Some Recent and Rediscovered Developments in Bayes Linear Statistics

Lachlan Astfalck
School of Physics, Mathematics and Computing
The University of Western Australia

With contributions from Danny Williamson and Cassie Bird

Statistics without probability

Defining Expectation Without Probability

We now define expectation (note, we still have not defined probability) of random quantity X, $\mathrm{E}[\mathrm{X}]$, as the value $\bar{x}$ you would choose if you must suffer penalty

$$L = \left(\frac{\mathrm{X} - \bar{x}}{k}\right)^2$$

once you observe X.

Assumption: Coherence. You do not have a preference for a given penalty if you have the option for one that is certainly smaller.

The Belief Structure

The Belief Structure

- Consider two random quantities X and D

The Belief Structure

- Consider two random quantities X and D
- X is our quantity of interest, and D is the quantity that we observe

The Belief Structure

- Consider two random quantities X and D
- X is our quantity of interest, and D is the quantity that we observe
- Quantities are defined on the Hilbert space defined by $\langle \mathrm{X}, \mathrm{Y} \rangle = \mathrm{E}[\mathrm{X}^\intercal \mathrm{Y}]$

The Belief Structure

- Consider two random quantities X and D
- X is our quantity of interest, and D is the quantity that we observe
- Quantities are defined on the Hilbert space defined by $\langle \mathrm{X}, \mathrm{Y} \rangle = \mathrm{E}[\mathrm{X}^\intercal \mathrm{Y}]$
- Define the *belief structure* $\mathscr{B} = \mathscr{X} \cup \mathscr{D} \cup 1$

The Belief Structure

- Consider two random quantities X and D
- X is our quantity of interest, and D is the quantity that we observe
- Quantities are defined on the Hilbert space defined by $\langle \mathrm{X}, \mathrm{Y} \rangle = \mathrm{E}[\mathrm{X}^\intercal \mathrm{Y}]$
- Define the *belief structure* $\mathscr{B} = \mathscr{X} \cup \mathscr{D} \cup 1$
- $\mathscr{B}$ requires the specification of $\mathrm{E}[\mathrm{X}]$, $\mathrm{E}[\mathrm{D}]$, $\mathrm{var}[\mathrm{X}]$, $\mathrm{var}[\mathrm{D}]$, and $\mathrm{cov}[\mathrm{X}, \mathrm{D}]$

The Belief Structure

- Consider two random quantities X and D
- X is our quantity of interest, and D is the quantity that we observe
- Quantities are defined on the Hilbert space defined by $\langle \mathrm{X}, \mathrm{Y} \rangle = \mathrm{E}[\mathrm{X}^\intercal \mathrm{Y}]$
- Define the *belief structure* $\mathscr{B} = \mathscr{X} \cup \mathscr{D} \cup 1$
- $\mathscr{B}$ requires the specification of $\mathrm{E}[\mathrm{X}]$, $\mathrm{E}[\mathrm{D}]$, $\mathrm{var}[\mathrm{X}]$, $\mathrm{var}[\mathrm{D}]$, and $\mathrm{cov}[\mathrm{X}, \mathrm{D}]$
- **Expectation is the fundamental unit of belief and $\mathscr{B}$ is the analogy of the joint probability measure in a standard Bayesian analysis.**

Adjusting belief structures

Adjusting belief structures

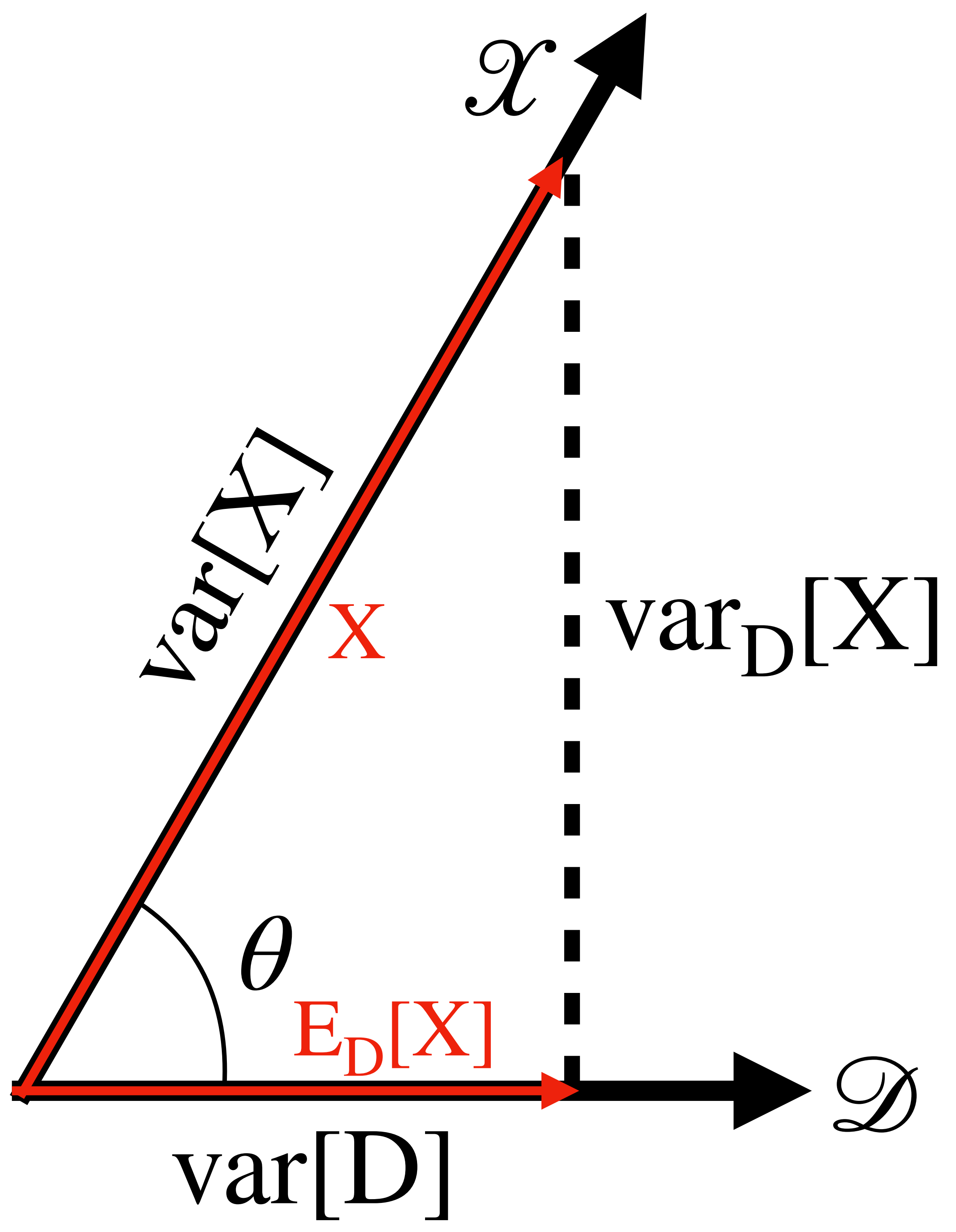
- The adjusted expectation, $\mathrm{E}_{\mathrm{D}}[\mathrm{X}]$, is the projection of X onto affine D, $\mathrm{h}_0 + \mathrm{H}_0\mathrm{D}$

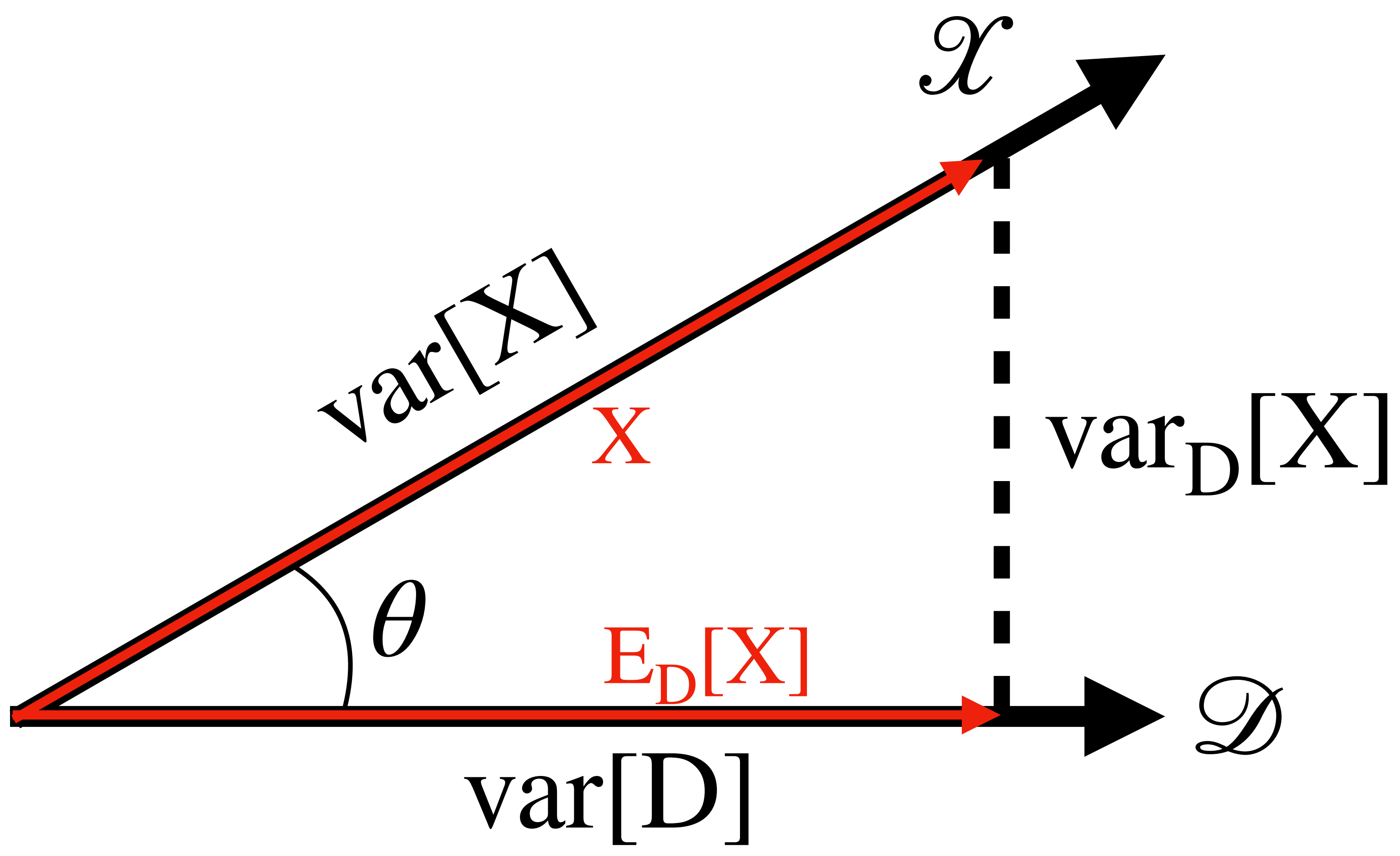
Adjusting belief structures

- The adjusted expectation, $\mathrm{E_D[X]}$, is the projection of X onto affine D, $\mathrm{h_0 + H_0 D}$
- The adjusted variance, $\mathrm{var_D[X]}$, is the squared length $\mathrm{||X - E_D[X]||^2}$

Adjusting belief structures

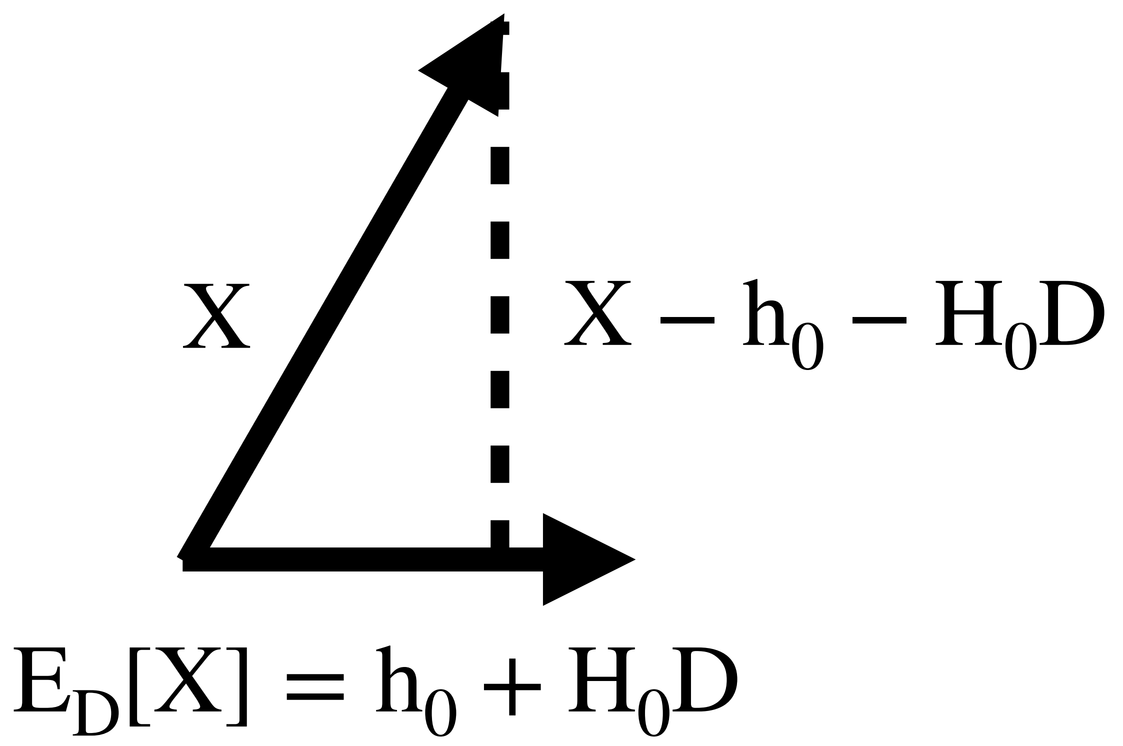
- The adjusted expectation, $\mathrm{E_D[X]}$, is the projection of X onto affine D, $\mathrm{h_0 + H_0 D}$
- The adjusted variance, $\mathrm{var_D[X]}$, is the squared length $||\mathrm{X} - \mathrm{E_D[X]}||^2$



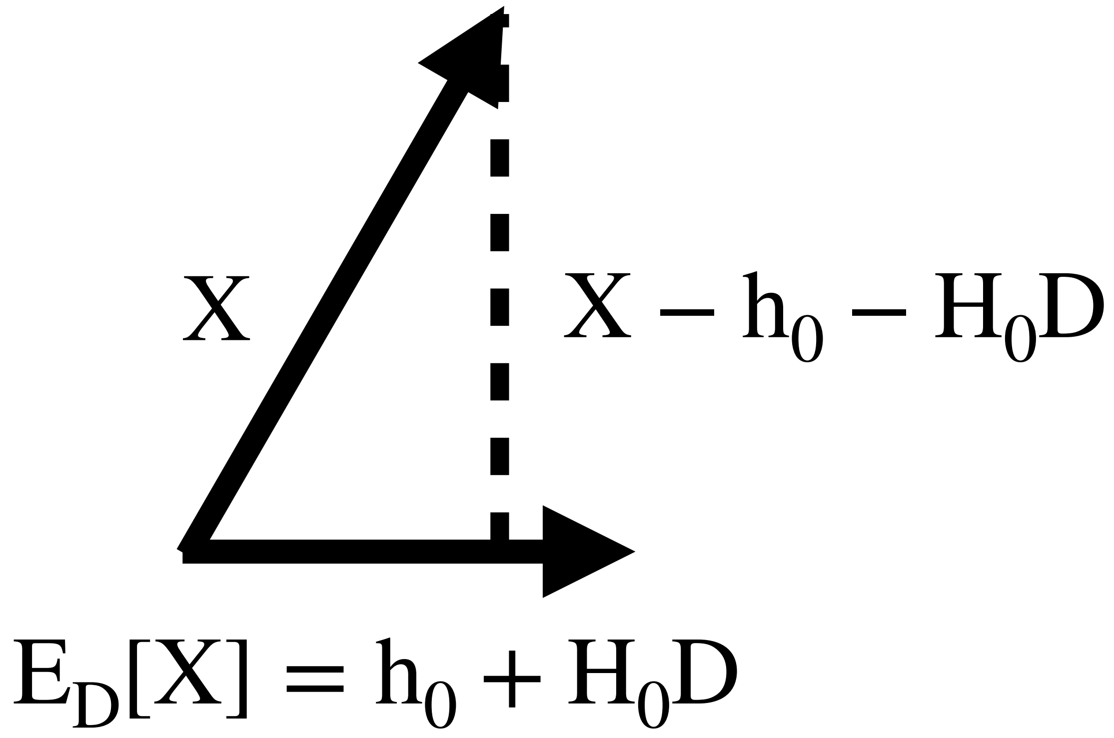


$$\theta = \cos^{-1}(\mathrm{cor[X, D]})$$

“the Bayes linear equations”



