Tunable Testbed for Detection and Attribution

IDAG Workshop 2018

Nathan Lenssen [1] Dorit Hammerling [2] Alexis Hannart [3]

[1]Columbia University, Lamont-Doherty Earth Observatory

[2]National Center for Atmospheric Resarch [3]Ouranos

March 14, 2018

Lamont-Doherty Earth Observatory
COLUMBIA UNIVERSITY | EARTH INSTITUTE

SAMSI Working Group/Collaborators

- The Statistical and Applied Mathematical Sciences Institute (SAMSI) Is conducting a year-long research program on Mathematical and Statistical Methods for Climate and the Earth System (CLIM)
- My research is as a member of the CLIM working group on Detection and Attribution led by Dorit Hammerling
- The testbed is joint with Alexis Hannart and a continuation of [Hannart, 2016]

samsi

NSF Duke NCSU UNC

Testbed Motivation

A flexible and tunable testbed will allow researchers working on detection and attribution methods to:

- Evaluate methods by comparing estimated and true parameter values
  - Performance scaling as a function of sample size/dimensionality
- Simulate real-world scenarios to enable testbed results to represent applications
  - Tunable variety of climate response patterns and climate variability covariances
- Determine robustness of methods through perturbations of testbed parameters
- Compare multiple D+A methods on variety of scenarios

Roadmap of Presentation

(I) Generative Statistical Model for Detection and Attribution
- Ordinary Least Squares
- Error-in-Variable Formulation
- Sources of Variability

(II) Major Testbed Components
- Observed vs. Latent Data/Parameters
- Simulation of Forced Responses
- Covariances of Climate and Observational Variability

(III) Results and Applications

Classical Formulation of Detection and Attribution

Ordinary Least Squares (OLS)

Observed Quantities:

- $\boldsymbol{y}$: The observed climate response of interest
- $\boldsymbol{X}^*$ The model-simulated forcing responses $\boldsymbol{X}^* = (\boldsymbol{x}_1^*, \ldots, \boldsymbol{x}_m^*, \ldots, \boldsymbol{x}_M^*)$

Classical Formulation of Detection and Attribution

Ordinary Least Squares (OLS)

Observed Quantities:

- $\boldsymbol{y}$: The observed climate response of interest
- $\boldsymbol{X}^*$ The model-simulated forcing responses $\boldsymbol{X}^* = (\boldsymbol{x}_1^*, \ldots, \boldsymbol{x}_m^*, \ldots, \boldsymbol{x}_M^*)$

$$\boldsymbol{y} = \boldsymbol{X}^* \beta + \boldsymbol{u}$$

Classical Formulation of Detection and Attribution

Ordinary Least Squares (OLS)

Observed Quantities:

- $\boldsymbol{y}$: The observed climate response of interest
- $\boldsymbol{X}^*$ The model-simulated forcing responses $\boldsymbol{X}^* = (\boldsymbol{x}_1^*, \ldots, \boldsymbol{x}_m^*, \ldots, \boldsymbol{x}_M^*)$

$$\boldsymbol{y} = \boldsymbol{X}^* \boldsymbol{\beta} + \boldsymbol{u}$$

Statistical Parameter of Interest:

- $\boldsymbol{\beta}$ Estimation of $\beta > 0$ provides detection, inference (CIs) gives us attribution

Climate Variability:

- $\boldsymbol{u}$ The error due to climate variability where $\boldsymbol{u} \sim \mathcal{N}(0, \mathbf{C})$
- $\mathbf{C}$ Estimated through model control runs

Statistical Formulation of Detection and Attribution

Error-in-Variable (EIV)

Observed Quantities:

- $\boldsymbol{y}$ The climate response of interest
- $\boldsymbol{X}$ The noisy responses to forcings $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m, \ldots, \boldsymbol{x}_M)$

Statistical Formulation of Detection and Attribution

Error-in-Variable (EIV)

Observed Quantities:

