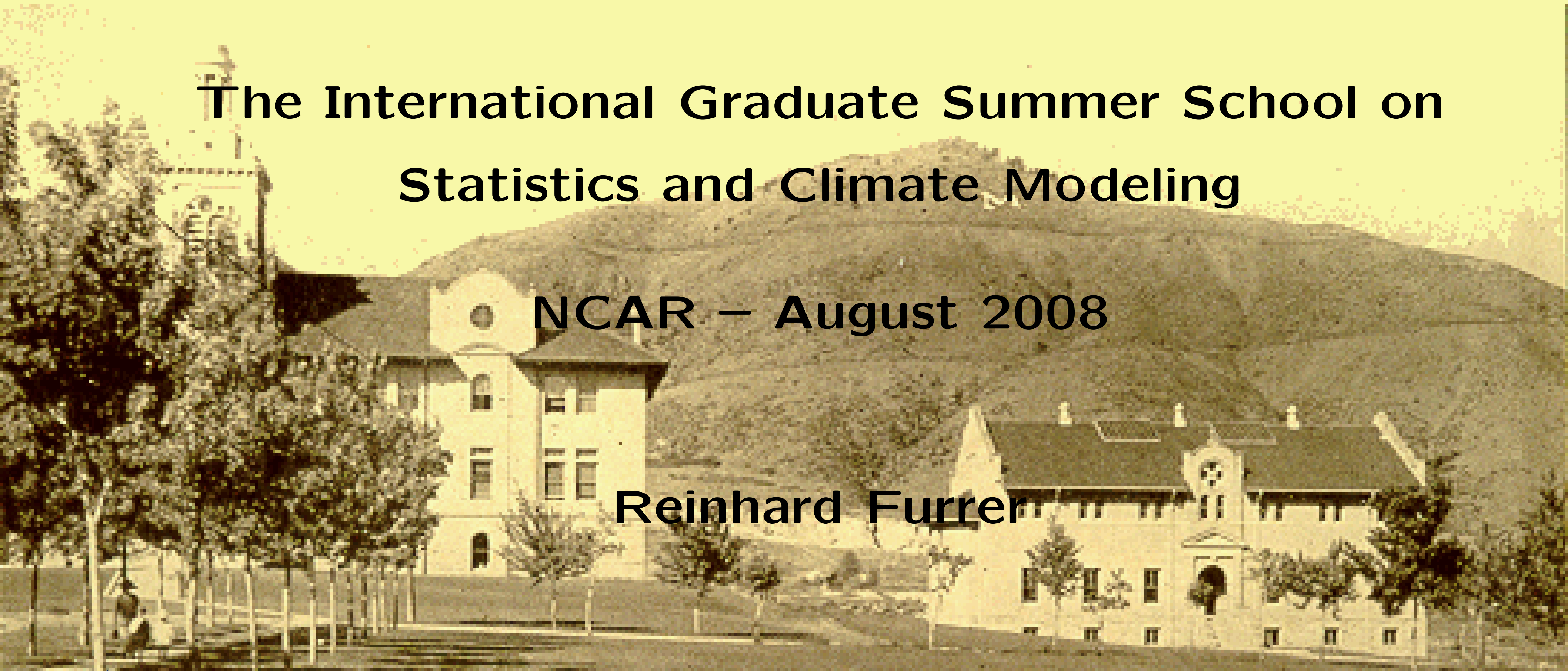
A Bayesian view of climate change: assessing uncertainties of general circulation model projections



The International Graduate Summer School on Statistics and Climate Modeling

NCAR – August 2008

Reinhard Furrer

We present probabilistic projections for spatial patterns of future temperature change using a hierarchical Bayesian model.

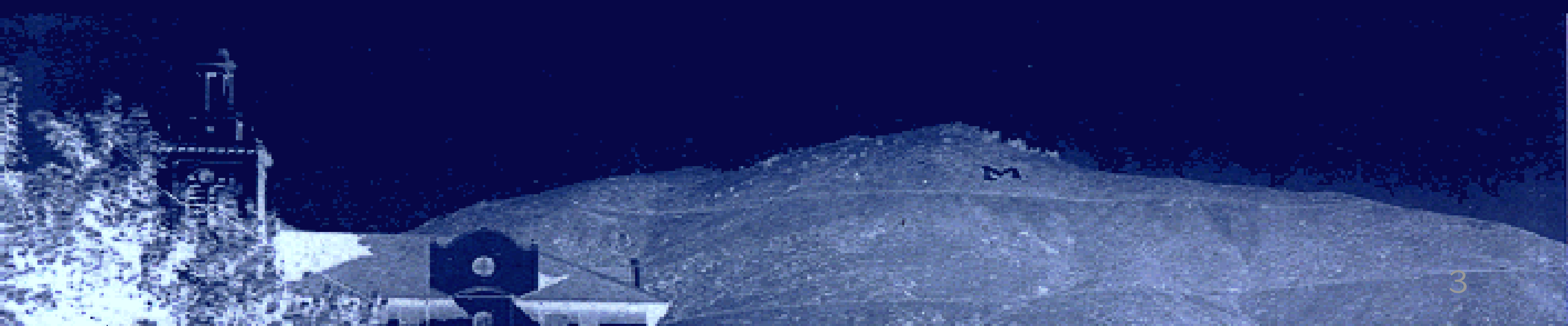
Collaboration with: Reto Knutti - ETHZ

Stephan Sain, Doug Nychka, Claudia Tebaldi, Jerry Meehl, Linda Mearns, . . . - NCAR

NSF DMS-0621118

Outline of the Talk

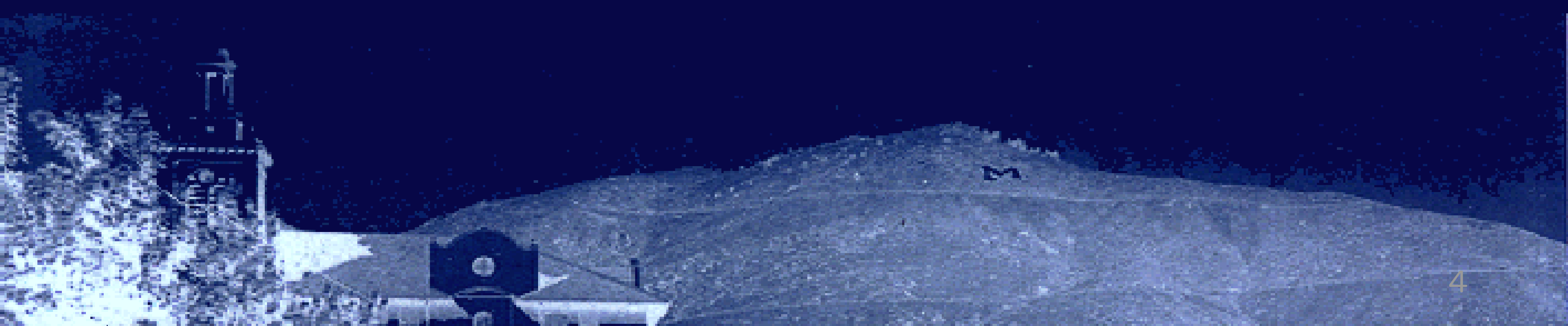
- Climate projection data
- A simple hierarchical Bayesian model
- Presenting uncertainty results
- Model extensions



Studying Climate with AOGCMs

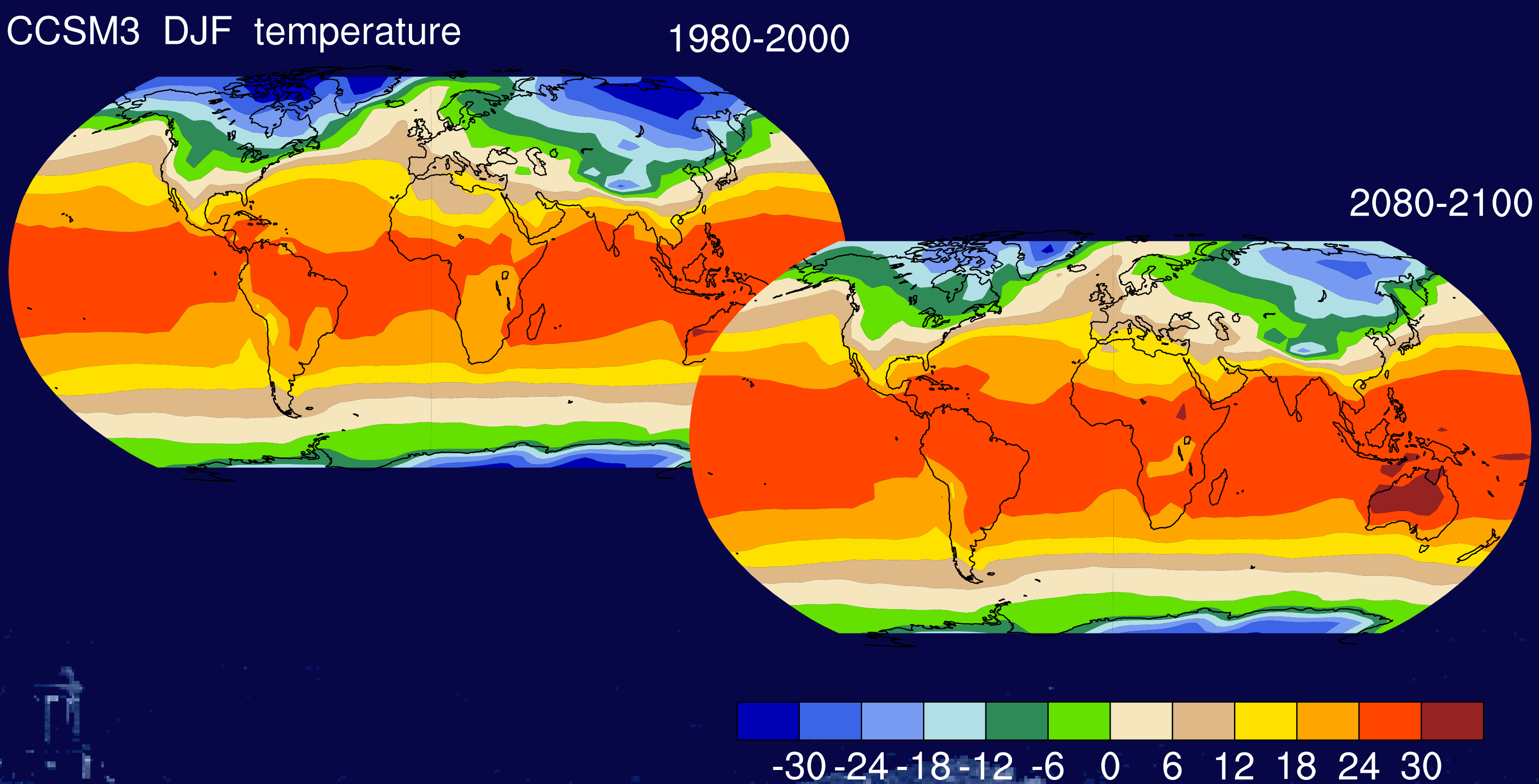
AOGCM: Atmosphere-Ocean General Circulation Models

> Numerical models that calculate the detailed large-scale motions of the atmosphere and the ocean explicitly from hydrodynamical equations.