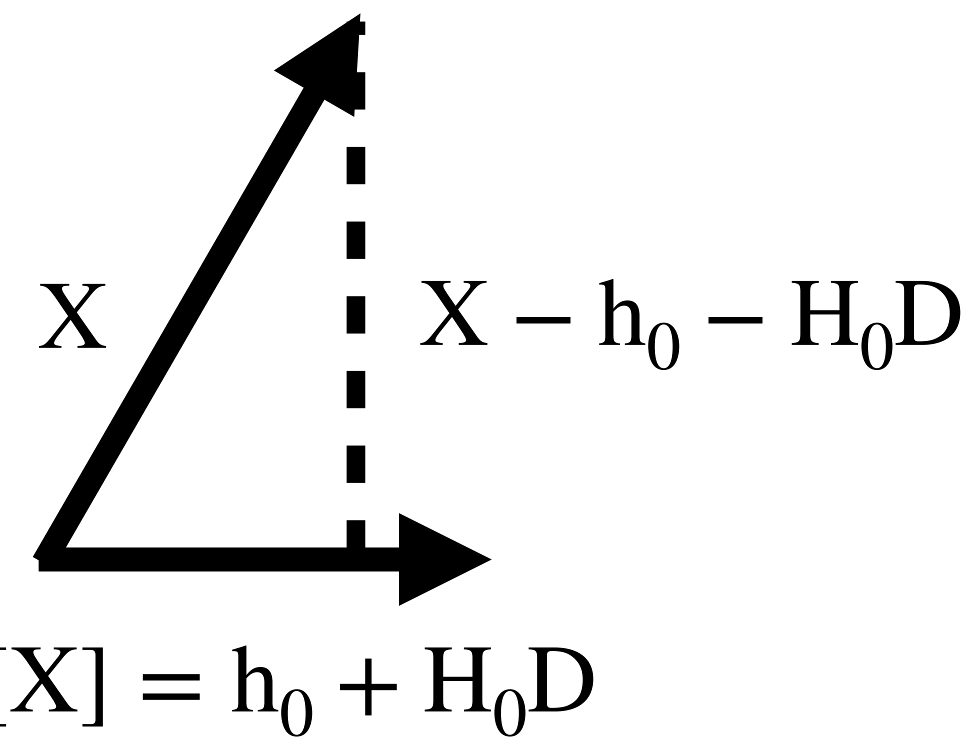
“the Bayes linear equations”



The orthogonal projection of X onto $h_0 + H_0 D$ solves:

$$\langle X - h_0 - H_0 D, h_0 + H_0 D \rangle = E[(X - h_0 - H_0 D)^\intercal (h_0 + H_0 D)] = 0,$$

“the Bayes linear equations”



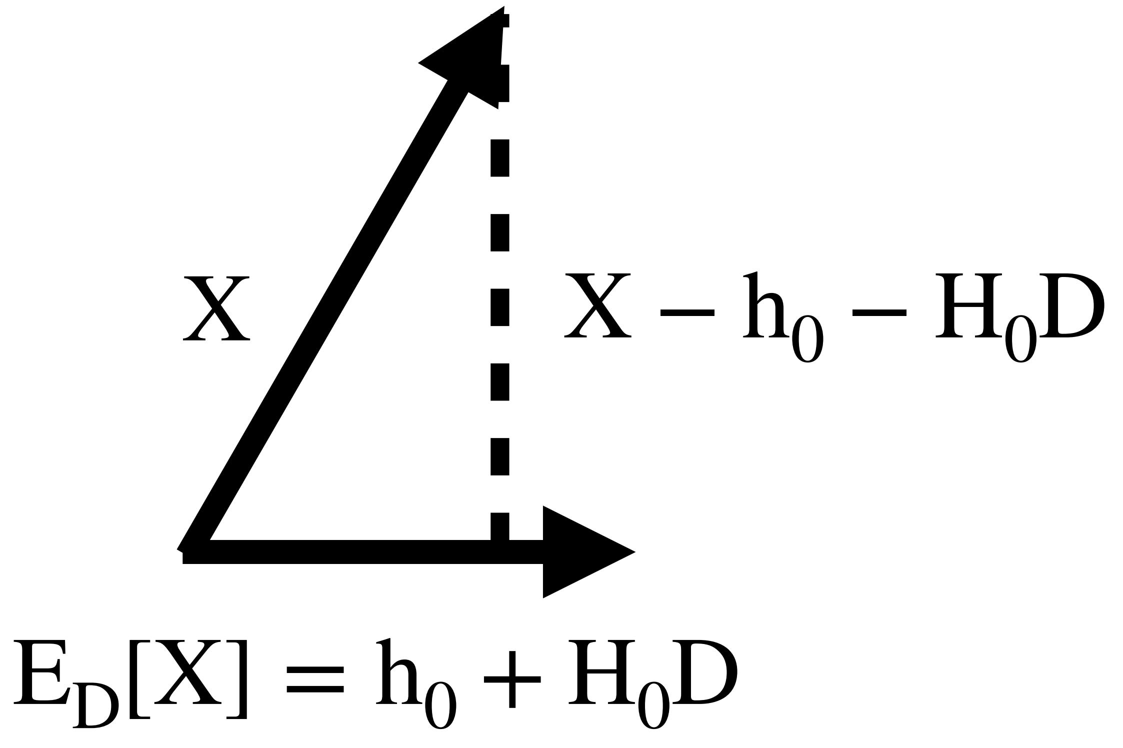
The orthogonal projection of X onto $h_0 + H_0 D$ solves:

$$\langle X - h_0 - H_0 D, h_0 + H_0 D \rangle = E[(X - h_0 - H_0 D)^\intercal (h_0 + H_0 D)] = 0,$$

so $h_0 = E[X] - H_0 E[D]$, $H_0 = \mathrm{cov}[X, D]\mathrm{var}[D]^{-1}$, and

$$E_D[X] = E[X] + \mathrm{cov}[X, D]\mathrm{var}[D]^{-1}(E[D] - D).$$

“the Bayes linear equations”



The orthogonal projection of X onto $h_0 + H_0 D$ solves:

$$\langle X - h_0 - H_0 D, h_0 + H_0 D \rangle = E[(X - h_0 - H_0 D)^\intercal (h_0 + H_0 D)] = 0,$$

so $h_0 = E[X] - H_0 E[D]$, $H_0 = \mathrm{cov}[X, D]\mathrm{var}[D]^{-1}$, and

$$E_D[X] = E[X] + \mathrm{cov}[X, D]\mathrm{var}[D]^{-1}(E[D] - D).$$

The length $\mathrm{var}_D[X] = \|X - E_D[X]\|^2 = \langle X - E_D[X], X - E_D[X] \rangle$, so

$$\mathrm{var}_D[X] = \mathrm{var}[X] - \mathrm{cov}[X, D]\mathrm{var}[D]^{-1}\mathrm{cov}[D, X]$$

Bayes Linear

Belief space $\mathscr{B}$

E[X], E[D], var[X], var[D], cov[X, D]

Probabilistic Bayes

Probability measure P

p(X), p(D | X)

Bayes Linear		Probabilistic Bayes
Belief space $\mathscr{B}$	$\longleftrightarrow$	Probability measure P
$E[X], E[D], var[X], var[D], cov[X, D]$		$p(X), p(D \mid X)$
Adjusted Expectation/Variance	$\longleftrightarrow$	Posterior Distribution
$E_D[X], var_D[X]$		$E[X \mid D], p(X \mid D)$

Bayes Linear		Probabilistic Bayes
Belief space $\mathscr{B}$	$\longleftrightarrow$	Probability measure P
$\mathrm{E}[\mathrm{X}], \mathrm{E}[\mathrm{D}], \mathrm{var}[\mathrm{X}], \mathrm{var}[\mathrm{D}], \mathrm{cov}[\mathrm{X}, \mathrm{D}]$		$\mathrm{p}(\mathrm{X}), \mathrm{p}(\mathrm{D} \mid \mathrm{X})$
Adjusted Expectation/Variance	$\longleftrightarrow$	Posterior Distribution
$\mathrm{E}_{\mathrm{D}}[\mathrm{X}], \mathrm{var}_{\mathrm{D}}[\mathrm{X}]$		$\mathrm{E}[\mathrm{X} \mid \mathrm{D}], p(\mathrm{X} \mid \mathrm{D})$
Probability from expectation	$\longleftrightarrow$	Expectation from probability
$\mathrm{P}(E) = \mathrm{E}[\mathbf{1}_E]$		$\mathrm{E}[\mathrm{X}] = \int x\mathrm{p}(x)\,\mathrm{d}x$

Normal without normality?

$$\mathrm{E_D[X] = E[X] + cov[X, D] var[D]^{-1}(E[D] - D)}$$

$$\mathrm{var_D[X] = var[X] - cov[X, D] var[D]^{-1} cov[D, X]}$$

$$E_D[X] = E[X] + \mathrm{cov}[X, D]\mathrm{var}[D]^{-1}(E[D] - D)$$

$$\mathrm{var}_D[X] = \mathrm{var}[X] - \mathrm{cov}[X, D]\mathrm{var}[D]^{-1}\mathrm{cov}[D, X]$$

Isn't this just the posterior equations for an update with normal prior and likelihood?

$$E[X \mid D] = E[X] + \mathrm{cov}[X, D]\mathrm{var}[D \mid X]^{-1}(E[D \mid X] - D)$$

$$\mathrm{var}[X \mid D] = \mathrm{var}[X] - \mathrm{cov}[X, D]\mathrm{var}[D \mid X]^{-1}\mathrm{cov}[D, X]$$

This isn’t the whole story…

(proof extended from the results of Hartigan, 1969)

This isn't the whole story…

Assume in a probabilistic Bayesian analysis that the posterior expectation is linear in D, $\mathrm{E}[\mathrm{X} \mid \mathrm{D}] = A\mathrm{D} + B$

(proof extended from the results of Hartigan, 1969)

This isn’t the whole story…

Assume in a probabilistic Bayesian analysis that the posterior expectation is linear in D, $\mathrm{E}[\mathrm{X} \mid \mathrm{D}] = A\mathrm{D} + B$

Law of iterated expectation: $\mathrm{E}[\mathrm{X}] = \mathrm{E}_{\mathrm{D}}\left[\mathrm{E}_{\mathrm{X}}[\mathrm{X} \mid \mathrm{D}]\right] = A\mathrm{E}[\mathrm{D}] + B$

(proof extended from the results of Hartigan, 1969)

This isn't the whole story…

Assume in a probabilistic Bayesian analysis that the posterior expectation is linear in D, $\mathrm{E}[\mathrm{X} \mid \mathrm{D}] = A\mathrm{D} + B$

Law of iterated expectation: $\mathrm{E}[\mathrm{X}] = \mathrm{E}_{\mathrm{D}}\left[\mathrm{E}_{\mathrm{X}}[\mathrm{X} \mid \mathrm{D}]\right] = A\mathrm{E}[\mathrm{D}] + B$

And again: $\mathrm{E}[\mathrm{D}\mathrm{X}^\intercal] = \mathrm{E}_{\mathrm{D}}\left[\mathrm{D}\mathrm{E}_{\mathrm{X}}[\mathrm{X} \mid \mathrm{D}]\right] = \mathrm{E}_{\mathrm{D}}\left[\mathrm{D}(A\mathrm{D} + B)^\intercal\right]$

$$= \mathrm{var}[\mathrm{D}]A^\intercal + \mathrm{E}[\mathrm{D}]\mathrm{E}[\mathrm{X}]^\intercal$$

(proof extended from the results of Hartigan, 1969)

This isn't the whole story…

Assume in a probabilistic Bayesian analysis that the posterior expectation is linear in D, $\mathrm{E}[\mathrm{X} \mid \mathrm{D}] = A\mathrm{D} + B$

Law of iterated expectation: $\mathrm{E}[\mathrm{X}] = \mathrm{E}_\mathrm{D}\left[\mathrm{E}_\mathrm{X}[\mathrm{X} \mid \mathrm{D}]\right] = A\mathrm{E}[\mathrm{D}] + B$

And again: $\mathrm{E}[\mathrm{D}\mathrm{X}^\intercal] = \mathrm{E}_\mathrm{D}\left[\mathrm{D}\mathrm{E}_\mathrm{X}[\mathrm{X} \mid \mathrm{D}]\right] = \mathrm{E}_\mathrm{D}\left[\mathrm{D}(A\mathrm{D} + B)^\intercal\right]$

$$= \mathrm{var}[\mathrm{D}]A^\intercal + \mathrm{E}[\mathrm{D}]\mathrm{E}[\mathrm{X}]^\intercal$$

Definition of covariance: $\mathrm{E}[\mathrm{D}\mathrm{X}^\intercal] = \mathrm{cov}[\mathrm{X}, \mathrm{D}] + \mathrm{E}[\mathrm{D}]\mathrm{E}[\mathrm{X}]^\intercal$

$$A = \mathrm{cov}[\mathrm{X}, \mathrm{D}]\mathrm{var}[\mathrm{D}]^{-1}, \qquad B = \mathrm{E}[\mathrm{X}] - A\mathrm{E}[\mathrm{D}]$$

(proof extended from the results of Hartigan, 1969)

This isn’t the whole story…

This isn't the whole story…

Substitute A and B into $\mathrm{E}[\mathrm{X} \mid \mathrm{D}] = A\mathrm{D} + B$

$$\mathrm{E}[\mathrm{X} \mid \mathrm{D}] = \mathrm{E}[\mathrm{X}] + \mathrm{cov}[\mathrm{X}, \mathrm{D}]\mathrm{var}[\mathrm{D}]^{-1}(\mathrm{E}[\mathrm{D}] - \mathrm{D})$$

This isn't the whole story…

Substitute A and B into $\mathrm{E}[\mathrm{X} \mid \mathrm{D}] = A\mathrm{D} + B$

$$\mathrm{E}[\mathrm{X} \mid \mathrm{D}] = \mathrm{E}[\mathrm{X}] + \mathrm{cov}[\mathrm{X}, \mathrm{D}]\mathrm{var}[\mathrm{D}]^{-1}(\mathrm{E}[\mathrm{D}] - \mathrm{D})$$

Now substitute this into $\mathrm{var}[\mathrm{X} \mid \mathrm{D}] = \mathrm{E}\left[(\mathrm{X} - \mathrm{E}[\mathrm{X} \mid \mathrm{D}])(\mathrm{X} - \mathrm{E}[\mathrm{X} \mid \mathrm{D}])^\intercal\right]$

$$\mathrm{var}[\mathrm{X} \mid \mathrm{D}] = \mathrm{var}[\mathrm{X}] - \mathrm{cov}[\mathrm{X}, \mathrm{D}]\mathrm{var}[\mathrm{D}]^{-1}\mathrm{cov}[\mathrm{D}, \mathrm{X}]$$

This isn't the whole story…

Substitute A and B into $\mathrm{E}[\mathrm{X} \mid \mathrm{D}] = A\mathrm{D} + B$

$$\mathrm{E}[\mathrm{X} \mid \mathrm{D}] = \mathrm{E}[\mathrm{X}] + \mathrm{cov}[\mathrm{X}, \mathrm{D}]\mathrm{var}[\mathrm{D}]^{-1}(\mathrm{E}[\mathrm{D}] - \mathrm{D})$$

Now substitute this into $\mathrm{var}[\mathrm{X} \mid \mathrm{D}] = \mathrm{E}\left[(\mathrm{X} - \mathrm{E}[\mathrm{X} \mid \mathrm{D}])(\mathrm{X} - \mathrm{E}[\mathrm{X} \mid \mathrm{D}])^\intercal\right]$

$$\mathrm{var}[\mathrm{X} \mid \mathrm{D}] = \mathrm{var}[\mathrm{X}] - \mathrm{cov}[\mathrm{X}, \mathrm{D}]\mathrm{var}[\mathrm{D}]^{-1}\mathrm{cov}[\mathrm{D}, \mathrm{X}]$$

We can recover the Bayes linear equations only with the assumption that the posterior expectation is linear in D

And when does this happen?

And when does this happen?

The exponential family of distributions with conjugate prior (Diaconis et al., 1979)

And when does this happen?

The exponential family of distributions with conjugate prior (Diaconis et al., 1979)

- Normal likelihood, Normal prior (real-valued)

And when does this happen?

The exponential family of distributions with conjugate prior (Diaconis et al., 1979)

- Normal likelihood, Normal prior (real-valued)
- Poisson likelihood, Gamma prior (counts)

And when does this happen?

The exponential family of distributions with conjugate prior (Diaconis et al., 1979)

- Normal likelihood, Normal prior (real-valued)
- Poisson likelihood, Gamma prior (counts)
- Bernoulli likelihood, Beta prior (probabilities)

And when does this happen?

The exponential family of distributions with conjugate prior (Diaconis et al., 1979)

- Normal likelihood, Normal prior (real-valued)
- Poisson likelihood, Gamma prior (counts)
- Bernoulli likelihood, Beta prior (probabilities)
- Gamma likelihood, Gamma prior (strictly positive)

And when does this happen?