- $\boldsymbol{y}$ The climate response of interest
- $\boldsymbol{X}$ The noisy responses to forcings $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m, \ldots, \boldsymbol{x}_M)$

$$\boldsymbol{y} = \boldsymbol{X}^*\beta + \boldsymbol{u}_y$$
$$\boldsymbol{X} = \boldsymbol{X}^* + \boldsymbol{U}$$

Statistical Formulation of Detection and Attribution

Error-in-Variable (EIV)

Observed Quantities:

- $\boldsymbol{y}$ The climate response of interest
- $\boldsymbol{X}$ The noisy responses to forcings $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m, \ldots, \boldsymbol{x}_M)$

$$\boldsymbol{y} = \boldsymbol{X}^* \beta + \boldsymbol{u}_y$$
$$\boldsymbol{X} = \boldsymbol{X}^* + \boldsymbol{U}$$

Latent Quantities:

- $\boldsymbol{y}^*$ The idealized climate response where $\boldsymbol{y}^* = \boldsymbol{X}^* \beta$
- $\boldsymbol{X}^*$ The idealized responses to forcings $\boldsymbol{X}^* = (\boldsymbol{x}_1^*, \ldots, \boldsymbol{x}_M^*)$

Climate Variability:

- $\boldsymbol{u}_y$ As OLS formulation with $\boldsymbol{u}_y \sim \mathcal{N}(0, \mathbf{C})$
- $\boldsymbol{U}$ The error on the forcing responses due to climate variability

$$\boldsymbol{U} = (\boldsymbol{u}_1, \ldots, \boldsymbol{u}_M) \overset{\text{iid}}{\sim} \mathcal{N}(0, \mathbf{C})$$

EIV model 2: Multimodel ensembles

Statistical Formulation of Detection and Attribution

Error-in-Variable (EIV): Multimember Ensembles

For a given forcing m, we run ensemble of size L_m

$$\boldsymbol{x}_m = \left(\boldsymbol{x}_m^{(1)}, \ldots \boldsymbol{x}_m^{(\ell)}, \ldots, \boldsymbol{x}_m^{(L_m)}\right), \qquad \boldsymbol{x}_m^{(\ell)} \overset{\text{iid}}{\sim} \mathcal{N}(\boldsymbol{x}_m^*, \mathbf{C})$$

The ensemble mean of the m^{th} forced response is

$$\overline{\boldsymbol{x}_m} = \frac{1}{L_m} \sum_{\ell=1}^{L_m} \boldsymbol{x}_m^{(\ell)}, \qquad \overline{\boldsymbol{x}_m} \sim \mathcal{N}\left(\boldsymbol{x}_m^*, L_m^{-1}\,\mathbf{C}\right)$$

Rewriting in the error-in-variable formulation

$$\overline{\boldsymbol{x}_m} = \boldsymbol{x}_m^* + L_m^{-1/2} \boldsymbol{u}_m$$

Or for all forcing responses with $\boldsymbol{L} = \operatorname{diag}(L_1, \ldots, L_M)$

$$\overline{\boldsymbol{X}} = \boldsymbol{X}^* + \boldsymbol{L}^{-1/2} \boldsymbol{U}$$

EIV Model 3: Incoporation of multimodel ensembles

Statistical Formulation of Detection and Attribution

Error-in-Variable (EIV): Multimember Ensembles

$$\overline{\boldsymbol{X}} = \boldsymbol{X}^* + \boldsymbol{L}^{-1/2}\boldsymbol{U}$$

Plugging into our full error in variable formulation:

$$\boldsymbol{y} = \boldsymbol{X}^*\beta + \boldsymbol{u}_y \qquad \boldsymbol{u}_y \sim \mathcal{N}(0, \mathbf{C})$$

$$\overline{\boldsymbol{X}} = \boldsymbol{X}^* + \boldsymbol{L}^{-1/2}\boldsymbol{U} \qquad \boldsymbol{u}_m \overset{\text{iid}}{\sim} \mathcal{N}(0, \mathbf{C})$$

