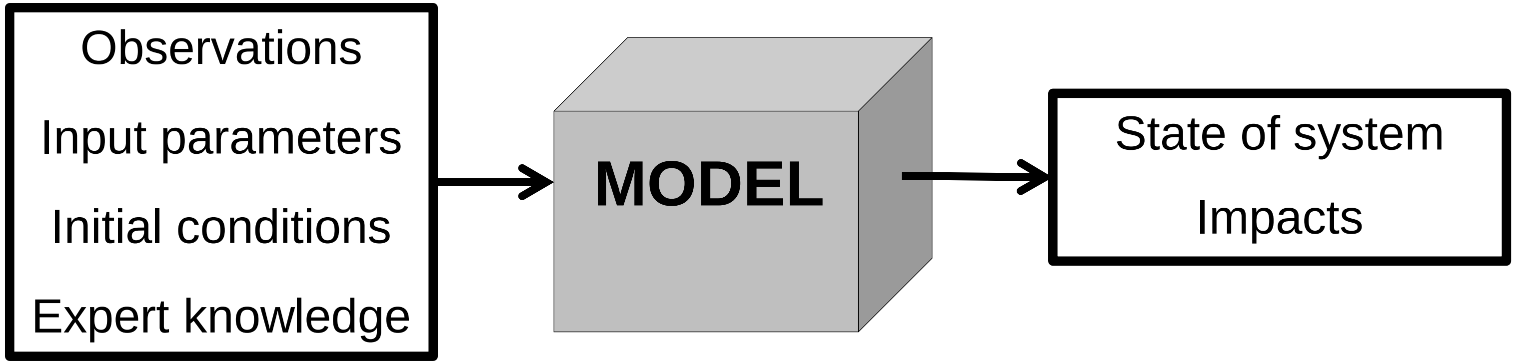
– **Quantitative**

- A range of plausible values…
- Standard deviation and variance…
- A statistical distribution…
- Confidence bounds (90%, 95%, 99%...).

Context: Models of complex systems

Models are used to **simulate our knowledge** of complex systems like the atmosphere and climate

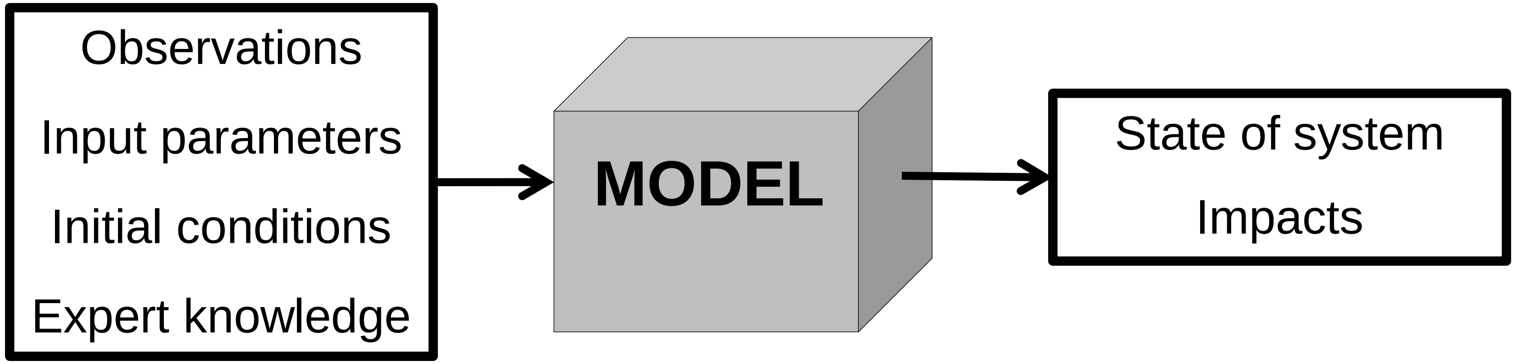
– **Aim:** to understand and predict the system's behaviour



Context: Models of complex systems

Models are used to **simulate our knowledge** of complex systems like the atmosphere and climate

– **Aim:** to understand and predict the system's behaviour

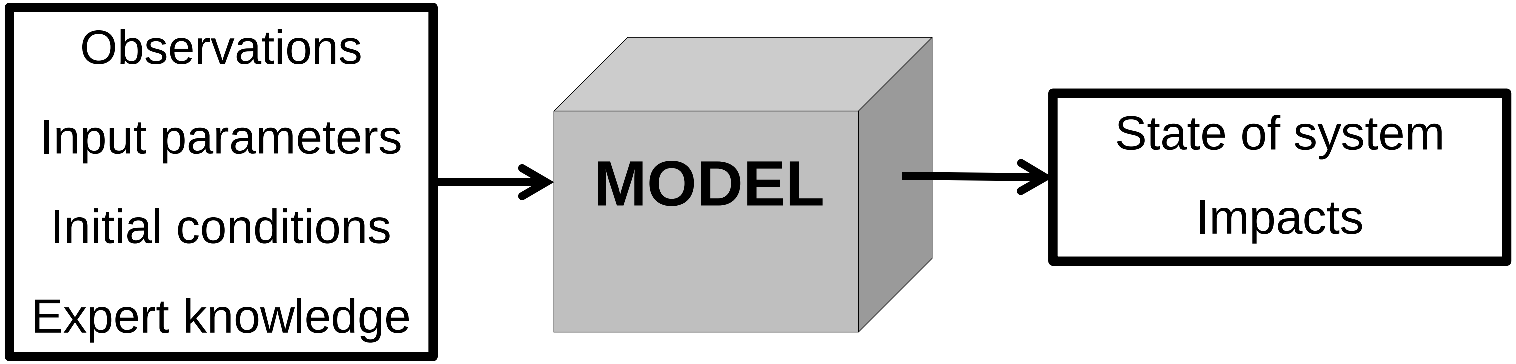


- A model is **inherently uncertain**
  - We **cannot include the full detail** of everything – we must make assumptions and simplifications (parameterisations)
  - There is **natural variability** in the system processes
  - Parts of the system are still **unknown / to be discovered**

Context: Models of complex systems

Models are used to **simulate our knowledge** of complex systems like the atmosphere and climate

– **Aim:** to understand and predict the system's behaviour



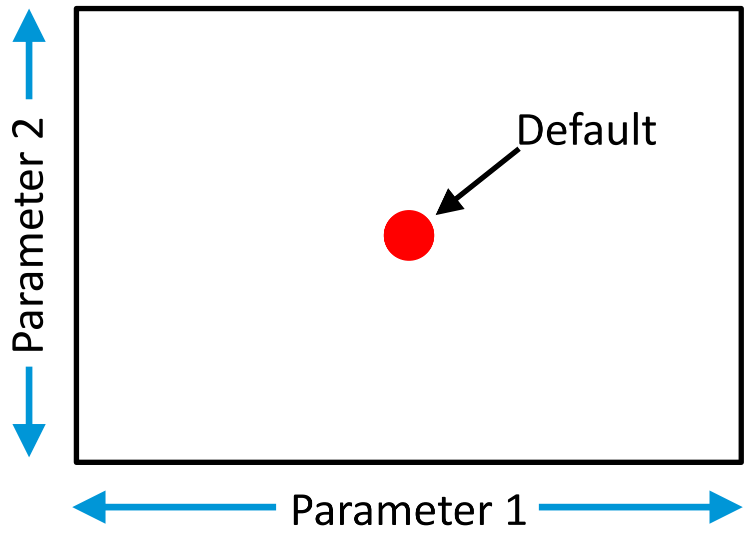
- A model is **inherently uncertain**
  – We **cannot include the full detail** of everything – we must make assumptions and simplifications (parameterisations)
  – There is **natural variability** in the system processes
  – Parts of the system are still **unknown / to be discovered**

- A model **has many uncertain inputs** (parameters) – these are what I'm interested in!
  – **How does the uncertainty in model parameters affect predictions of system behaviour?**

Context: How to explore the effects of parameter uncertainty on predictions of system behaviour?

Uncertainty ranges on model parameters generate a (multi-dimensional) **parameter uncertainty space** to explore

2-D Example:



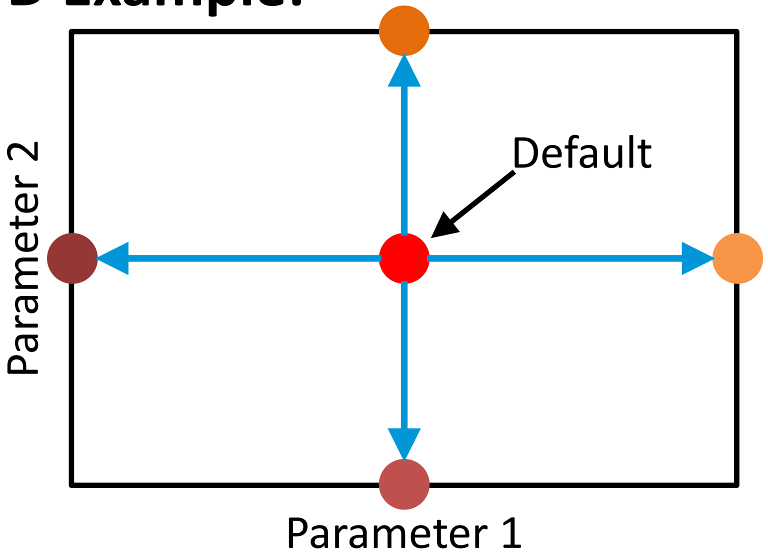
Context: How to explore the effects of parameter uncertainty on predictions of system behaviour?

Uncertainty ranges on model parameters generate a (multi-dimensional) **parameter uncertainty space** to explore

For large models, **'one-at-a-time' perturbations** are often used to explore the effects of parameter uncertainty, but they provide **minimal coverage** of 'uncertainty space'

- What is the model behaviour in the rest of the space?

2-D Example:



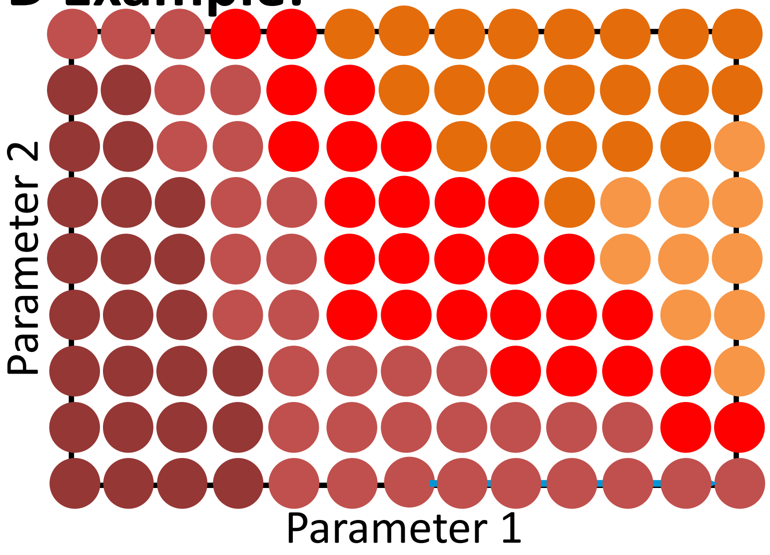
Context: How to explore the effects of parameter uncertainty on predictions of system behaviour?

Uncertainty ranges on model parameters generate a (multi-dimensional) **parameter uncertainty space** to explore

For large models, **'one-at-a-time' perturbations** are often used to explore the effects of parameter uncertainty, but they provide **minimal coverage** of 'uncertainty space'

- What is the model behaviour in the rest of the space?

2-D Example:



To fully understand the system behaviour, we need to **densely sample** the space

BUT, running a complex model requires significant computational resource → NOT FEASIBLE with the model itself

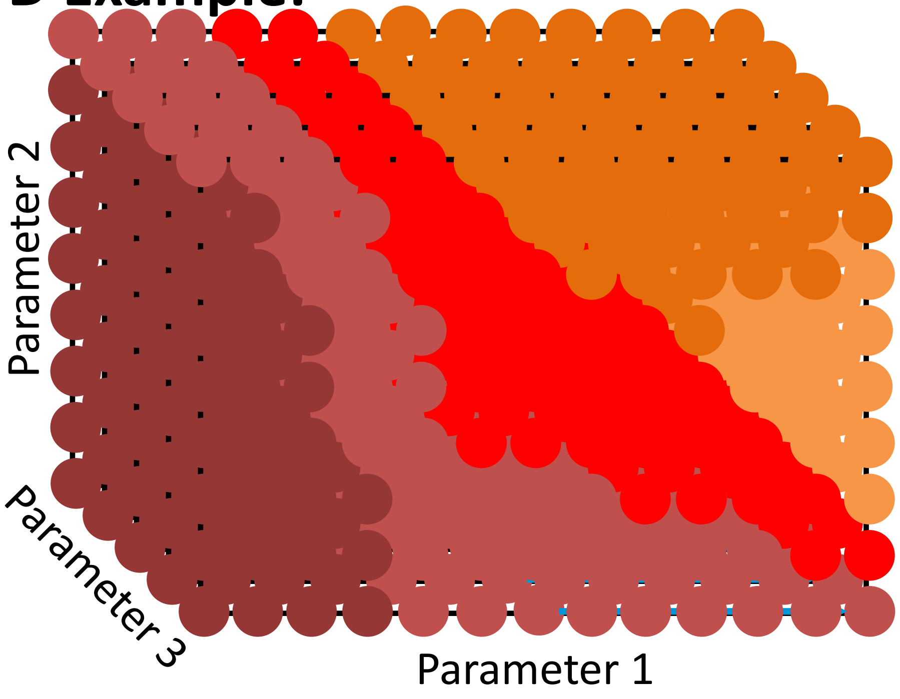
Context: How to explore the effects of parameter uncertainty on predictions of system behaviour?

Uncertainty ranges on model parameters generate a (multi-dimensional) **parameter uncertainty space** to explore

For large models, **'one-at-a-time' perturbations** are often used to explore the effects of parameter uncertainty, but they provide **minimal coverage** of 'uncertainty space'

- What is the model behaviour in the rest of the space?

3-D Example:



To fully understand the system behaviour, we need to **densely sample** the space

BUT, running a complex model requires significant computational resource → NOT FEASIBLE with the model itself, **especially as the number of input dimensions increase**

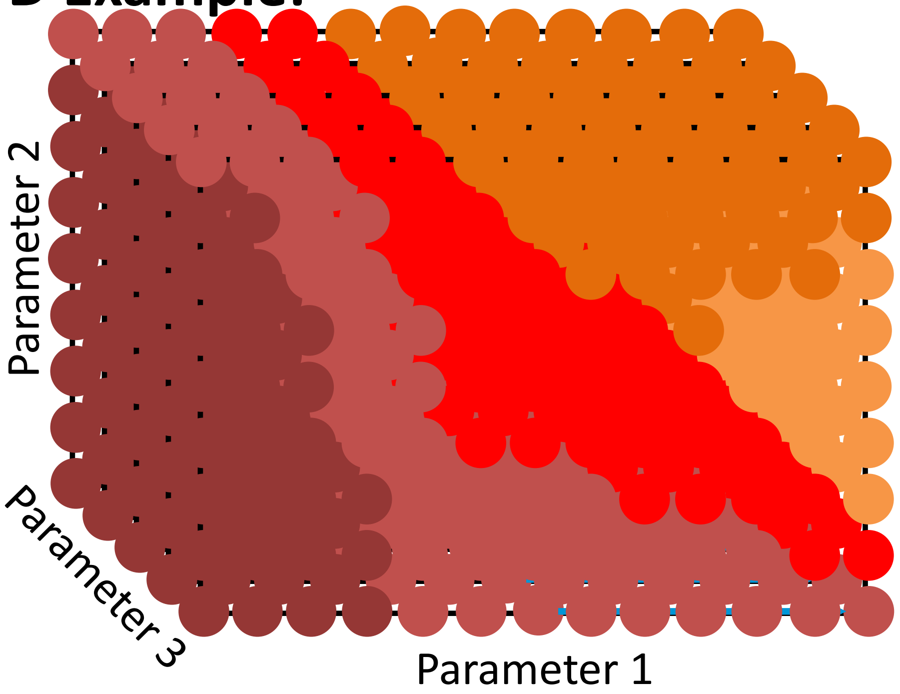
Context: How to explore the effects of parameter uncertainty on predictions of system behaviour?

Uncertainty ranges on model parameters generate a (multi-dimensional) **parameter uncertainty space** to explore

For large models, **'one-at-a-time' perturbations** are often used to explore the effects of parameter uncertainty, but they provide **minimal coverage** of 'uncertainty space'

- What is the model behaviour in the rest of the space?

3-D Example:



To fully understand the system behaviour, we need to **densely sample** the space

BUT, running a complex model requires significant computational resource → NOT FEASIBLE with the model itself, **especially as the number of input dimensions increase**

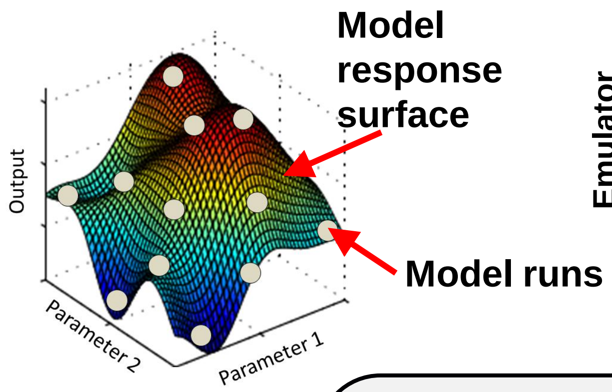
We need a **statistical framework** to enable **dense sampling at a low computational cost**, so to explore the model behaviour over the uncertainty

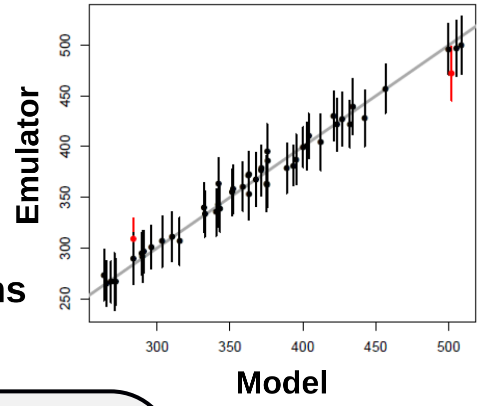
A statistical framework for UQ in complex models

Oakley and O'Hagan (2004); Lee et al. (2013); Johnson et al. (2015)

Expert elicitation

(choose parameters and their ranges)





Experimental design

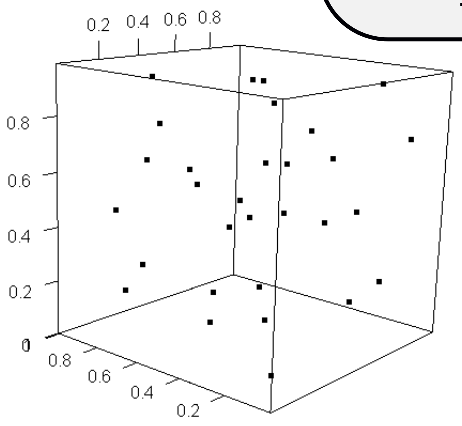
(select points in parameter space). e.g. Latin Hypercube

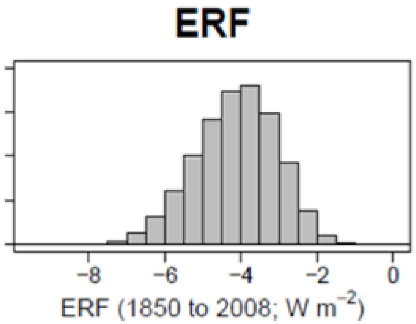
Run perturbed parameter ensemble

Build emulator (surrogate model)

Test emulator against simulator

(additional validation runs)





Uncertainty Analysis and Model Evaluation

- **Explore** model outputs and their responses
- **Identify key sources** of uncertainty (variance-based sensitivity analysis)
- **Constrain** the model output (history matching)

A statistical framework for UQ in complex models

Oakley and O'Hagan (2004); Lee et al. (2013); Johnson et al. (2015)

Expert elicitation

(choose parameters
and their ranges)

Expert Elicitation

Bring together
experts in the field...