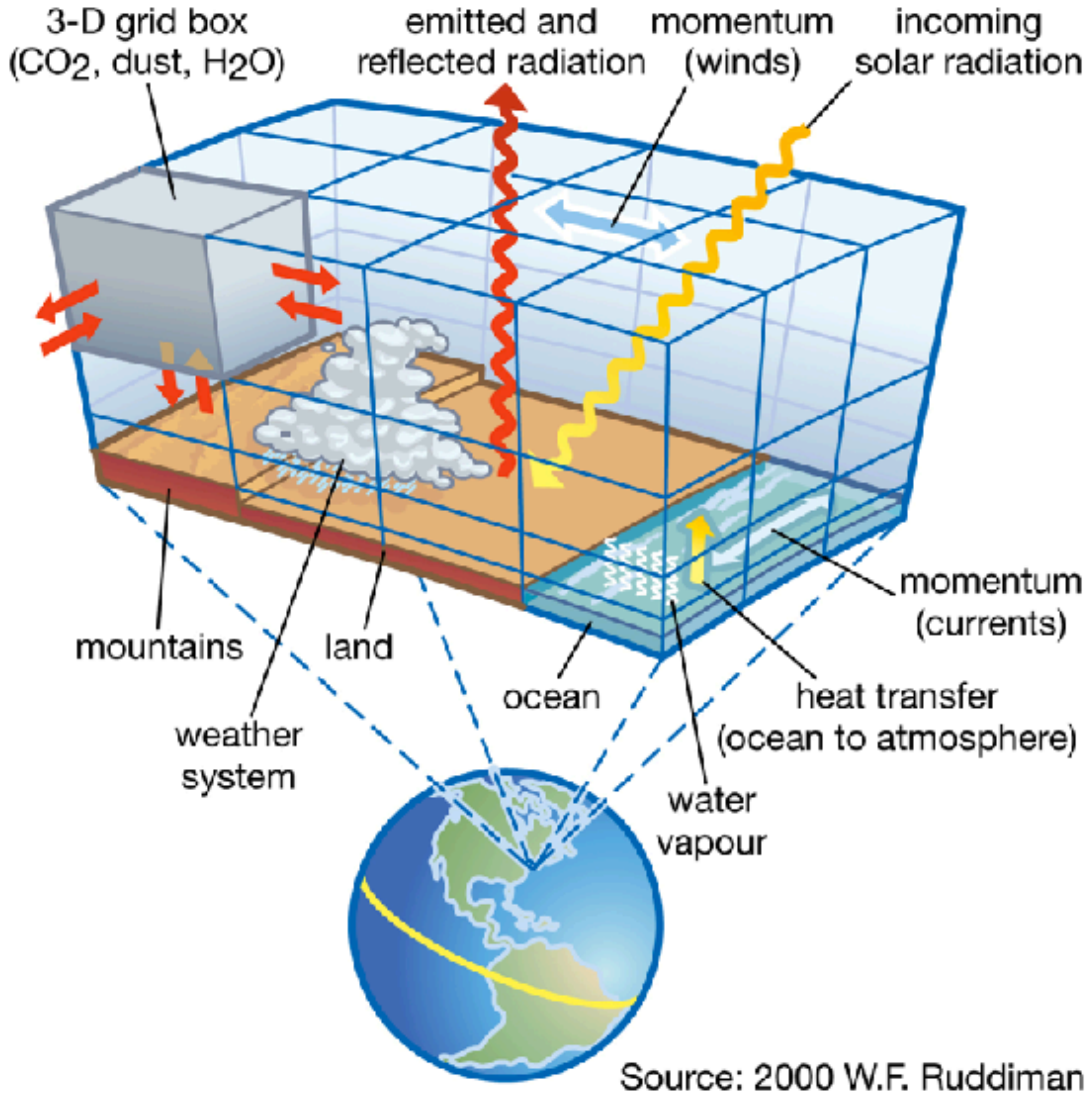
The exponential family of distributions with conjugate prior (Diaconis et al., 1979)

- Normal likelihood, Normal prior (real-valued)
- Poisson likelihood, Gamma prior (counts)
- Bernoulli likelihood, Beta prior (probabilities)
- Gamma likelihood, Gamma prior (strictly positive)

And some mixture models (Ericson, 1969)

Hierarchical Bayes Linear

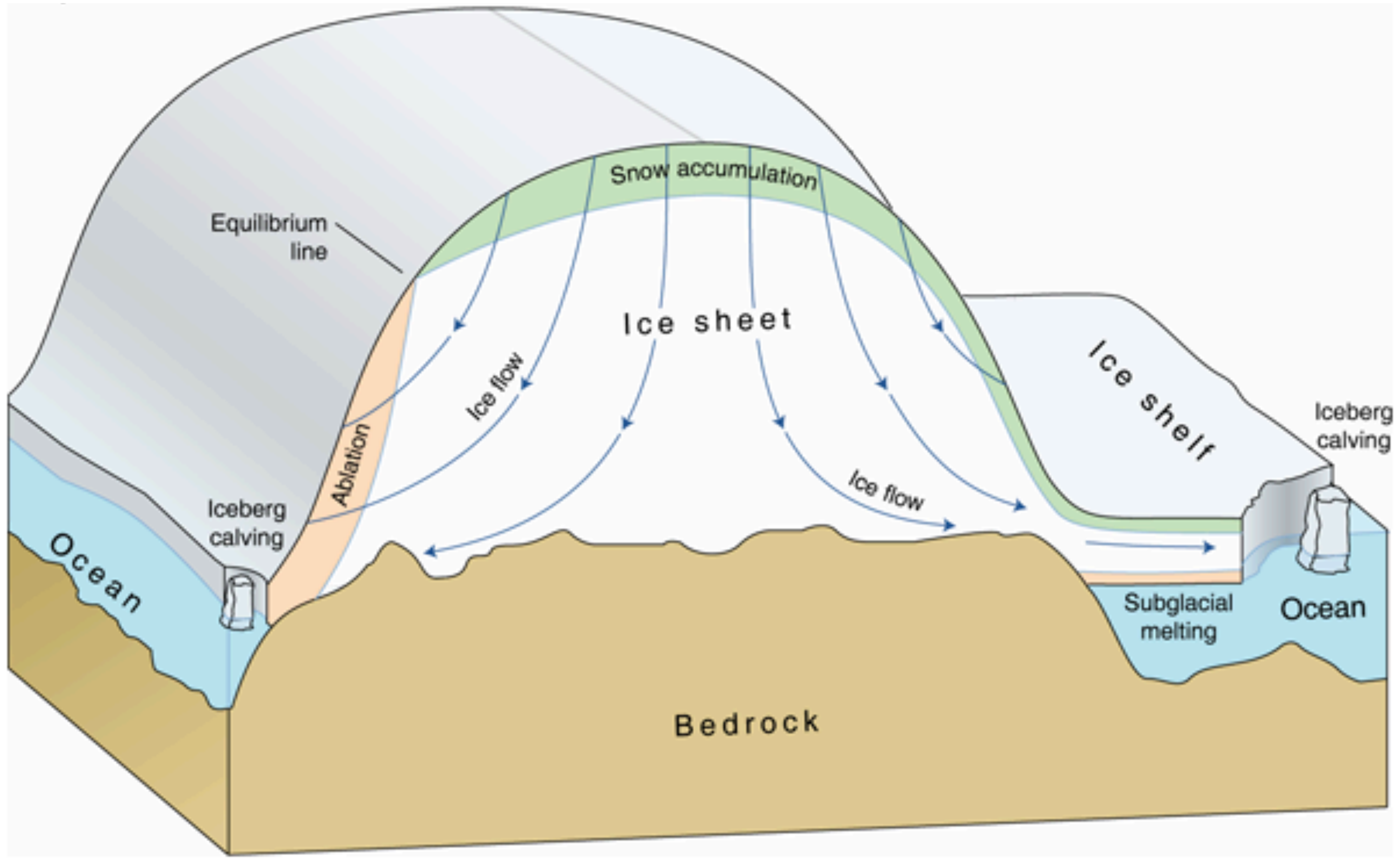
Modelling glacier dynamics is hard…



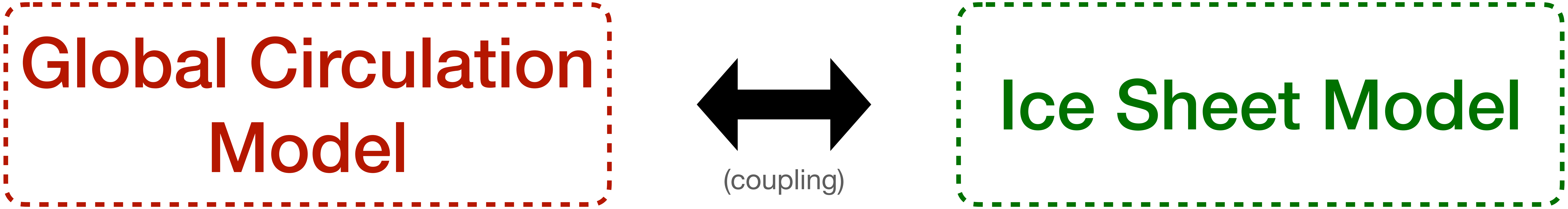
Global Circulation Models (GCM)

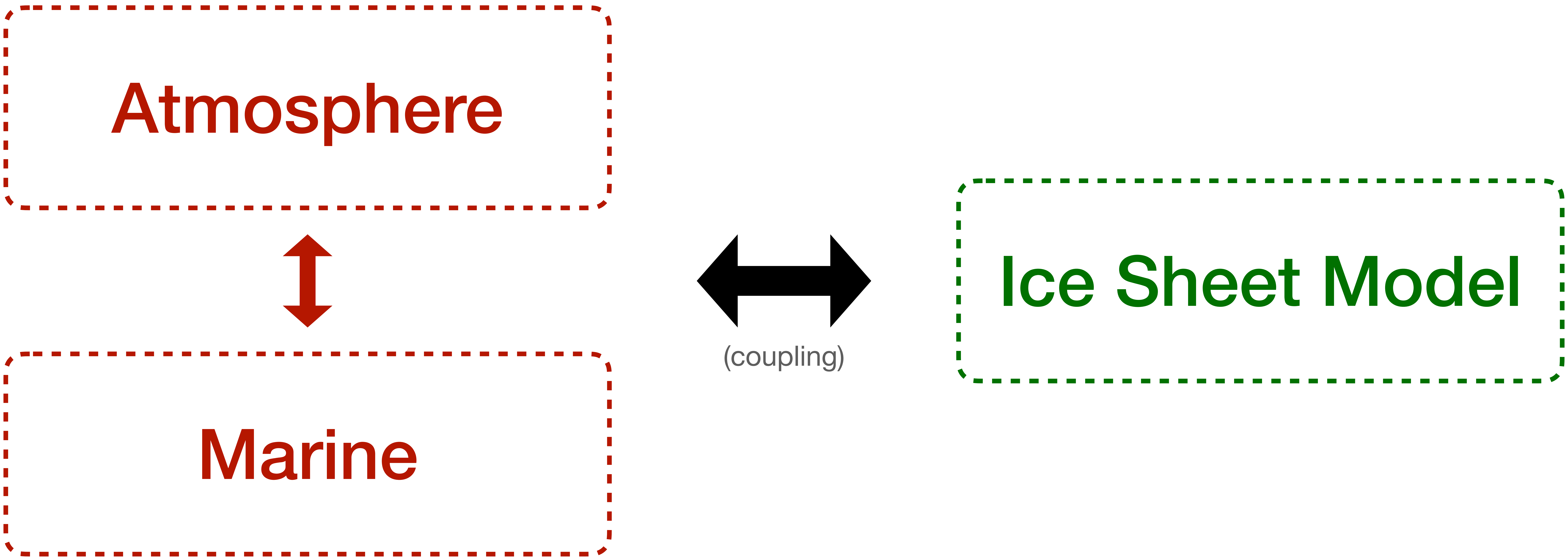
⟷

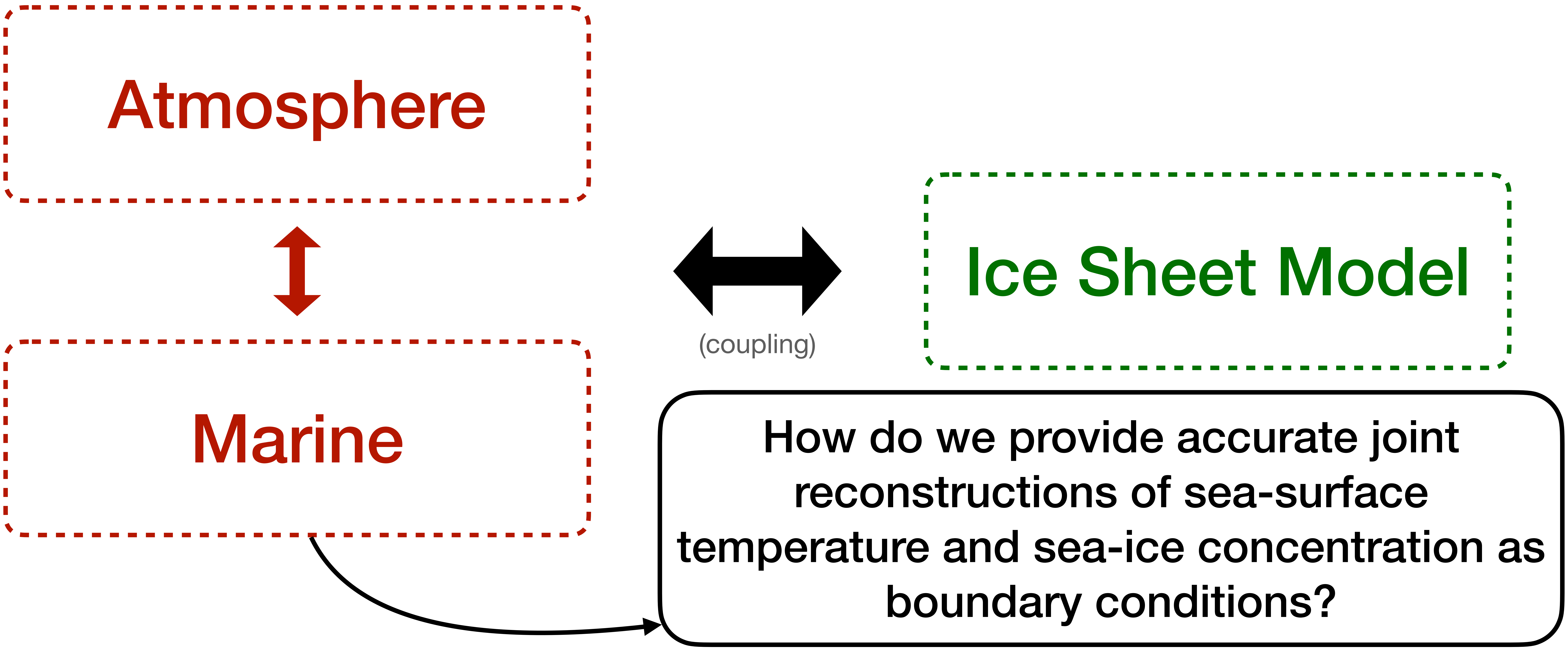
(coupling)