- $\overline{\boldsymbol{X}}$ The ensemble means $\overline{\boldsymbol{X}} = (\overline{\boldsymbol{x}_1}, \ldots, \overline{\boldsymbol{x}_M})$
- $\boldsymbol{L}$ The ensemble sizes $\boldsymbol{L} = (L_1, \ldots, L_M)$
- $\boldsymbol{U}$ The forcing variability matrix $\boldsymbol{U} = (\boldsymbol{u}_1, \ldots, \boldsymbol{u}_M)$

Statistical Formulation of Detection and Attribution

Observed Response Variability

The incomplete expression for the climate response $\boldsymbol{y}$ is

$$\boldsymbol{y} = \boldsymbol{X}^*\beta + \boldsymbol{u}_y\,, \qquad \boldsymbol{u}_y \sim \mathcal{N}(0, \mathbf{C})$$

We propose three different $\boldsymbol{y}$ states to fully incorporate all of the sources of variability.

- $\boldsymbol{y}^*$ The idealized climate response (latent) $\quad \boldsymbol{y}^* = \boldsymbol{X}^*\beta$
- $\boldsymbol{y}_{rel}$ The realized climate response (latent) $\quad \boldsymbol{y}_{rel} = \boldsymbol{y}^* + \boldsymbol{u}_Y$
- $\boldsymbol{y}_{obs}$ The observed climate response $\quad \boldsymbol{y}_{obs} = \boldsymbol{y}_{rel} + \varepsilon_Y$
  - With observational error $\varepsilon_Y \sim \mathcal{N}(0, \mathbf{W})$

$$\begin{aligned} \boldsymbol{y}_{rel} &= \boldsymbol{X}^*\beta + \boldsymbol{u}_y & \qquad \boldsymbol{u}_y &\sim \mathcal{N}(0, \mathbf{C}) \\ \boldsymbol{y}_{obs} &= \boldsymbol{y}_{rel} + \varepsilon_y & \qquad \varepsilon_y &\sim \mathcal{N}(0, \mathbf{W}) \end{aligned}$$

Statistical Formulation of Detection and Attribution

Full Model

Full Error-in-Variable Model:

$$
\begin{aligned}
\boldsymbol{y}_{obs} &= \boldsymbol{y}_{rel} + \varepsilon_y & \varepsilon_y &\sim \mathcal{N}(0, \mathbf{W}) \\
\boldsymbol{y}_{rel} &= \boldsymbol{X}^* \beta + \boldsymbol{u}_y & \boldsymbol{u}_y &\sim \mathcal{N}(0, \mathbf{C}) \\
\overline{\boldsymbol{X}} &= \boldsymbol{X}^* + \boldsymbol{L}^{-1/2} \boldsymbol{U} & \boldsymbol{u}_m &\overset{\text{iid}}{\sim} \mathcal{N}(0, \mathbf{C})
\end{aligned}
$$

Statistical Formulation of Detection and Attribution

Full Model

Full Error-in-Variable Model:

$$\begin{aligned} \boldsymbol{y}_{obs} &= \boldsymbol{y}_{rel} + \varepsilon_y & \varepsilon_y &\sim \mathcal{N}(0, \mathbf{W}) \\ \boldsymbol{y}_{rel} &= \boldsymbol{X}^*\beta + \boldsymbol{u}_y & \boldsymbol{u}_y &\sim \mathcal{N}(0, \mathbf{C}) \\ \overline{\boldsymbol{X}} &= \boldsymbol{X}^* + \boldsymbol{L}^{-1/2}\boldsymbol{U} & \boldsymbol{u}_m &\overset{\text{iid}}{\sim} \mathcal{N}(0, \mathbf{C}) \end{aligned}$$

Scale-Variant Error-in-Variable Model:

$$\begin{aligned} \boldsymbol{y}_{obs} &= \boldsymbol{y}_{rel} + \varepsilon_y & \varepsilon_y &\sim \mathcal{N}(0, \mathbf{W}) \\ \boldsymbol{y}_{rel} &= \boldsymbol{X}^*\beta + \boldsymbol{u}_y & \boldsymbol{u}_y &\sim \mathcal{N}(0, \alpha^{-1}\,\mathbf{C}) \\ \overline{\boldsymbol{X}} &= \boldsymbol{X}^* + \boldsymbol{L}^{-1/2}\boldsymbol{U} & \boldsymbol{u}_m &\overset{\text{ind}}{\sim} \mathcal{N}(0, \gamma_m^{-1}\,\mathbf{C}) \end{aligned}$$

Dorit will talk about fitting this model!