‘We think these are the **uncertain parameters**, and their values are very unlikely to fall outside of these **ranges**’

Through **expert elicitation**, we:

– **Identify** the uncertain parameters to consider

– Determine a **range** (min → max) for each one

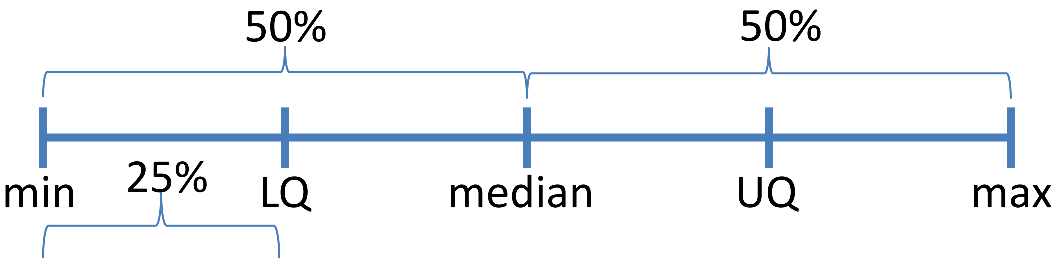
Expert Elicitation

Bring together experts in the field...

‘We think these are the uncertain parameters, and their values are very unlikely to fall outside of these ranges’

Through **expert elicitation**, we:

- **Identify** the uncertain parameters to consider
- Determine a **range** (min → max) for each one
- Obtain a **probability distribution** over the parameter range through evaluation of the median and different quantiles (LQ, UQ) over the range.



A statistical framework for UQ in complex models

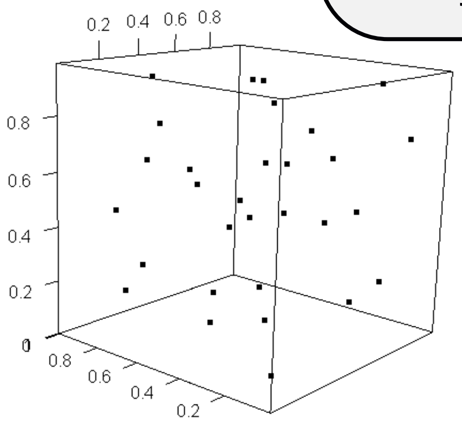
Oakley and O'Hagan (2004); Lee et al. (2013); Johnson et al. (2015)

Expert elicitation

(choose parameters and their ranges)

Experimental design

(select points in parameter space). e.g. Latin Hypercube



Experiment Design

The experiment design is a critical stage in this approach…

We want to obtain the maximum information over our parameter unc. space from the fewest possible model runs

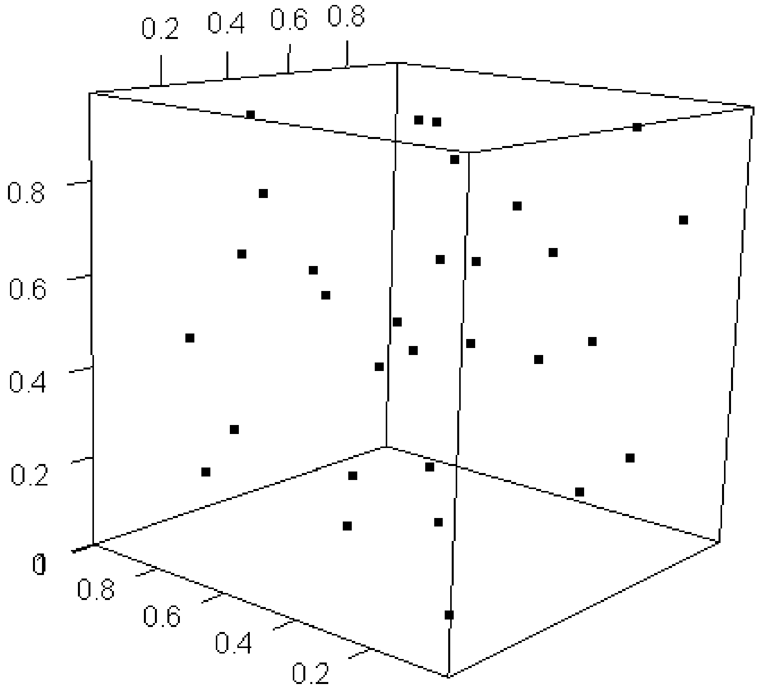
Experiment Design

The experiment design is a critical stage in this approach…

We want to obtain the maximum information over our parameter unc. space from the fewest possible model runs

Maximin Latin Hypercube

- Good marginal coverage.
- Good space-filling properties.
- Here, the minimum distance between any two points is maximised.
- Number of runs depends on 'active' parameters and function smoothness.
  - General rule: 10 x p



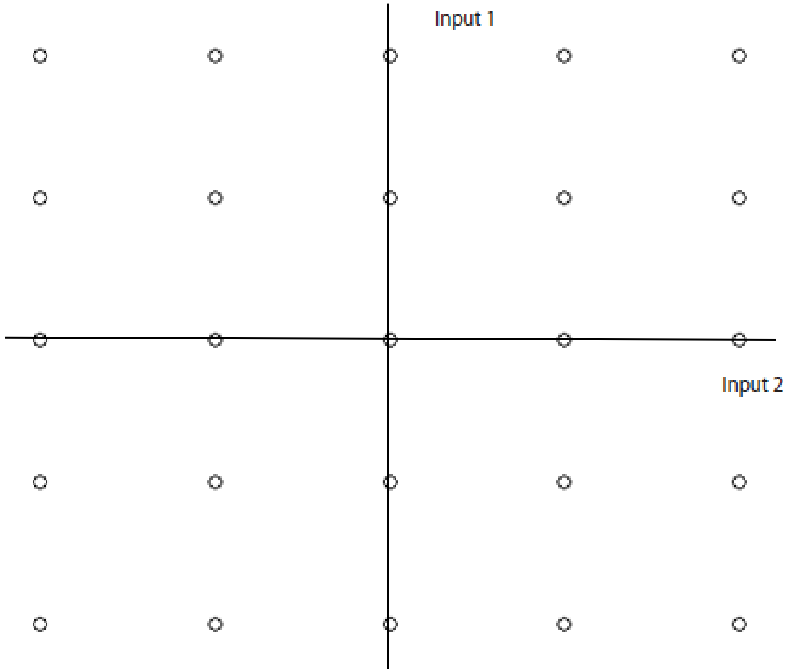
Extend to *N* dimensions for *N* important uncertainties

Experiment Design

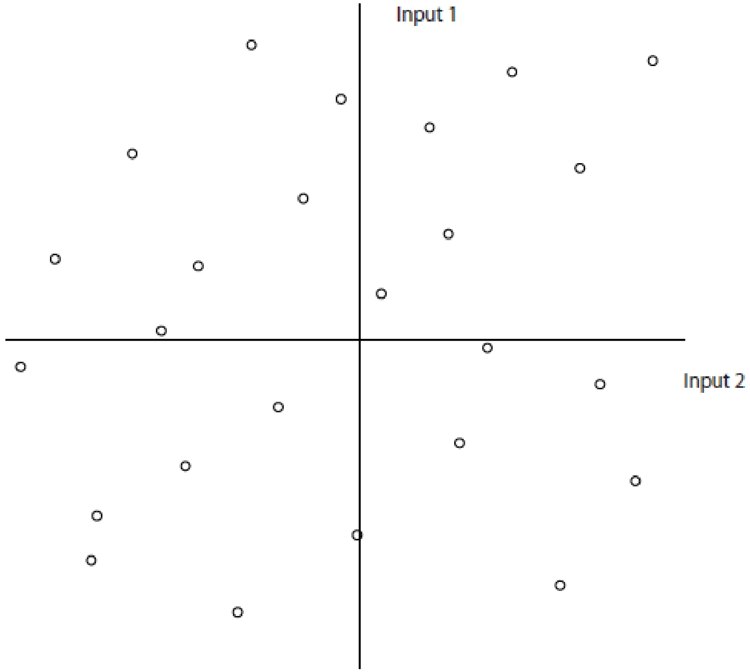
The experiment design is a critical stage in this approach…

We want to obtain the maximum information over our parameter unc. space from the fewest possible model runs

Factorial (gridded) designs



Maximin Latin Hypercube

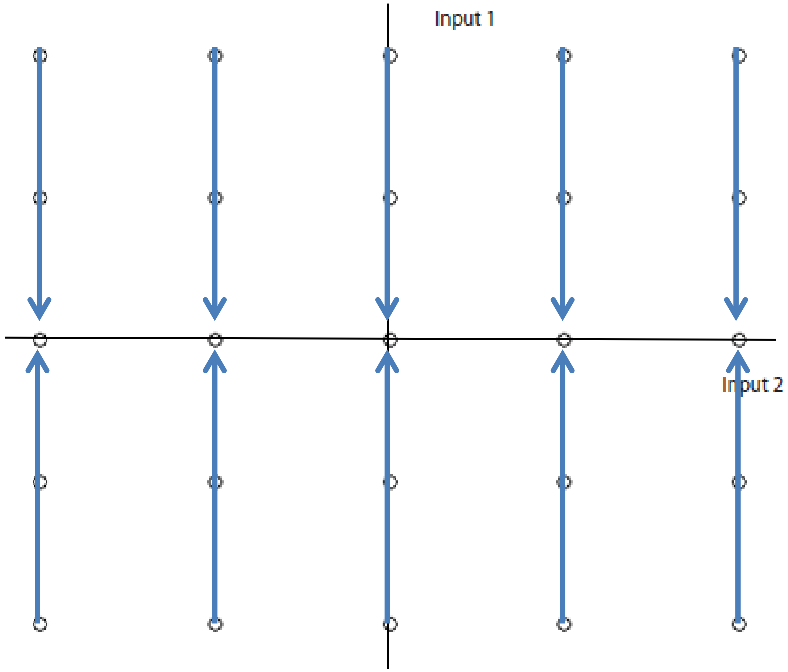


Experiment Design

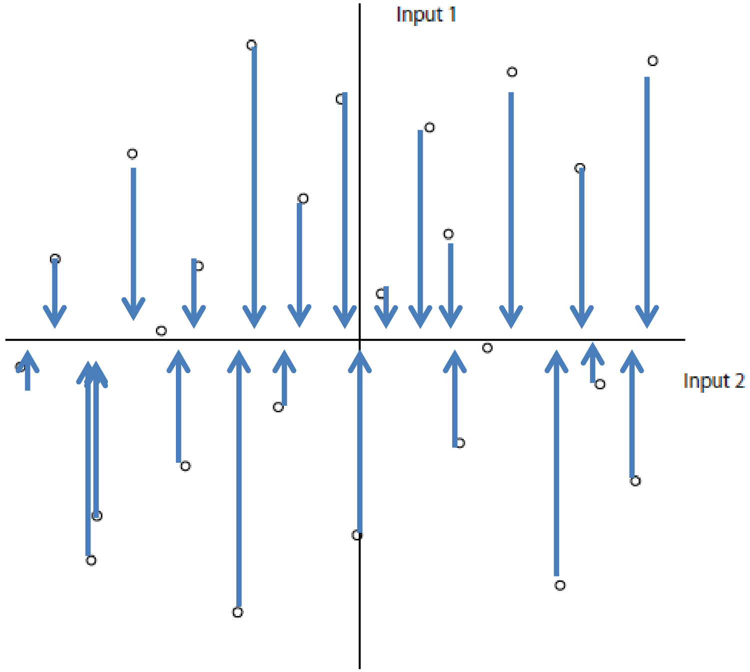
The experiment design is a critical stage in this approach…

We want to obtain the maximum information over our parameter unc. space from the fewest possible model runs

Factorial (gridded) designs



Maximin Latin Hypercube

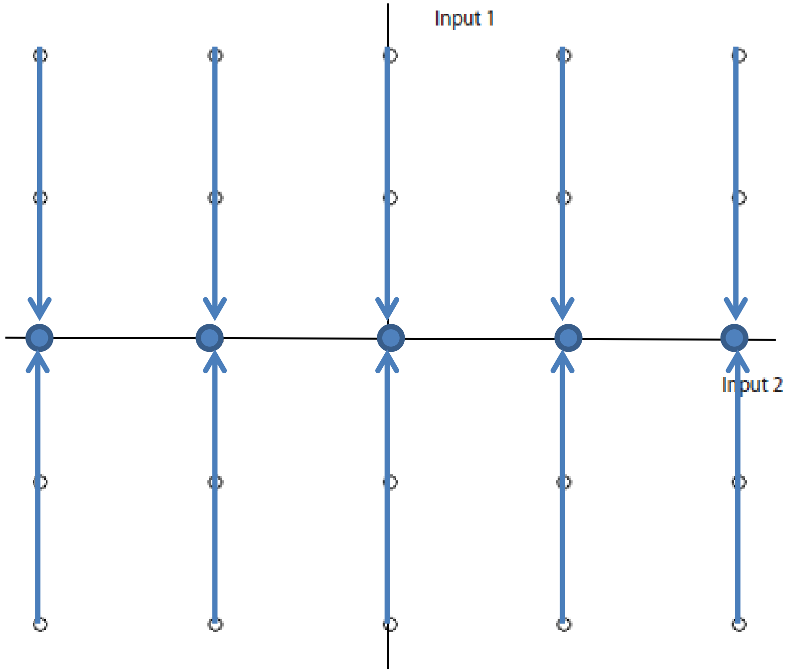


Experiment Design

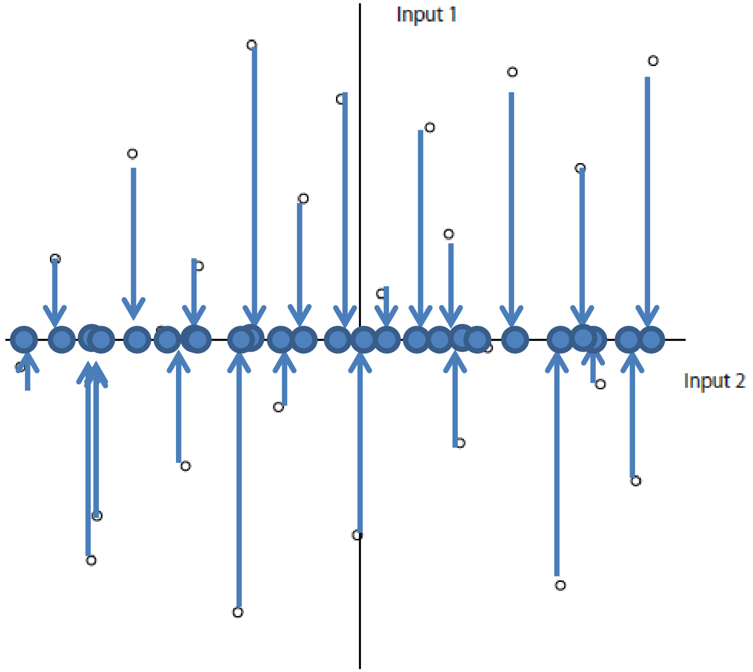
The experiment design is a critical stage in this approach…

We want to obtain the maximum information over our parameter unc. space from the fewest possible model runs

Factorial (gridded) designs



Maximin Latin Hypercube



A statistical framework for UQ in complex models

Oakley and O'Hagan (2004); Lee et al. (2013); Johnson et al. (2015)

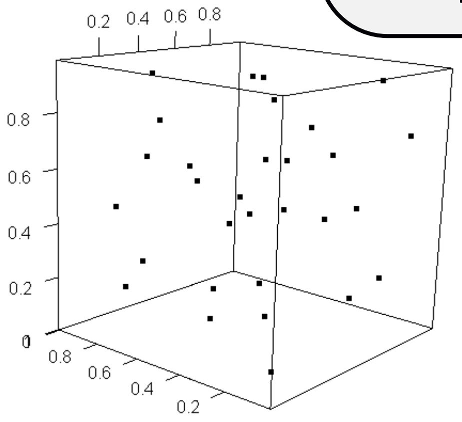
Expert elicitation

(choose parameters and their ranges)

Experimental design

(select points in parameter space). e.g. Latin Hypercube

Run perturbed parameter ensemble



A statistical framework for UQ in complex models

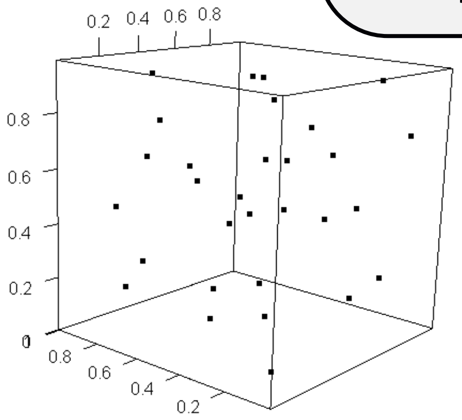
Oakley and O'Hagan (2004); Lee et al. (2013); Johnson et al. (2015)

Expert elicitation

(choose parameters and their ranges)

Experimental design

(select points in parameter space). e.g. Latin Hypercube



Run perturbed parameter ensemble

https://www.epcc.ed.ac.uk/hpc-services/archer2

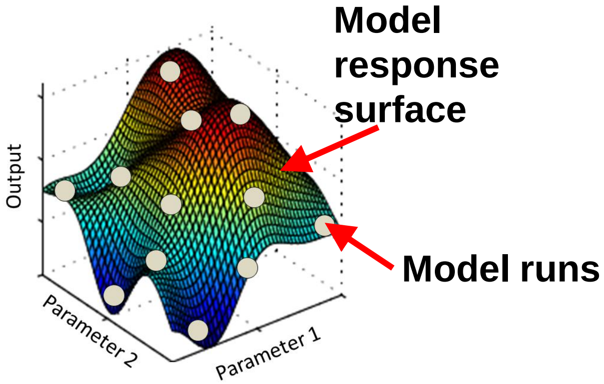
We collate the model outputs for each selected input combination

A statistical framework for UQ in complex models

Oakley and O'Hagan (2004); Lee et al. (2013); Johnson et al. (2015)

Expert elicitation

(choose parameters and their ranges)



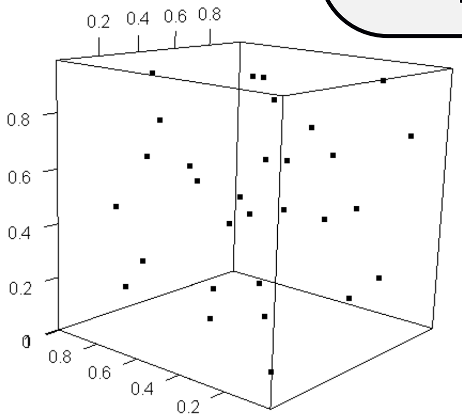
Experimental design

(select points in parameter space). e.g. Latin Hypercube

Run perturbed parameter ensemble

Build emulator

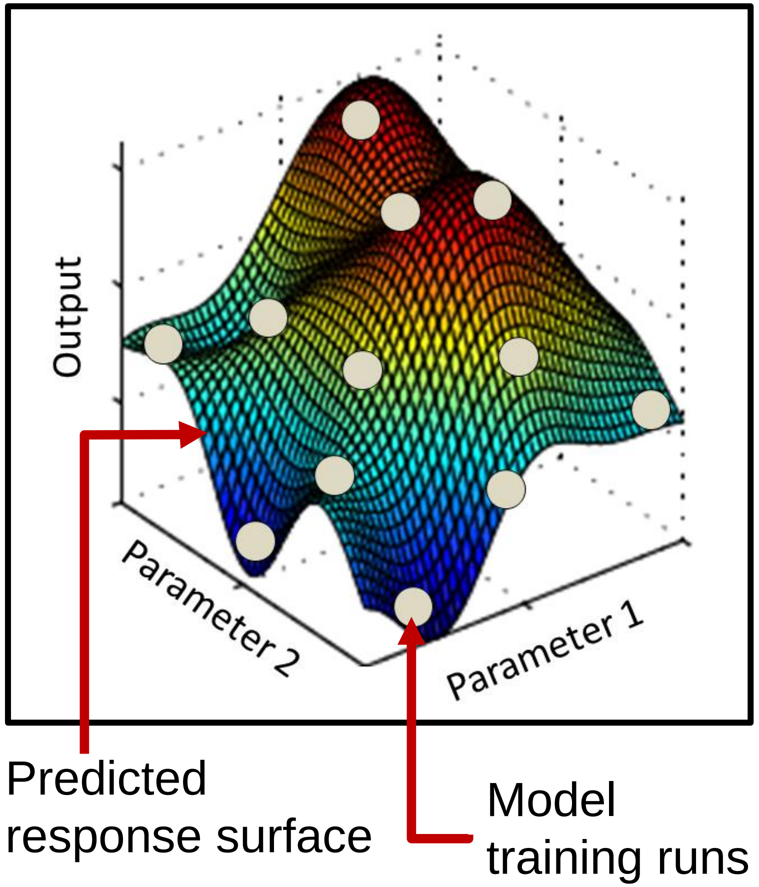
(surrogate model)



What is an emulator?