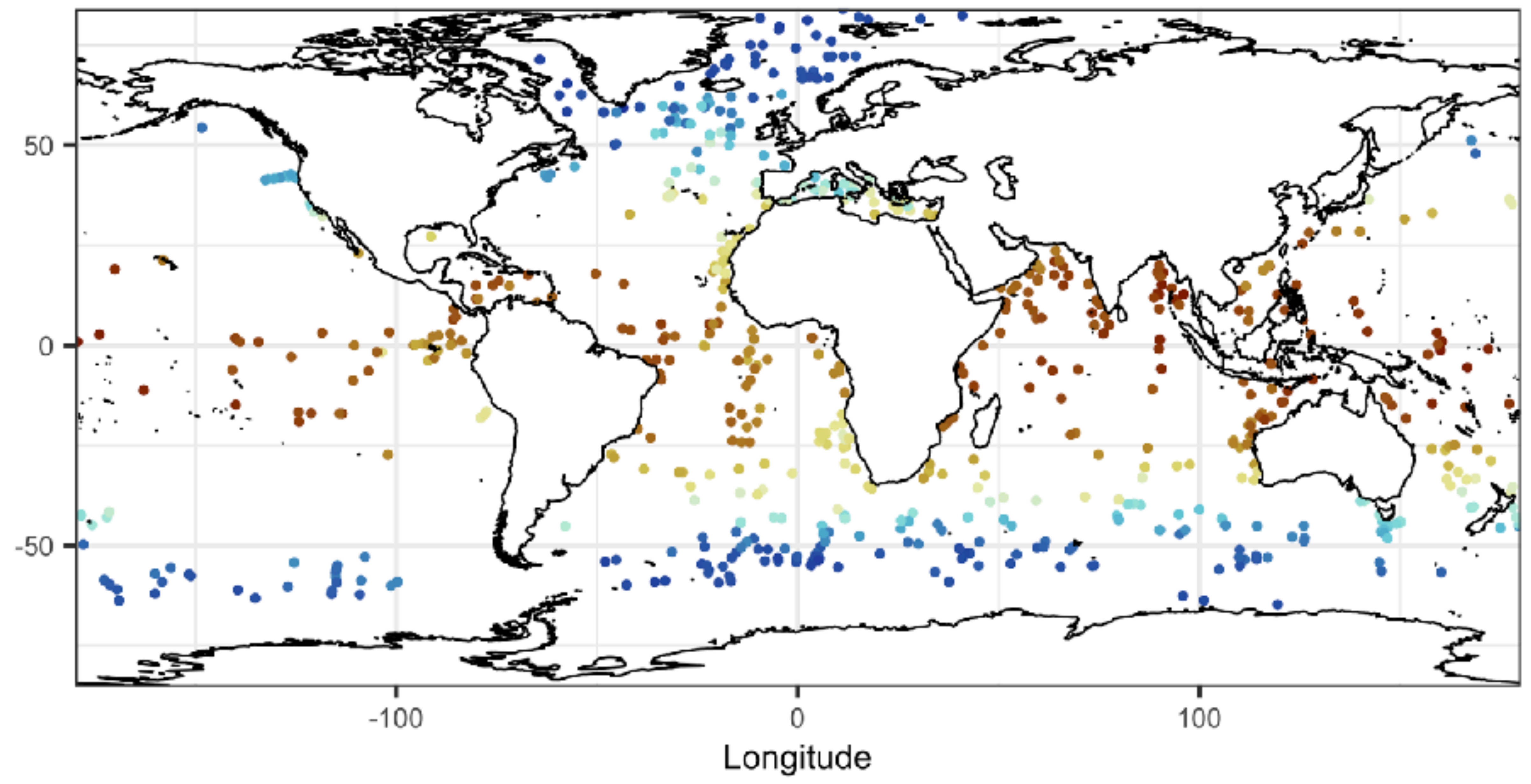
Regional Ice Sheet Models


Global Circulation Model
(coupling)
Ice Sheet Model


Atmosphere
Marine
(coupling)
Ice Sheet Model

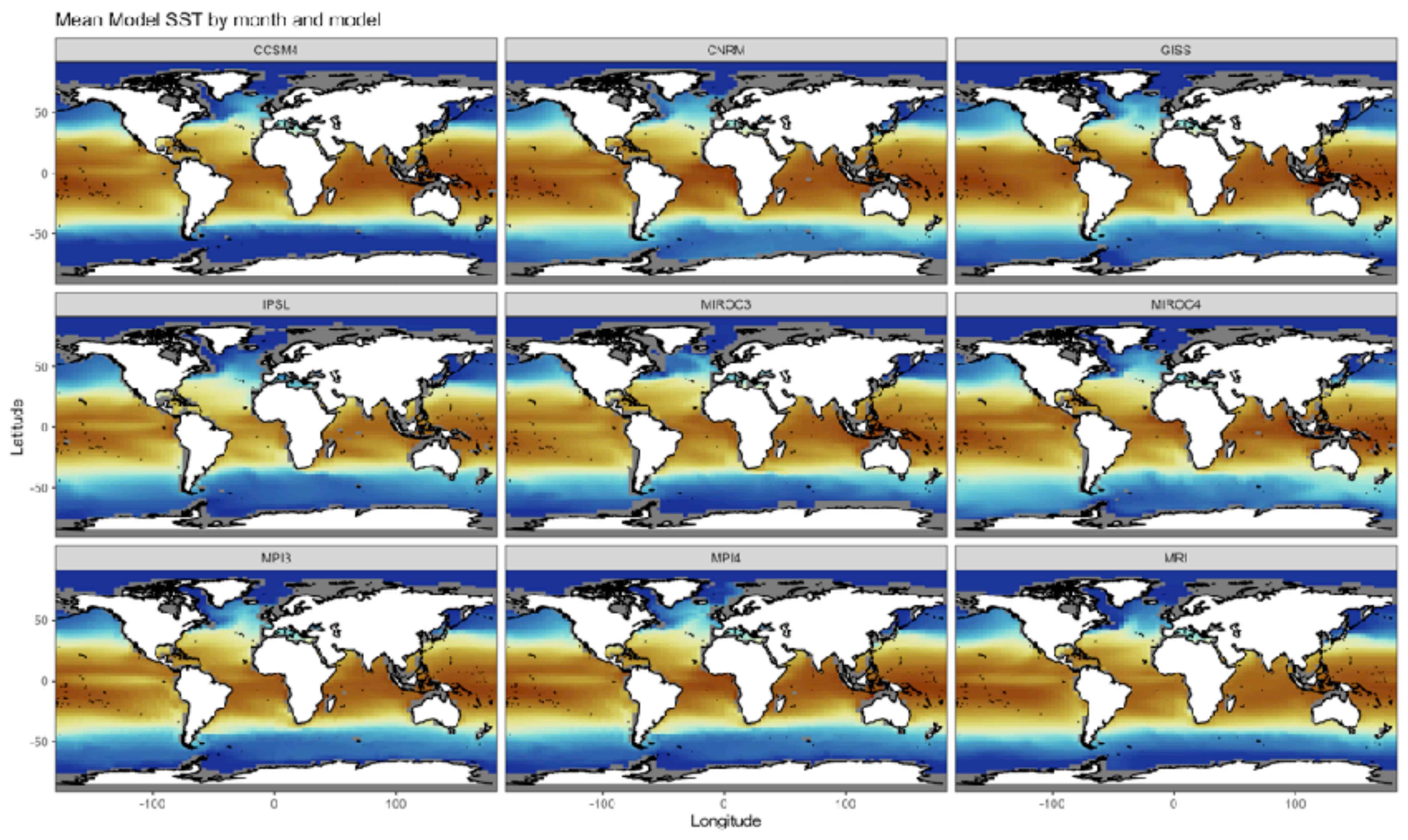

Atmosphere
Marine
(coupling)
Ice Sheet Model
How do we provide accurate joint reconstructions of sea-surface temperature and sea-ice concentration as boundary conditions?