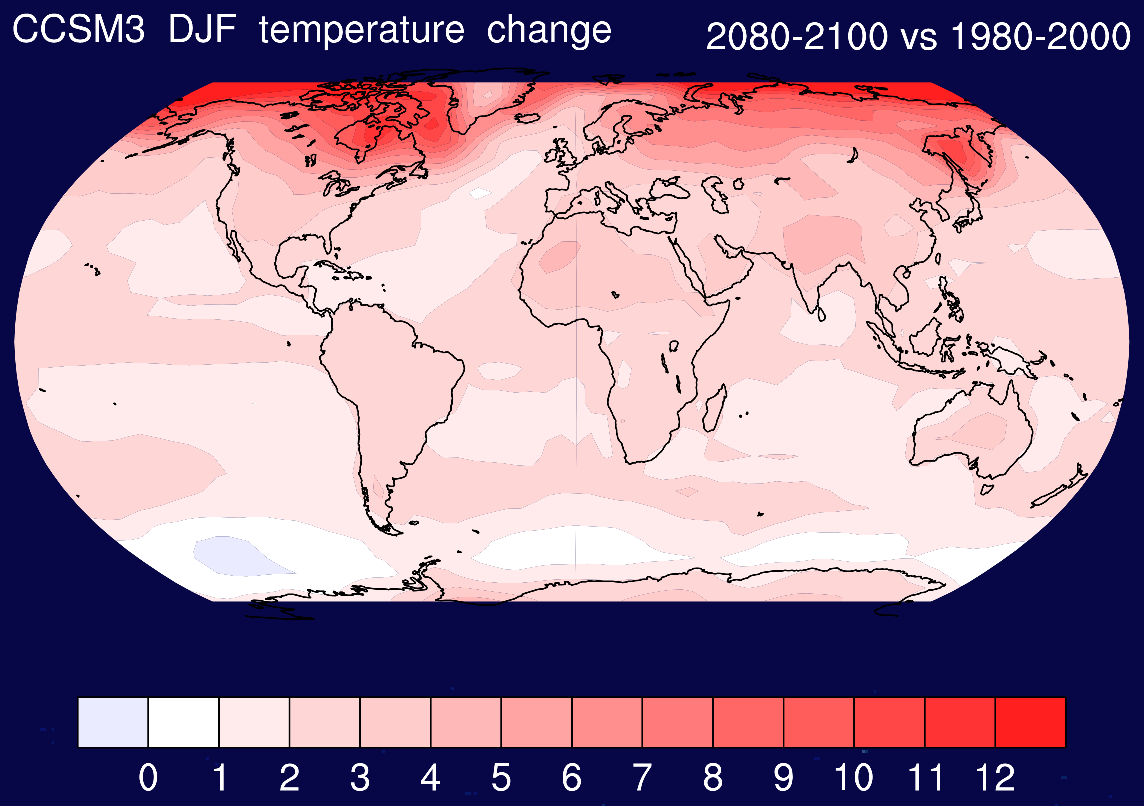
Studying Climate with AOGCMs

AOGCM: Atmosphere-Ocean General Circulation Models

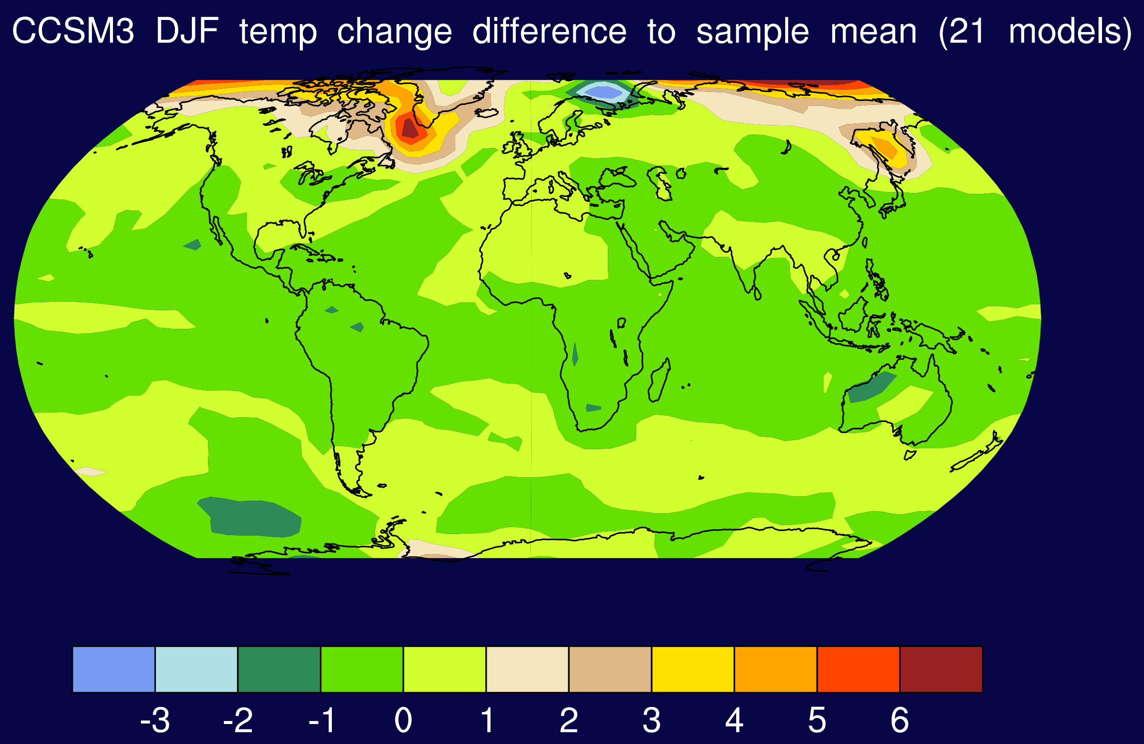


Studying Climate with AOGCMs

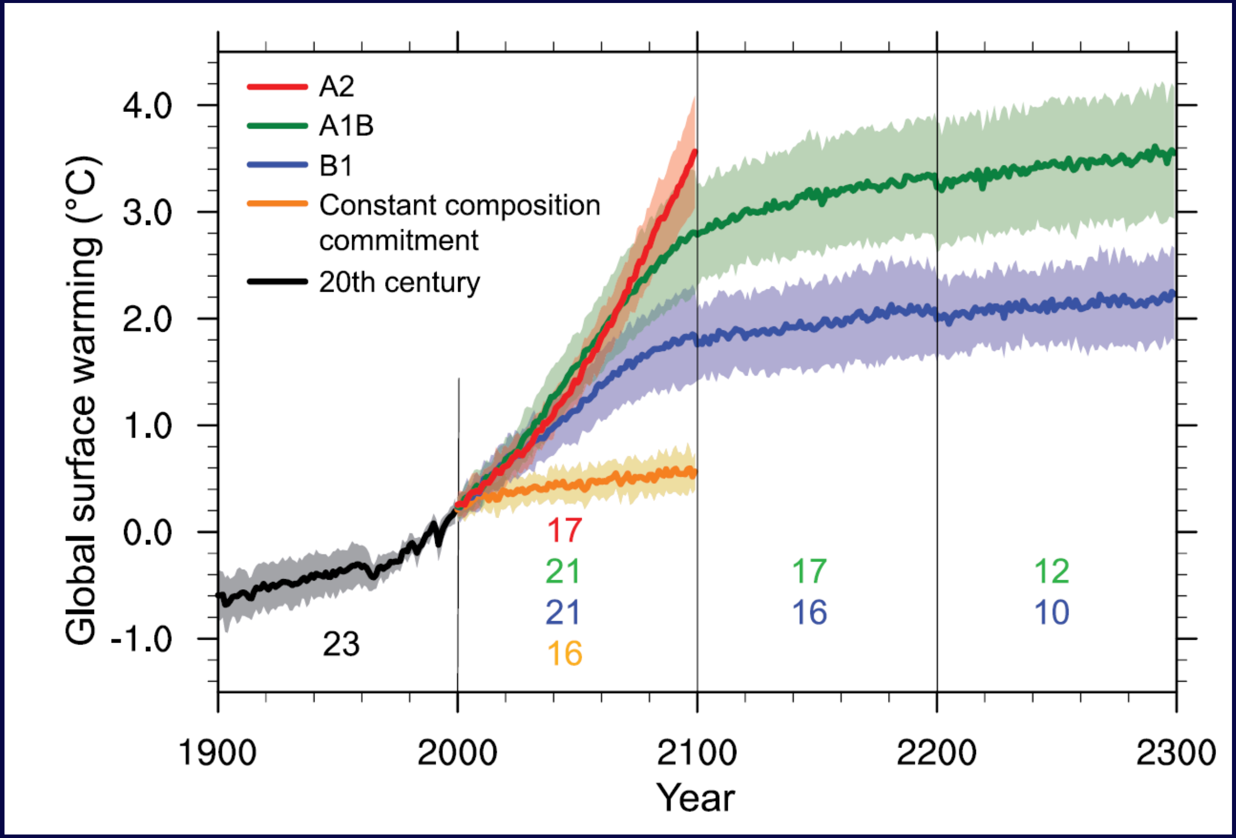
AOGCM: Atmosphere-Ocean General Circulation Models