O'Hagan (2006)

An emulator is a **statistical representation** that **maps** the relationship between a set of uncertain inputs and a model output of interest.

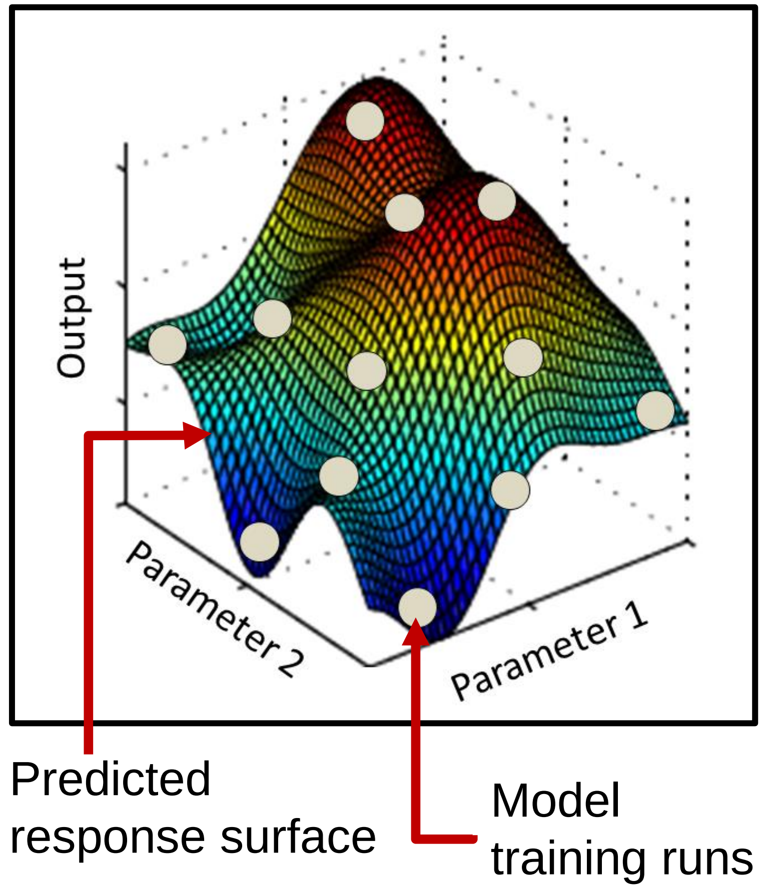


Extend to *N* dimensions for *N* important uncertainties

Gaussian process emulation

O'Hagan (2006)

An emulator is a **statistical representation** that **maps** the relationship between a set of uncertain inputs and a model output of interest.



Extend to *N* dimensions for *N* important uncertainties

– Based on the **Gaussian Process** (GP)

Key assumptions:

– The model output Y is **smooth**

– Model output at specific input parameter settings X gives information about model behaviour close by in parameter space.

a priori: $Y = g(X) \sim \mathrm{GP}[m(X)\,, k(X, X')]$

Applied within a **Bayesian statistical framework** that exploits conditional probability: Posterior emulator ∝ Prior × Likelihood

$$Y^*|Y, X, X^*, \theta \sim \mathrm{GP}[m^*(X)\,, k^*(X, X')]$$

General formulation of the emulator:

- $Y = g(X)$ represents the model simulator
- Parameters: $P_1, P_2, \dots, P_d$ that form an d-dimensional parameter space
- Training Inputs: $\boldsymbol{X} = \{X_1, X_2, \dots, X_n\}$, where $X_i = (p_{1i}, p_{2i}, \dots, p_{di})$
- Training outputs: $\boldsymbol{Y} = \{Y_1 = g(X_1), Y_2 = g(X_2), \dots, Y_n = g(X_n)\}$

A priori, we assume:

$$g(X) \sim \mathrm{GP}[m(X), k(X, X')]$$

where $m(X)$ and $k(X, X')$ are the mean and covariance functions of the GP, resp.

Given $\boldsymbol{X}$, the model for the training data is:

$$\boldsymbol{Y} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$$

with $\mu_i = m(X_i)$ and $\Sigma_{ij} = k\left(X_i, X_j\right)$, $i, j \in \{1, 2, \dots, n\}$

General formulation of the emulator: Mean and Covariance function choices

Mean Function:

– For a D-dimensional input vector X_i, the most popular choices are:

- Constant: $m(X_i) = \beta_0$
- Linear: $m(X_i) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_D X_{iD}$

Covariance Structure:

– The most popular choice is the "Squared Exponential" covariance function. For input vectors X_p and X_q of D-dimensions, this takes the form:

$$k(X_p, X_q) = Cov\left(g(X_p), g(X_q)\right) = \sigma_f^2 \prod_{d=1}^{D} \left[exp\left\{-\eta_d \left|X_{pd} - X_{qd}\right|^2\right\}\right] + \sigma_n^2 \delta_{pq},$$

where η_d, $d = \{1, \ldots, D\}$ are roughness parameters, σ_f^2 corresponds to the signal variance, and σ_n^2 corresponds to a noise (nugget) effect.

General formulation of the emulator:

– Test Inputs: $\boldsymbol{X}_* = \{X_{*1}, X_{*2}, \dots, X_{*s}\}$, not contained in $\boldsymbol{X}$, at which we wish to predict Y.

– Let $\boldsymbol{Y}_*$ be the corresponding vector of predictions.

– By the prior, the joint distribution of $(\boldsymbol{Y},\boldsymbol{Y}_*)$ is

$$\begin{bmatrix} \boldsymbol{Y} \\ \boldsymbol{Y}_* \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \boldsymbol{\mu} \\ \boldsymbol{\mu}_* \end{bmatrix}, \begin{bmatrix} \Sigma & \Sigma_* \\ {\Sigma_*}^T & \Sigma_{**} \end{bmatrix} \right)$$

where $\mu_{*i} = m(X_{*i})$, $i \in \{1,2, \dots, s\}$, Σ_* contains the training-test set covariances and Σ_{**} contains the test set covariances, given $k(X, X')$.

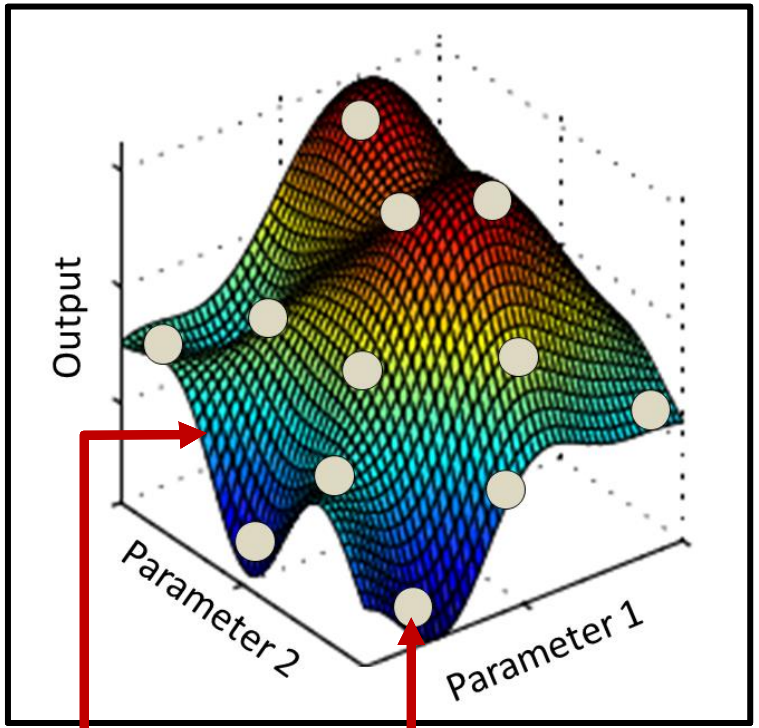
– By conditioning on the information in the training dataset, we obtain the following posterior distribution for $\boldsymbol{Y}_*$, from which we can predict $\boldsymbol{Y}_*$:

$$\boldsymbol{Y}_* | \boldsymbol{Y}, \boldsymbol{X}_*, \boldsymbol{X} \sim \mathcal{N}\left(\boldsymbol{\mu}_* + {\Sigma_*}^T \Sigma^{-1} (\boldsymbol{Y} - \boldsymbol{\mu}) , \Sigma_{**} - {\Sigma_*}^T \Sigma^{-1} \Sigma_* \right).$$

Gaussian process emulation

O'Hagan (2006)

An emulator is a **statistical representation** that **maps** the relationship between a set of uncertain inputs and a model output of interest.



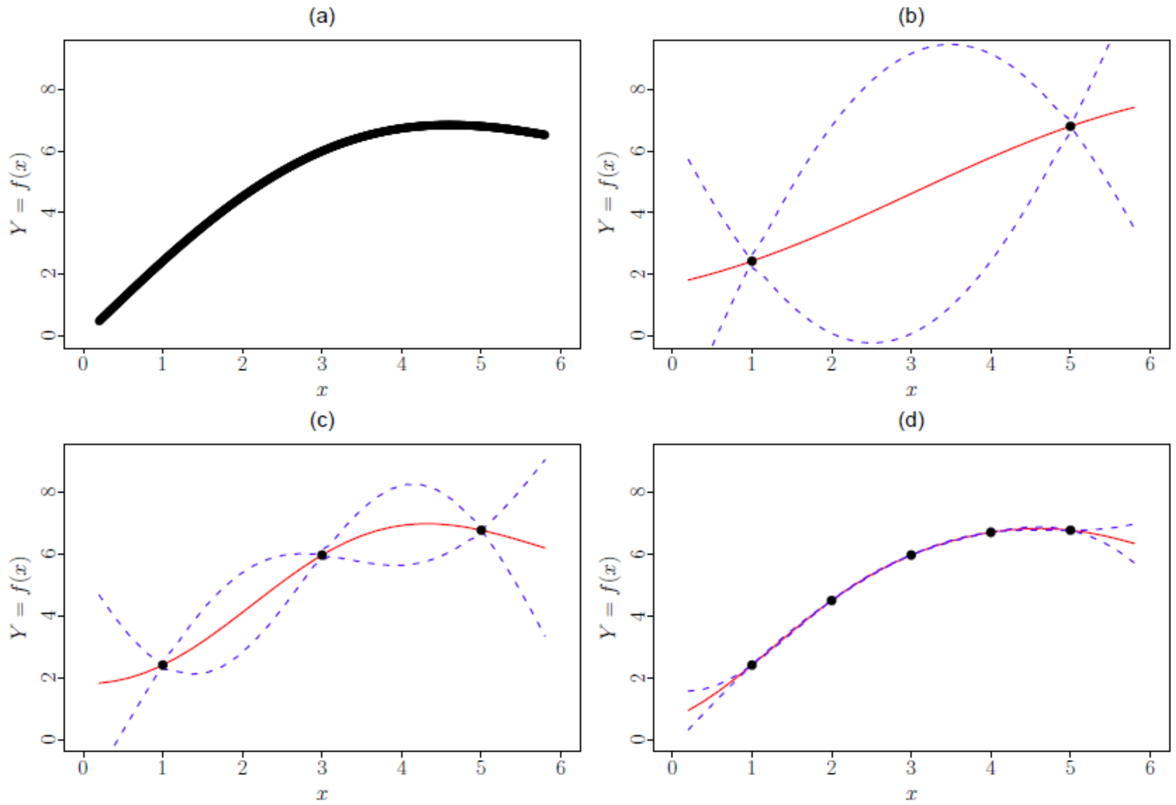
Predicted response surface

Model training runs

Extend to *N* dimensions for *N* important uncertainties

Example: O'Hagan (2006)

(a): The true function; **(b) - (d):** forming the emulator model, adding further data points until the true function is recovered.

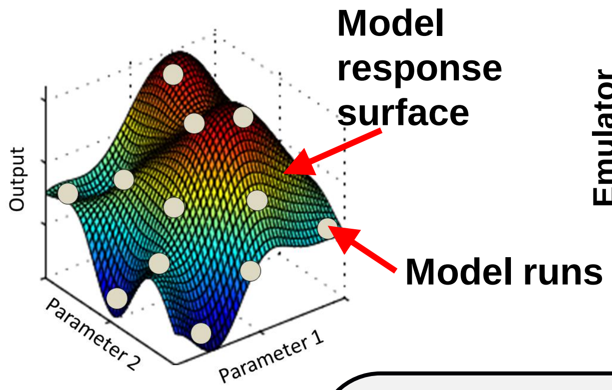


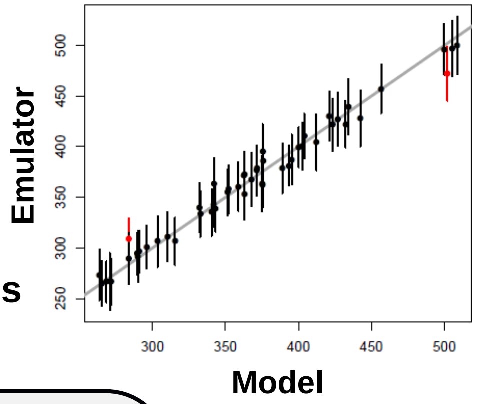
A statistical framework for UQ in complex models

Oakley and O'Hagan (2004); Lee et al. (2013); Johnson et al. (2015)

Expert elicitation

(choose parameters and their ranges)





Experimental design

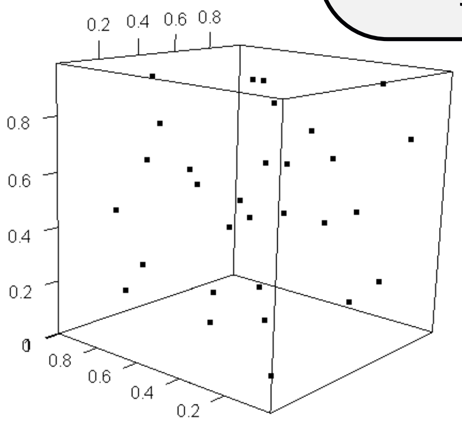
(select points in parameter space). e.g. Latin Hypercube

→ **Run perturbed parameter ensemble**

→ **Build emulator** (surrogate model)

→ **Test emulator against simulator**

(additional validation runs)

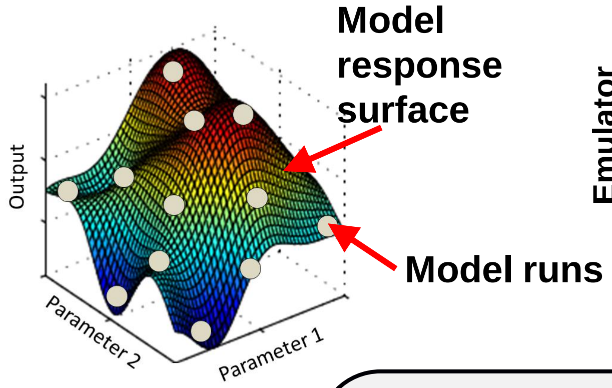


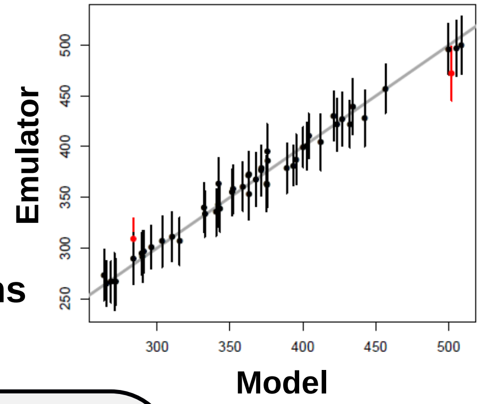
A statistical framework for UQ in complex models

Oakley and O'Hagan (2004); Lee et al. (2013); Johnson et al. (2015)

Expert elicitation

(choose parameters and their ranges)





Experimental design

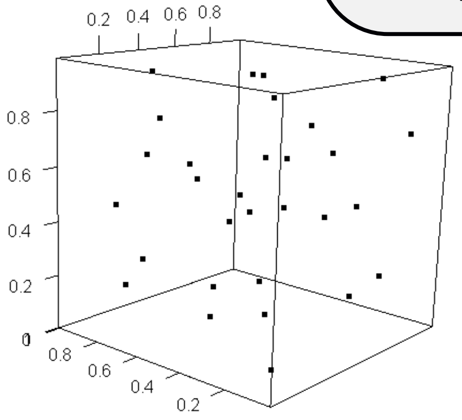
(select points in parameter space). e.g. Latin Hypercube

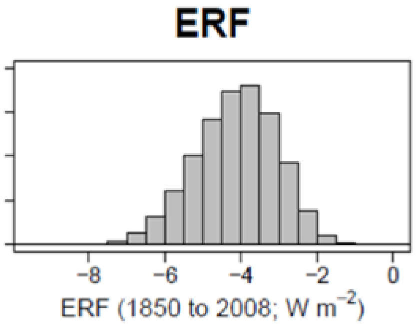
→ **Run perturbed parameter ensemble**

→ **Build emulator** (surrogate model)

→ **Test emulator against simulator**

(additional validation runs)

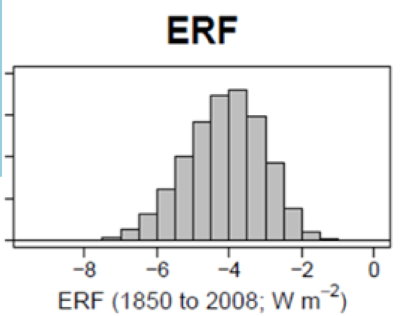




Uncertainty Analysis and Model Evaluation

- **Explore** model outputs and their responses
- **Identify key sources** of uncertainty (variance-based sensitivity analysis)
- **Constrain** the model output (history matching)

Variance-based Sensitivity Analysis (Saltelli et al., 2000)



– How does parameter uncertainty affect model predictions?

– Using **Variance decomposition**, we decompose the variance in the model output due to the parametric uncertainty, $\mathrm{V} = \mathrm{Var}(E[Y|\boldsymbol{X}])$, to its contribution sources:

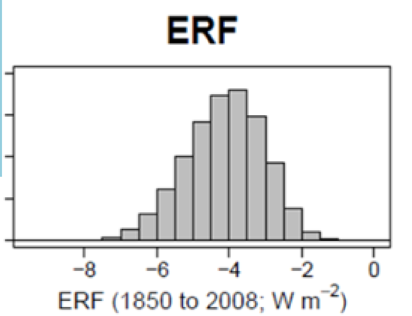
$$\mathrm{V} = \mathrm{Var}(E[Y|\boldsymbol{X}]) = \sum_{i=1}^{d} V_i + \sum_{i<j} W_{ij} + \cdots + W_{1,2,\ldots,d}$$

where:

- $V_i = \mathrm{Var}_{X_i}(E_{X_{-i}}[Y|X_i])$ represents the **expected amount by which the uncertainty in the model output Y will be reduced if the parameter $\boldsymbol{X_i}$ were known exactly**

- $V_{ij} = \mathrm{Var}_{X_{ij}}\left(E_{X_{-ij}}[Y|X_{ij}]\right) = V_i + V_j + W_{ij}$ represents… if X_i, X_j known exactly…

Variance-based Sensitivity Analysis (Saltelli et al., 2000)



– How does parameter uncertainty affect model predictions?