Roadmap of Presentation

(I) Generative Statistical Model for Detection and Attribution
- Ordinary Least Squares
- Error-in-Variable Formulation
- Sources of Variability

(II) Major Testbed Components
- Observed vs. Latent Data/Parameters
- Simulation of Forced Responses
- Covariances of Climate and Observational Variability

(III) Results and Applications

Data and Parameters of Interest

$$
\begin{aligned}
\boldsymbol{y}_{obs} &= \boldsymbol{y}_{rel} + \varepsilon_y & \varepsilon_y &\sim \mathcal{N}(0, \mathbf{W}) \\
\boldsymbol{y}_{rel} &= \boldsymbol{X}^* \beta + \boldsymbol{u}_y & \boldsymbol{u}_y &\sim \mathcal{N}(0, \alpha^{-1}\,\mathbf{C}) \\
\overline{\boldsymbol{X}} &= \boldsymbol{X}^* + \boldsymbol{L}^{-1/2}\boldsymbol{U} & \boldsymbol{u}_m &\overset{\text{ind}}{\sim} \mathcal{N}(0, \gamma_m^{-1}\,\mathbf{C})
\end{aligned}
$$

Observed Objects:

$\boldsymbol{X}$ Observed forcing response ensembles

$$
\begin{aligned}
\boldsymbol{X} &= (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_M) \\
&= \left([\boldsymbol{x}_1^{(1)}, \ldots, \boldsymbol{x}_1^{(L_1)}], \ldots, [\boldsymbol{x}_M^{(1)}, \ldots, \boldsymbol{x}_M^{(L_M)}] \right)
\end{aligned}
$$

$\boldsymbol{Y}_{obs}$ Observed climate response ensemble

$$
\boldsymbol{Y}_{obs} = (\boldsymbol{y}_{obs}^{(1)}, \ldots, \boldsymbol{y}_{obs}^{(L_y)})
$$

$\boldsymbol{X}_0$ Control runs from the climate model used in forcing response experiments

$$
\boldsymbol{X}_0 = (\boldsymbol{X}_0^{(1)}, \ldots, \boldsymbol{X}_0^{(L_0)})
$$

Data and Parameters of Interest

$$\boldsymbol{y}_{obs} = \boldsymbol{y}_{rel} + \boldsymbol{\varepsilon}_y \qquad \boldsymbol{\varepsilon}_y \sim \mathcal{N}(0, \mathbf{W})$$
$$\boldsymbol{y}_{rel} = \boldsymbol{X}^*\beta + \boldsymbol{u}_y \qquad \boldsymbol{u}_y \sim \mathcal{N}(0, \alpha^{-1}\,\mathbf{C})$$
$$\overline{\boldsymbol{X}} = \boldsymbol{X}^* + \boldsymbol{L}^{-1/2}\boldsymbol{U} \qquad \boldsymbol{u}_m \overset{\text{ind}}{\sim} \mathcal{N}(0, \gamma_m^{-1}\,\mathbf{C})$$

Latent Objects

θ The statistical parameters in the model

$$\theta = \{\beta, \alpha, \boldsymbol{\gamma}, \mathbf{C}, \mathbf{W}\}$$

$\boldsymbol{X}^*$ True forcing response $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_M)$

$\boldsymbol{y}^*$ True climate response $\boldsymbol{y}^* = \boldsymbol{X}\beta$