Models Do Not Agree

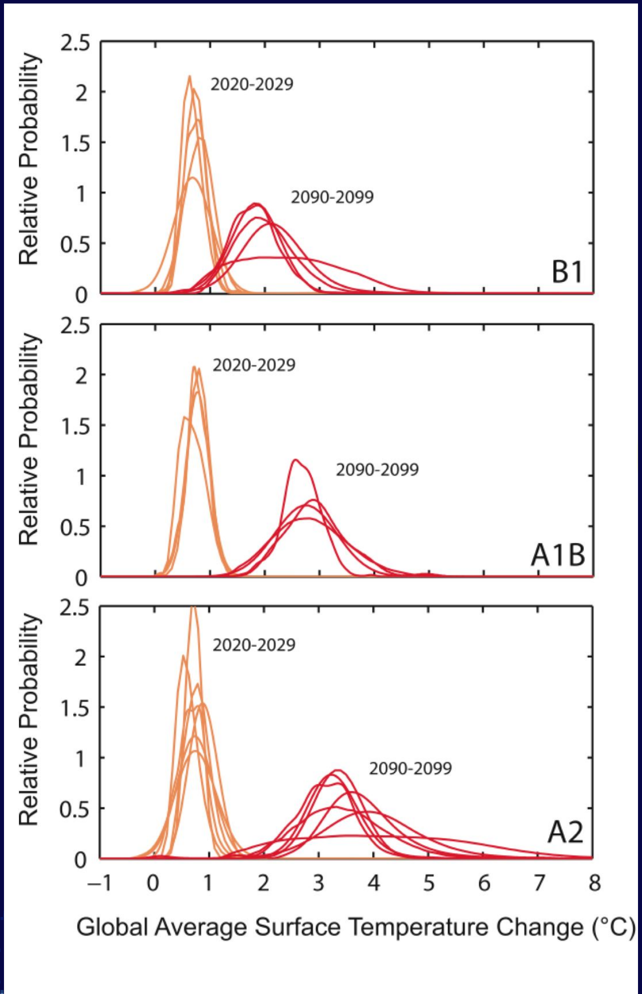


Models Do Not Agree



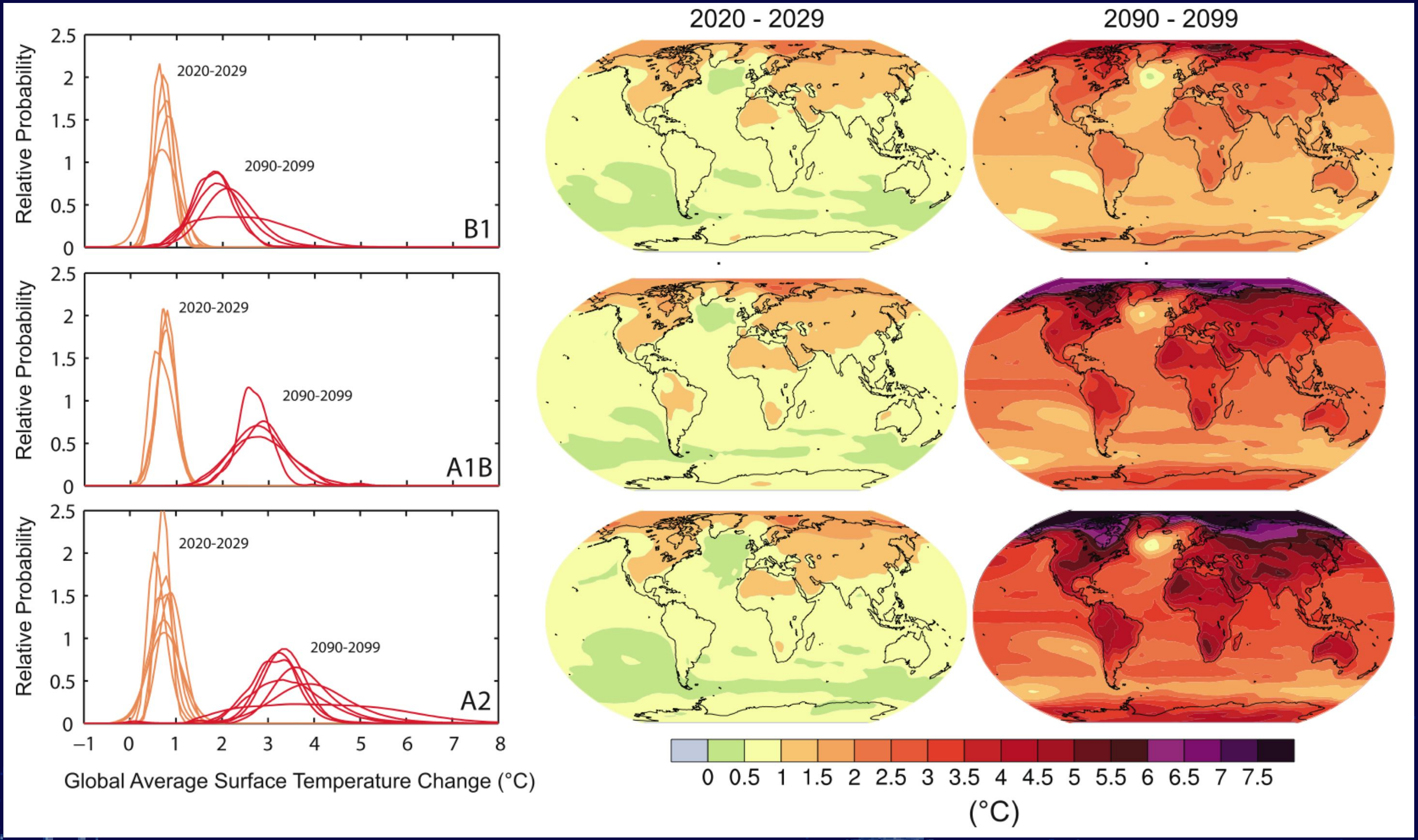
Source: AR4, IPCC

Quantifying Uncertainty



Source: AR4, IPCC

Quantifying Uncertainty

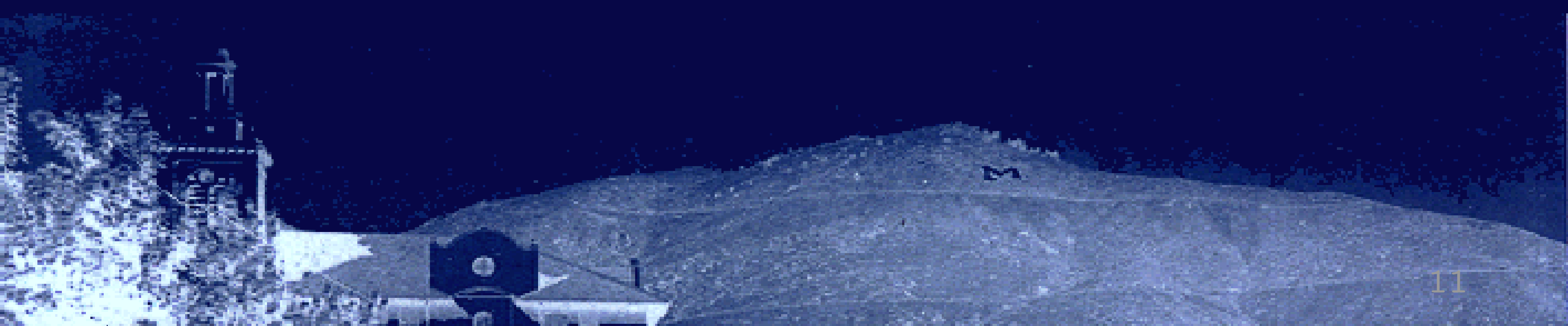


Source: AR4, IPCC

Data

Data provided for the Fourth Assessment Report of IPCC:

- 21 models (CCSM, GFDL, HADCM, PCM, . . .)
- Around $2.8^\circ \times 2.8^\circ$ resolution (8192 data points, T42)
- Different scenarios (A2: “business as usual”, A1B, B1)
- Temperature, precipitation, pressure, winds. . .



Data

Data provided for the Fourth Assessment Report of IPCC:

- 21 models (CCSM, GFDL, HADCM, PCM, . . .)
- Around $2.8^\circ \times 2.8^\circ$ resolution (8192 data points, T42)
  aggregate to $5^\circ \times 5^\circ$ and omit the "poles" (3264 points).
- Different scenarios (A2: "business as usual", A1B, B1)
- Temperature, precipitation, pressure, winds. . .
  seasonal averages over years 1980–1999 and 2080–2099

Statistical Model

Given AOGCM output construct a statistical model to describe climate change probabilistically while accounting for all (most?) underlying uncertainties.

Statistical Model

Given AOGCM output construct a statistical model to describe climate change probabilistically while accounting for all (most?) underlying uncertainties.

For models $i = 1, \ldots, N$, stack the gridded seasonal temperature into vectors:

$$\begin{aligned} \mathbf{X}_i &= \text{simulated present climate}_i \\ \mathbf{Y}_i &= \text{simulated future climate}_i \end{aligned}$$

PDF and probabilistic description of climate change

$$\mathbf{D}_i = \mathbf{Y}_i - \mathbf{X}_i$$

Hierarchical Model

Separate the statistical modeling of a complex process into different levels consisting of:

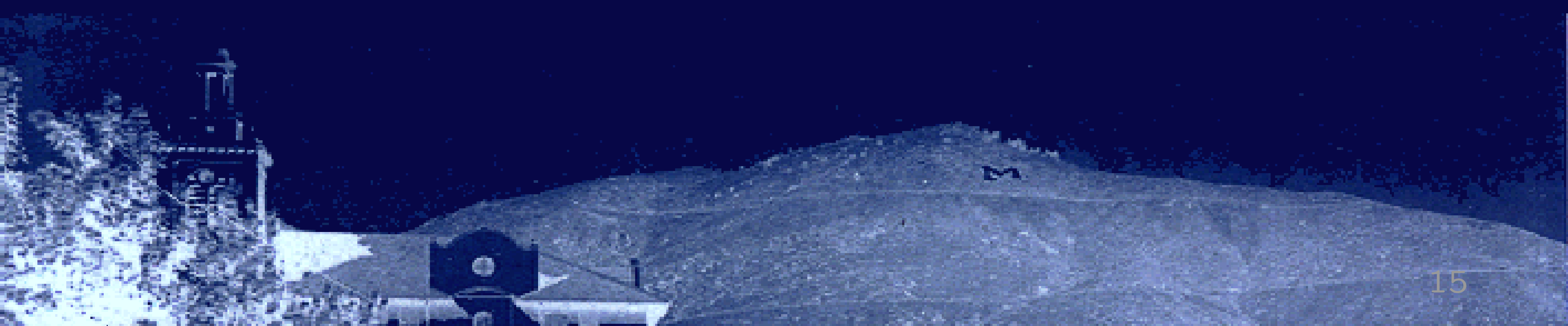
Data level:	Classical geostatistics	(variogram, kriging)
Process level:	Multivariate analysis	(EOF, PCA)
Prior level:	Bayesian statistics	(priors, MCMC)

$\rightsquigarrow$ hierarchical Bayesian modeling

Data Level

{ Data level | Process level | Prior level }

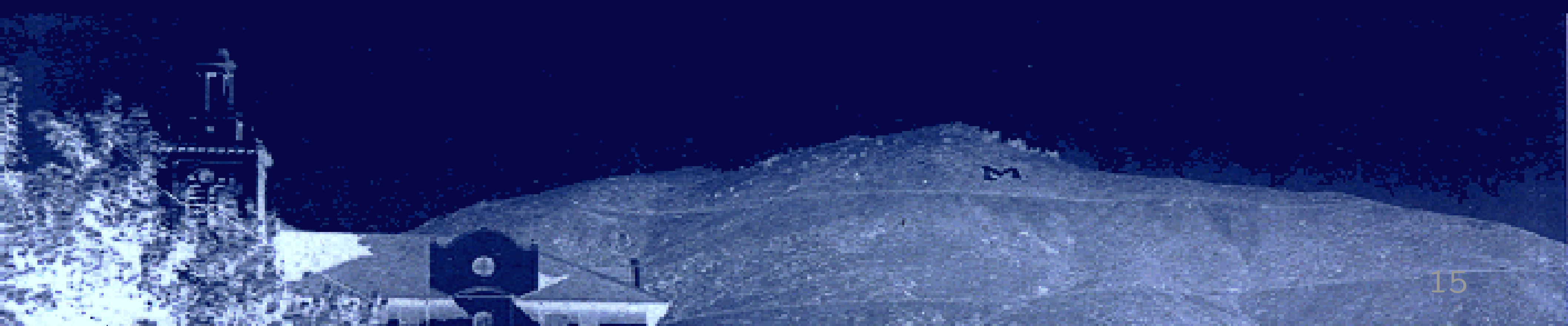
$$\mathbf{D}_i = \mathbf{Y}_i - \mathbf{X}_i = \text{simulated climate change}$$



Data Level

{ Data level | Process level | Prior level }

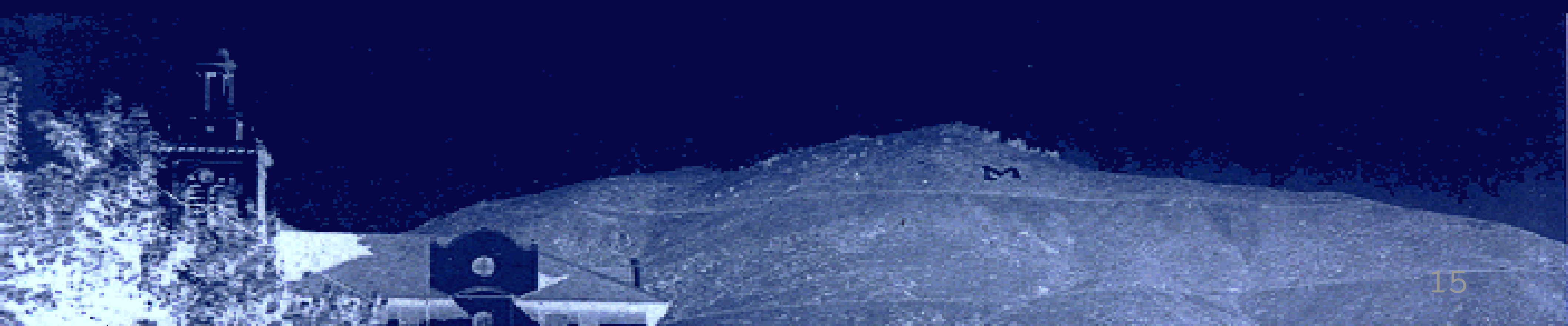
$$\mathbf{D}_i = \mathbf{Y}_i - \mathbf{X}_i = \text{simulated climate change}$$
$$= \text{large scale structure} + \text{small scale structure}$$



Data Level

{ Data level | Process level | Prior level }

$$\begin{aligned}\mathbf{D}_i = \mathbf{Y}_i - \mathbf{X}_i &= \text{simulated climate change} \\ &= \text{large scale structure} + \text{small scale structure} \\ &= \text{climate signal} + \text{model bias and internal variability}\end{aligned}$$