– Using **Variance decomposition**, we decompose the variance in the model output due to the parametric uncertainty, $\mathrm{V} = \mathrm{Var}(E[Y|\boldsymbol{X}])$, to its contribution sources:

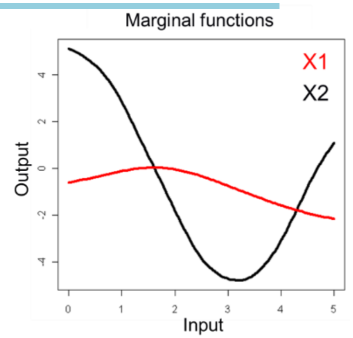
$$\mathrm{V} = \mathrm{Var}(E[Y|\boldsymbol{X}]) = \sum_{i=1}^{d} V_i + \sum_{i<j} W_{ij} + \cdots + W_{1,2,\ldots,d}$$

where:

- $V_i = \mathrm{Var}_{X_i}\left(E_{X_{-i}}[Y|X_i]\right)$ represents the **expected amount by which the uncertainty in the model output Y will be reduced if the parameter $\boldsymbol{X_i}$ were known exactly**

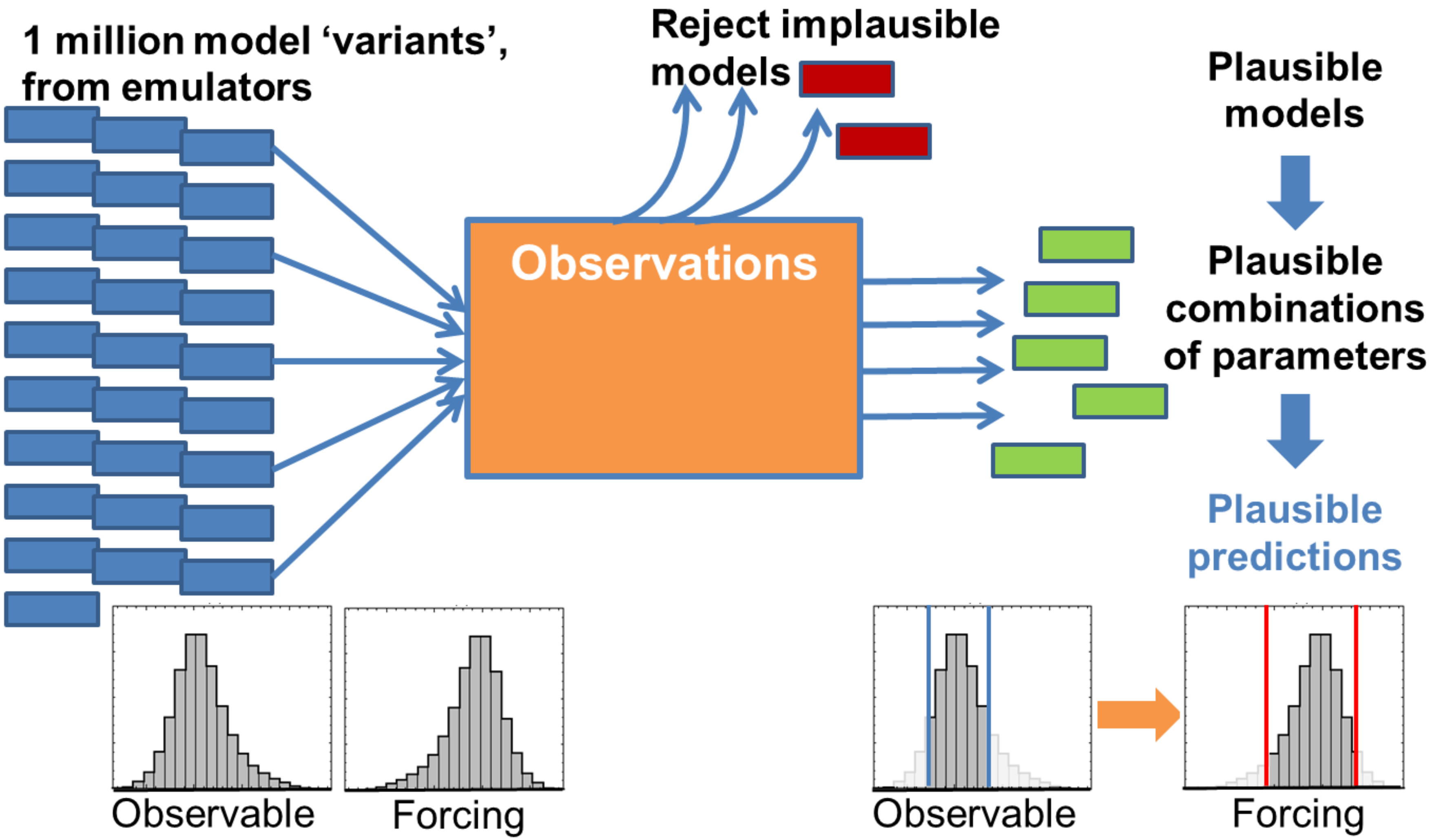
- $V_{ij} = \mathrm{Var}_{X_{ij}}\left(E_{X_{-ij}}\left[Y|X_{ij}\right]\right) = V_i + V_j + W_{ij}$ represents… if X_i, X_j known exactly…

Sensitivity Indices: The individual main effects (%age contributions to *V*) are given by: $\boldsymbol{S_i = \frac{V_i}{V}}$, and $\sum_{i=1}^{d} S_i + \sum_{i<j} S_{ij} + \cdots + S_{1,2,\ldots,d} = 1$



A ‘History Matching’ approach to reduce uncertainty

History Matching (Craig et al., 1996) rules out regions of parameter space that are not consistent with observations:



Johnson et al. (2018, 2020)

Example 1: Volcanic aerosol study – Marshall et al., 2019

(doi: 10.1029/2018JD028675)

Exploring the radiative forcing caused by a volcanic eruption….

Model: UM-UKCA (Met-office general circulation model (GCM) coupled to the UK Chemistry and Aerosol scheme; Based on 'Global-Atmosphere 4' configuration)

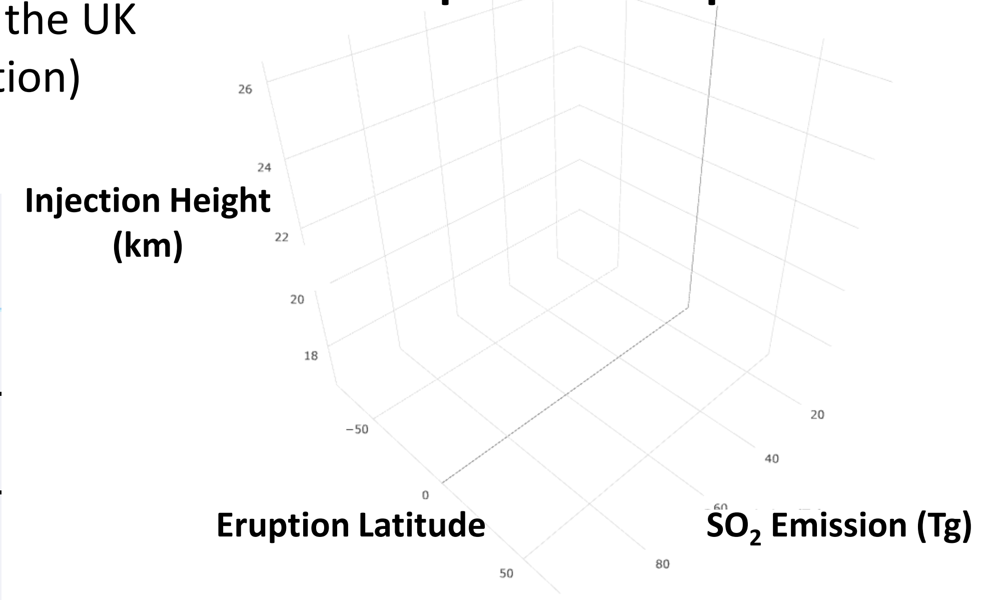
PPE: 30 training runs; 11 validation runs

Table 1

Eruption Source Parameters and Range in Values That are Perturbed in This Study

Parameter	Parameter range
SO_2 emission magnitude	10–100 Tg SO_2
Injection height (plume bottom)	15–25 km
Latitude	80°S to 80°N

3-d parameter space



Considered 3 outputs in total:

We'll concentrate on 'Integrated global mean net radiative forcing'

– Ranges from -68 to -692 MJ/m^2 in the simulations
  (more negative = stronger forcing effect)

Example 1: Volcanic aerosol study – Marshall et al., 2019

(doi: 10.1029/2018JD028675)

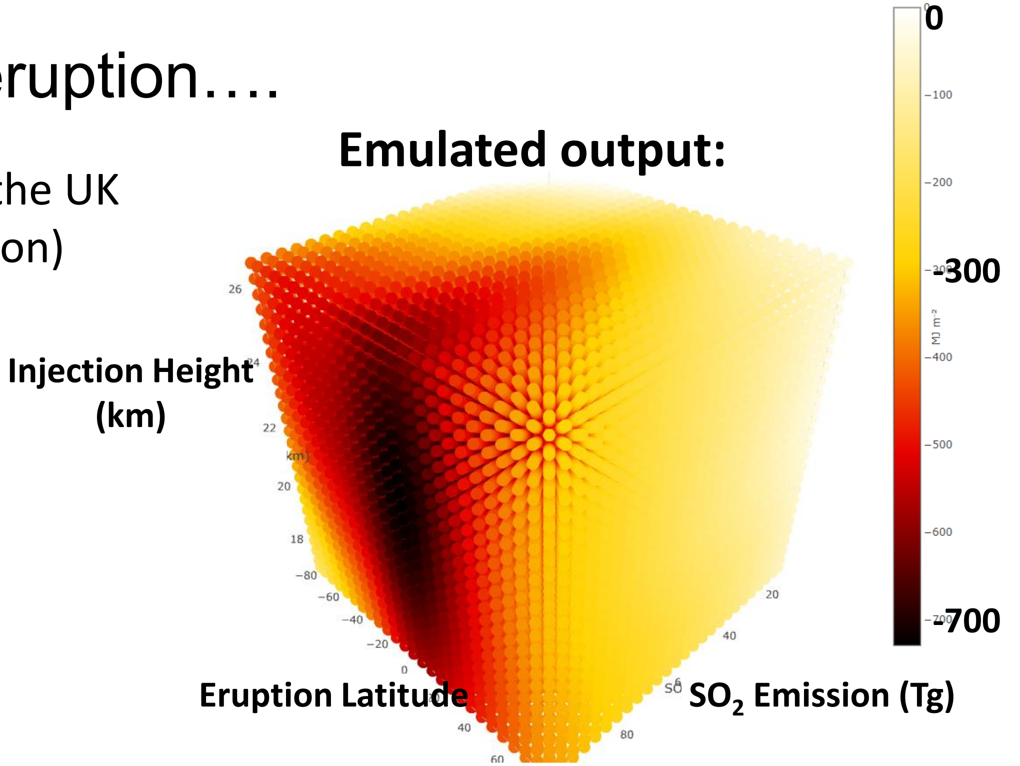
Exploring the radiative forcing caused by a volcanic eruption….

Model: UM-UKCA (Met-office general circulation model (GCM) coupled to the UK Chemistry and Aerosol scheme; Based on 'Global-Atmosphere 4' configuration)
PPE: 30 training runs; 11 validation runs

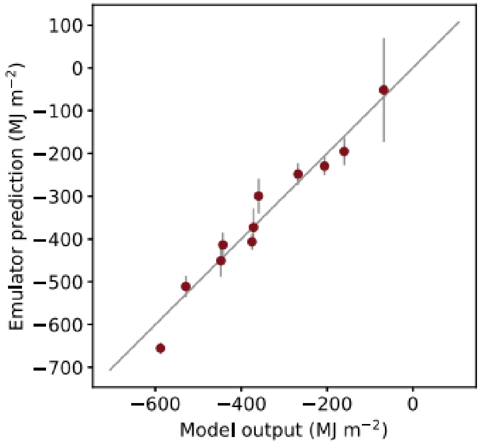
Table 1
Eruption Source Parameters and Range in Values That are Perturbed in This Study

Parameter	Parameter range
SO_2 emission magnitude	10–100 Tg SO_2
Injection height (plume bottom)	15–25 km
Latitude	80°S to 80°N

Emulated output:



Validation:



Considered 3 outputs in total:
We'll concentrate on 'Integrated global mean net radiative forcing'
- Ranges from -68 to -692 MJ/m^2 in the simulations
  (more negative = stronger forcing effect)

Example 1: Volcanic aerosol study – Marshall et al., 2019

(doi: 10.1029/2018JD028675)

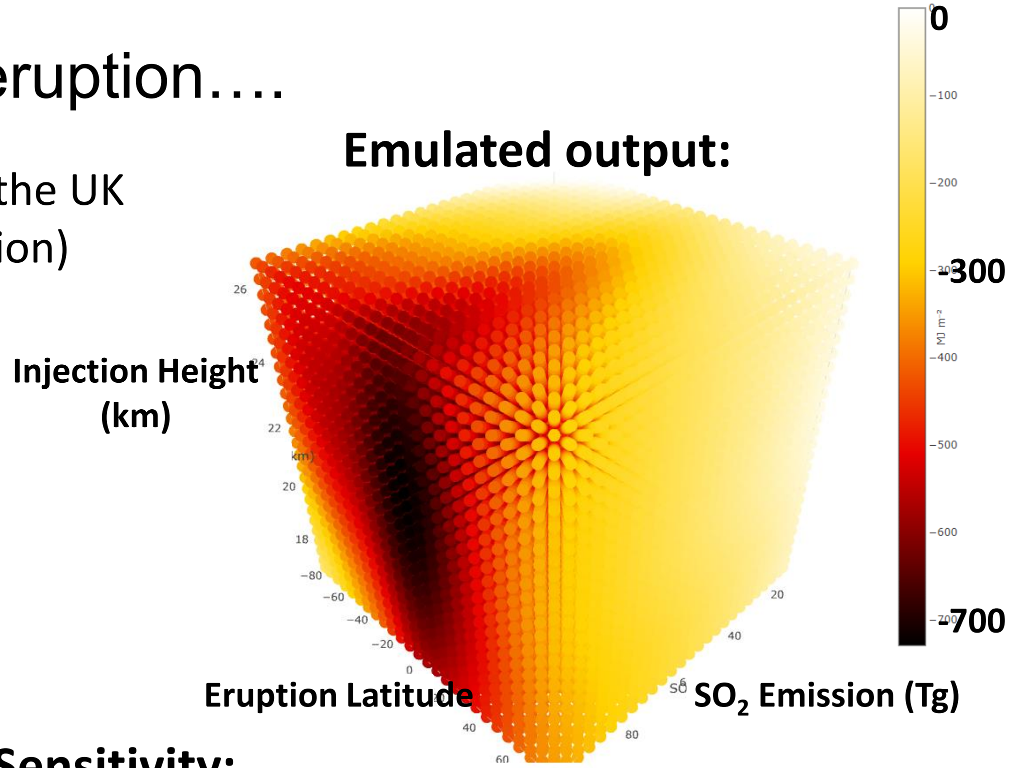
Exploring the radiative forcing caused by a volcanic eruption….

Model: UM-UKCA (Met-office general circulation model (GCM) coupled to the UK Chemistry and Aerosol scheme; Based on 'Global-Atmosphere 4' configuration)
PPE: 30 training runs; 11 validation runs

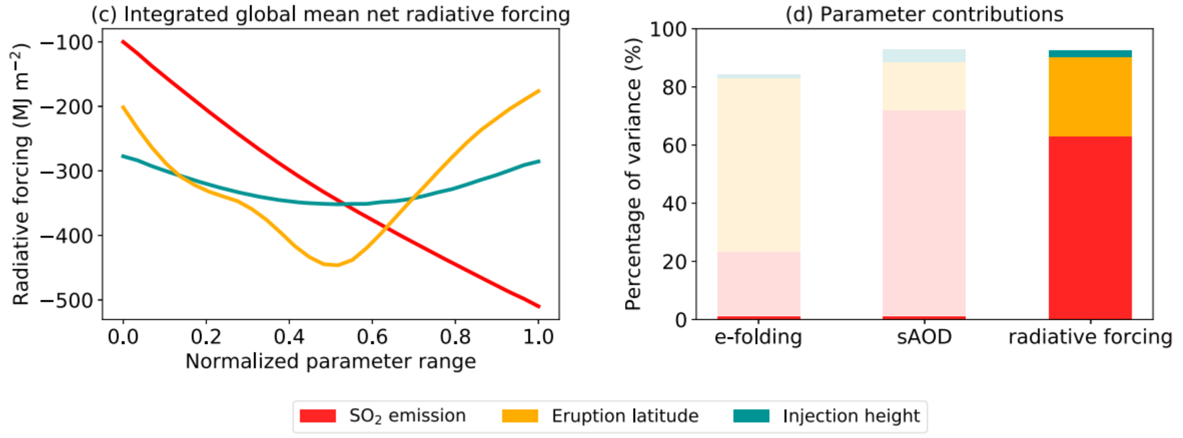
Table 1
Eruption Source Parameters and Range in Values That are Perturbed in This Study

Parameter	Parameter range
SO_2 emission magnitude	10–100 Tg SO_2
Injection height (plume bottom)	15–25 km
Latitude	80°S to

Emulated output:



Sensitivity:



Considered 3 outputs in total:

We'll concentrate on 'Integrated global mean net radiative f

- Ranges from -68 to -692 MJ/m^2 in the simulations
  (more negative = stronger forcing effect)

Example 1: Volcanic aerosol study – Marshall *et al*., 2019

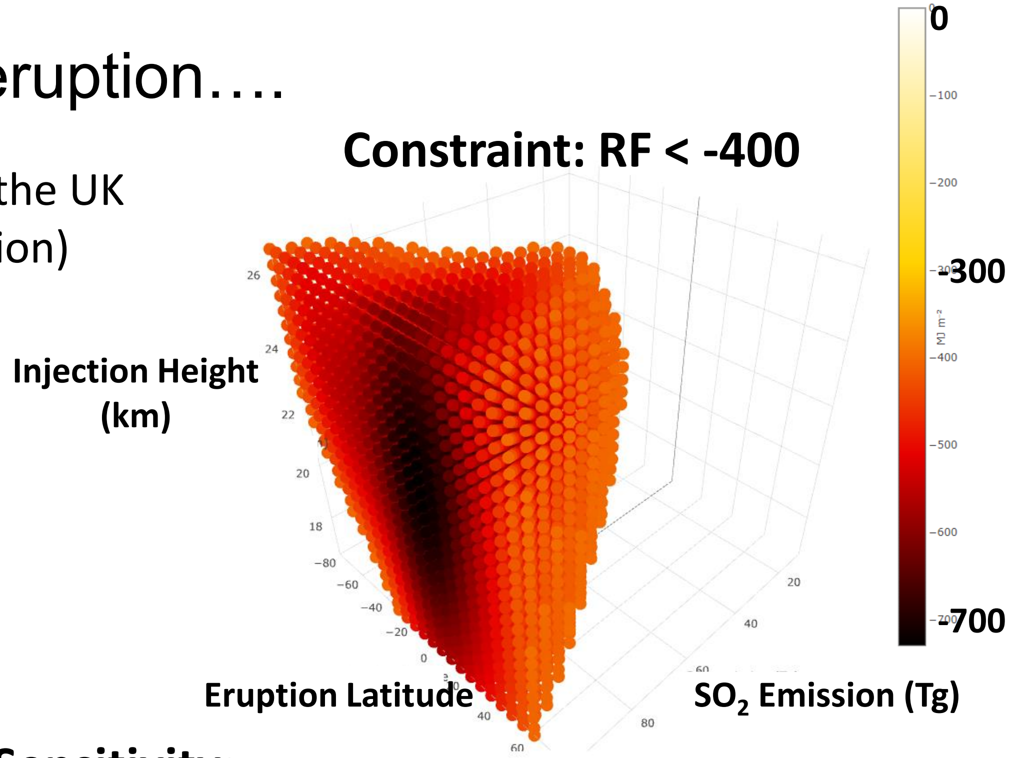
(doi: 10.1029/2018JD028675)

Exploring the radiative forcing caused by a volcanic eruption….