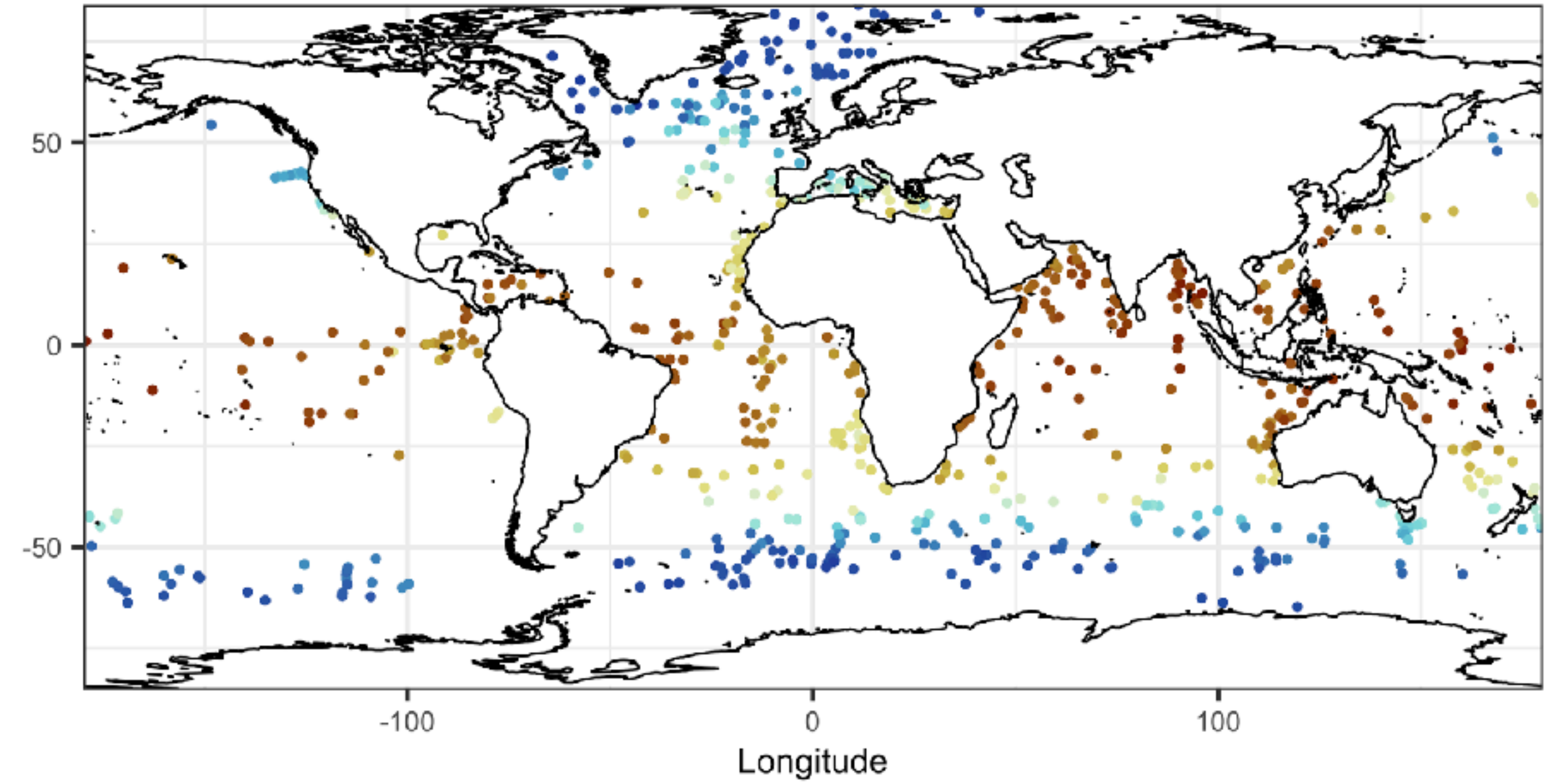
Data

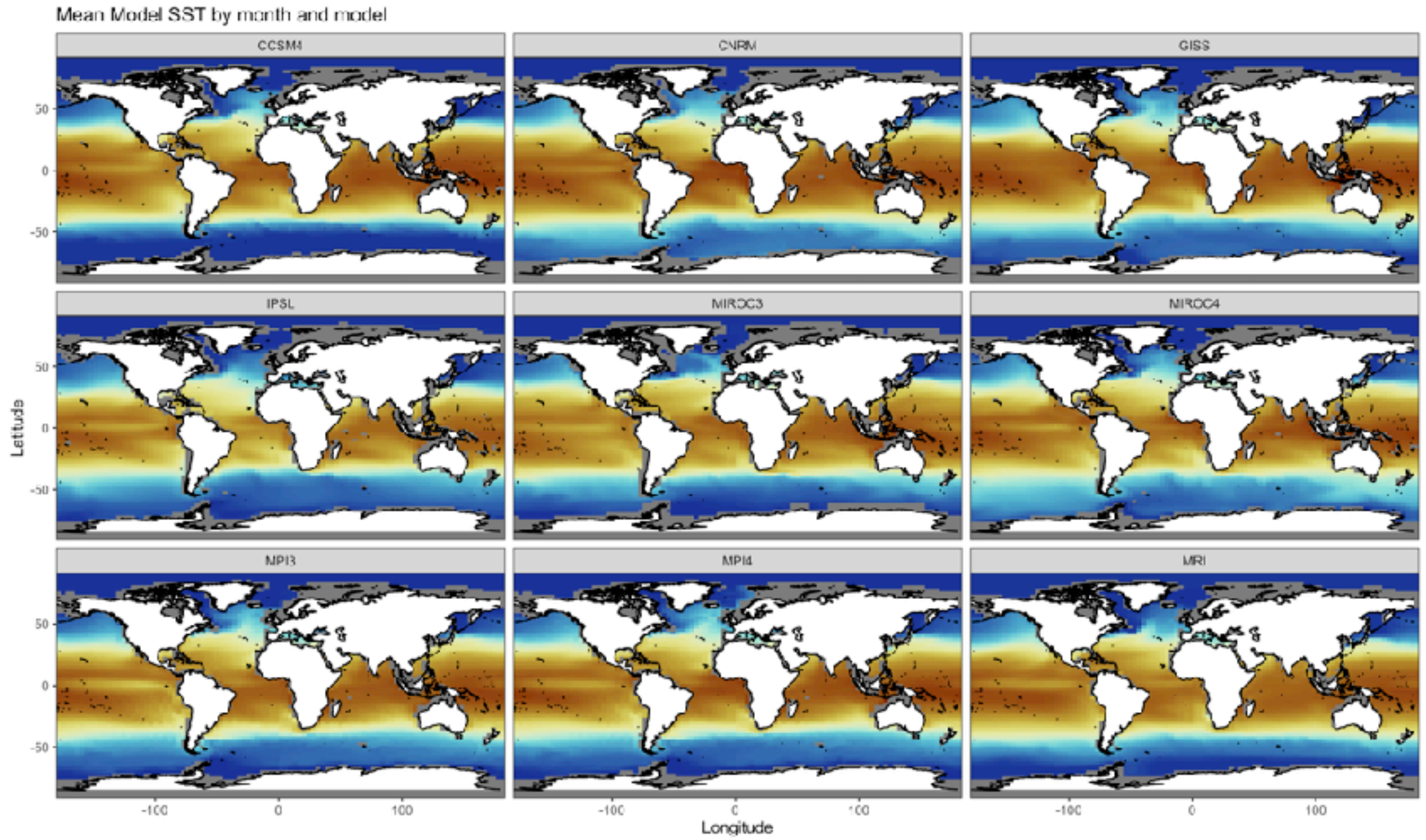


+

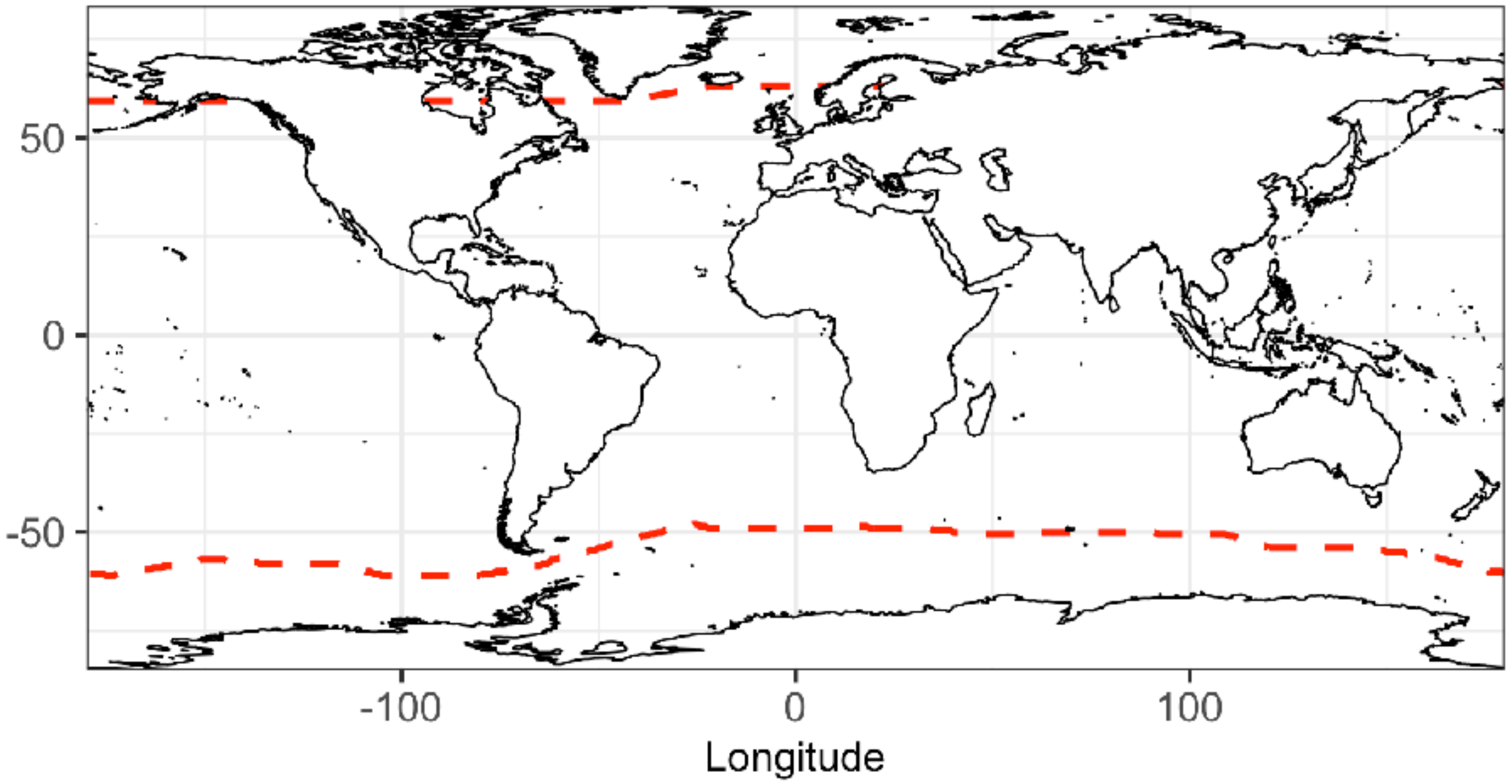
Model Runs

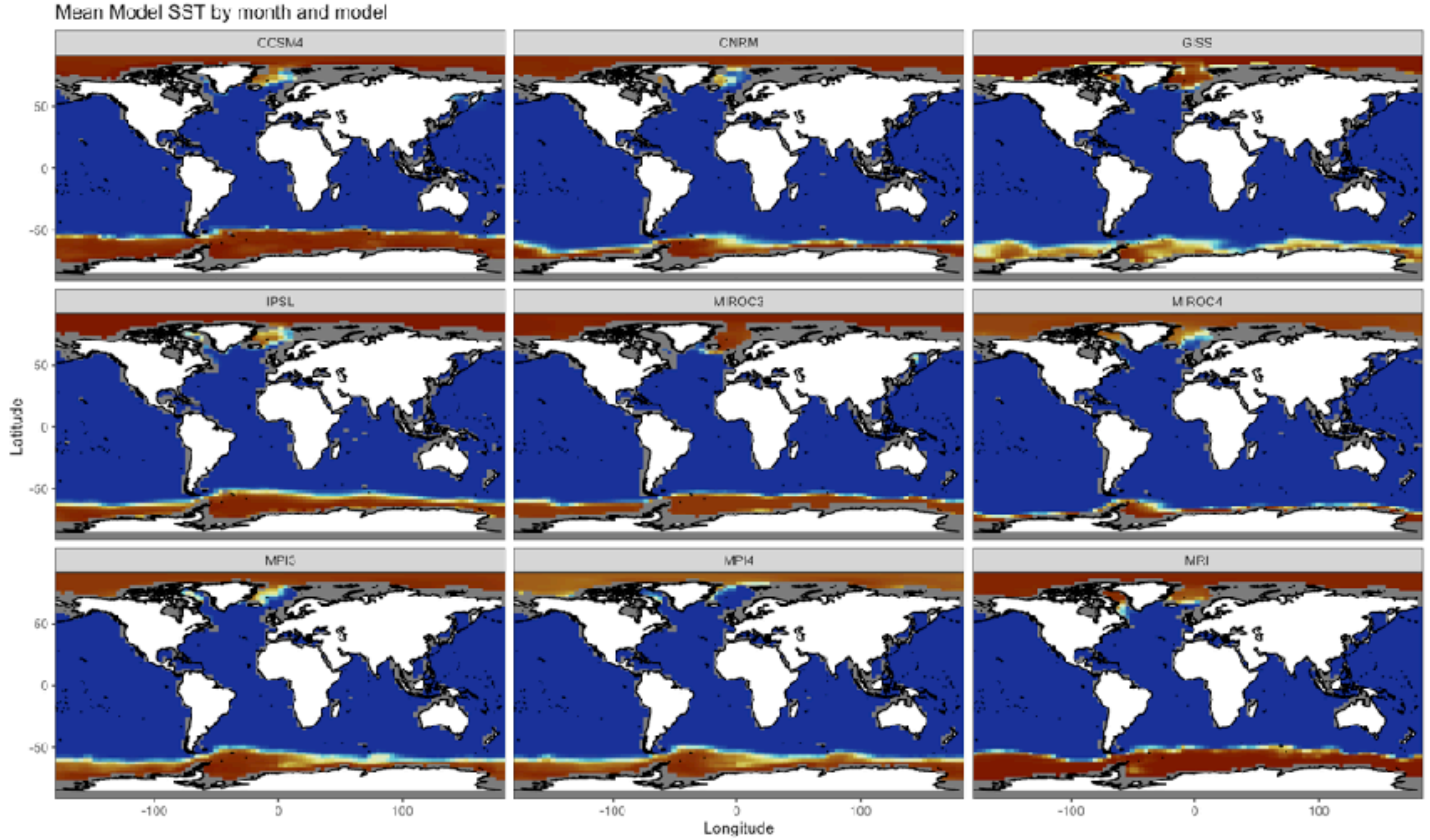



50
0
-50
-100
0
100
Longitude

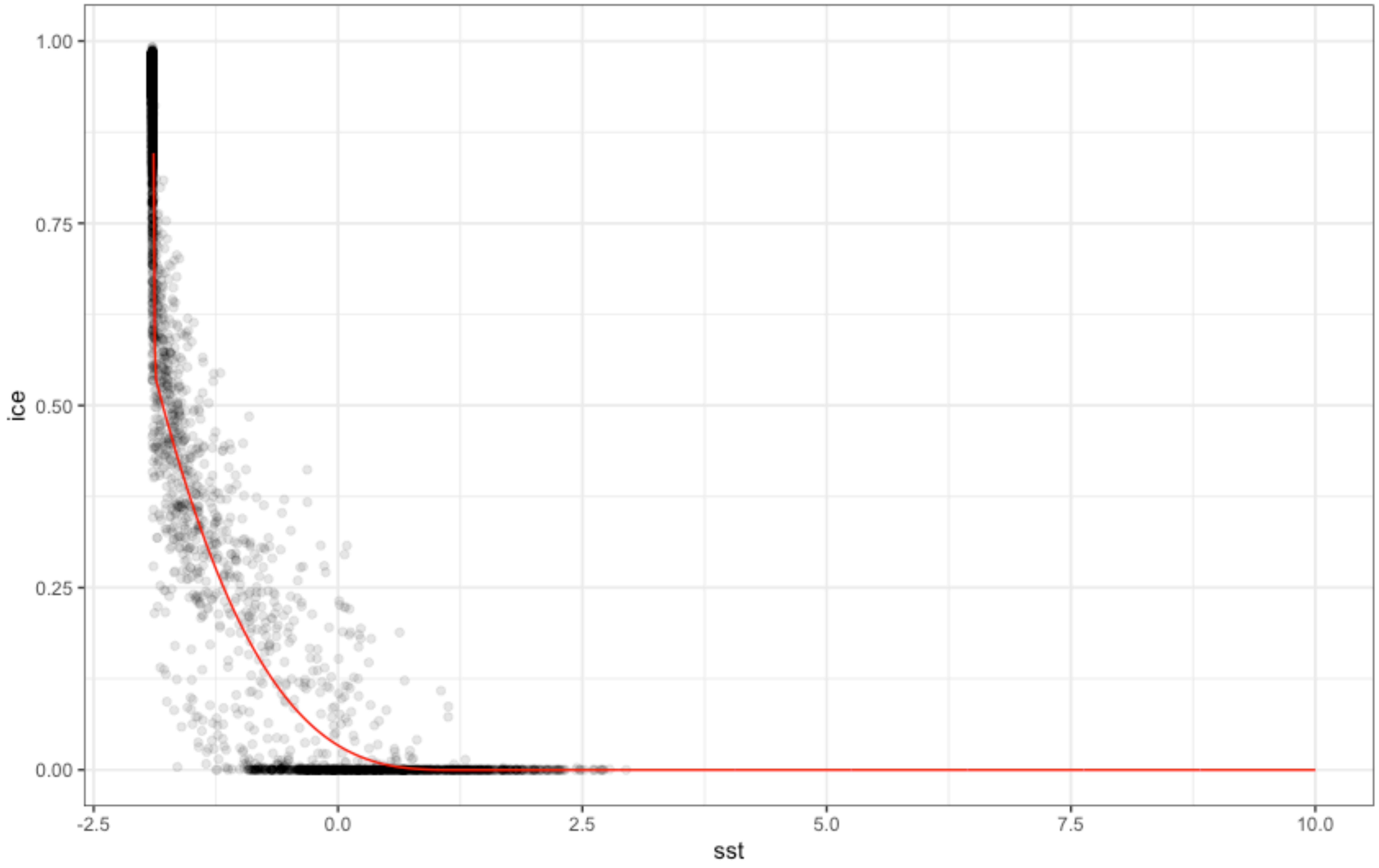

Mean Model SST by month and model
CCSM4
CNRM
GISS
IPSL
MIROC3
MIROC4
MPI3
MPI4
MRI
Latitude
Longitude

+

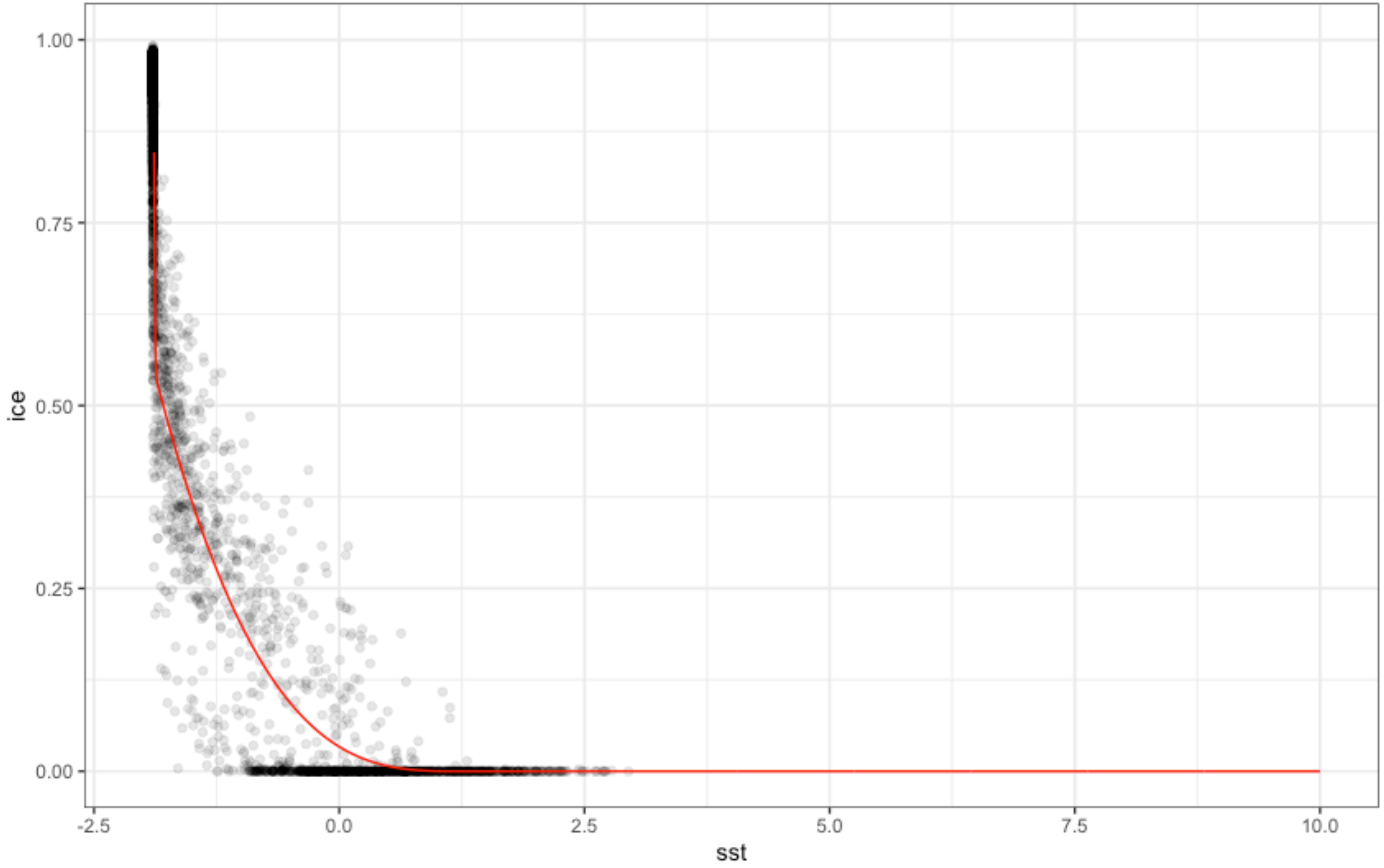

50
0
-50
-100
0
100
Longitude


Mean Model SST by month and model
CCSM4
CNRM
GISS
IPSL
MIROC3
MIROC4
MPI3
MPI4
MRI
Latitude
Longitude

Joint behaviour of SST and SIC



Joint behaviour of SST and SIC



We have this at every grid cell in the model

$$
\begin{tikzcd}
\textbf{SIC:} & & Y_1 & & \cdots & & Y_m & \\
& & \Phi(X_1)\beta_1 \arrow[u] & & \cdots \arrow[u] & & \Phi(X_m)\beta_m \arrow[u] & \\
\textbf{SST:} & X_1 \arrow[ur] & & & \cdots \arrow[u] & & & X_m \arrow[ul] \\
& & & \beta_1 \arrow[uul] & & \beta_m \arrow[uur] & & \\
& & & & \mathrm{M}(\beta) \arrow[ul] \arrow[uu] \arrow[ur] & & &
\end{tikzcd}
$$

Adjusting Beliefs of $M(\beta)$

$$\begin{bmatrix} Y_1 \\ \vdots \\ Y_m \\ \beta_1 \\ \vdots \\ \beta_m \end{bmatrix} = \left[\begin{array}{ccc|c} \Phi_1 & \cdots & 0 & \\ \vdots & \ddots & \vdots & 0 \\ 0 & \cdots & \Phi_m & \\ \hline & 0 & & \mathbf{J}_{m\times 1} \otimes \mathbf{I} \end{array}\right] \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_m \\ \mathcal{M}(\beta) \end{bmatrix} + \begin{bmatrix} \mathcal{R}_1(Y) \\ \vdots \\ \mathcal{R}_m(Y) \\ \mathcal{R}_1(\beta) \\ \vdots \\ \mathcal{R}_m(\beta) \end{bmatrix}$$

Adjusting Beliefs of $M(\beta)$

$$\begin{bmatrix} Y_1 \\ \vdots \\ Y_m \\ \beta_1 \\ \vdots \\ \beta_m \end{bmatrix} = \left[\begin{array}{ccc|c} \Phi_1 & \cdots & 0 & \\ \vdots & \ddots & \vdots & 0 \\ 0 & \cdots & \Phi_m & \\ \hline & 0 & & \mathbf{J}_{m\times 1} \otimes \mathbf{I} \end{array}\right] \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_m \\ \mathcal{M}(\beta) \end{bmatrix} + \begin{bmatrix} \mathcal{R}_1(Y) \\ \vdots \\ \mathcal{R}_m(Y) \\ \mathcal{R}_1(\beta) \\ \vdots \\ \mathcal{R}_m(\beta) \end{bmatrix}$$

Adjusting Beliefs of $M(\beta)$

Following Hodges (1998), note that $0 = M(\beta) - \beta_i + R_i(\beta)$, and so with some manipulation

$$\begin{bmatrix} Y_1 \\ \vdots \\ Y_m \\ 0_{km\times 1} \end{bmatrix} = \left[\begin{array}{ccc|c} \Phi_1 & \cdots & 0 & \\ \vdots & \ddots & \vdots & 0 \\ 0 & \cdots & \Phi_m & \\ \hline & -\mathbf{I}_{km} & & \mathbf{J}_{m\times 1} \otimes \mathbf{I}_k \end{array}\right] \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_m \\ \mathcal{M}(\beta) \end{bmatrix} + \begin{bmatrix} \mathcal{R}_1(Y) \\ \vdots \\ \mathcal{R}_m(Y) \\ \mathcal{R}_1(\beta) \\ \vdots \\ \mathcal{R}_m(\beta) \end{bmatrix}$$

Now let’s make it fast

Define $\hat{\Phi}_i = (\Phi_i^\intercal \Phi_i)^{-1} \Phi_1^\intercal Y_1$ as the projection of Y_i onto the column space of Φ_i

$$
\begin{bmatrix} \hat{\Phi}_1 \\ \vdots \\ \hat{\Phi}_m \\ 0 \end{bmatrix}
=
\left[\begin{array}{ccc|c}
\mathbf{I}_k & \cdots & 0 & \\
\vdots & \ddots & \vdots & 0 \\
0 & \cdots & \mathbf{I}_k & \\
\hline
\multicolumn{3}{c|}{-\mathbf{I}_{km}} & \mathbf{J}_{m\times 1} \otimes \mathbf{I}_k
\end{array}\right]
\begin{bmatrix} \beta_1 \\ \vdots \\ \beta_m \\ \mathcal{M}(\beta) \end{bmatrix}
+
\begin{bmatrix} \mathcal{R}_1(Y) \\ \vdots \\ \mathcal{R}_m(Y) \\ \mathcal{R}_1(\beta) \\ \vdots \\ \mathcal{R}_m(\beta) \end{bmatrix}
$$