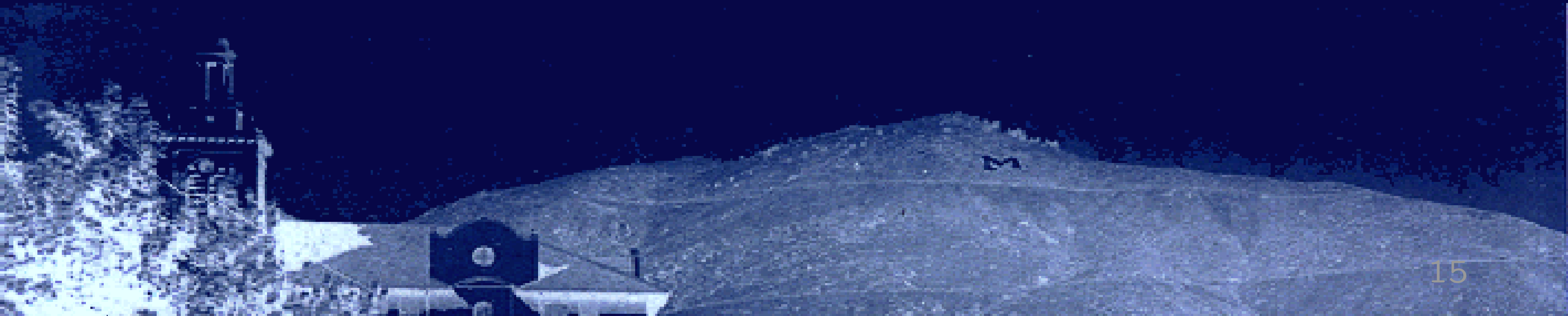
Data Level

{ Data level | Process level | Prior level }

$$
\begin{aligned}
\mathbf{D}_i = \mathbf{Y}_i - \mathbf{X}_i &= \text{simulated climate change} \\
&= \text{large scale structure} + \text{small scale structure} \\
&= \text{climate signal} + \text{model bias and internal variability} \\
&= \qquad \boldsymbol{\mu}_i \qquad\qquad + \qquad\qquad \boldsymbol{\varepsilon}_i
\end{aligned}
$$

$$
\mathbf{D}_i \mid \boldsymbol{\mu}_i, \phi_i \overset{\text{iid}}{\sim} \mathcal{N}_n(\boldsymbol{\mu}_i, \phi_i \boldsymbol{\Sigma}) \qquad \phi_i > 0 \qquad i = 1, \ldots, N
$$

for given $\boldsymbol{\Sigma}$

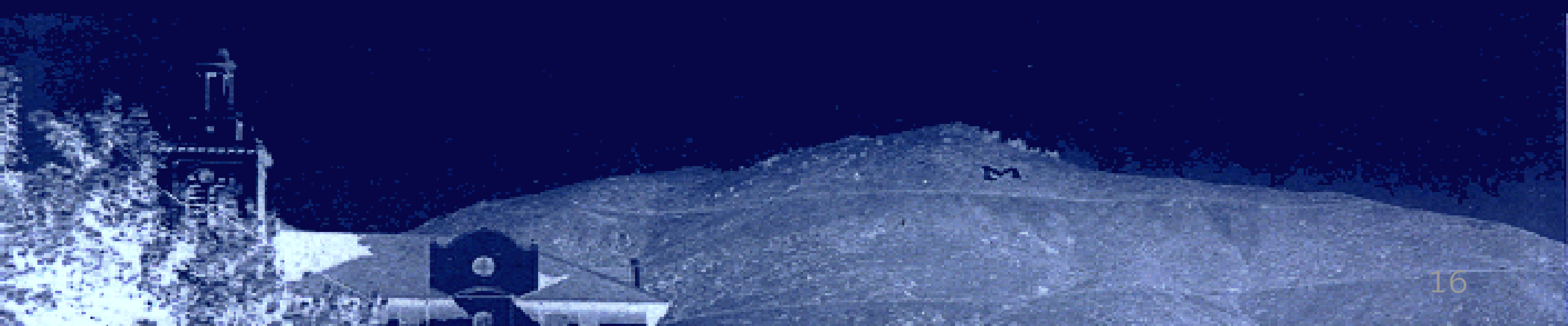


Process Level

{ Data level | Process level | Prior level }

$$\boldsymbol{\mu}_i = \mathbf{M}\boldsymbol{\theta}_i$$

for given $\mathbf{M}$



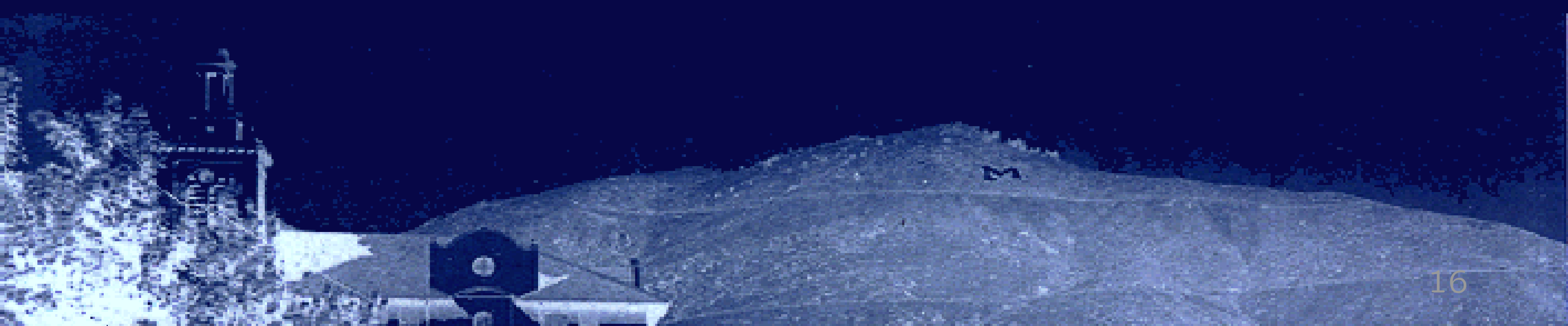
Process Level

{ Data level | Process level | Prior level }

$$\boldsymbol{\mu}_i = \mathbf{M}\boldsymbol{\theta}_i$$

for given $\mathbf{M}$

$$\boldsymbol{\theta}_i \mid \boldsymbol{\nu},\ \psi_i \overset{\text{iid}}{\sim} \mathcal{N}_p(\, \boldsymbol{\nu},\ \psi_i \mathbf{I}\,) \qquad \psi_i > 0 \qquad i = 1, \ldots, N$$

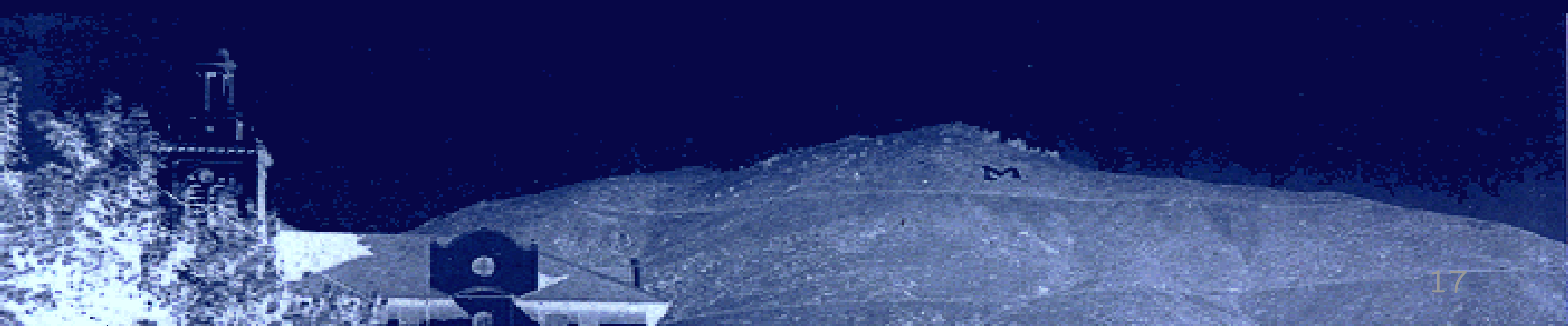


Prior Level

{ Data level | Process level | Prior level }

$$\begin{array}{lll}
\phi_i \overset{\text{iid}}{\sim} \mathrm{I}\Gamma(\,\xi_1,\,\xi_2\,) & \xi_1,\,\xi_2 > 0 & i = 1,\ldots,N \\
\psi_i \overset{\text{iid}}{\sim} \mathrm{I}\Gamma(\,\xi_3,\,\xi_4\,) & \xi_3,\,\xi_4 > 0 & i = 1,\ldots,N \\
\boldsymbol{\nu} \sim \mathcal{N}_p(\,\mathbf{0},\,\xi_5\mathbf{I}\,) & \xi_5 > 0 &
\end{array}$$

for given $\xi_1,\ldots,\xi_5$



Initial Parameters

For the different levels we need to specify:

Data level Covariance model for $\phi_i \mathbf{\Sigma}$:
spatial coherence of internal variability and bias

Process level Basis functions used in **M**:
practical decompostion of possible signals,
dimension reduction

Prior level Hyperparameters $\xi_1, \xi_2, \quad \xi_3, \xi_4, \quad \xi_5$:
tuning parameters

Covariance Model for $\phi_i\Sigma$

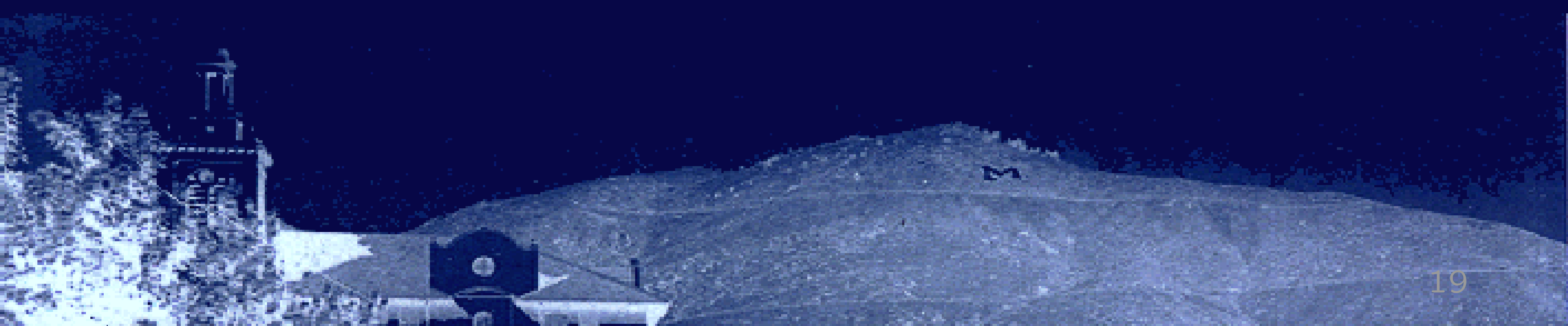
{ Data level | Process level | Prior level }

For the covariance matrices $\phi_i\mathbf{\Sigma}$, we need positive definite functions on the sphere (by restricting one on $\mathbb{R}^3$ to $\mathbb{S}^2$):

$$c(h; \phi_i, \tau) = \phi_i \exp\Big(-\tau \sin(h/2)\Big)$$

Individual variances ϕ_i are modelled.

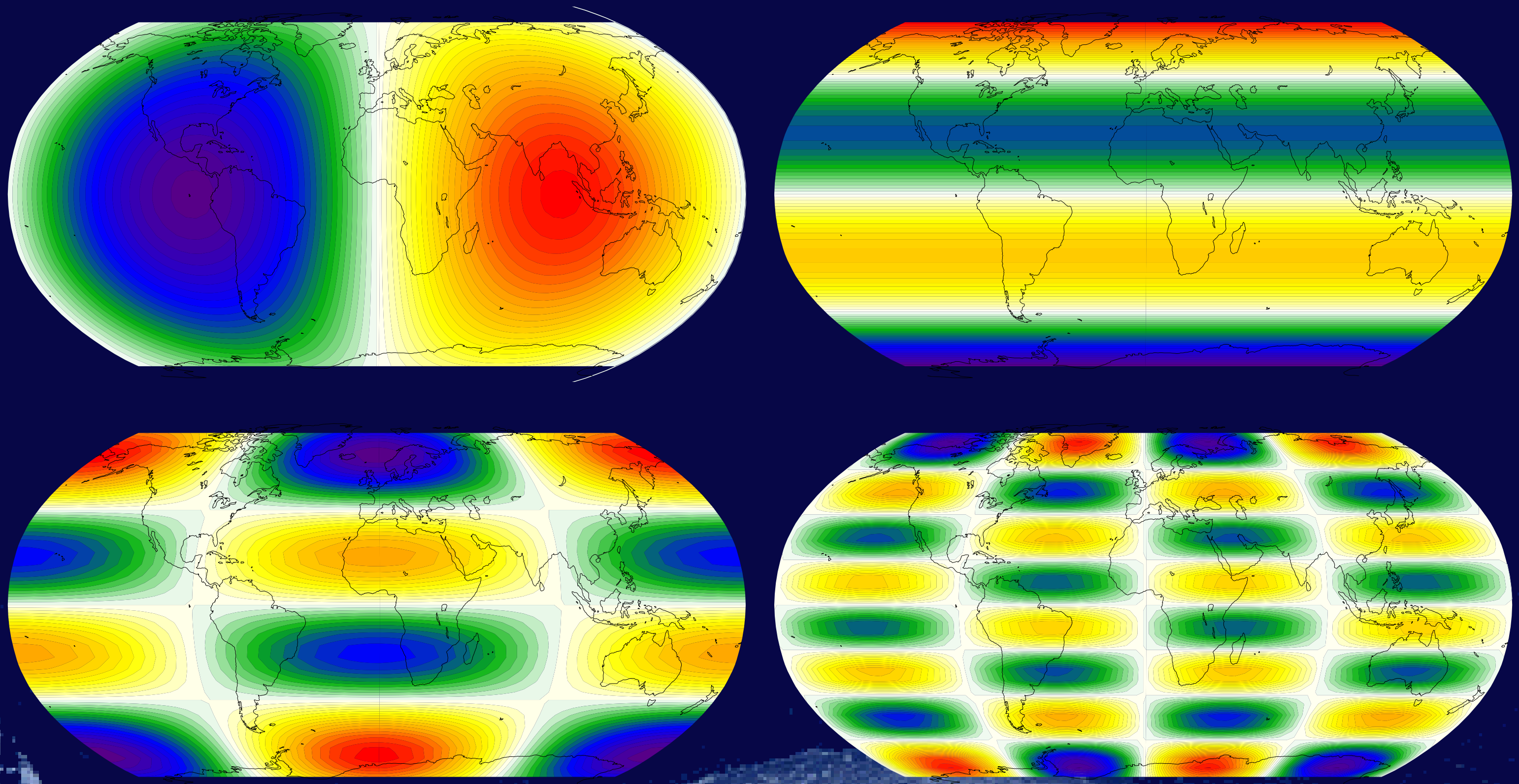
Common range τ is choosen according to an "empirical Bayes" approach.