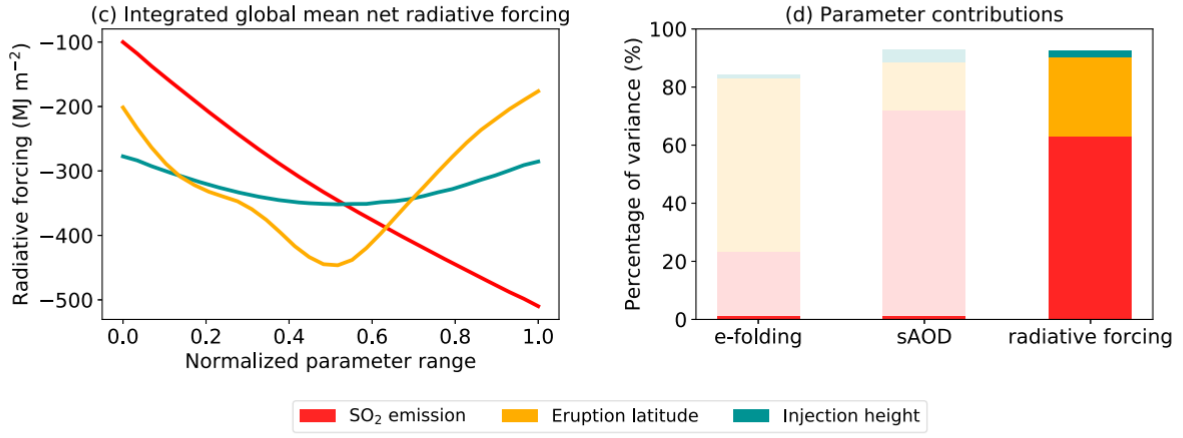
Model: UM-UKCA (Met-office general circulation model (GCM) coupled to the UK Chemistry and Aerosol scheme; Based on 'Global-Atmosphere 4' configuration)
PPE: 30 training runs; 11 validation runs

Table 1
Eruption Source Parameters and Range in Values That are Perturbed in This Study

Parameter	Parameter range
SO_2 emission magnitude	10–100 Tg SO_2
Injection height (plume bottom)	15–25 km
Latitude	80°S to



Sensitivity:



Considered 3 outputs in total:
We'll concentrate on 'Integrated global mean net radiative f
- Ranges from -68 to -692 MJ/m^2 in the simulations (more negative = stronger forcing effect)

Example 2: UK Met Office Climate Model – Johnson *et al*., 2020

(doi: 10.5194/acp-20-9491-2020)

Exploring the effects of parameter uncertainties on predictions of aerosol radiative forcing

Atmos. Chem. Phys., 20, 9491–9524, 2020
https://doi.org/10.5194/acp-20-9491-2020

Atmospheric Chemistry and Physics
Open Access

Robust observational constraint of uncertain aerosol processes and emissions in a climate model and the effect on aerosol radiative forcing

Jill S. Johnson[1], Leighton A. Regayre[1], Masaru Yoshioka[1], Kirsty J. Pringle[1], Steven T. Turnock[2], Jo Browse[3], David M. H. Sexton[2], John W. Rostron[2], Nick A. J. Schutgens[4], Daniel G. Partridge[5], Dantong Liu[6,a], James D. Allan[6,7], Hugh Coe[6], Aijun Ding[8], David D. Cohen[9], Armand Atanacio[9], Ville Vakkari[10,11], Eija Asmi[10], and Ken S. Carslaw[1]

[1]Institute for Climate and Atmospheric Science, School of Earth and Environment, University of Leeds, Leeds, UK
[2]Met Office Hadley Centre, Exeter, UK
[3]Centre for Geography and Environmental Science, University of Exeter, Penryn, UK

HadGEM3 ... one...

26 aerosol

- natu
- anth burn
- aero
- pH
- mod
- star

1850 (pre-in

1 year per p

2008 meteorology

Nudged horizontal winds and temperatures

Lee et al, 2013, ACP

Example 2: UK Met Office Climate Model – Johnson *et al*., 2020

(doi: 10.5194/acp-20-9491-2020)

Exploring the effects of parameter uncertainties on predictions of aerosol radiative forcing

HadGEM3-UKCA (vn8.4) Total simulations: 235
26 aerosol parameters and processes perturbed, including: ▪ natural emissions (e.g. Sea Spray, DMS, Volcanic, Dust) ▪ anthropogenic emissions (e.g. SO_2, Fossil fuel, Biomass burning, residential) ▪ aerosol removal properties ▪ pH of cloud droplets ▪ modal width for aerosol size (Aitken and accumulation) ▪ standard deviation of updraft velocity
1850 (pre-industrial) and 2008 (present-day) emissions
1 year per period
2008 meteorology
Nudged horizontal winds and temperatures

Several studies led to this one…

Lee et al, 2013, ACP

Example 2: UK Met Office Climate Model – Johnson *et al*., 2020

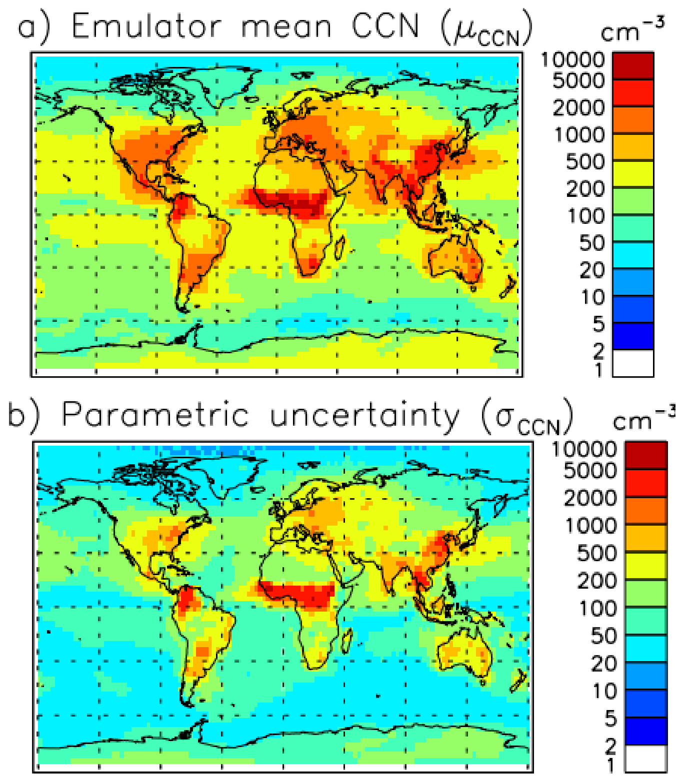
(doi: 10.5194/acp-20-9491-2020)

Exploring the effects of parameter uncertainties on predictions of aerosol radiative forcing

HadGEM3-UKCA (vn8.4)	Total simulations: 235
26 aerosol parameters and processes perturbed, including: ▪ natural emissions (e.g. Sea Spray, DMS, Volcanic, Dust) ▪ anthropogenic emissions (e.g. SO_2, Fossil fuel, Biomass burning, residential) ▪ aerosol removal properties ▪ pH of cloud droplets ▪ modal width for aerosol size (Aitken and accumulation) ▪ standard deviation of updraft velocity	
1850 (pre-industrial) and 2008 (present-day) emissions	
1 year per period	
2008 meteorology	
Nudged horizontal winds and temperatures	

Several studies led to this one…

Emulation at the 'grid-box' level for comparison to obs.



Lee et al, 2013, ACP

Example 2: UK Met Office Climate Model – Johnson *et al*., 2020

(doi: 10.5194/acp-20-9491-2020)

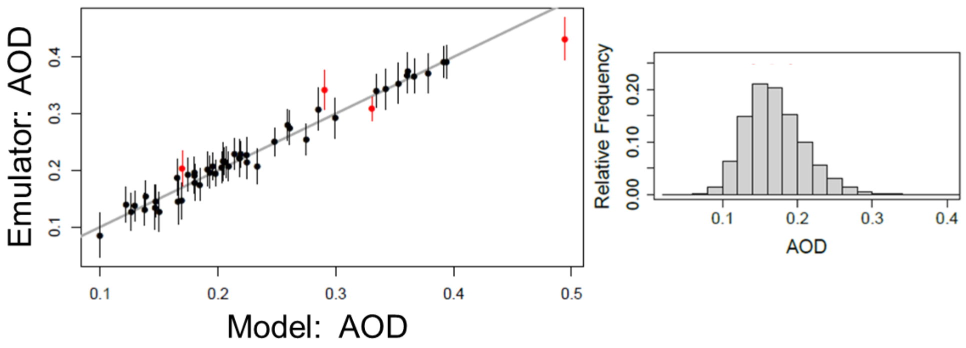
Exploring the effects of parameter uncertainties on predictions of aerosol radiative forcing

HadGEM3-UKCA (vn8.4)	Total simulations: 235
26 aerosol parameters and processes perturbed, including: ▪ natural emissions (e.g. Sea Spray, DMS, Volcanic, Dust) ▪ anthropogenic emissions (e.g. SO_2, Fossil fuel, Biomass burning, residential) ▪ aerosol removal properties ▪ pH of cloud droplets ▪ modal width for aerosol size (Aitken and accumulation) ▪ standard deviation of updraft velocity	
1850 (pre-industrial) and 2008 (present-day) emissions	
1 year per period	
2008 meteorology	
Nudged horizontal winds and temperatures	

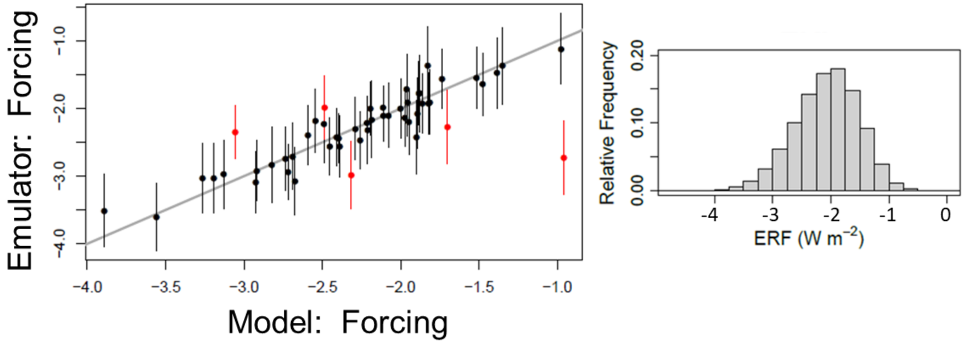
Several studies led to this one…

Emulation at the 'grid-box' level for comparison to obs.

Grid-box level AOD



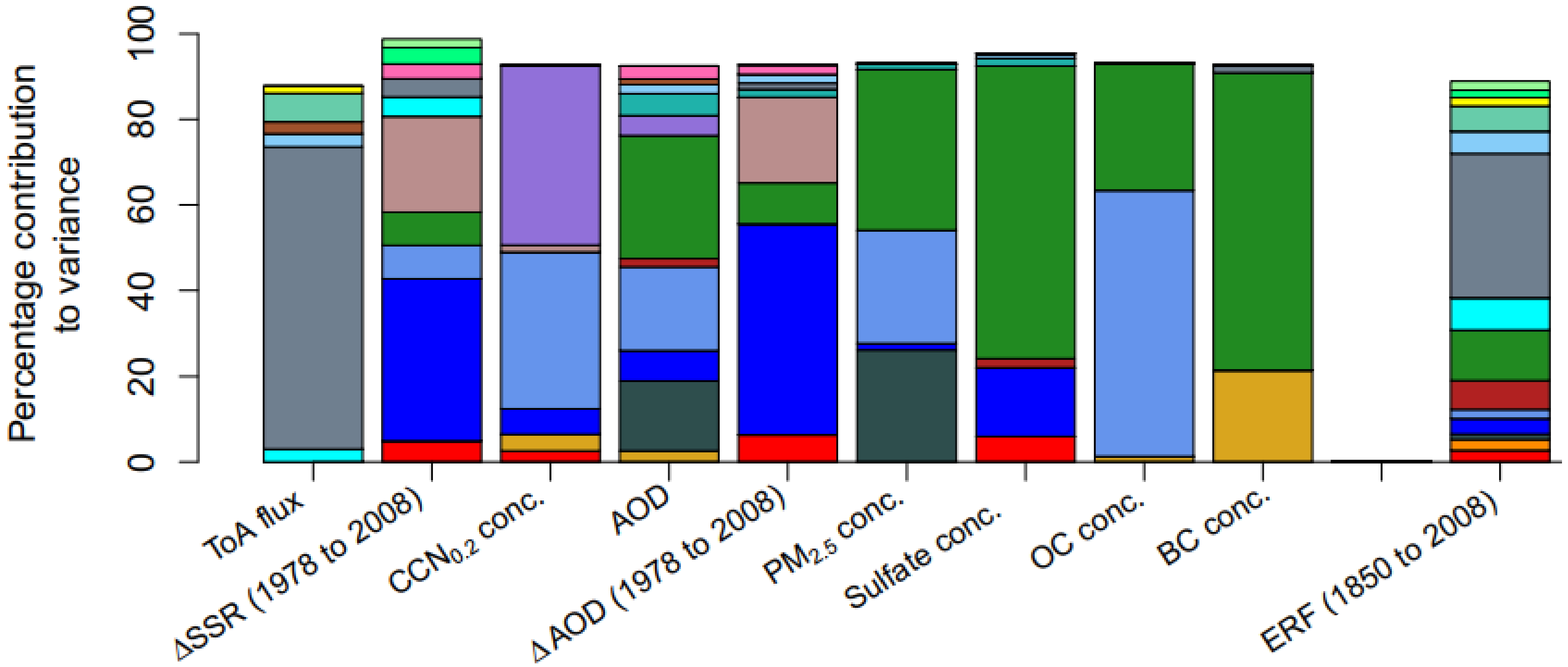
Global Mean Aerosol Radiative Forcing



Sensitivity analysis

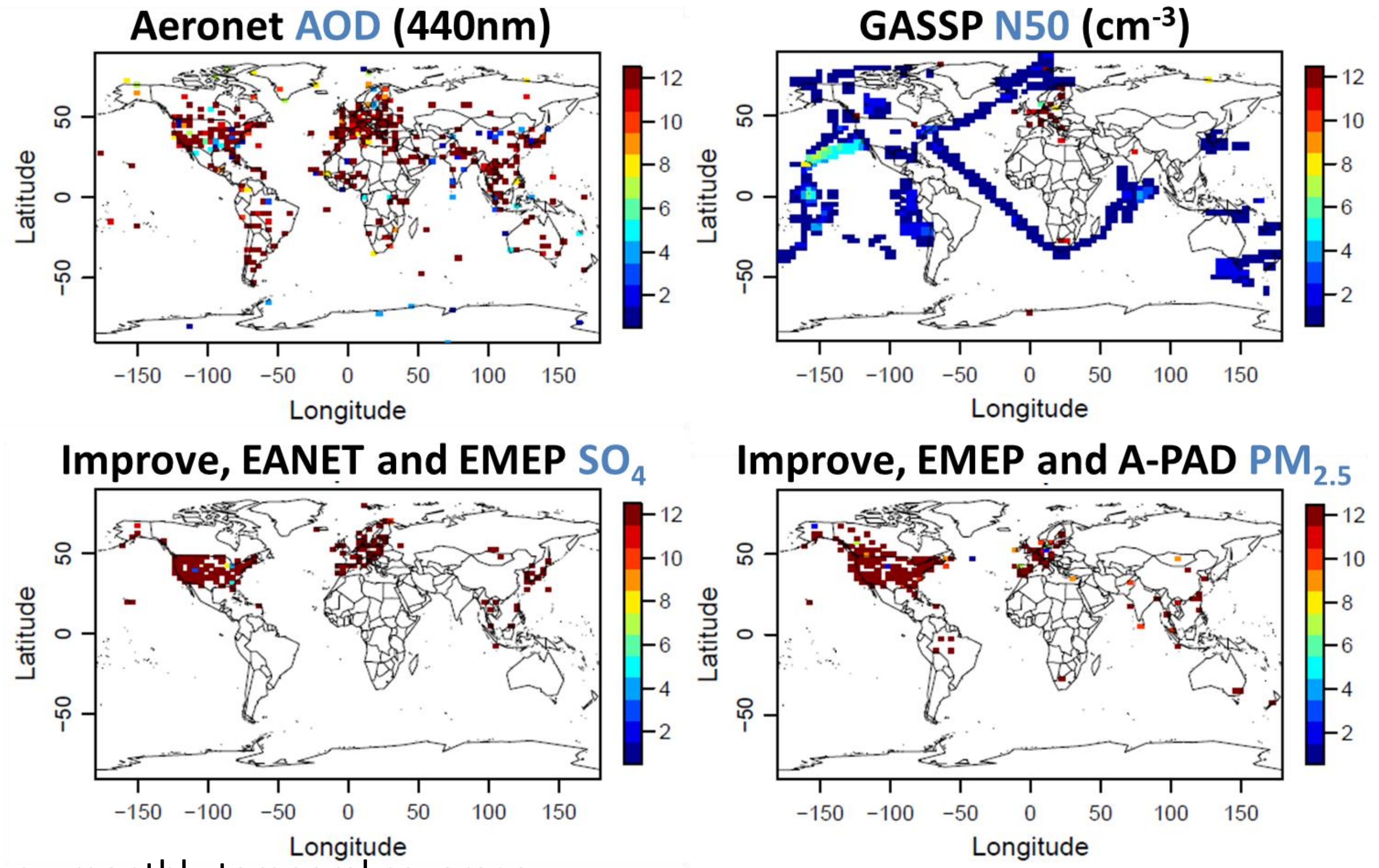
We can use sensitivity analysis to determine common causes of uncertainty between observable quantities and aerosol radiative forcing:

[European averages; **Johnson *et al*., 2018, ACP**]



Aerosol observations

An extensive set of aerosol observations was used to constrain the model's uncertainty… and so our uncertainty in predictions of aerosol radiative forcing…



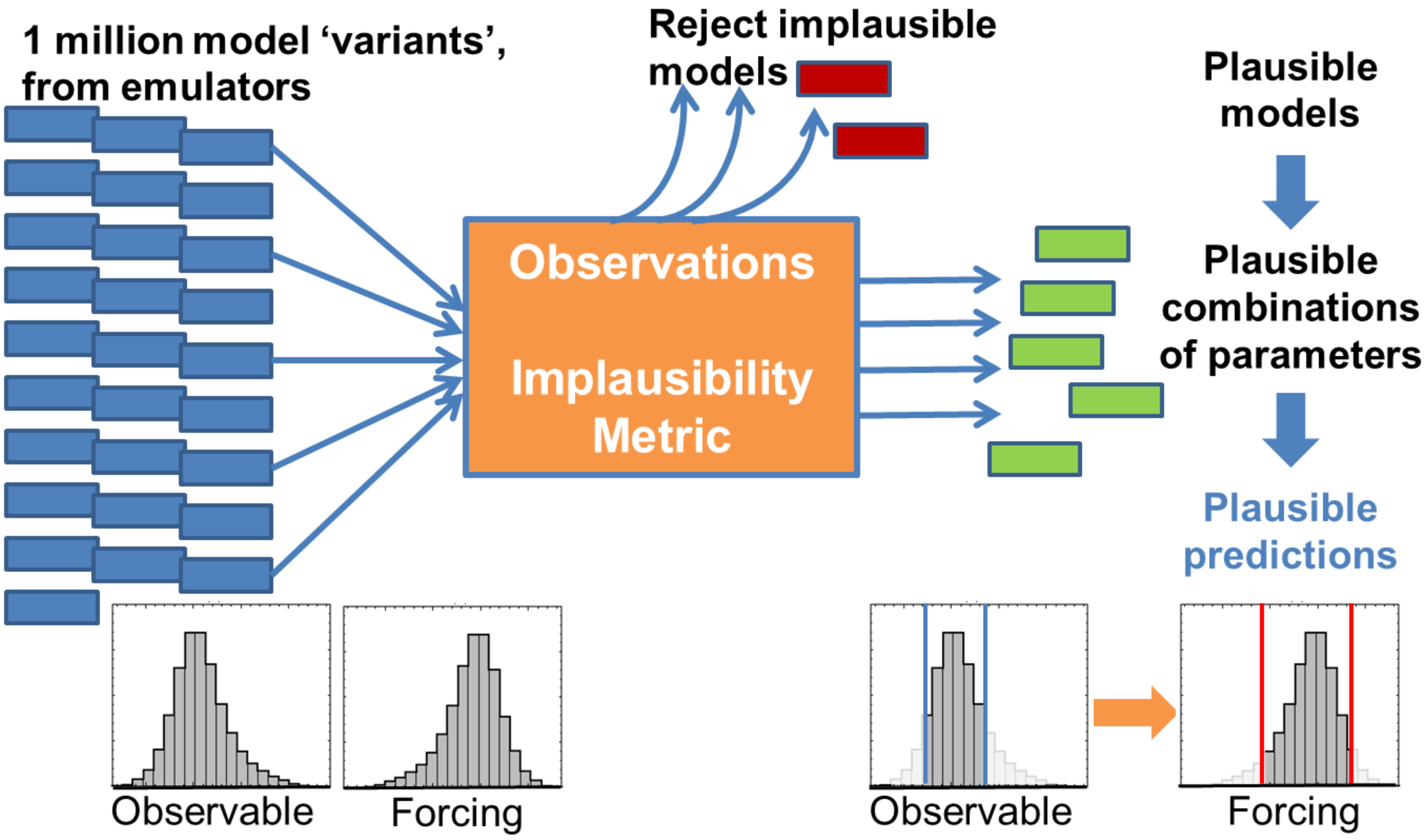
colour = monthly temporal coverage

- Varied global coverage: Spatially/temporally sparse
- Data from large networks (e.g. AERONET)
- Data from ship and Aircraft campaigns
- **9000+** observations

Johnson et al. (2020)