$\boldsymbol{y}_{rel}$ Realized climate response

Simulation Procedure

$$\begin{aligned} \boldsymbol{y}_{obs} &= \boldsymbol{y}_{rel} + \varepsilon_y & \varepsilon_y &\sim \mathcal{N}(0, \mathbf{W}) \\ \boldsymbol{y}_{rel} &= \boldsymbol{X}^*\boldsymbol{\beta} + \boldsymbol{u}_y & \boldsymbol{u}_y &\sim \mathcal{N}(0, \alpha^{-1}\,\mathbf{C}) \\ \overline{\boldsymbol{X}} &= \boldsymbol{X}^* + \boldsymbol{L}^{-1/2}\boldsymbol{U} & \boldsymbol{u}_m &\overset{\text{ind}}{\sim} \mathcal{N}(0, \gamma_m^{-1}\,\mathbf{C}) \end{aligned}$$

For a simulation of fixed dimensionality:

1) Set/Simulate fixed objects $\theta = \{\beta, \alpha, \boldsymbol{\gamma}, \mathbf{C}, \mathbf{W}\}$ and $\boldsymbol{X}^*$
   - $\boldsymbol{X}^*$, $\mathbf{C}$, and $\mathbf{W}$ according to simulation modules

2a) Simulate the observed forcing response ensembles $\boldsymbol{x}_m$

$$\boldsymbol{x}_m^{(\ell)} \overset{\text{iid}}{\sim} \mathcal{N}(\boldsymbol{x}_m^*, \mathbf{C})$$

2b) Simulate the realized climate response $\boldsymbol{y}_{rel}$

$$\boldsymbol{y}_{rel} \sim \mathcal{N}(\boldsymbol{X}^*\boldsymbol{\beta}, \mathbf{C})$$

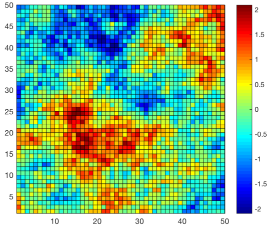
3) Simulate the observed climate response ensemble $\boldsymbol{Y}_{obs}$

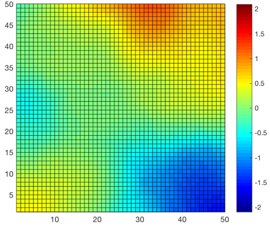
$$\boldsymbol{y}_{obs}^{(\ell)} \overset{\text{iid}}{\sim} \mathcal{N}(y_{rel}, \mathbf{W})$$

(M1) True Forcing Response $\boldsymbol{X}^*$

Simulated Matérn Patterns

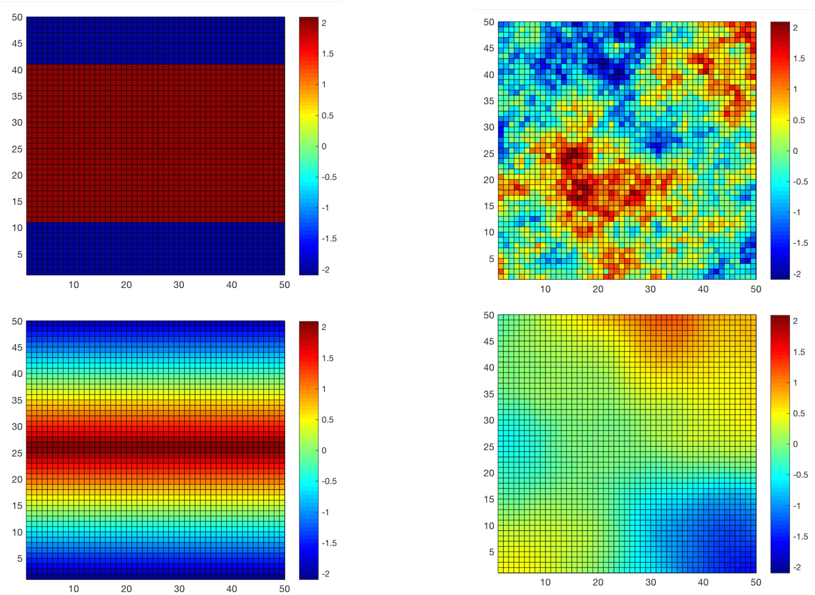
- Fields simulated with covariance matrices according to the Matérn covariance function
- Flexible statistical model that can be fit to replicate climate fields
  - Simulate land-sea interface
- Random generation of $\boldsymbol{X}^*$ allows for robustness testing of methods