Basis Functions Used in M

{ Data level | Process level | Prior level }

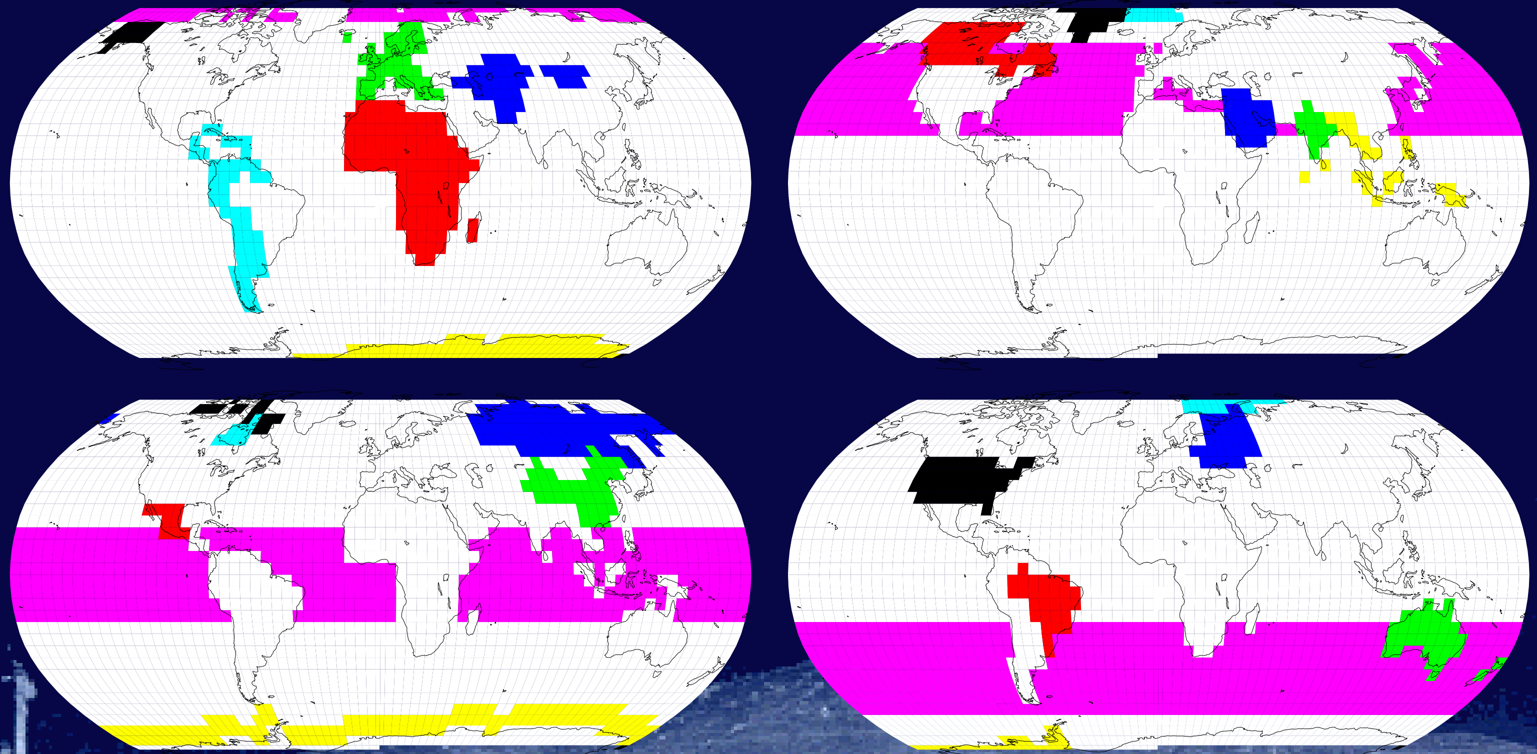
1. Spherical harmonics (here shown 4 out of 121)



Basis Functions Used in M

{ Data level | Process level | Prior level }

1. Spherical harmonics
2. Indicator functions (28)



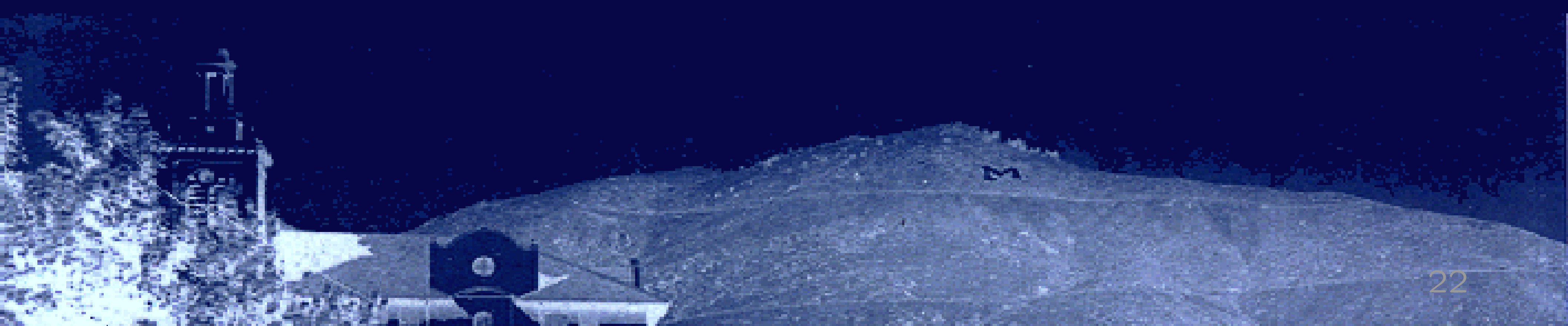
Hyperparameters $\xi_1, \ldots, \xi_5$

{ Data level | Process level | Prior level }

To make sure that variability around the truth is smaller than bias and internal variability

$$\phi_i > \psi_i$$

Choose ξ_1, ξ_2, ξ_3 small, $\xi_4 \in [1, 2.5]$, ξ_5 large.

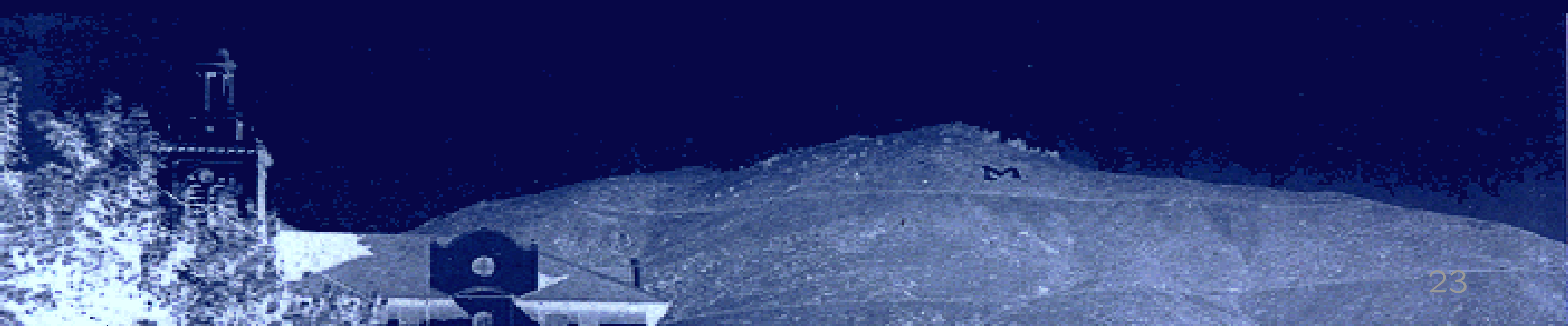


PDF of Climate Change

The goal is the (posterior) PDF of the climate change signal given the AOGCM data and model parameters:

$$[\,\text{climate change} \mid \text{AOGCM data}, \text{model parameters} \ldots]$$

$$[\quad \mathbf{M}\nu \quad \mid \quad \mathbf{D}_1, \ldots, \mathbf{D}_N \;, \quad \ldots \quad]$$



PDF of Climate Change

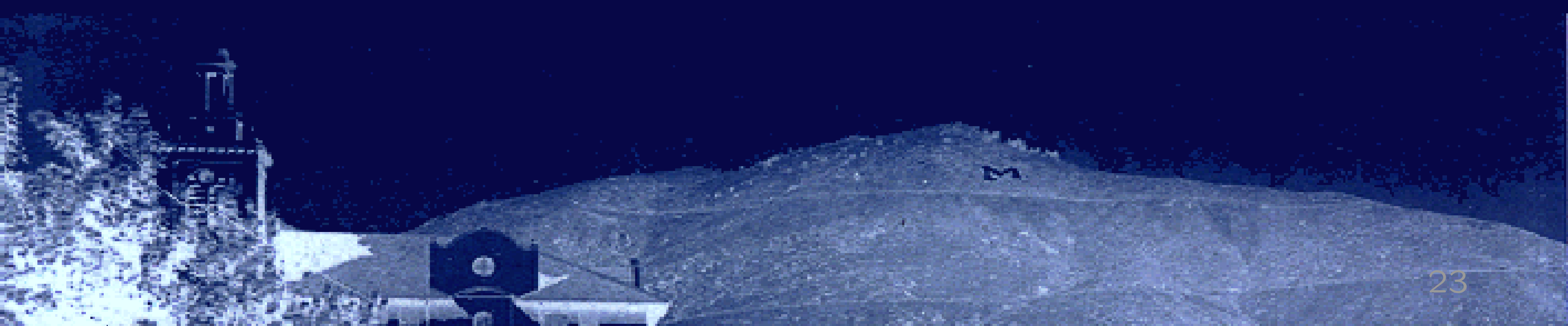
The goal is the (posterior) PDF of the climate change signal given the AOGCM data and model parameters:

$$[\,\text{climate change} \mid \text{AOGCM data}, \text{model parameters} \ldots]$$

$$[\qquad \mathbf{M}\nu \qquad \mid \; \mathbf{D}_1, \ldots, \mathbf{D}_N \;, \qquad \ldots \qquad]$$

Via Bayes' theorem, the (posterior) PDF is

$$\begin{aligned} &[\text{ process} \mid \text{data, parameters }] \\ &\qquad \propto [\text{ data} \mid \text{process, parameters }] \\ &\qquad\qquad \cdot [\text{ process} \mid \text{parameters }] \cdot [\text{ parameters }] \end{aligned}$$



Computational Approach

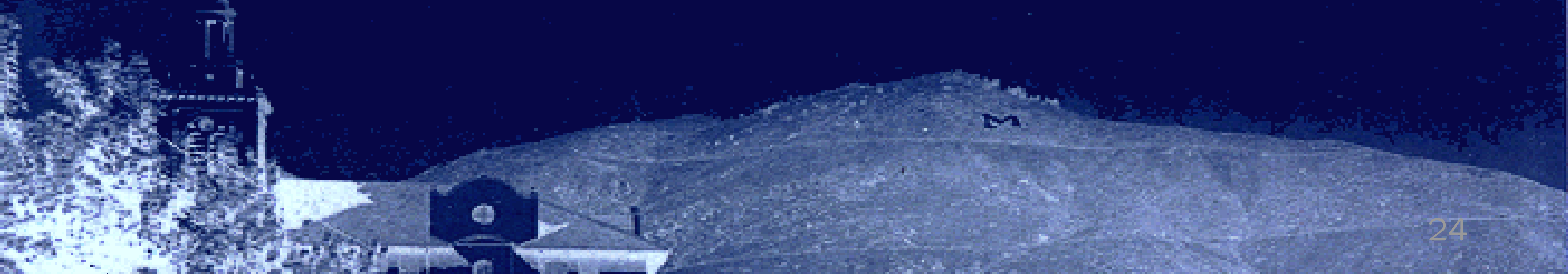
No closed form of the posterior density.