A ‘History Matching’ approach to reduce uncertainty

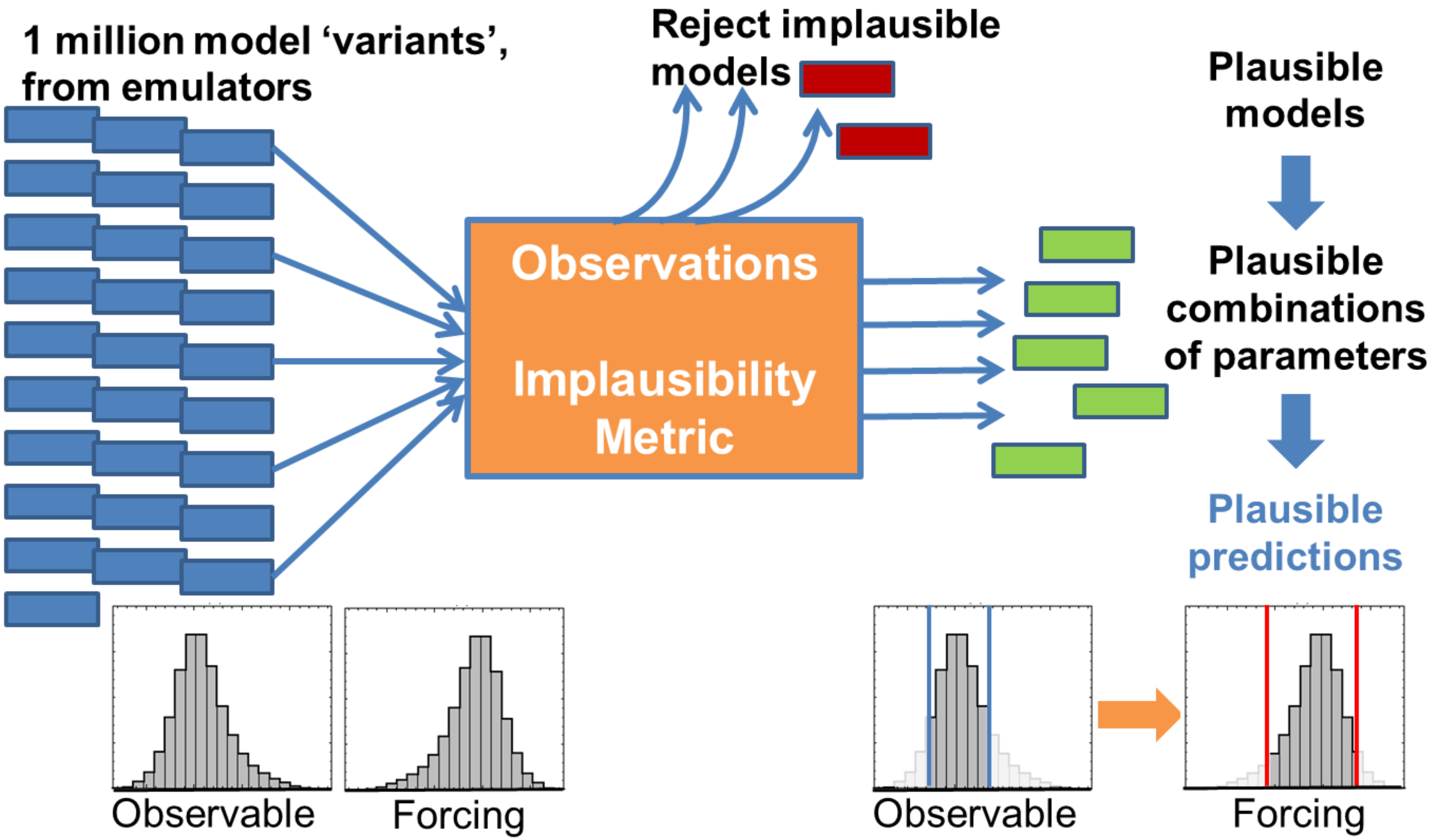
History Matching (Craig et al., 1996) rules out regions of parameter space that are not consistent with observations using an implausibility metric:



Johnson et al. (2018, 2020)

A 'History Matching' approach to reduce uncertainty

History Matching (Craig et al., 1996) rules out regions of parameter space that are not consistent with observations using an implausibility metric:



Approx. scale of the analysis:

Model
- **26 perturbed parameters**
- **1 million model variants**

Observations
- **9000 in situ measurements (AOD, particle number, N_{50}, $PM_{2.5}$, SO_4, OC)**

Derivation of an implausibility metric to rule out poor models on comparison to real aerosol measurements

Johnson et al. (2020): 1 million model variants, compared to **9000+ gridded observations** using an **implausibility metric** → Accounting for all uncertainties in the comparison process

Model output [Prediction from emulator]

Observation

Implausibility

$$I = \frac{|O - M|}{\sqrt{[Var(M) + Var(O) + Var(R) + Var(S)]}}$$

Emulator prediction uncertainty

Measurement uncertainty

Representation uncertainty

Structural uncertainty

Smaller I implies a variant is more plausible w.r.t. the observation

For **a single aerosol property** in a particular month, we **rule out variants** if I is large for >T observations

Joint Constraint: Rule out a variant if it is ruled out for **ANY** individual month/observation type

Spatial and temporal differences in resolution between the observations and the model

(Schutgens et al., 2016)

Components of representation uncertainty

- Spatial co-location, $\boldsymbol{R_{SP}}$
  - Comparing point measurements with the model grid
  - Where in the grid-box (central / edge) the observation lies

(Schutgens et al., 2016a)

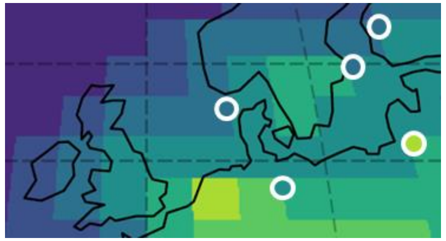
- Temporal co-location, $\boldsymbol{R_{T}}$
  - Comparing campaign data (measured over a few hours/days) to monthly mean model output

(Schutgens et al., 2016b)

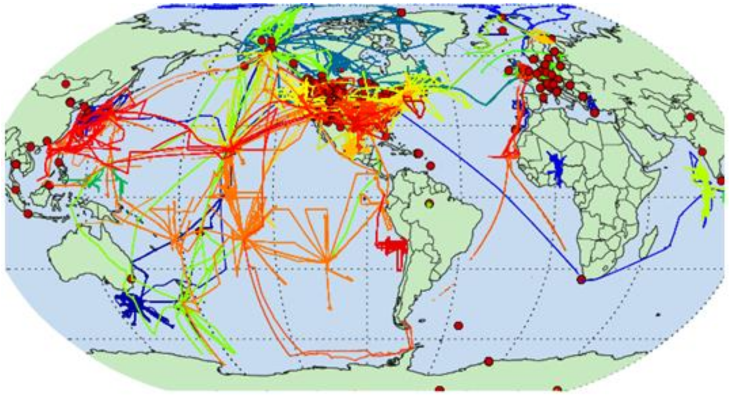
- Inter-annual variability, $\boldsymbol{R_{IAV}}$
  - Campaigns are ‘one-off’ studies
  - Comparing observations taken in a particular year to model output of a different year

These uncertainties can vary between different aerosol properties

Model v's Observations resolution

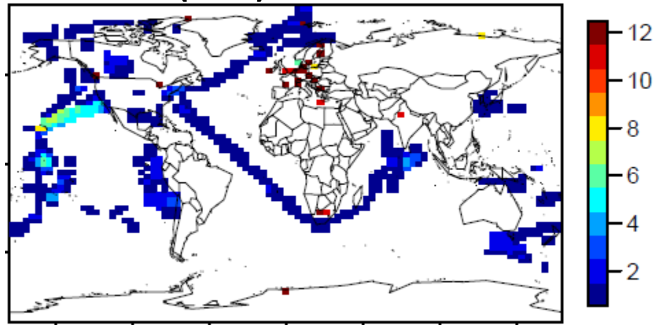


GASSP observations



Reddington et al., 2016

GASSP N50 (cm^{-3}) at model resolution



Quantifying the uncertainty terms

Var(O): **Instrument measurement uncertainty**

- Information that observations are measured to an accuracy of within +/- p % of the true value
- Assume Gaussian approximation for uncertainty (+/- p% = +/- 2σ)

$\boldsymbol{p}$ **= 10%**

Var(M): **Emulator uncertainty**

- Comes from the fitted emulator model for each prediction

Var(R_{SP}) **and** ***Var(R_T)***: **Spatial/temporal co-location**

- Similarly to *Var(O)*: % error on observed value, using information from Schutgens et al. (2016)

$\boldsymbol{p}$ **= 20%** and $\boldsymbol{p}$ **= 10%**

Var(S): **Structural uncertainty**

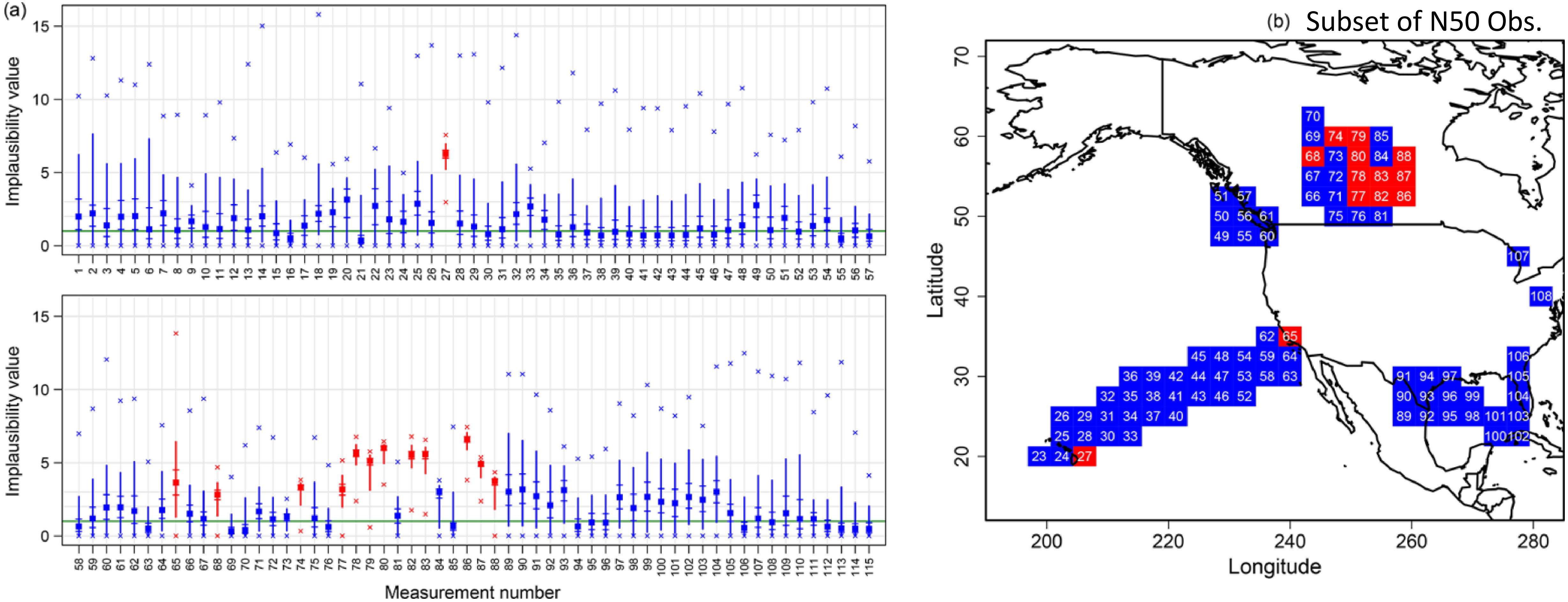
- We assume **NO** structural uncertainty term [$\boldsymbol{Var}(\boldsymbol{S})$=0].
- We allow the implausibility measure to inform us about any potential structural errors.

Var(R_{IAV}): **Inter-annual variability**

- Estimated from an analysis of the trend and variation of gridded aerosol properties in a UKCA hindcast simulation over the period of 1980–2009 (Turnock et al., 2015)

Identifying observations that do not compare well

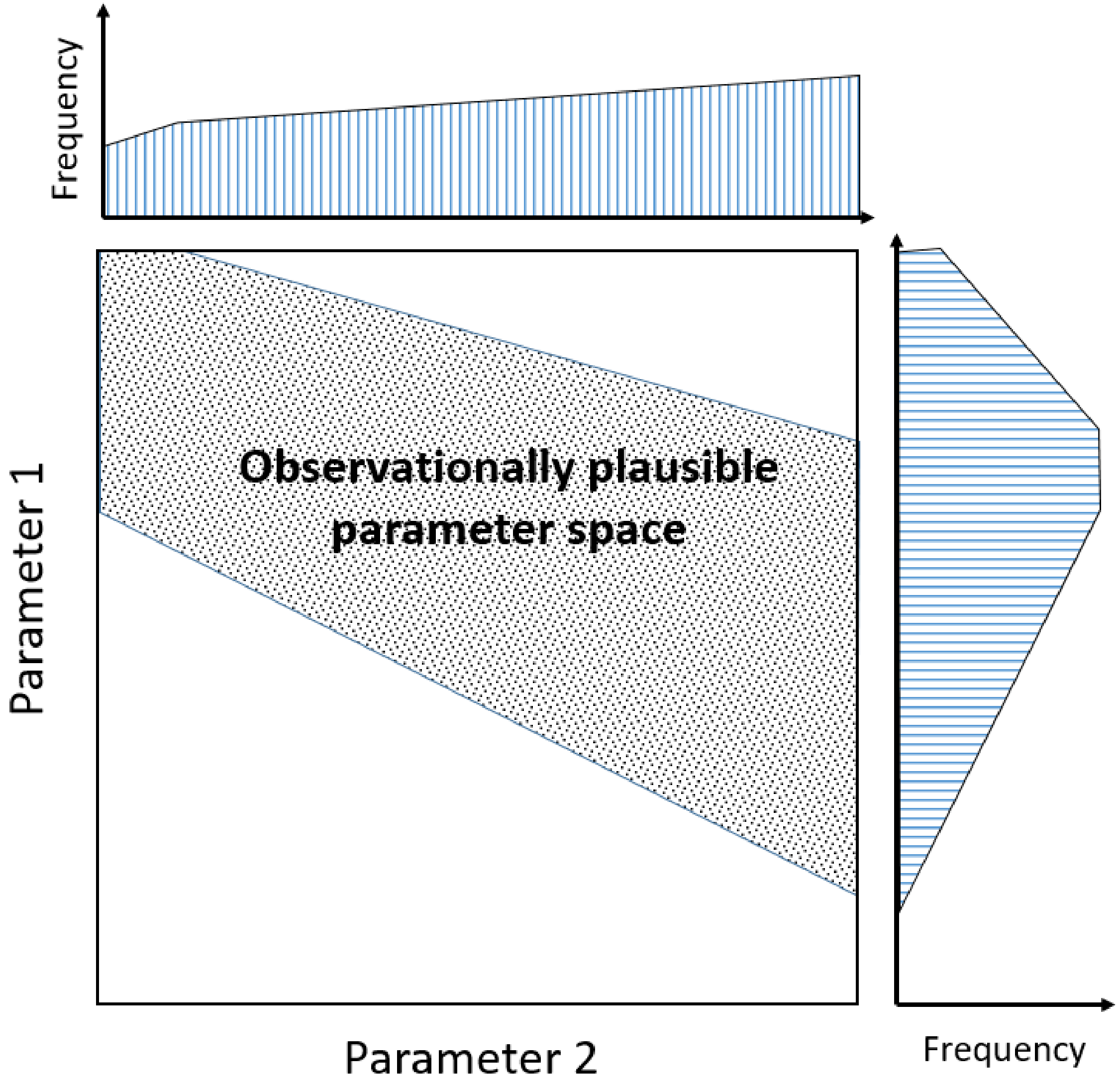
We remove observations if the lower 95% credible interval bound on *I* (across variants) is >1



It can be difficult to pin-point the cause: Are the mis-matches due to **representation errors**?
Or, are they **indicators of structural errors** in the model?

Fig 4, Johnson et al. (2020)

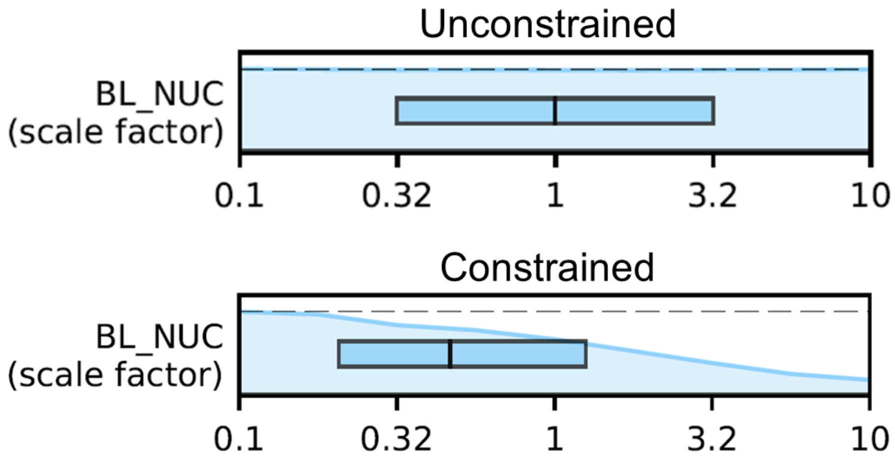
Parameter Constraint



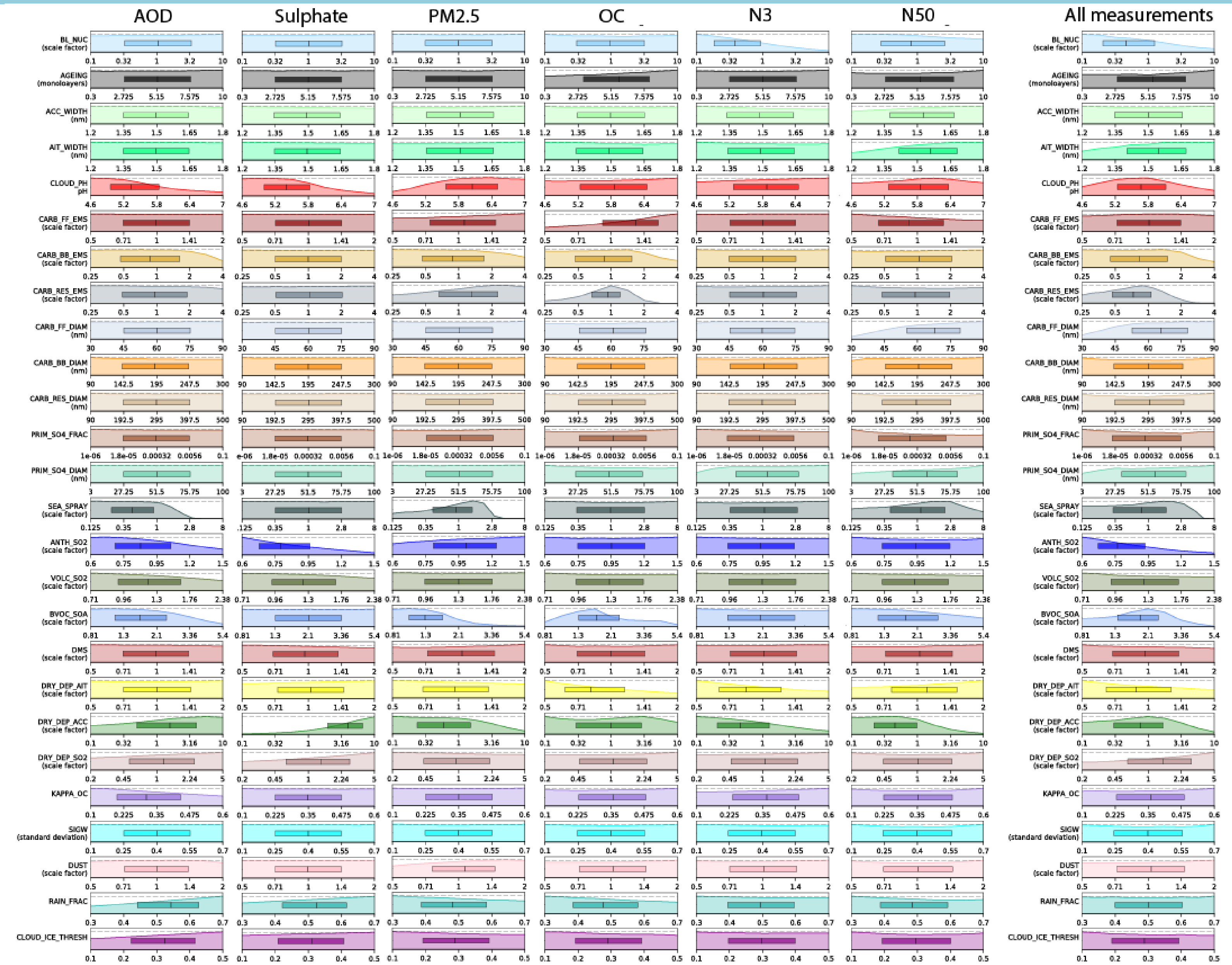
It is only possible to constrain joint parameter distributions (in 26 dimensions)