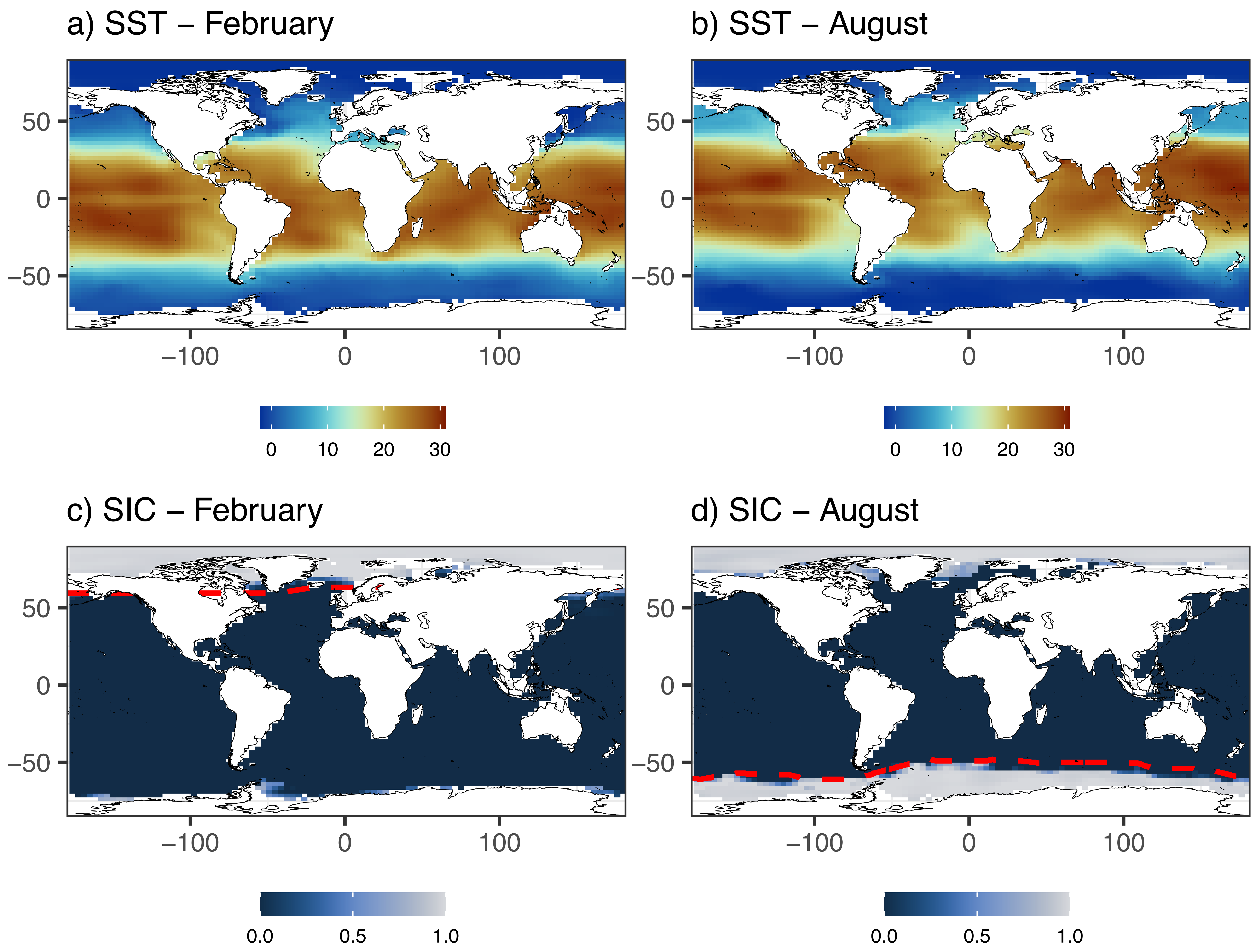
The statistical model

SST	SIC
$\mathbf{X}_i = \mathcal{M}(\mathbf{X}) + \mathcal{R}_i(\mathbf{X})$	$\mathbf{Y}_i = \Phi_{\mathbf{X}_i}\beta_i + \epsilon_i$ $\beta_i = \mathcal{M}(\beta) + \mathcal{R}_i(\beta)$
$\mathbf{T}_\mathbf{X} = \mathcal{M}(\mathbf{X}) + \mathbf{U}_\mathbf{X}$	$\mathbf{T}_\mathbf{Y} = \Phi_{\mathbf{T}_\mathbf{X}}\mathcal{M}(\beta) + \mathbf{U}_\mathbf{Y}$
$\mathbf{Z} = \mathbf{H}\mathbf{T}_\mathbf{X} + \mathbf{W}$	

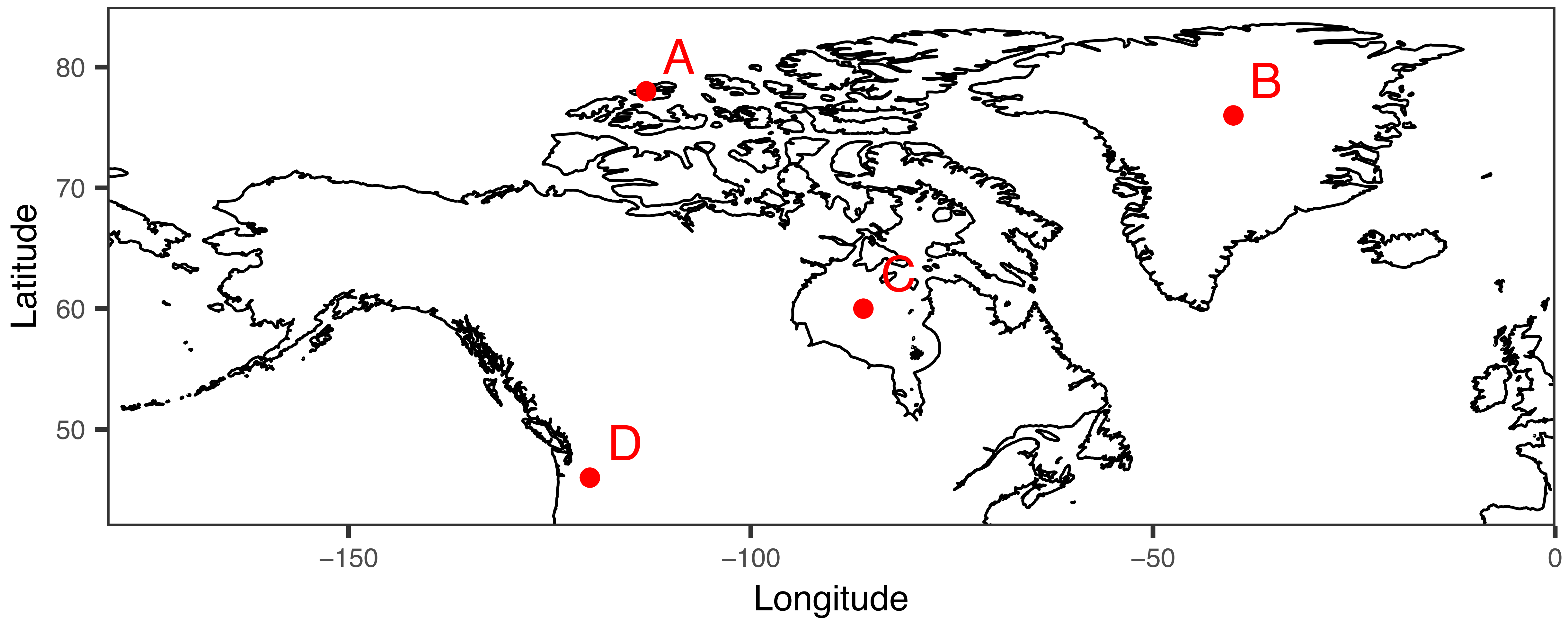
The coexchangeable model of Rougier et al. (2013)

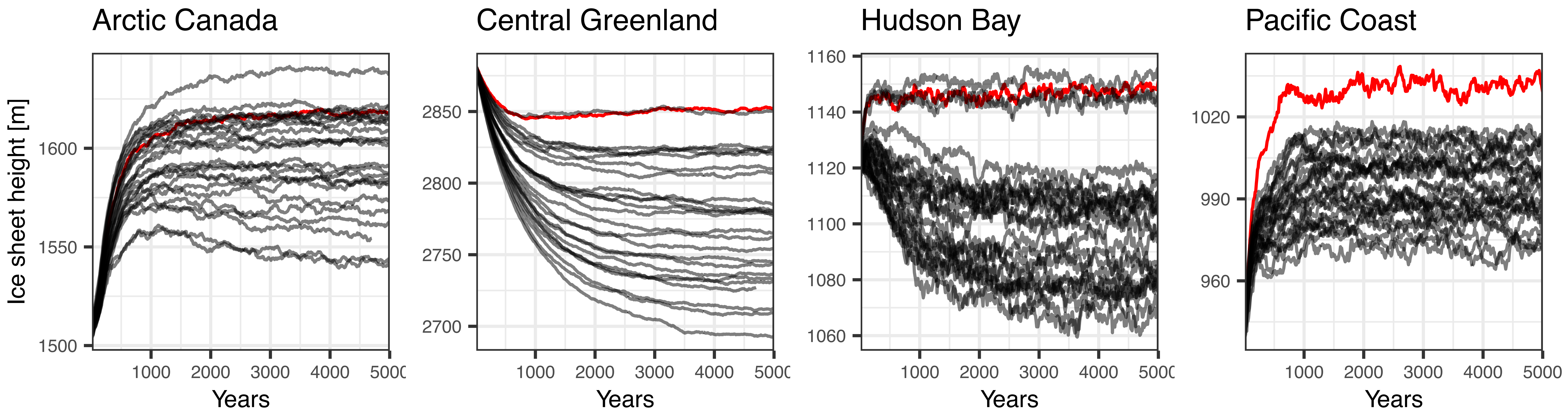
The coexchangeable process model of Astfalck et al. (2024)

Reconstructions of SST and SIC



Effects on the ice-sheet





Generalising Bayes Linear

A generalised Bayes inference

Property 1: An underlying geometry $\mathscr{G}$, establishing the space in which inference takes place

Property 2: A notion of closeness between objects in $\mathscr{G}$ to relate beliefs and data

Property 3: An optimisation, within solution space C, for the closest belief representation to the data generating process

Bayes Linear

Probabilistic Bayes

Bayes Linear

The product inner product

$$\langle X, Y \rangle = E[X^{\intercal} Y]$$

Probabilistic Bayes

The $\mathscr{L}_2$ inner product

$$\langle f(\theta), g(\theta) \rangle = \int f(\theta) g(\theta) \, \mathrm{d}\mu$$

Property 1: An underlying geometry $\mathscr{G}$, establishing the space in which inference takes place

Bayes Linear

Probabilistic Bayes

The product inner product

$$\langle \mathrm{X}, \mathrm{Y} \rangle = \mathrm{E}[\mathrm{X}^{\intercal}\mathrm{Y}]$$

↔

The $\mathscr{L}_2$ inner product

$$\langle f(\theta), g(\theta) \rangle = \int f(\theta) g(\theta) \, \mathrm{d}\mu$$

The belief structure, $\mathscr{B}$ ↔ Probability measure P

Property 2: A notion of closeness between objects in $\mathscr{G}$ to relate beliefs and data

Bayes Linear		Probabilistic Bayes
The product inner product $\langle \mathrm{X}, \mathrm{Y} \rangle = \mathrm{E}[\mathrm{X}^\intercal \mathrm{Y}]$	$\longleftrightarrow$	The $\mathscr{L}_2$ inner product $\langle f(\theta), g(\theta) \rangle = \int f(\theta) g(\theta)\, \mathrm{d}\mu$
The belief structure, $\mathscr{B}$	$\longleftrightarrow$	Probability measure P
Affine space of D $\mathrm{E}_\mathrm{D}[\mathrm{X}] = \mathrm{h} + \mathrm{HD}$	$\longleftrightarrow$	Posterior distributions Π

Property 3: An optimisation, within solution space C, for the closest belief representation to the data generating process

Bayes as optimisation

Bayes as optimisation

Bissiri et al. (2016) recast probabilistic Bayes as the solution to

$$q^*(\theta) = \arg\min_{q \in \Pi} \left\{ \mathrm{E}_{q(\theta)} \left[\sum_{i=1}^{n} l(\theta, x_i) \right] + \mathrm{KLD}(q \| \pi) \right\}$$

Bayes as optimisation

Bissiri et al. (2016) recast probabilistic Bayes as the solution to

$$q^*(\theta) = \arg\min_{q \in \Pi} \left\{ \mathrm{E}_{q(\theta)} \left[\sum_{i=1}^{n} l(\theta, x_i) \right] + \mathrm{KLD}(q \| \pi) \right\}$$

The Bayes update is the solution of an optimisation that seeks the posterior distribution in Π that minimises the divergence from the data generating process.

Bayes as optimisation

Bissiri et al. (2016) recast probabilistic Bayes as the solution to

$$q^*(\theta) = \arg\min_{q \in \Pi} \left\{ \mathrm{E}_{q(\theta)} \left[\sum_{i=1}^{n} l(\theta, x_i) \right] + \mathrm{KLD}(q \| \pi) \right\}$$

The Bayes update is the solution of an optimisation that seeks the posterior distribution in Π that minimises the divergence from the data generating process.

This provides an immediate connection to Bayes linear methods being the solution of a (different) optimisation problem.

Bayes as optimisation

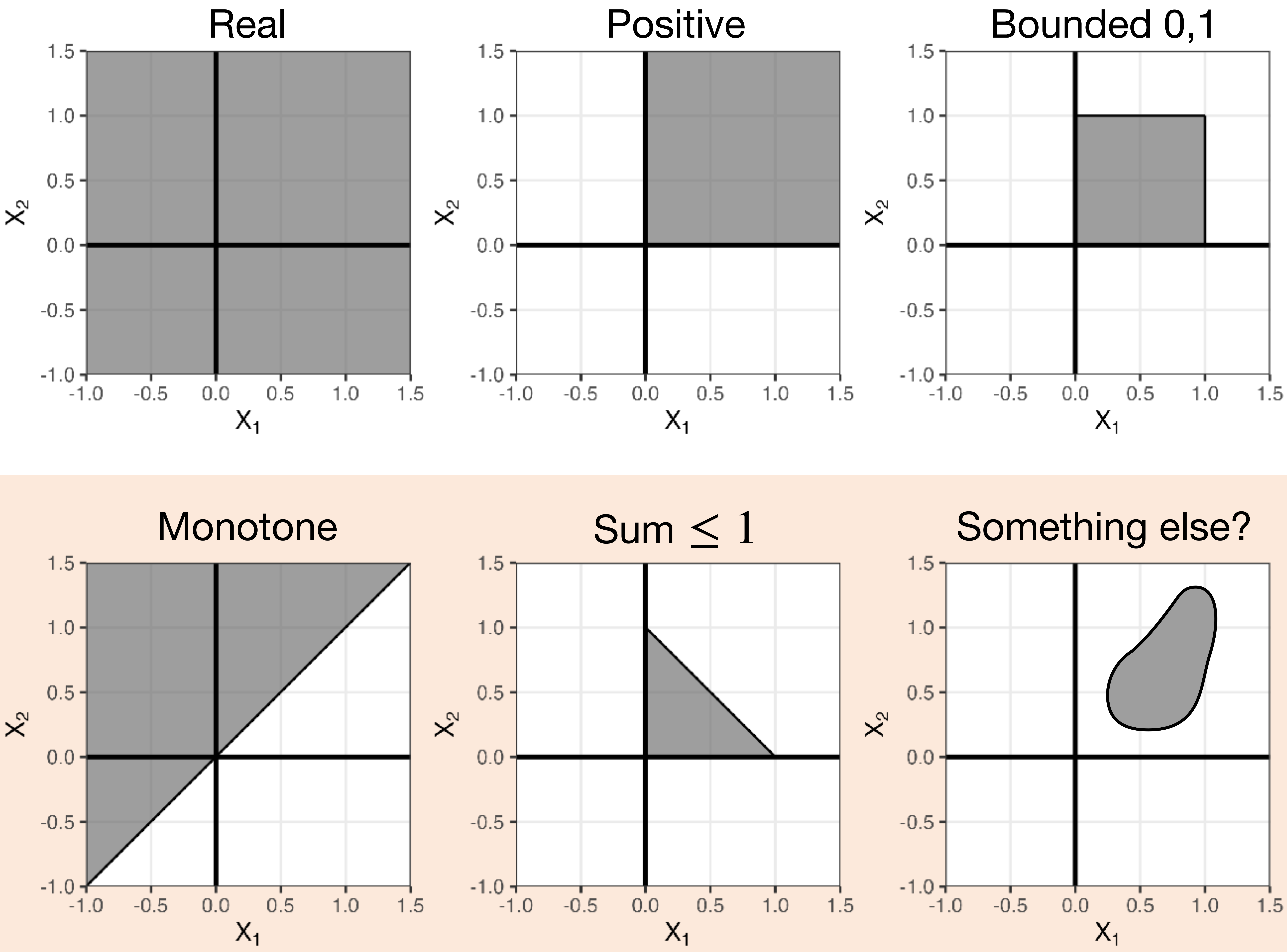
Bissiri et al. (2016) recast probabilistic Bayes as the solution to

$$q^*(\theta) = \arg\min_{q \in \Pi} \left\{ \mathrm{E}_{q(\theta)} \left[\sum_{i=1}^{n} l(\theta, x_i) \right] + \mathrm{KLD}(q \| \pi) \right\}$$

The Bayes update is the solution of an optimisation that seeks the posterior distribution in Π that minimises the divergence from the data generating process.

This provides an immediate connection to Bayes linear methods being the solution of a (different) optimisation problem.

What do we achieve by playing with Π?