Use a computational approach: Markov Chain Monte Carlo (MCMC), here a Gibbs sampler.

1. Express the distribution of each parameter conditional on everything else (full conditionals).

2. Cycle through the parameters: draw a new value based on the full conditional and the current values of the other parameters.

3. Repeat, . . .



Full Conditionals

Full conditionals for all parameters have been derived:

$$\boldsymbol{\nu} \mid \ldots \sim \mathcal{N}_p(\mathbf{A}^{-1}\mathbf{b},\ \mathbf{A}^{-1})$$

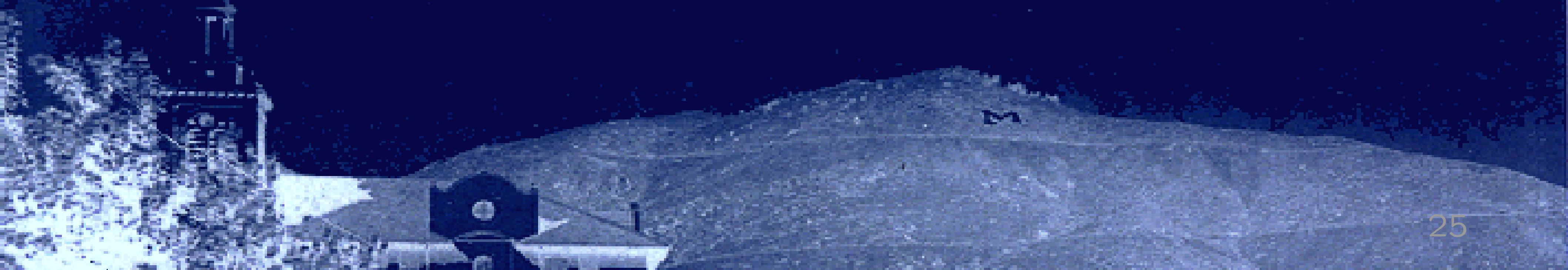
$$\mathbf{A} = \frac{1}{\xi_5}\mathbf{I} + \sum_{i=1}^{N} \frac{1}{\psi_i}\mathbf{I} \qquad \mathbf{b} = \sum_{i=1}^{N} \frac{1}{\psi_i}\boldsymbol{\theta}_i$$

$$i = 1, \ldots, N:\ \boldsymbol{\theta}_i \mid \ldots \sim \mathcal{N}_p(\mathbf{A}^{-1}\mathbf{b},\ \mathbf{A}^{-1})$$

$$\mathbf{A} = \frac{1}{\psi_i}\mathbf{I} + \frac{1}{\phi_i}\mathbf{M}^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}\mathbf{M} \quad \mathbf{b} = \frac{1}{\psi_i}\boldsymbol{\nu} + \frac{1}{\phi_i}\mathbf{M}^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}\mathbf{D}_i$$

$$i = 1, \ldots, N:\ \phi_i \mid \ldots \sim \mathrm{I}\Gamma\left(\xi_1 + \frac{n}{2}, \xi_2 + \frac{1}{2}(\mathbf{D}_i - \mathbf{M}\boldsymbol{\theta}_i)^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{D}_i - \mathbf{M}\boldsymbol{\theta}_i)\right)$$

$$i = 1, \ldots, N:\ \psi_i \mid \ldots \sim \mathrm{I}\Gamma\left(\xi_3 + \frac{p}{2}, \xi_4 + \frac{1}{2}(\boldsymbol{\theta}_i - \boldsymbol{\nu})^{\mathsf{T}}(\boldsymbol{\theta}_i - \boldsymbol{\nu})\right)$$



Full Conditionals

Full conditionals for all parameters have been derived:

$$\boldsymbol{\nu} \mid \ldots \sim \mathcal{N}_p(\quad , \quad)$$

$$i = 1, \ldots, N: \ \boldsymbol{\theta}_i \mid \ldots \sim \mathcal{N}_p(\quad , \quad)$$

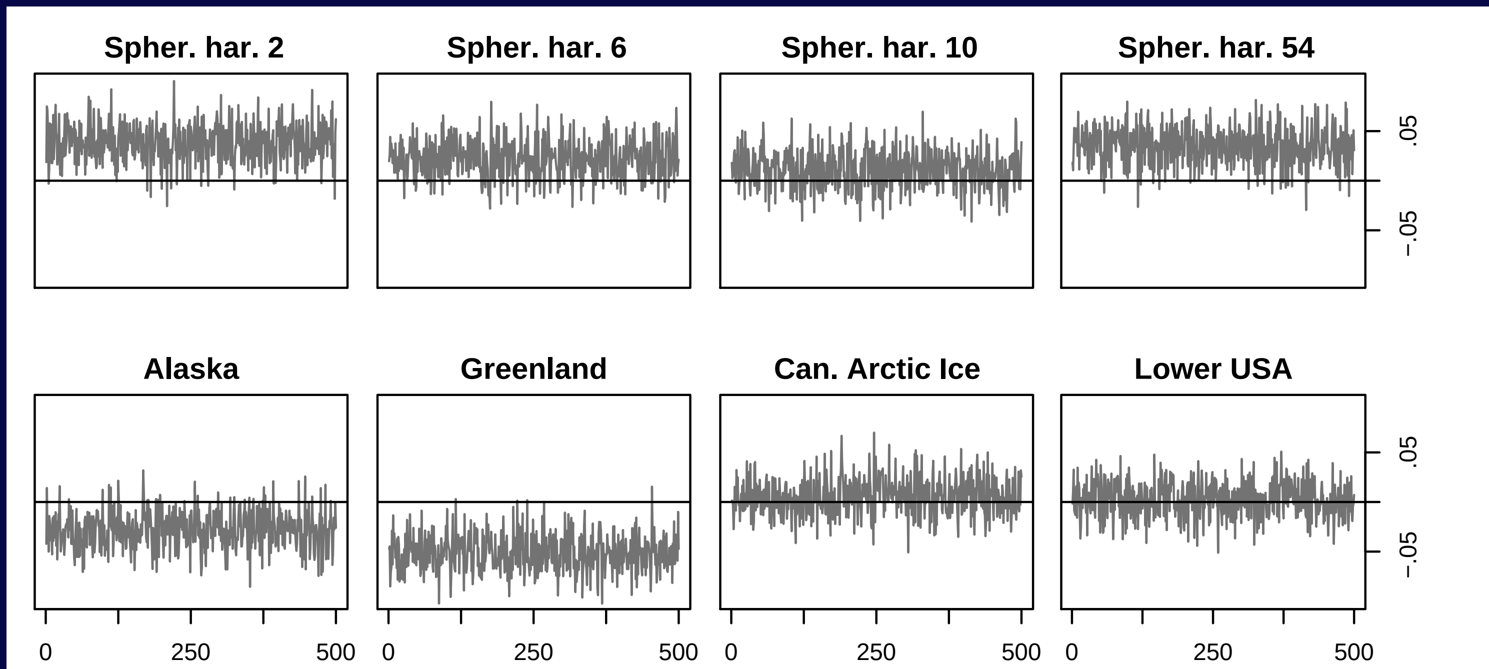
$$i = 1, \ldots, N: \ \phi_i \mid \ldots \sim \mathrm{I}\Gamma\left(\quad , \quad\right)$$

$$i = 1, \ldots, N: \ \psi_i \mid \ldots \sim \mathrm{I}\Gamma\left(\quad , \quad\right)$$

Computational Aspects

- Gibbs sampler programmed in R
  free software environment for statistical computing and graphics
- Run 20000 iterations
  10000 burn-in, keep every 20th, takes a few hours
- Visual/primitive inspection of convergence