We show **marginal distributions**

Example: Marginal distribution of constrained boundary layer nucleation rate using all measurements



Results: Marginal parameter constraints from constraint with individual variables and the joint constraint



We **rule out model variants for each variable** (columns) and combine (last column) to **quantify the effect of the constraint on parameter values**

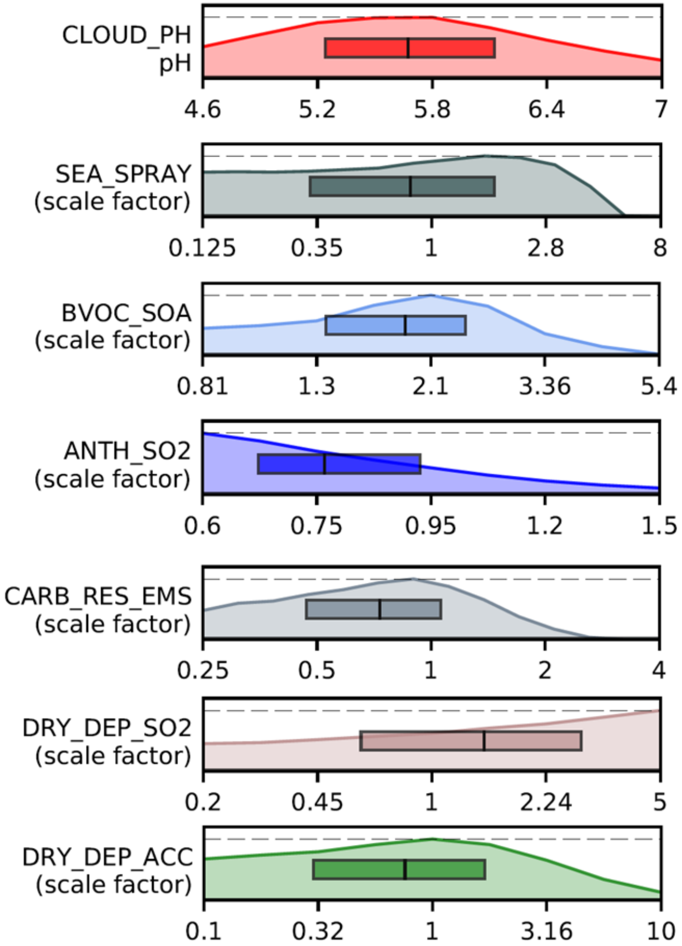
These marginal parameter constraint plots show **where parameter values are more / less likely within the constrained samples**

Johnson et al. (2020)

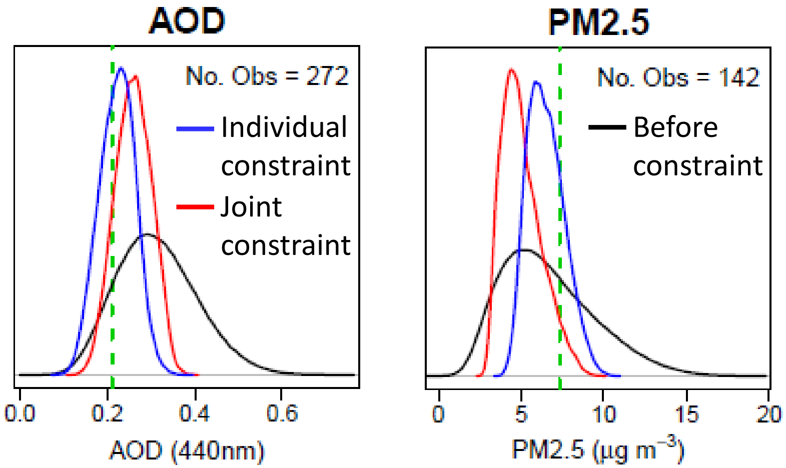
Results: Joint constraint effect from using observations of multiple aerosol variables

1 million model variants, compared to **9000+ gridded observations**

Constraint on parameters

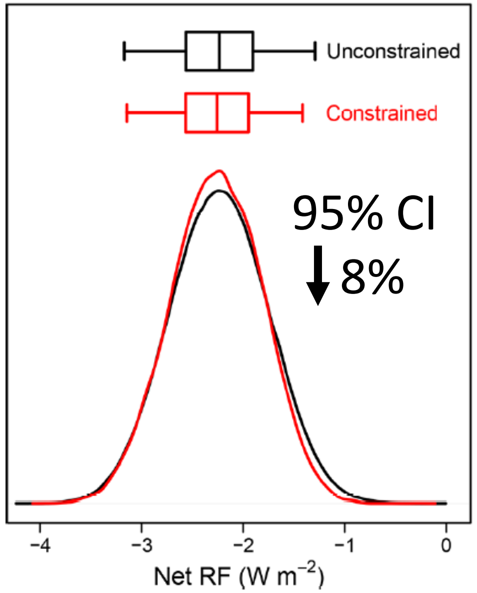


Constraint on observables



Around **2%** of variants are **retained**

Constraint on forcing: Net RF



Constraint achieved on aerosol forcing is **weak**

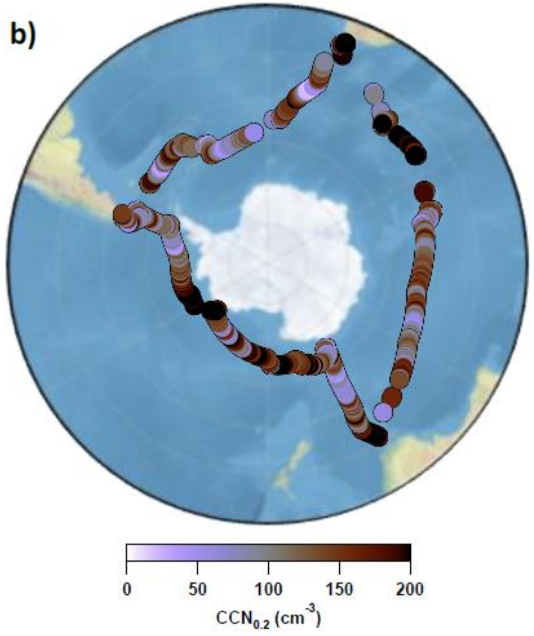
[the effect of compensating errors – '**equifinality**']

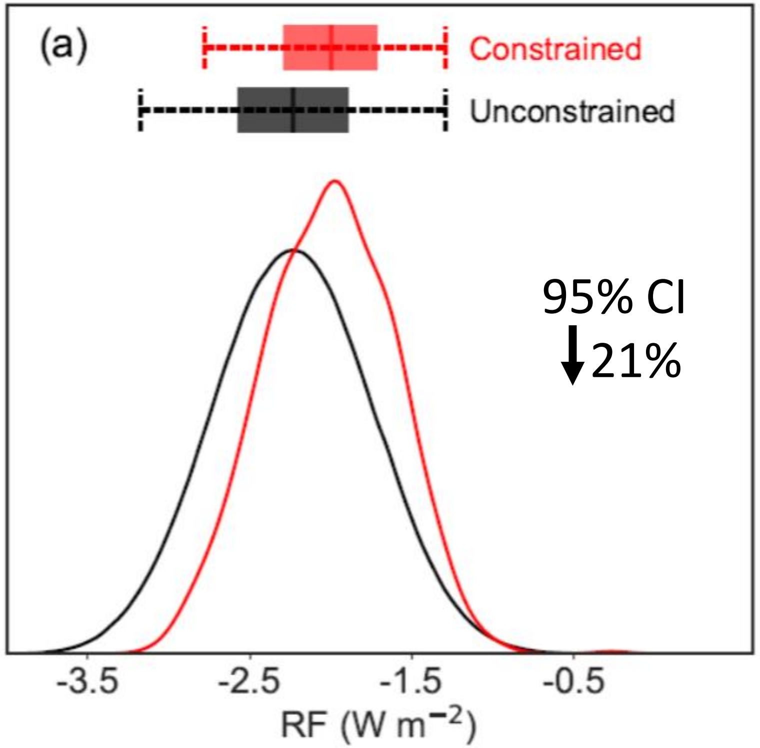
Johnson et al. (2020)

Results: Joint constraint effect from using observations of multiple aerosol variables – improved with targeted observations

1 million model variants, compared to **9000+ gridded observations**

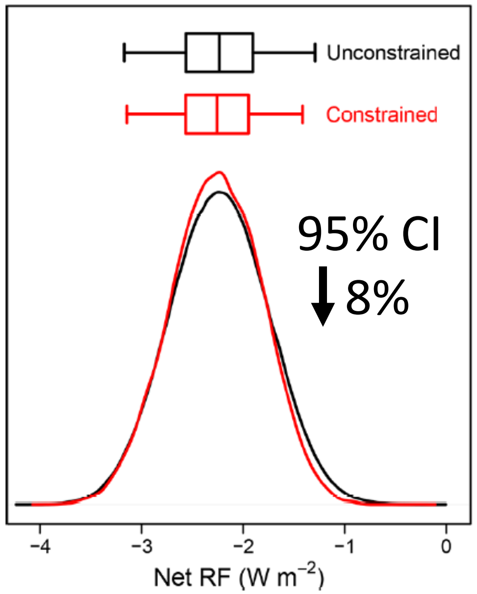
Extension of the constraint using **additional targeted observations** over the Southern Ocean from the ACE-SPACE campaign (Dec 2016 – Mar 2017)





Regayre et al. (2020)

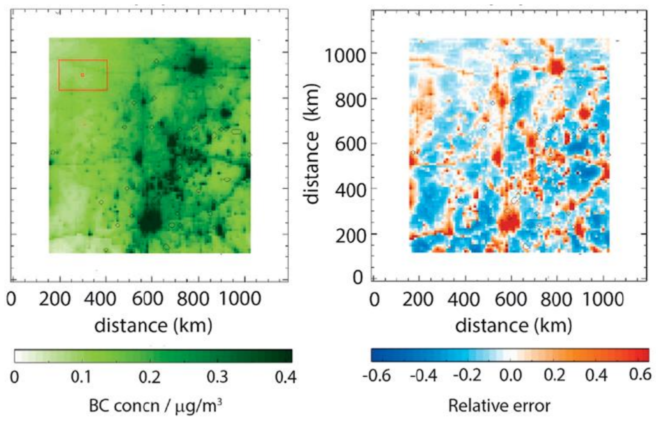
Constraint on forcing: Net RF



The constraint on aerosol forcing is improved but still relatively **weak**

Our work highlights several key challenges in the model-observation comparison process

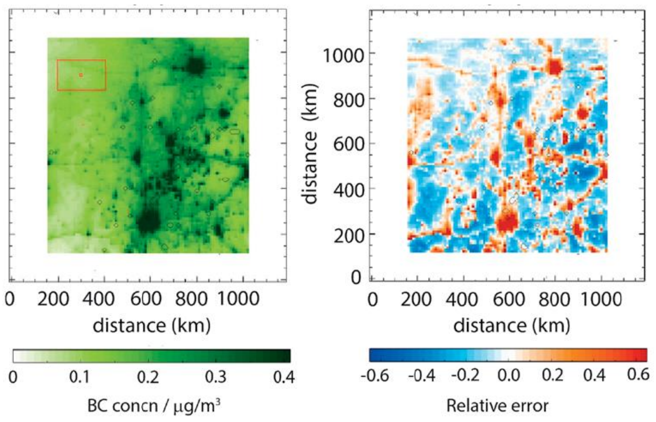
Some key statistical challenges to address in future research



[Reddington et al. (2017); Fig. 5]

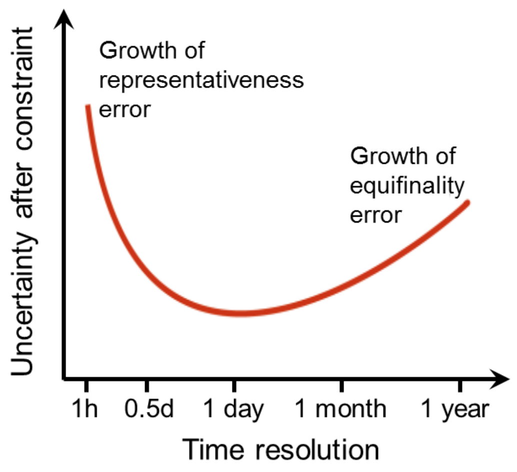
Improve the accuracy and realism of spatial and temporal representation error estimates

Some key statistical challenges to address in future research



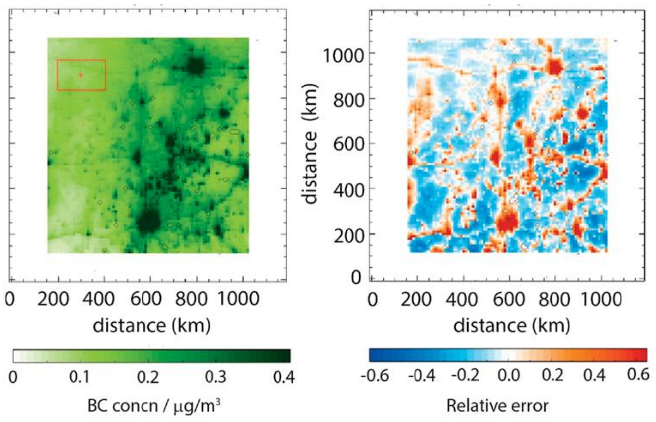
[Reddington et al. (2017); Fig. 5]

Improve the accuracy and realism of spatial and temporal representation error estimates



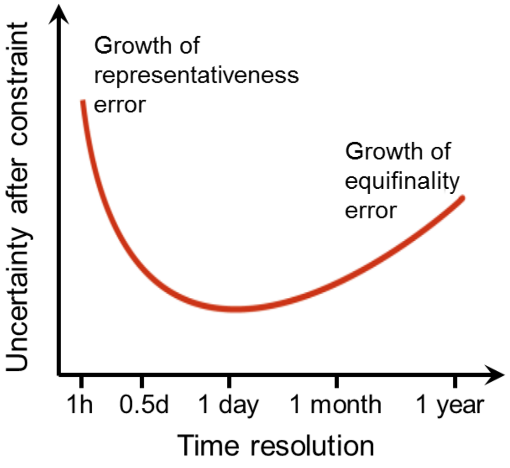
To explore the optimal resolution for model-observation comparisons [Reduce ‘equifinality’]

Some key statistical challenges to address in future research

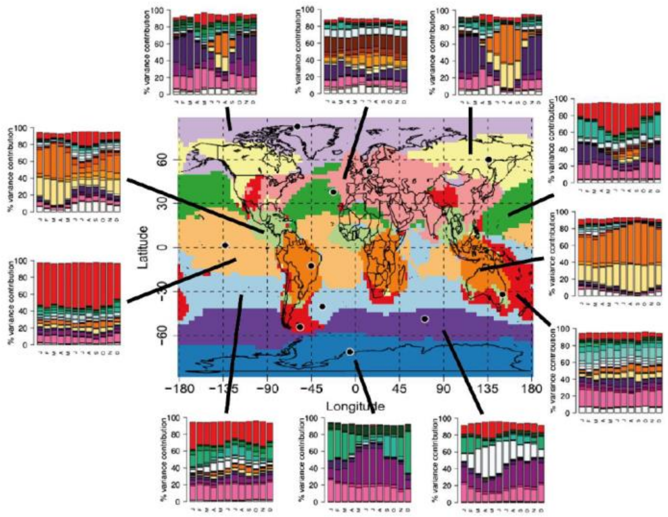


[Reddington et al. (2017); Fig. 5]

Improve the accuracy and realism of spatial and temporal representation error estimates



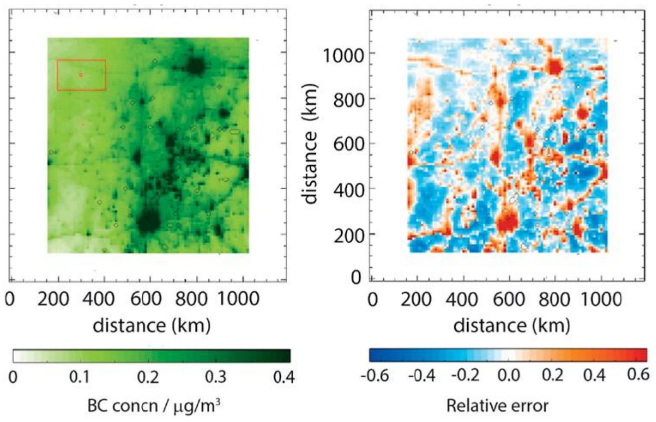
To explore the optimal resolution for model-observation comparisons [Reduce 'equifinality']



Utilise key patterns, regimes and relationships in model errors and observations as metrics for constraint

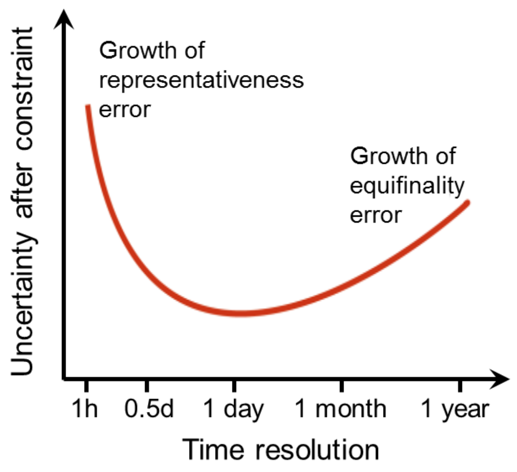
[Reddington et al. (2017)]

Some key statistical challenges to address in future research

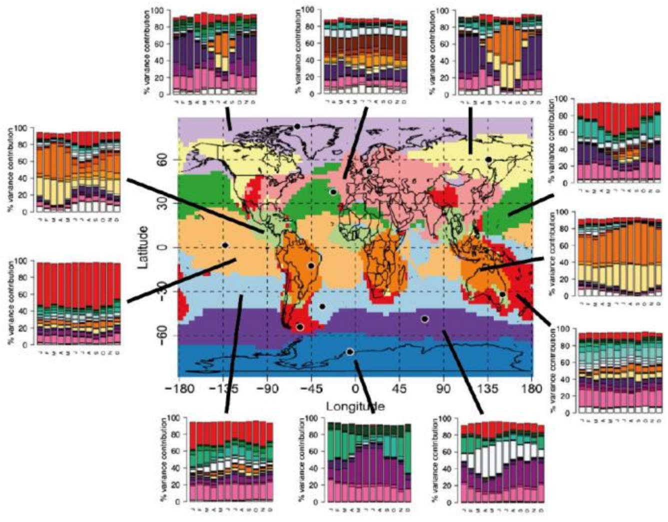


[Reddington et al. (2017); Fig. 5]

Improve the accuracy and realism of spatial and temporal representation error estimates

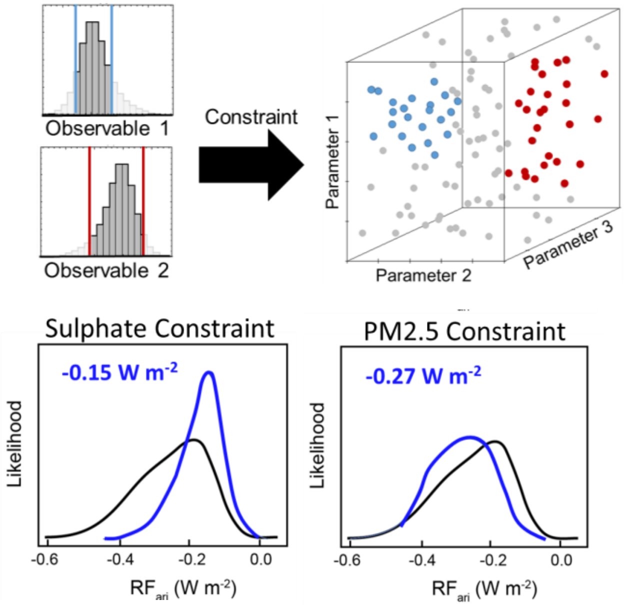


To explore the optimal resolution for model-observation comparisons [Reduce 'equifinality']



[Reddington et al. (2017)]

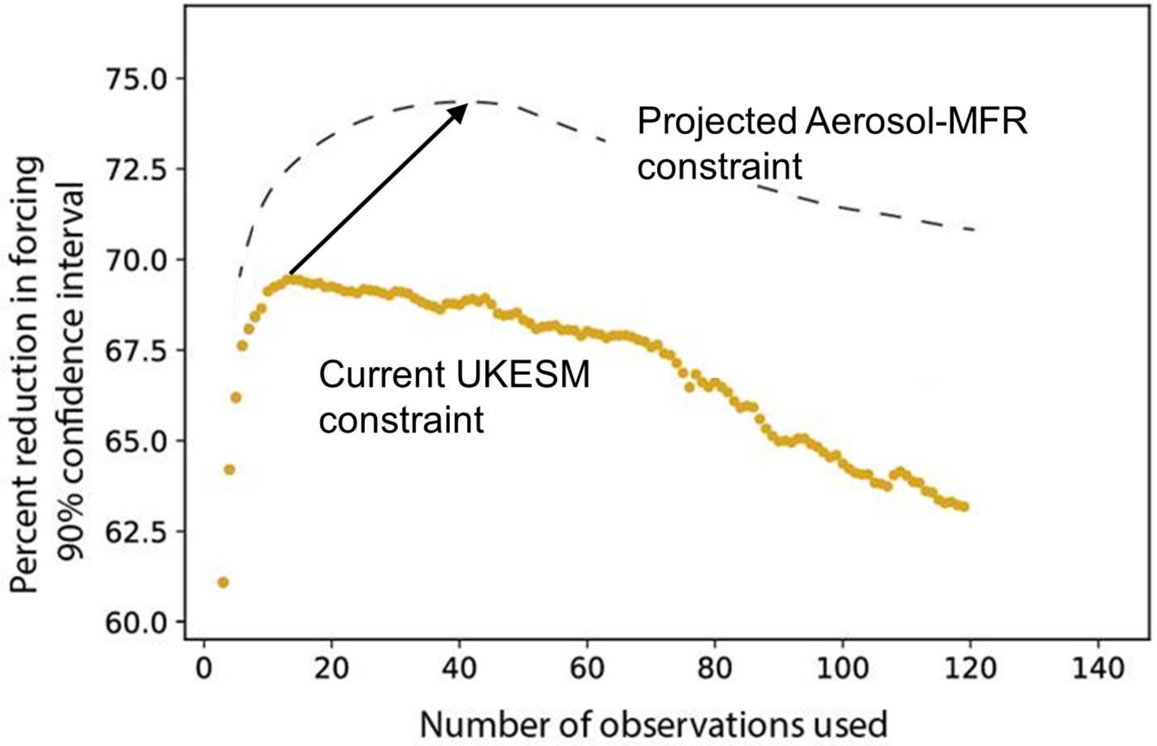
Utilise key patterns, regimes and relationships in model errors and observations as metrics for constraint



Can we exploit the dense sampling to identify and address structural model errors, to improve model performance?