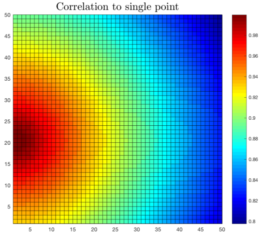
(M1) True Forcing Response $\boldsymbol{X}^*$

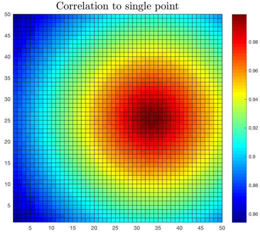
Simulated Patterns



(M2) Climate Variability Covariance **C**

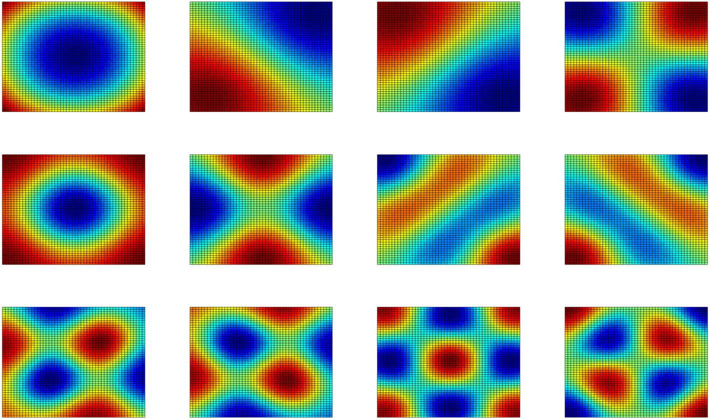
- Exponential covariance function too smooth and regular for climate applications like precipitation
- Goal: Generate non-stationary, non-isotropic covariance matrix
- Issue: Difficult to create and guarantee invertibility!
- Solution: Modify the eigenvalues of a decomposed exponential covariance matrix





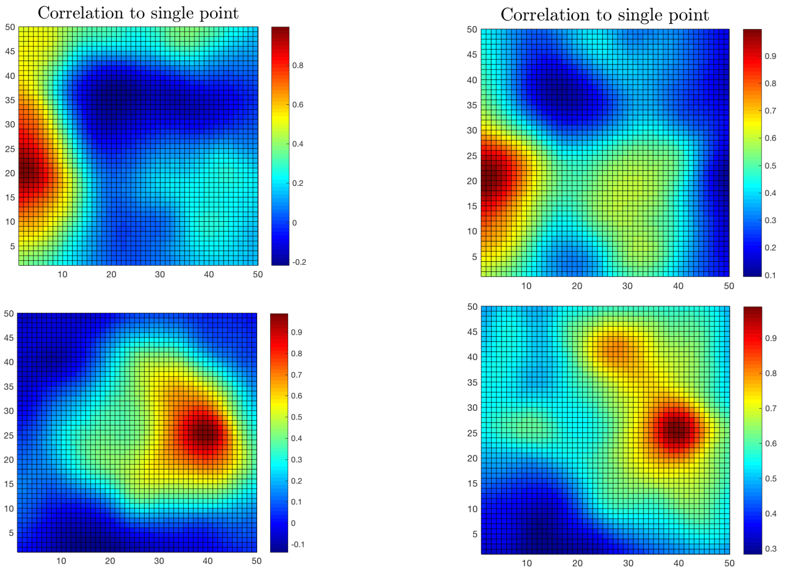
Eigenfunctions of Exponential Covariance Function

$\Sigma_{exp} = V_{exp} \Lambda V_{exp}^{-1}$



(M2) Climate Variability Covariance **C**

Modified eigenfunction covariance $\mathbf{C} = V_{exp}\tilde{\Lambda}V_{exp}^{-1}$