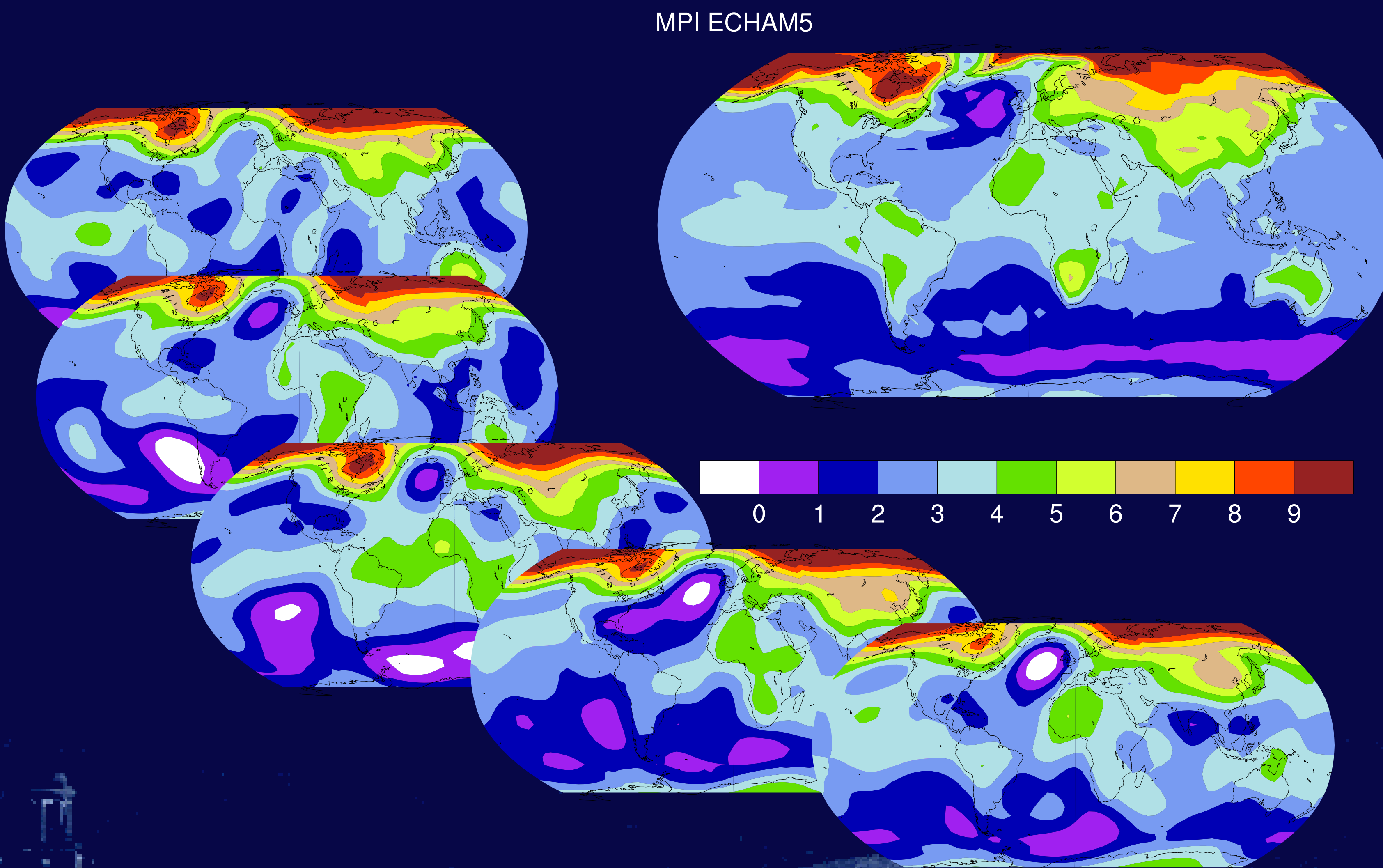
Computational Aspects



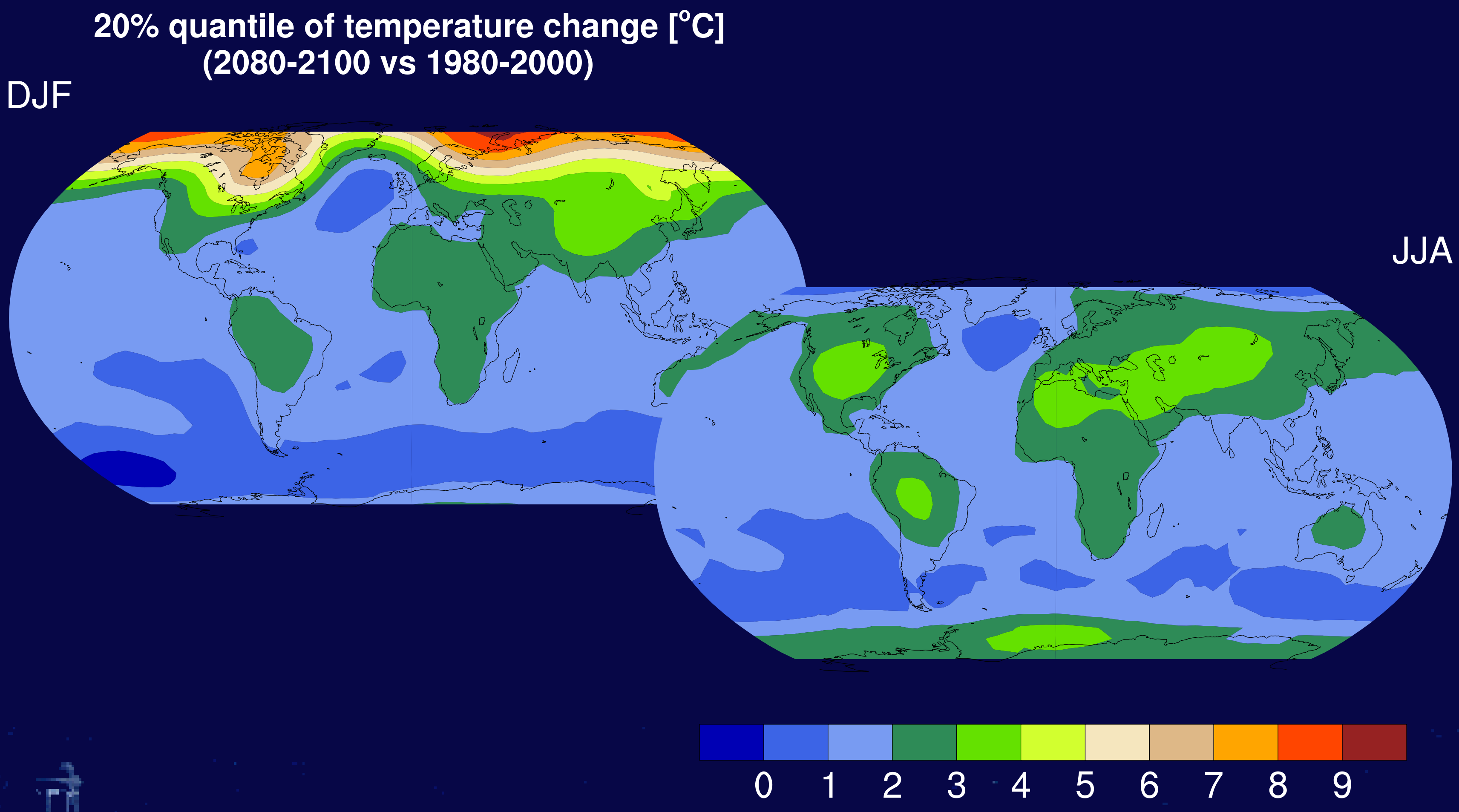
and graphics

- Visual/primitive inspection of convergence

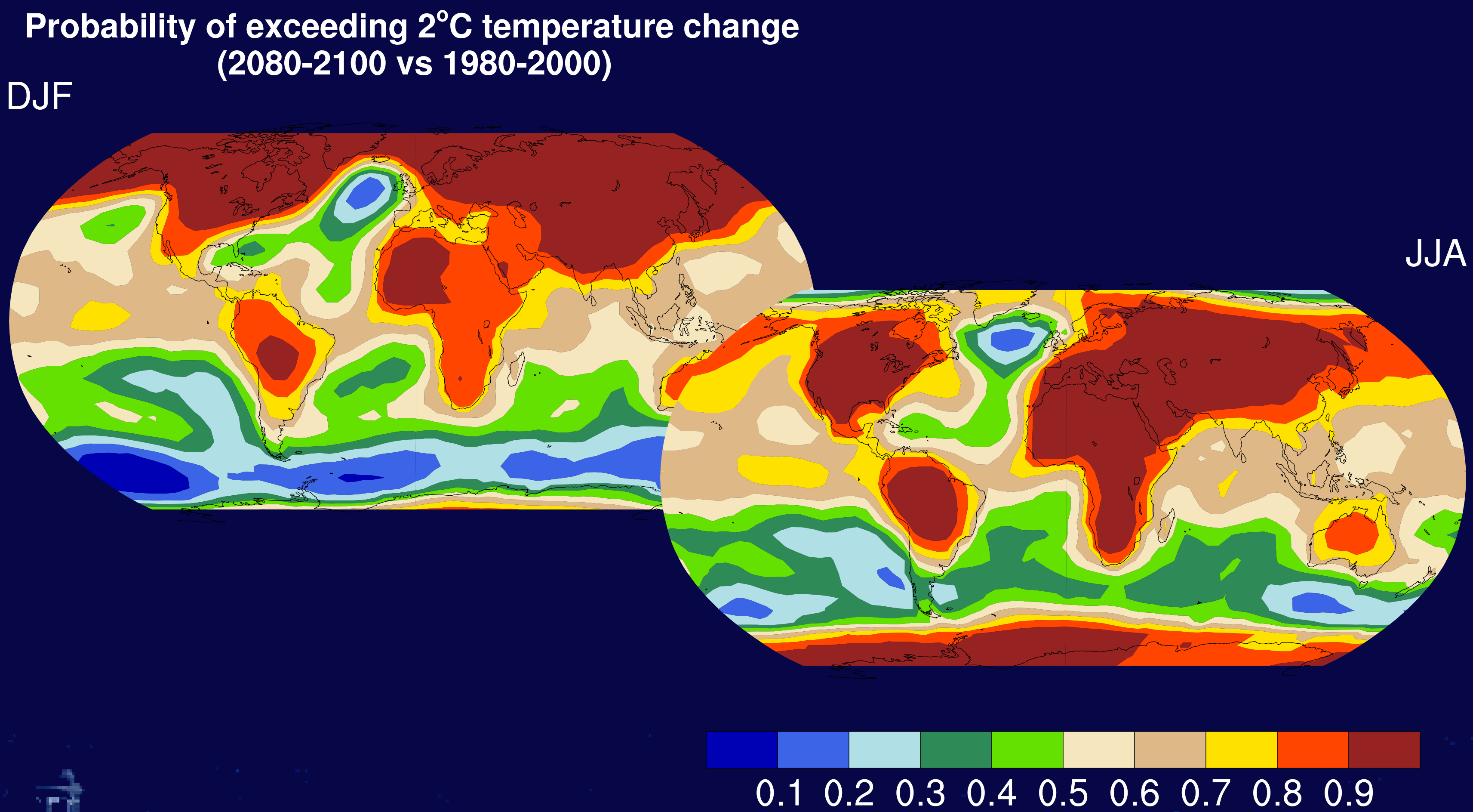
Posterior Draws



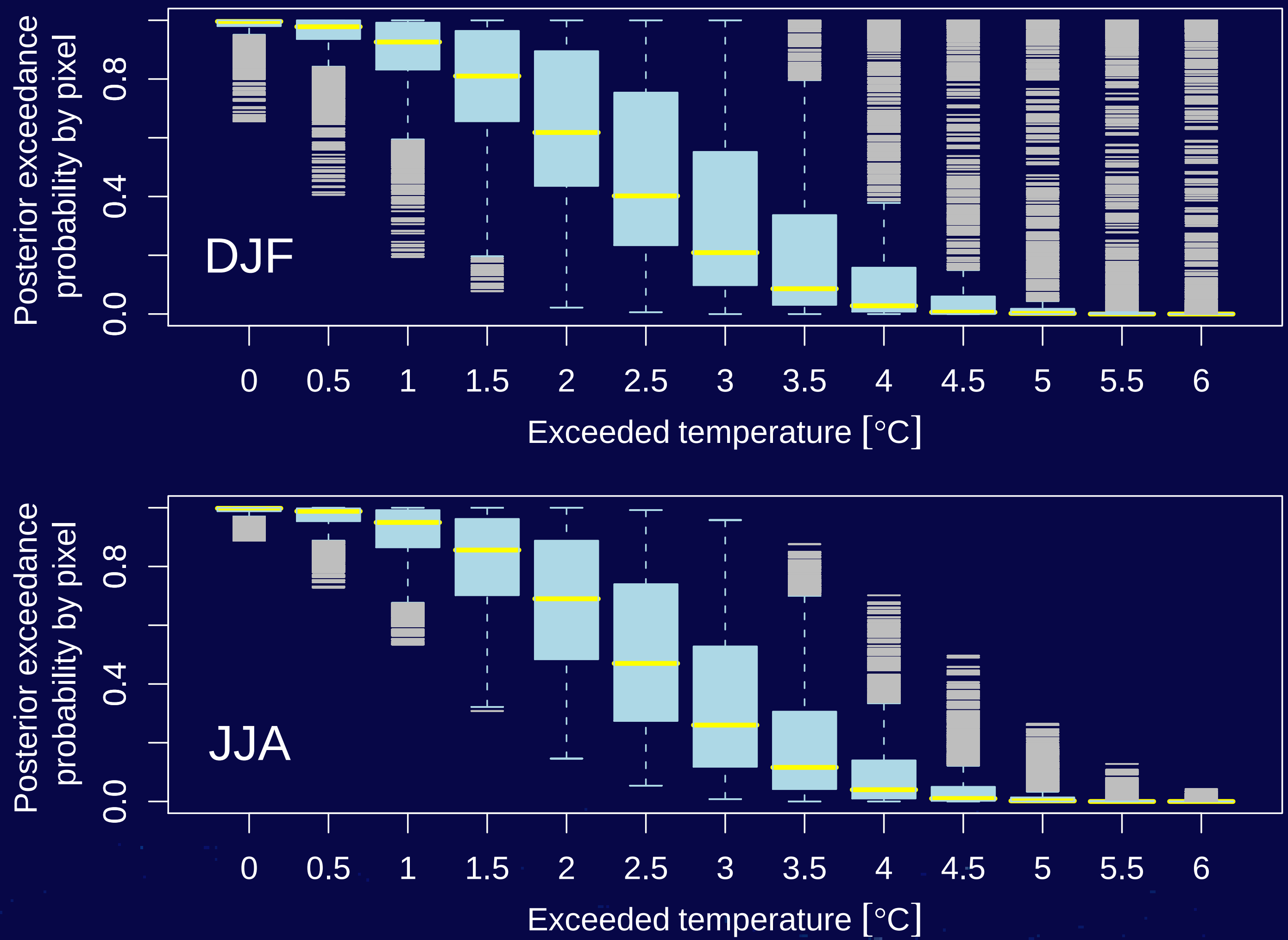
Temperature Change Quantiles



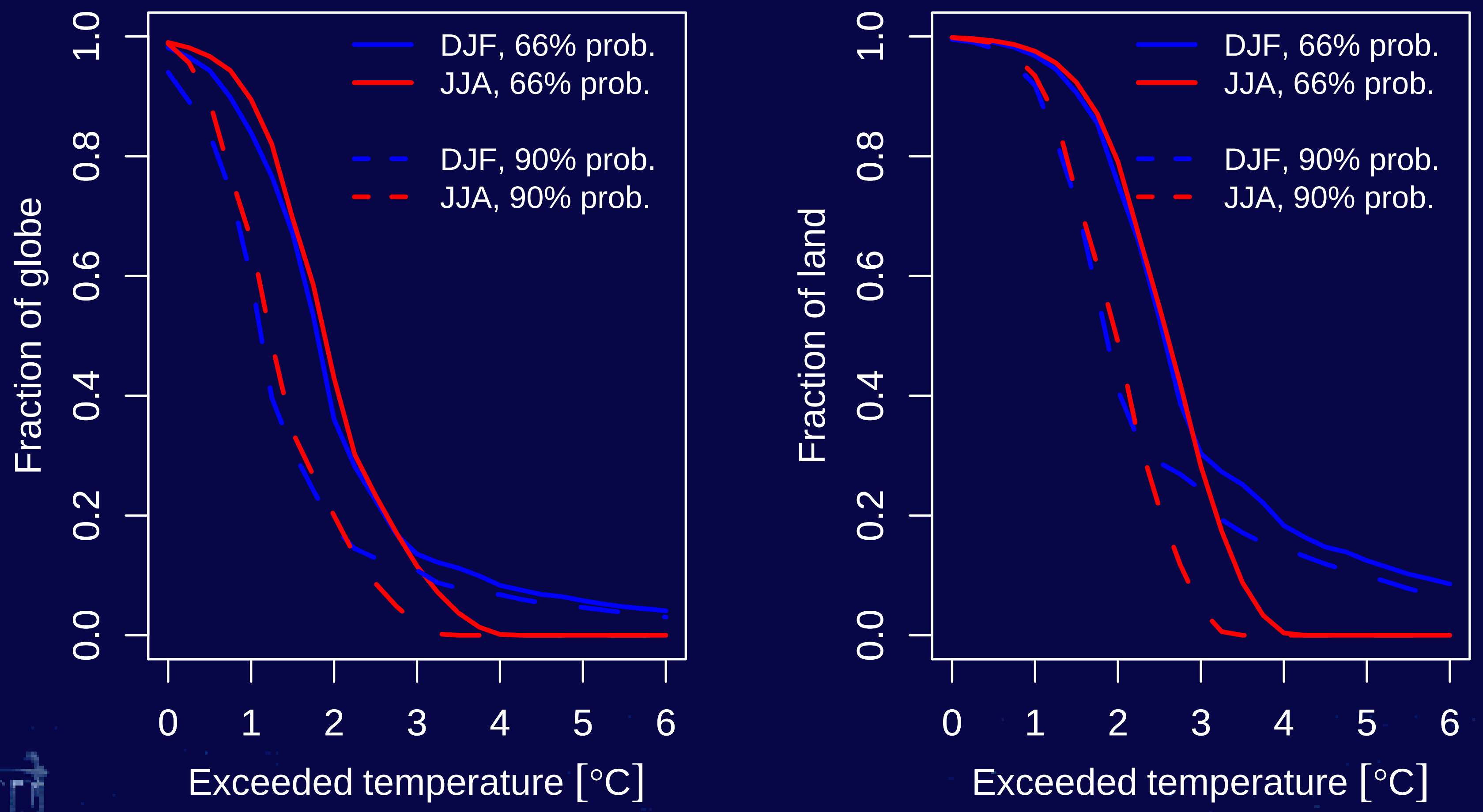
Exceedance Probabilities



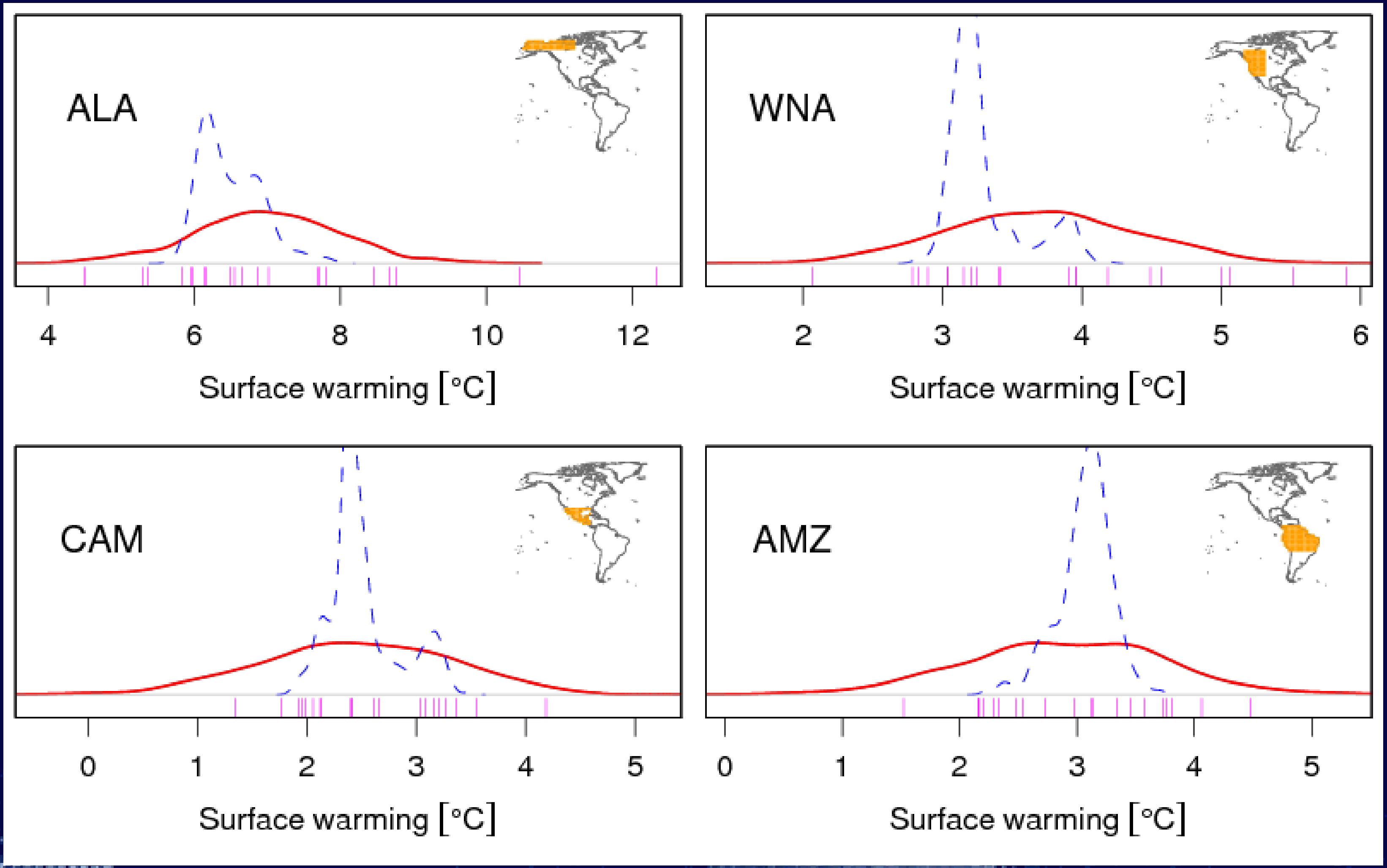
Exceedance Probabilities



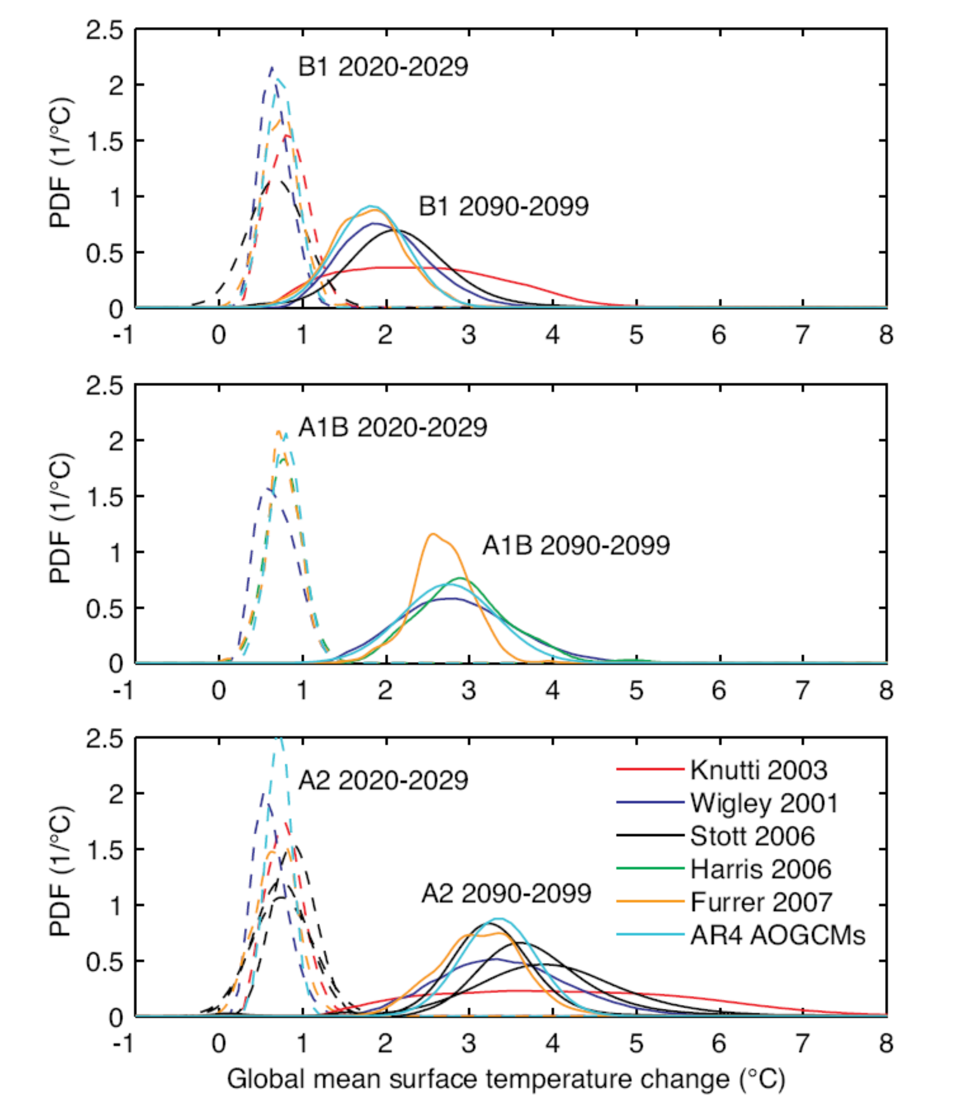
Exceedance Fractions



Regional Assessment



Global Assessment



Source: AR4, IPCC

Model Extensions

- Use "more" data
  - $\rightsquigarrow$ ensemble runs, model bias and internal variability, model present and future individually, . . .
- Use AOGCM specific weighting
  - $\rightsquigarrow$ performance, "core" simililarities, . . .
- Parameterize covariance matrices
  - $\rightsquigarrow$ built in range, nonstationarity, . . .
- Building bi-/multivariate models
  - $\rightsquigarrow$ use temperature for precipitation prediction, . . .
- Address computational complexity
  - $\rightsquigarrow$ sparsity, GMRF, Metropolis-Hastings steps, . . .

References

Furrer, Knutti, Sain, Nychka, Meehl, (2007). Spatial patterns of probabilistic temperature change projections from a multivariate Bayesian analysis, *Geophys. Res. Lett.*, 34, L06711, doi:10.1029/2006GL027754.

Furrer, Sain, Nychka, Meehl, (2007). Multivariate Bayesian Analysis of Atmosphere-Ocean General Circulation Models, to appear in *Environmental and Ecological Statistics*.

Furrer, Sain, (2007). Spatial Model Fitting for Large Datasets with Applications to Climate and Microarray Problems, submitted to *Statistics and Computing*.

Sain, Furrer, Cressie, (2007). Combining Regional Climate Model Output via a Multivariate Markov Random Field Model. 56th Session of the International Statistical Institute, Lisbon, Portugal.