Real
Positive
Bounded 0,1
Monotone
Sum ≤ 1
Something else?
X_1
X_2
-1.0
-0.5
0.0
0.5
1.0
1.5

Inference with constrained solutions

Inference with constrained solutions

In a probabilistic Bayesian analysis we generally handle this in two ways:

1. Assign zero weight to regions in the prior (or equivalently, add a rejection step into the MCMC).

2. Transform your data/model.

Inference with constrained solutions

In a probabilistic Bayesian analysis we generally handle this in two ways:

1. Assign zero weight to regions in the prior (or equivalently, add a rejection step into the MCMC).

2. Transform your data/model.

Bayes linear inference orthogonally projects X into the affine subspace of D

$$E_D[X] = \underset{h+HD}{\arg\min} \left\{ \langle X - h - HD, h + HD \rangle \right\}$$

Inference with constrained solutions

In a probabilistic Bayesian analysis we generally handle this in two ways:

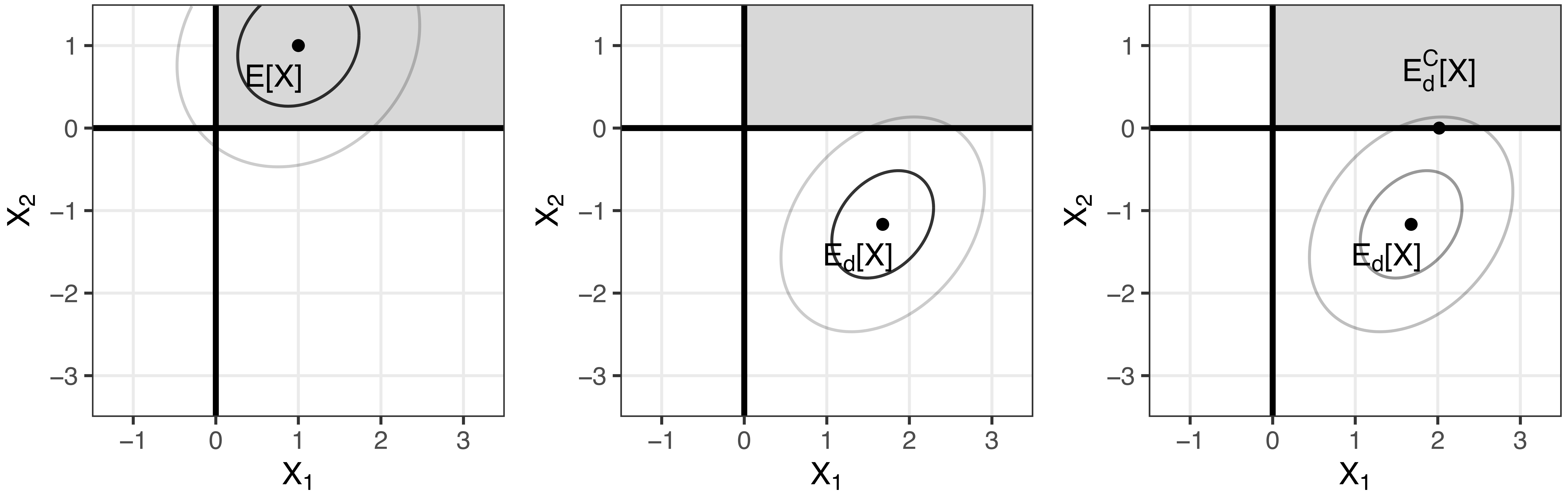
1. Assign zero weight to regions in the prior (or equivalently, add a rejection step into the MCMC).

2. Transform your data/model.

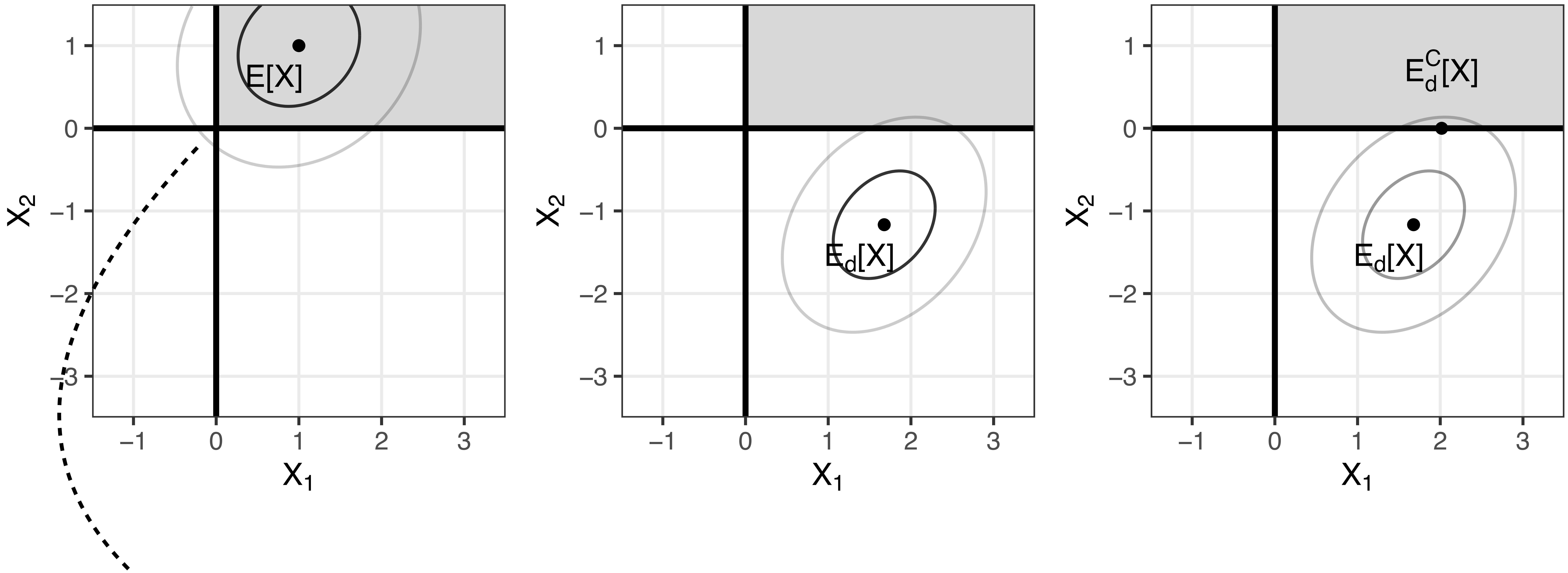
Bayes linear inference orthogonally projects X into the affine subspace of D

$$\mathrm{E_D[X]} = \underset{\mathrm{h+HD}}{\arg\min} \left\{ \langle \mathrm{X} - \mathrm{h} - \mathrm{HD}, \mathrm{h} + \mathrm{HD} \rangle \right\}$$

Constrain the solution to lie in some subset C and call this quantity $\mathrm{E}_{\mathrm{d}}^{C}[\mathrm{X}]$. Note, $\mathrm{E}_{\mathrm{d}}^{C}[\mathrm{X}]$ is not necessarily affine in D.

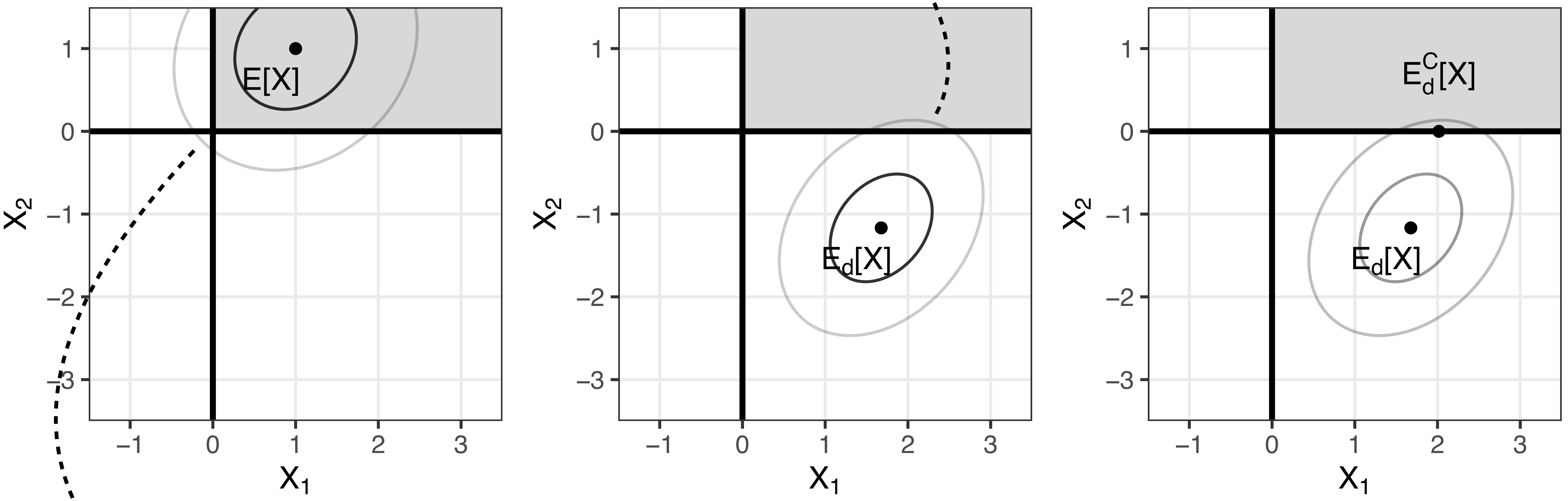


E[X]
$E_d[X]$
$E_d^C[X]$
X_1
X_2



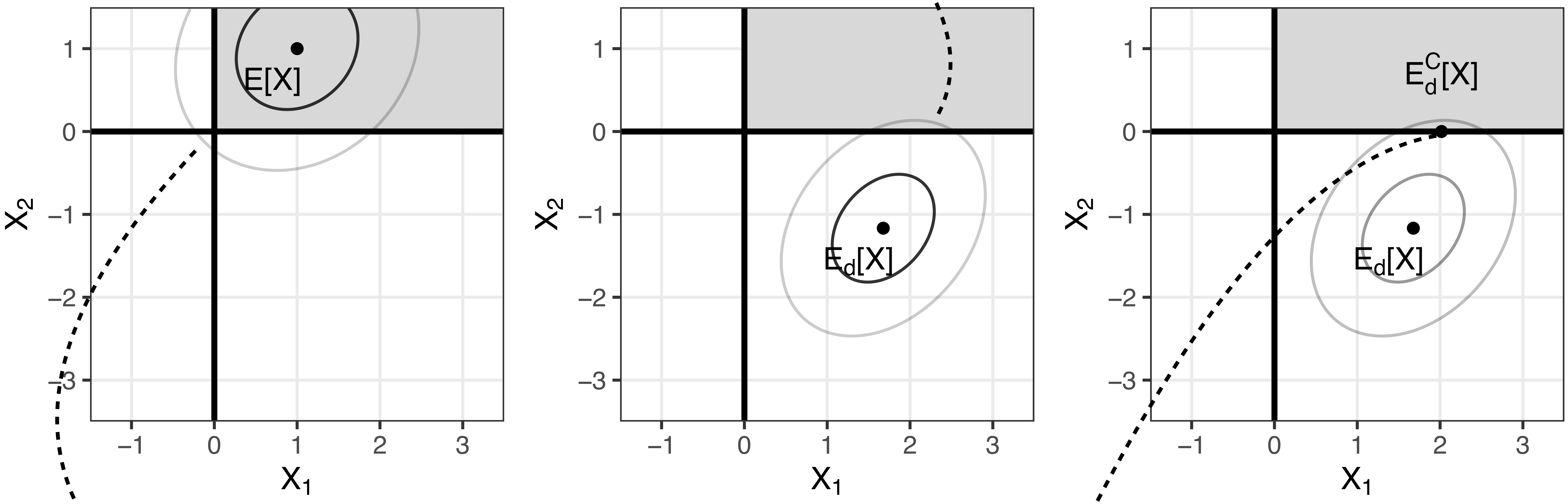
$$\langle X, Y \rangle = E[X^\intercal Y]$$

$$\langle X, Y\rangle_D = \langle X - E_D[X], Y - E_D[Y]\rangle$$



$$\langle X, Y\rangle = E[X^\intercal Y]$$

$$\langle X, Y\rangle_D = \langle X - E_D[X], Y - E_D[Y]\rangle$$



$$\langle X, Y\rangle = E[X^\intercal Y]$$

$$E_d^C[X] = \underset{q\in C}{\arg\min}\|E_d[X] - q\|$$

Generalised Adjusted Variance

Generalised Adjusted Variance

Adjusted variance is calculated from an outer product assuming affine $\mathrm{E}_{\mathrm{D}}[\mathrm{X}]$. If $\mathrm{E}_{\mathrm{d}}^{C}[\mathrm{X}] \neq \mathrm{E}_{\mathrm{d}}[\mathrm{X}]$, we break the affine assumption.

Generalised Adjusted Variance

Adjusted variance is calculated from an outer product assuming affine $\mathrm{E}_{\mathrm{D}}[\mathrm{X}]$. If $\mathrm{E}^{C}_{\mathrm{d}}[\mathrm{X}] \neq \mathrm{E}_{\mathrm{d}}[\mathrm{X}]$, we break the affine assumption.

Define L as a square-root decomposition $\mathrm{var}_{\mathrm{D}}[\mathrm{X}] = \mathrm{L}\mathrm{L}^{\intercal}$ (I like $\mathrm{L} = \mathrm{Q}\sqrt{\Lambda}$) and the *constraint discrepancy* $\mathrm{z} = \mathrm{L}^{-1}(\mathrm{E}^{C}_{\mathrm{d}}[\mathrm{X}] - \mathrm{E}_{\mathrm{d}}[\mathrm{X}])$.

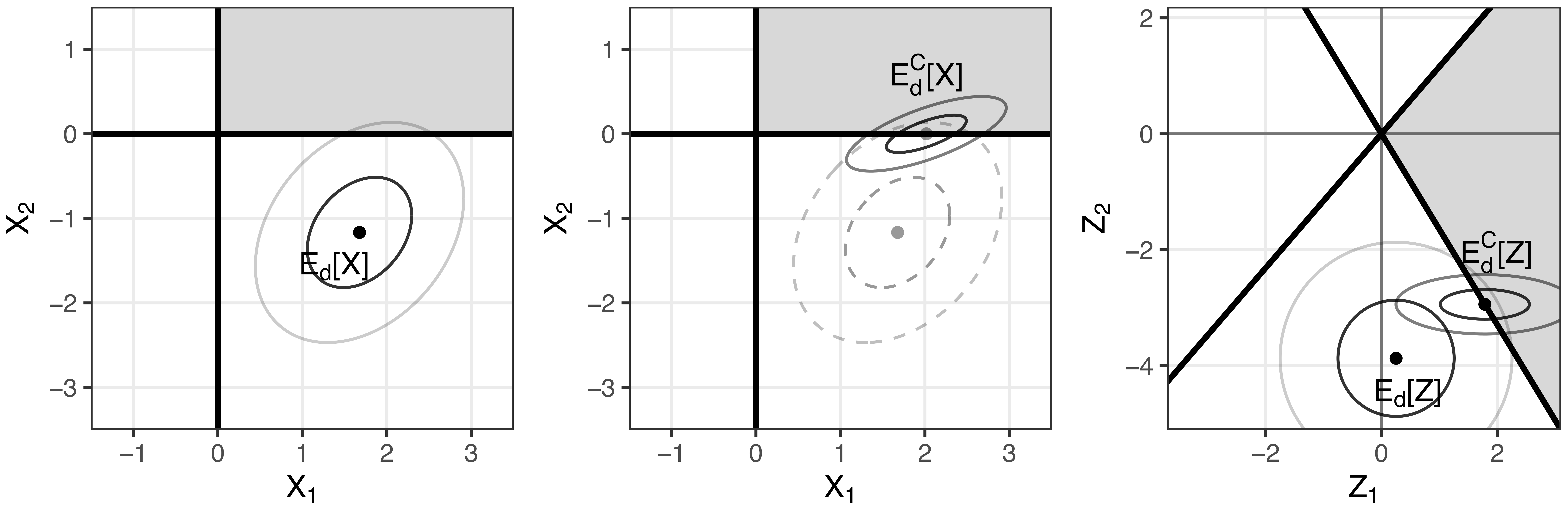
Generalised Adjusted Variance

Adjusted variance is calculated from an outer product assuming affine $\mathrm{E_D}[\mathrm{X}]$. If $\mathrm{E}_\mathrm{d}^{C}[\mathrm{X}] \neq \mathrm{E_d}[\mathrm{X}]$, we break the affine assumption.