Next steps – tackling some of these challenges in ‘Aerosol-MFR’

Aerosol-MFR – Towards Maximum Feasible Reduction in Aerosol Forcing Unc.



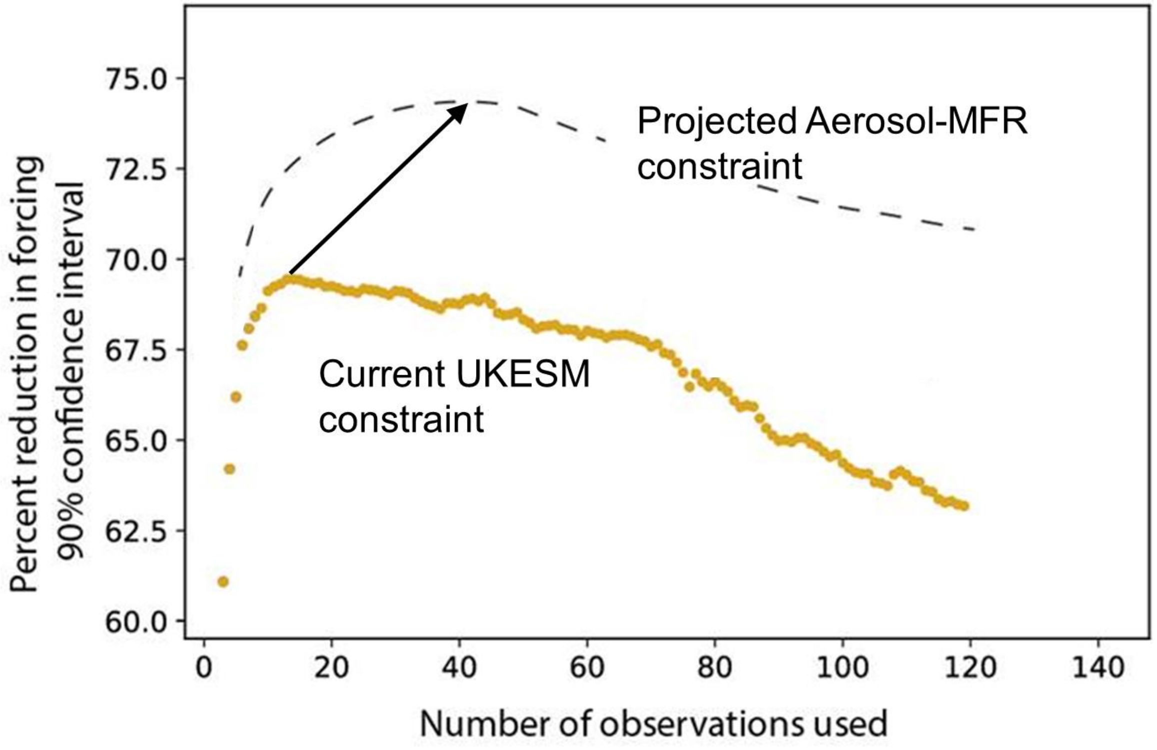
Recent work has shown that the forcing **uncertainty range can be reduced** as more observations are used to constrain the model processes – **but only up to a point**.

- **Beyond ~15 this constraint weakens** – we’ve found that the model cannot match more observations than this simultaneously.

Regayre et al, (2023) [ACP]

Next steps – tackling some of these challenges in ‘Aerosol-MFR’

Aerosol-MFR – Towards Maximum Feasible Reduction in Aerosol Forcing Unc.



Regayre et al, (2023), ACP

Recent work has shown that the forcing **uncertainty range can be reduced** as more observations are used to constrain the model processes – **but only up to a point**.

- **Beyond ~15 this constraint weakens** – we’ve found that the model cannot match more observations than this simultaneously.

We hypothesise (hope!) that the uncertainty in aerosol radiative forcing will be significantly reduced in this project!

– Aims: To tackle structural errors and the effects of parametric uncertainty (particularly, compensating errors) together within a single framework…

Challenges with our emulation approach…
Example 3 - Uncertainty in Modelling a Cloud Field

Exploring properties of a cloud field: Stratocumulus to cumulus transition

[Sansom et al. (2023), In Prep]

Here, natural variability affects the simulation output.

- Cloud properties are sensitive to any small variation in initial conditions.
- Each training simulation is just 1 possible cloud state for the selected parameter settings.

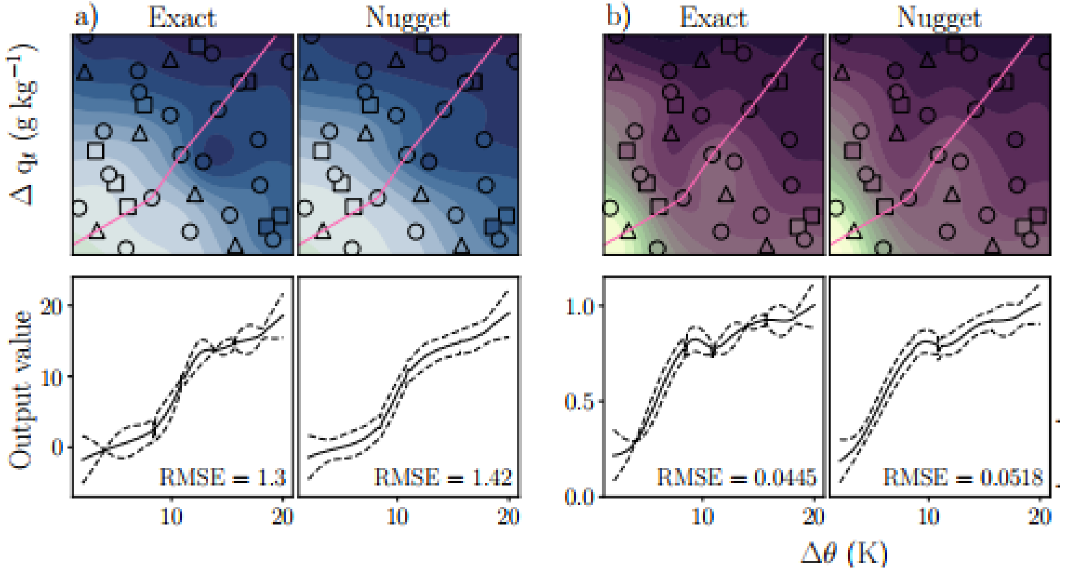
Challenges with our emulation approach…
Example 3 - Uncertainty in Modelling a Cloud Field

Exploring properties of a cloud field: Stratocumulus to cumulus transition

[Sansom et al. (2023), In Prep]

Here, natural variability affects the simulation output.

- Cloud properties are sensitive to any small variation in initial conditions.
- Each training simulation is just 1 possible cloud state for the selected parameter settings.



The GP emulator needs **a nugget term** in the covariance function:

$$k(X_p, X_q) = Cov\left(g(X_p), g(X_q)\right)$$

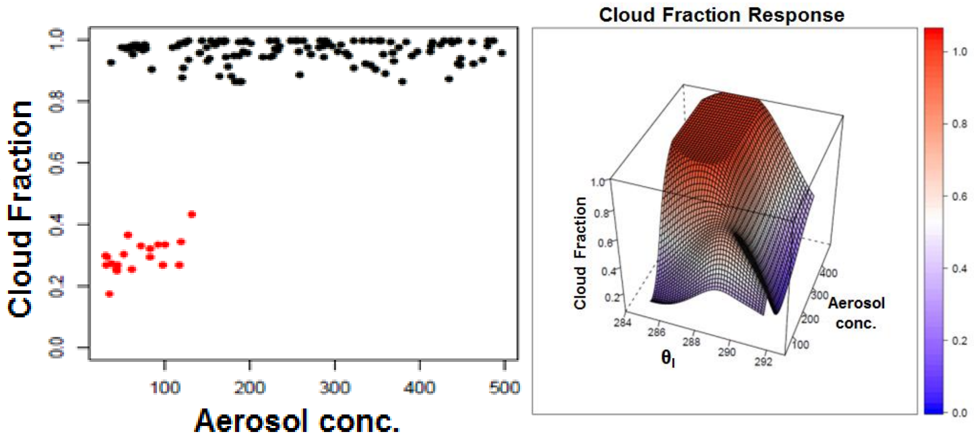
$$= \sigma_f^2 \prod_{d=1}^{D} \left[exp\left\{-\eta_d |X_{pd} - X_{qd}|^2\right\}\right] + \sigma_n^2 \delta_{pq}$$

How can we estimate it?

Challenges with our emulation approach…
Example 3: we must be careful – the emulation doesn't always work!

Properties of a cloud field are a prime example of non-stationary behaviour in the natural world (NOAA collaboration)

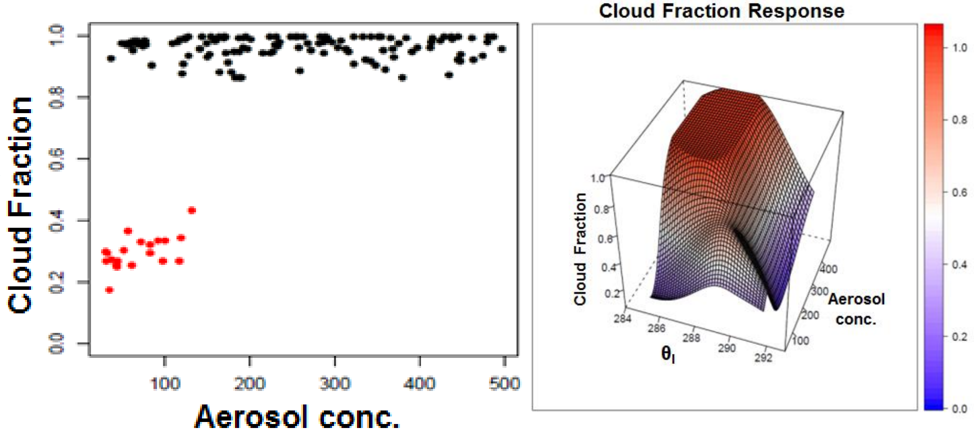
Sharp and distinct changes in cloud fraction response

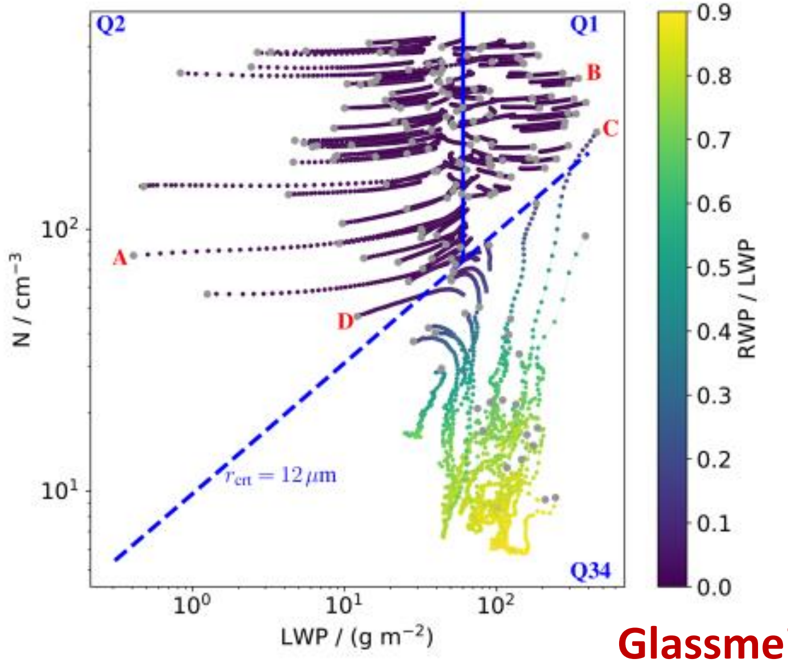


Challenges with our emulation approach…
Example 3: we must be careful – the emulation doesn't always work!

Properties of a cloud field are a prime example of non-stationary behaviour in the natural world (NOAA collaboration)

Sharp and distinct changes in cloud fraction response





In this study, 6 'initial condition' parameters were perturbed.

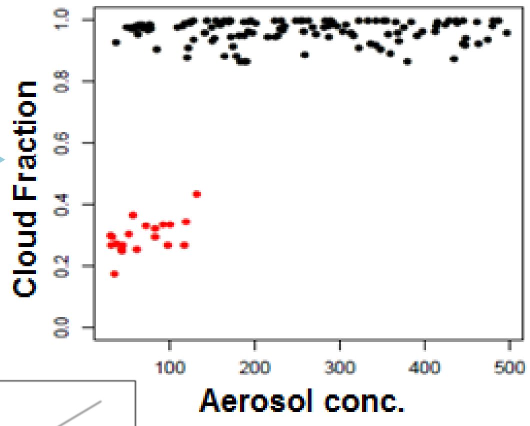
We need to be careful how we select our outputs for emulation – The clouds evolve through time at different rates…

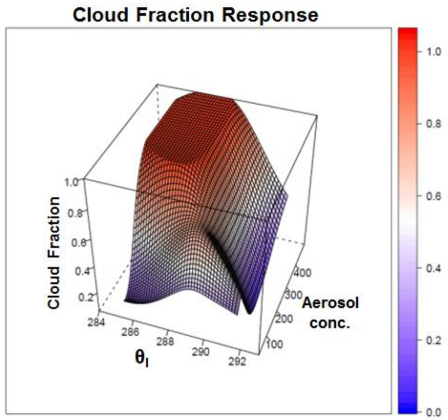
Glassmeier et al., 2019, ACP

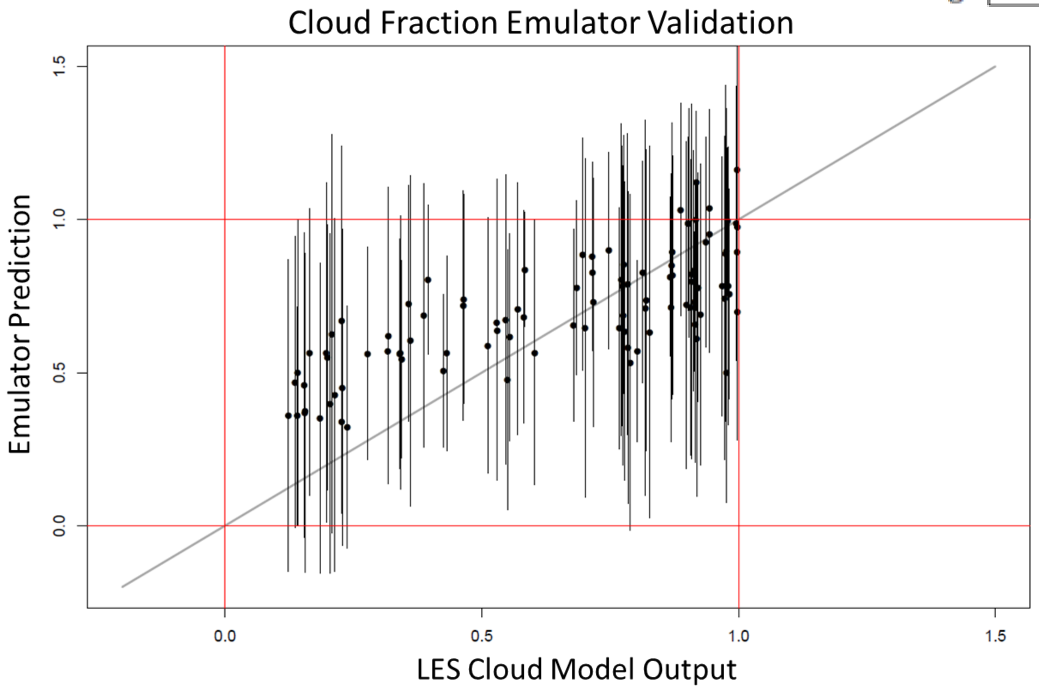
Challenges with our emulation approach…
Example 3: we must be careful – the emulation doesn't always work!

Properties of a cloud field are a prime example of non-stationary behaviour in the natural world (NOAA collaboration)

Sharp and distinct changes in cloud fraction response

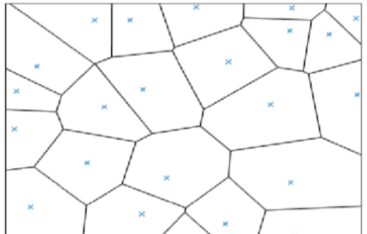






– We need a more complex way to approximate the output response surface (e.g. Treed GP [Gramacy & Lee, 2008]; use of Voronoi tessellations [Pope *et al*., 2021]; mixture covariance [Volodina & Williamson, 2018])

– It's an active area of research! ☺



10.1137/15M1049166

Summary

- **Complex models of natural systems** like the atmosphere & climate **are inherently uncertain**.
- It’s important to quantify uncertainty in model predictions, in order to have confidence in them.
- To fully understand a system’s behaviour, we must **densely sample** it over the key input uncertainties – the statistical framework enables this via **Gaussian Process emulation**.
- Once quantified, we can use real-world **observations** to try and reduce output uncertainties via ‘**History Matching**’.

Summary

- **Complex models of natural systems** like the atmosphere & climate **are inherently uncertain**.
- It's important to quantify uncertainty in model predictions, in order to have confidence in them.
- To fully understand a system's behaviour, we must **densely sample** it over the key input uncertainties – the statistical framework enables this via **Gaussian Process emulation**.
- Once quantified, we can use real-world **observations** to try and reduce output uncertainties via '**History Matching**'.

For the Met Office's aerosol-climate model, HadGEM-UKCA:

- **Robust constraint** of the model is achievable using multiple types of aerosol measurements.
  - Significant constraint of parameter space and aerosol properties leads to **some** constraint on aerosol forcing
  - **Large representation errors** and **equifinality** can **limit** the forcing constraint
- There are **several challenges** in these applications that require more research! ☺


Presentation Outline
Any Questions?