Simulation Procedure

$$\boldsymbol{y}_{obs} = \boldsymbol{y}_{rel} + \varepsilon_y \qquad \varepsilon_y \sim \mathcal{N}(0, \mathbf{W})$$
$$\boldsymbol{y}_{rel} = \boldsymbol{X}^*\beta + \boldsymbol{u}_y \qquad \boldsymbol{u}_y \sim \mathcal{N}(0, \alpha^{-1}\,\mathbf{C})$$
$$\overline{\boldsymbol{X}} = \boldsymbol{X}^* + \boldsymbol{L}^{-1/2}\boldsymbol{U} \qquad \boldsymbol{u}_m \overset{\text{ind}}{\sim} \mathcal{N}(0, \gamma_m^{-1}\,\mathbf{C})$$

For a simulation of fixed dimensionality:

1) Set/Simulate fixed objects $\theta = \{\beta, \alpha, \gamma, \mathbf{C}, \mathbf{W}\}$ and $\boldsymbol{X}^*$
   - $\boldsymbol{X}^*$, $\mathbf{C}$, and $\mathbf{W}$ according to simulation modules

2a) Simulate the observed forcing response ensembles $\boldsymbol{x}_m$

$$\boldsymbol{x}_m^{(\ell)} \overset{\text{iid}}{\sim} \mathcal{N}(\boldsymbol{x}_m^*, \mathbf{C})$$

2b) Simulate the realized climate response $\boldsymbol{y}_{rel}$

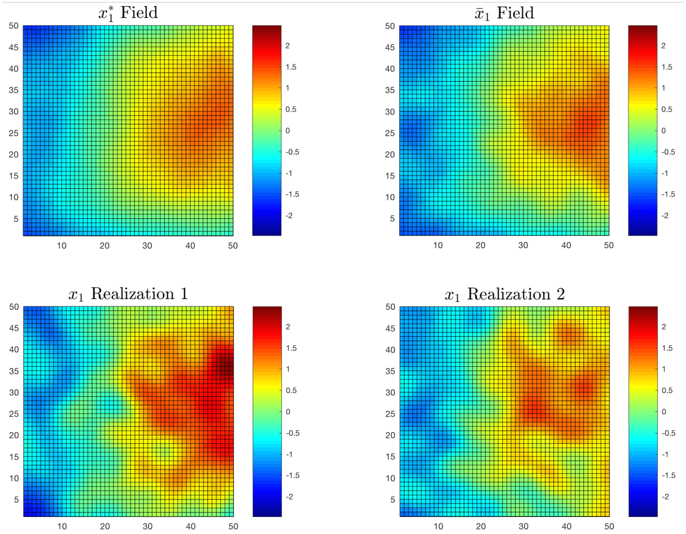
$$\boldsymbol{y}_{rel} \sim \mathcal{N}(\boldsymbol{X}^*\beta, \mathbf{C})$$

3) Simulate the observed climate response ensemble $\boldsymbol{Y}_{obs}$

$$\boldsymbol{y}_{obs}^{(\ell)} \overset{\text{iid}}{\sim} \mathcal{N}(y_{rel}, \mathbf{W})$$

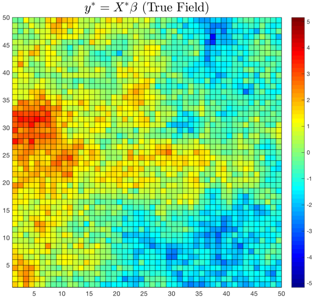
2a) Simulate the Observed Forcing Responses

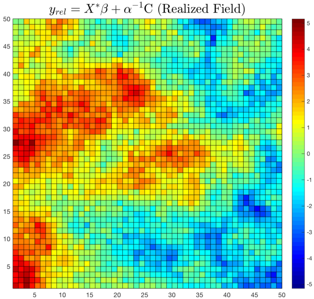
$\boldsymbol{x}_m^{(\ell)} \overset{\text{iid}}{\sim} \mathcal{N}(\boldsymbol{x}_m^*, \gamma_m^{-1}\mathbf{C})$