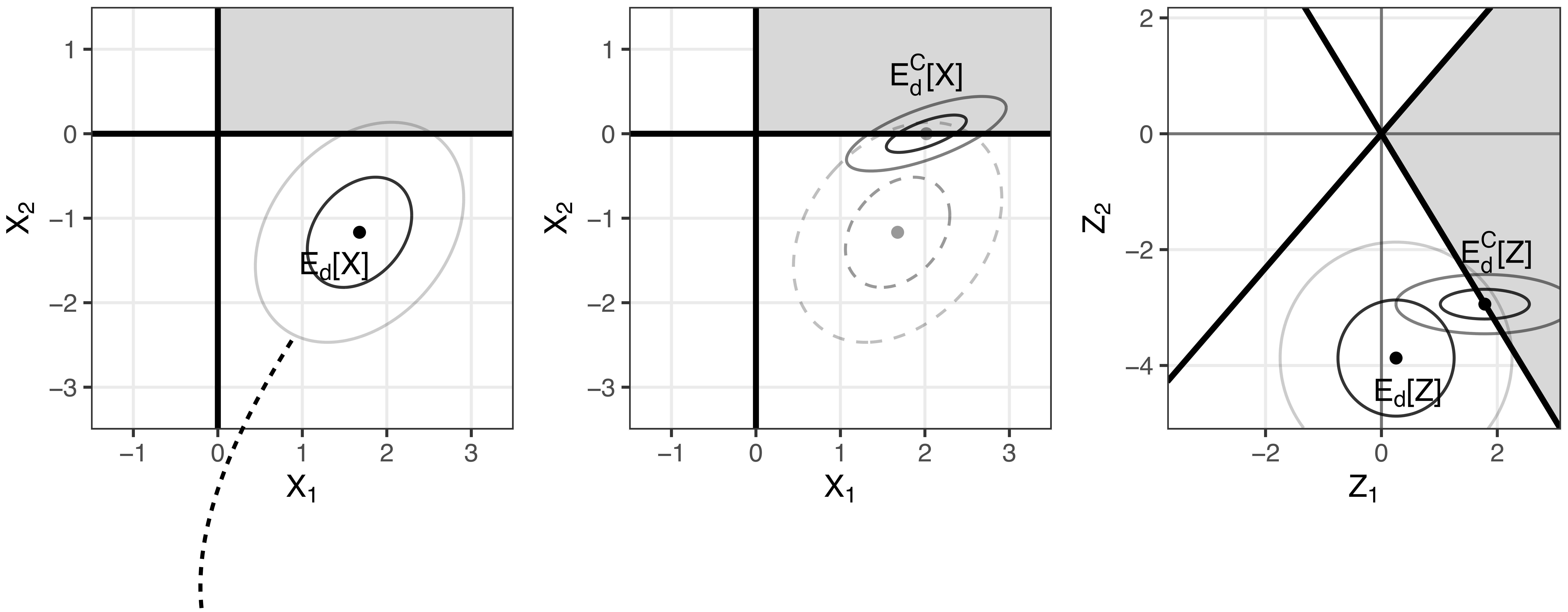
Define L as a square-root decomposition $\mathrm{var_D}[\mathrm{X}] = \mathrm{LL}^\intercal$ (I like $\mathrm{L} = \mathrm{Q}\sqrt{\Lambda}$) and the *constraint discrepancy* $\mathrm{z} = \mathrm{L}^{-1}(\mathrm{E}_\mathrm{d}^{C}[\mathrm{X}] - \mathrm{E_d}[\mathrm{X}])$.

The generalised adjusted variance is $\mathrm{var}_\mathrm{d}^{C}[\mathrm{X}] = \mathrm{LSL}^\intercal$, where

1. The limit $\lim_{|z_i|\to 0} \{\mathrm{S}_{ii}\} = 1$
2. The limit $\lim_{|z_i|\to \infty} \{\mathrm{S}_{ii}\} = 0$
3. $\mathrm{S}_{ii} = f(z_i)$ is non-increasing in z_i

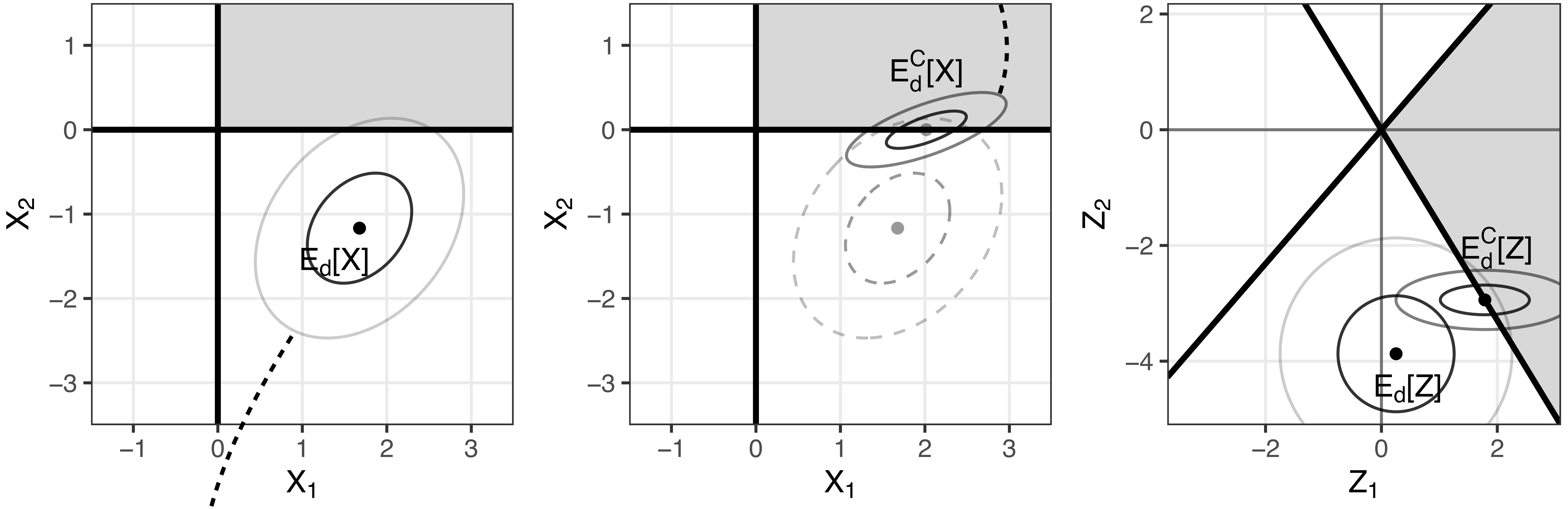


$E_d[X]$
$E_d^C[X]$
$E_d^C[Z]$
$E_d[Z]$
X_1
X_2
Z_1
Z_2



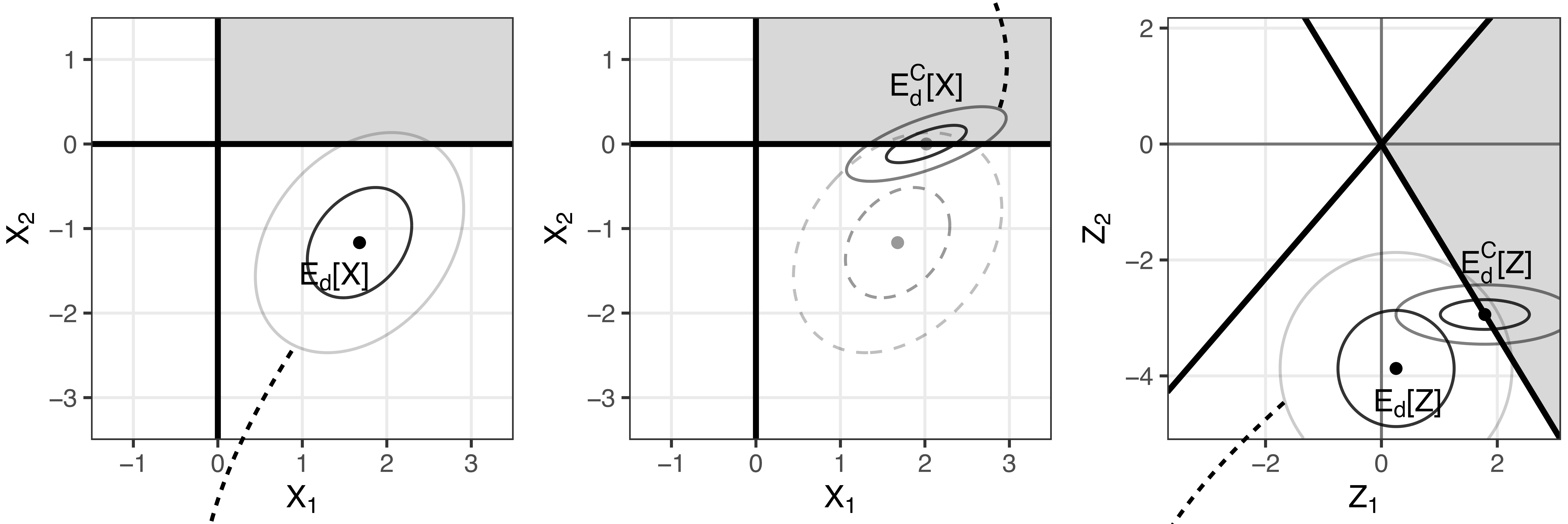
$\mathrm{var}_D[X] = LL^\intercal$, and $L = Q\sqrt{\Lambda}$

$$\mathrm{var}_{\mathrm{d}}^{C}[\mathrm{X}] = \mathrm{LSL}^{\intercal}$$



$\mathrm{var}_{\mathrm{D}}[\mathrm{X}] = \mathrm{LL}^{\intercal}$, and $\mathrm{L} = \mathrm{Q}\sqrt{\Lambda}$

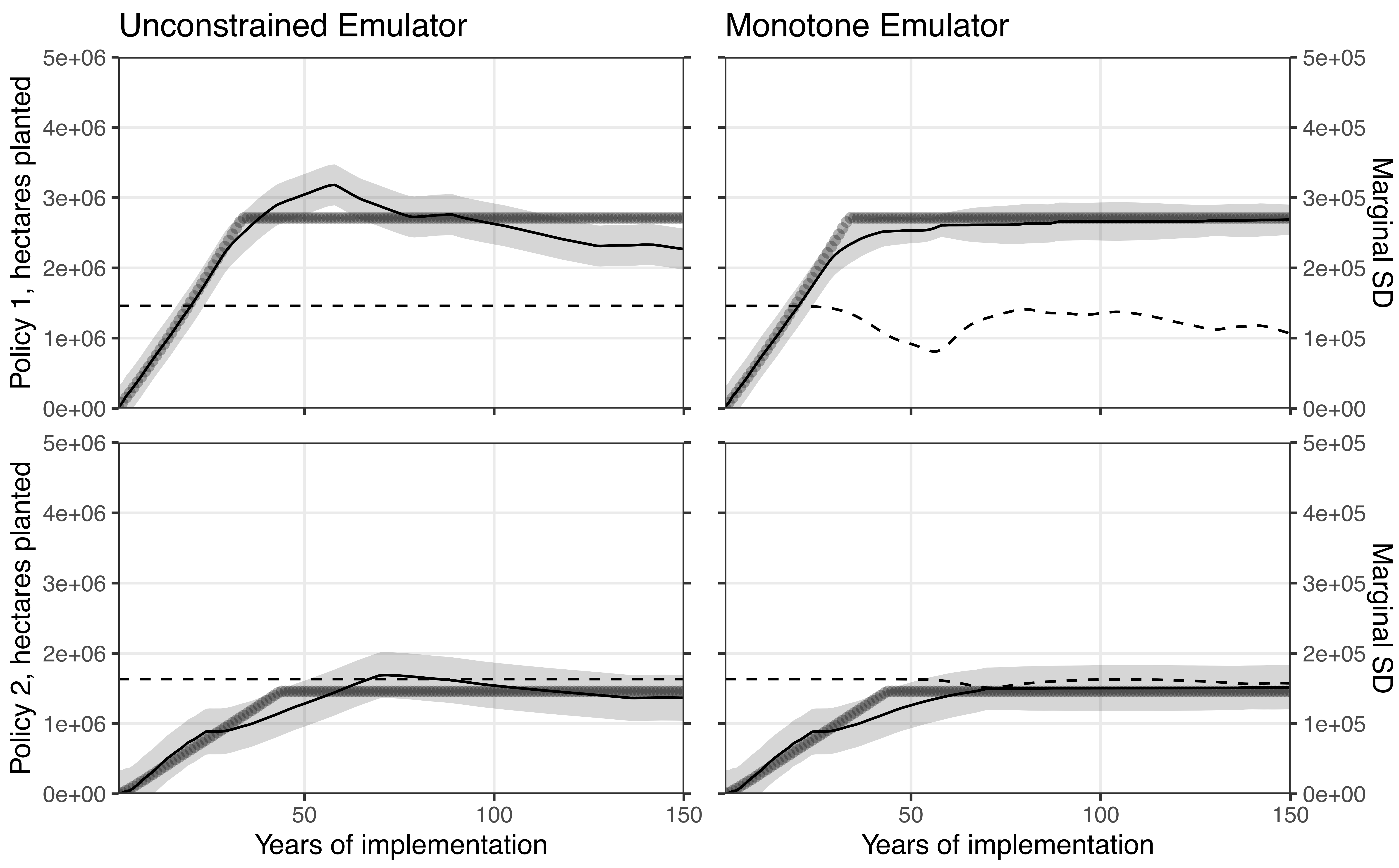
$$\mathrm{var}_{\mathrm{d}}^{C}[\mathrm{X}] = \mathrm{LSL}^{\intercal}$$



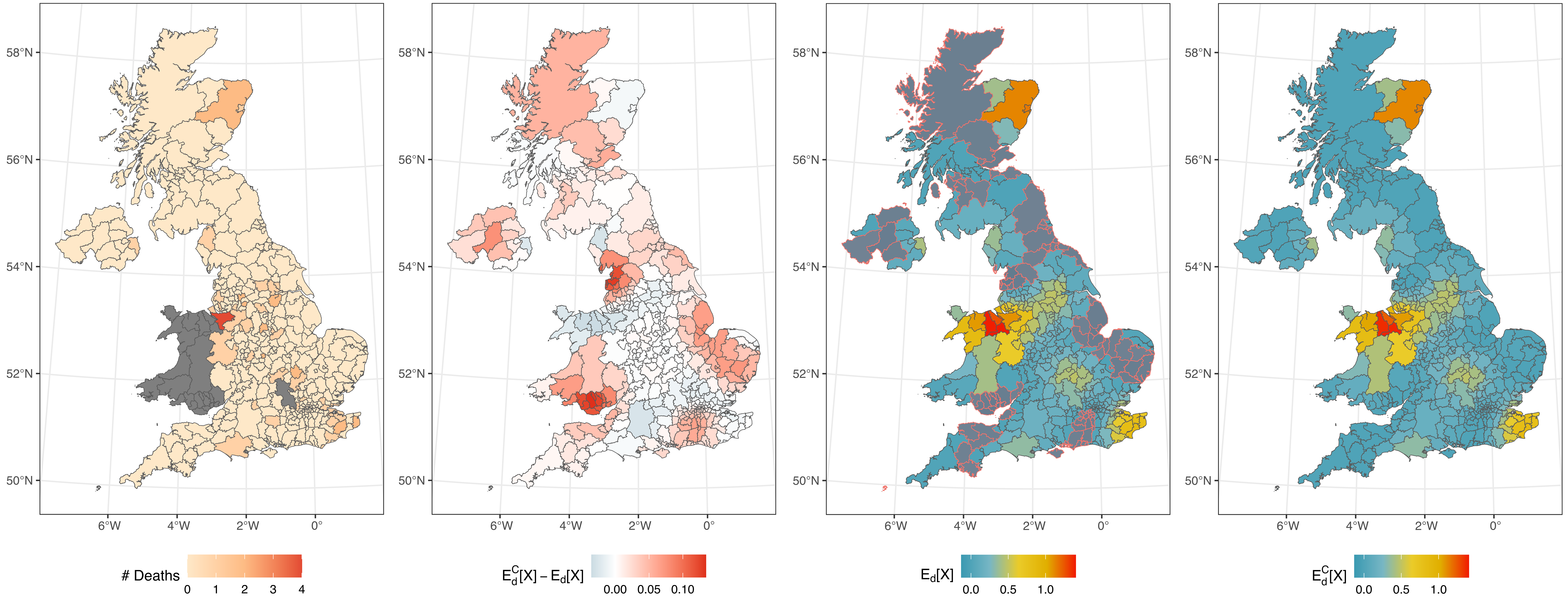
$\mathrm{var}_{\mathrm{D}}[\mathrm{X}] = \mathrm{LL}^{\intercal}$, and $\mathrm{L} = \mathrm{Q}\sqrt{\Lambda}$

Rotate into $\mathscr{Z}$ via L^{-1}

An Afforestation Uptake Model



One day of COVID19 Deaths



- Astfalck, L., Williamson, D., Gandy, N., Gregoire, L., & Ivanovic, R. (2024). Coexchangeable Process Modeling for Uncertainty Quantification in Joint Climate Reconstruction. *Journal of the American Statistical Association*, 1-14.
- Astfalck, L., Bird, C., & Williamson, D. (2024). Generalised Bayes Linear Inference. *arXiv preprint arXiv:2405.14145*.
- Diaconis, P., & Ylvisaker, D. (1979). Conjugate priors for exponential families. *The Annals of statistics*, 269-281.
- De Finetti, B. (1975). Theory of probability: A critical introductory treatment. John Wiley & Sons
- Ericson, W. A. (1969). A note on the posterior mean of a population mean. *Journal of the Royal Statistical Society: Series B (Methodological)*, *31*(2), 332-334.
- Goldstein, M., & Wooff, D. (2007). Bayes linear statistics: Theory and methods. John Wiley & Sons.
- Hartigan, J. A. (1969). Linear bayesian methods. Journal of the Royal Statistical Society: Series B (Methodological), 31(3), 446-454.
- Hodges, J. S. (1998). Some algebra and geometry for hierarchical models, applied to diagnostics. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *60*(3), 497-536.
- Rougier, J., Goldstein, M., & House, L. (2013). Second-order exchangeability analysis for multimodel ensembles. *Journal of the American Statistical Association*, *108*(503), 852-863.

astfalckl.github.io/presentations