2b) Simulate the Realized Climate Response

$\boldsymbol{y}_{rel} \sim \mathcal{N}(\boldsymbol{X}^*\boldsymbol{\beta}, \alpha^{-1}\mathbf{C})$





Simulation Procedure

$$\begin{aligned} \boldsymbol{y}_{obs} &= \boldsymbol{y}_{rel} + \varepsilon_y & \varepsilon_y &\sim \mathcal{N}(0, \mathbf{W}) \\ \boldsymbol{y}_{rel} &= \boldsymbol{X}^*\boldsymbol{\beta} + \boldsymbol{u}_y & \boldsymbol{u}_y &\sim \mathcal{N}(0, \alpha^{-1}\,\mathbf{C}) \\ \overline{\boldsymbol{X}} &= \boldsymbol{X}^* + \boldsymbol{L}^{-1/2}\boldsymbol{U} & \boldsymbol{u}_m &\overset{\text{ind}}{\sim} \mathcal{N}(0, \gamma_m^{-1}\,\mathbf{C}) \end{aligned}$$

For a simulation of fixed dimensionality:

1) Set/Simulate fixed objects $\theta = \{\beta, \alpha, \boldsymbol{\gamma}, \mathbf{C}, \mathbf{W}\}$ and $\boldsymbol{X}^*$
   - $\boldsymbol{X}^*$, $\mathbf{C}$, and $\mathbf{W}$ according to simulation modules

2a) Simulate the observed forcing response ensembles $\boldsymbol{x}_m$

$$\boldsymbol{x}_m^{(\ell)} \overset{\text{iid}}{\sim} \mathcal{N}(\boldsymbol{x}_m^*, \mathbf{C})$$

2b) Simulate the realized climate response $\boldsymbol{y}_{rel}$

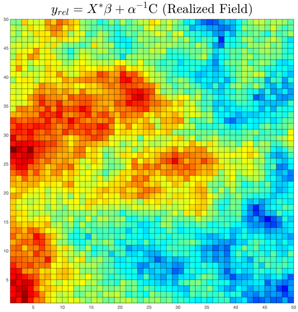
$$\boldsymbol{y}_{rel} \sim \mathcal{N}(\boldsymbol{X}^*\boldsymbol{\beta}, \mathbf{C})$$

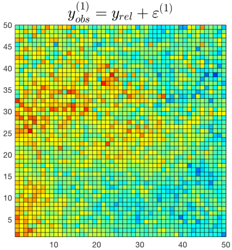
3) Simulate the observed climate response ensemble $\boldsymbol{Y}_{obs}$

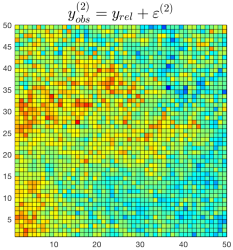
$$\boldsymbol{y}_{obs}^{(\ell)} \overset{\text{iid}}{\sim} \mathcal{N}(y_{rel}, \mathbf{W})$$

3) Simulate the observed climate response

$\boldsymbol{y}_{obs}^{(\ell)} \overset{\text{iid}}{\sim} \mathcal{N}(y_{rel}, \mathbf{W})$







Discussion and Next Steps

- Current Testbed
  - Flexible, fast generation of EIV Detection and Attribution data
  - Non-isotropic climate variability
- First Applications
  - Used in testing new method from Smith, Hammerling and Johnson
- Next Steps
  - Fit forcing responses and sources of variability to real applications
  - Compare traditional OLS and TLS methods as a function of testbed parameters
  - Evaluate models from [Hannart, 2016] and [Katzfuss et al., 2017]
- Thank you/Input/Requests

References

Hannart, A. (2016).
Integrated optimal fingerprinting: Method description and illustration.
Journal of Climate, 29(6):1977–1998.

Katzfuss, M., Hammerling, D., and Smith, R. L. (2017).
A bayesian hierarchical model for climate change detection and attribution.
Geophysical Research Letters, 44(11):5720–5728.
2017GL073688.