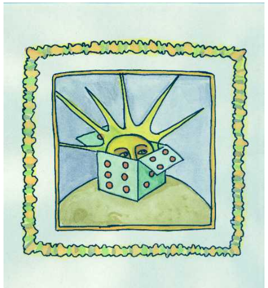
Probabilities for climate projections

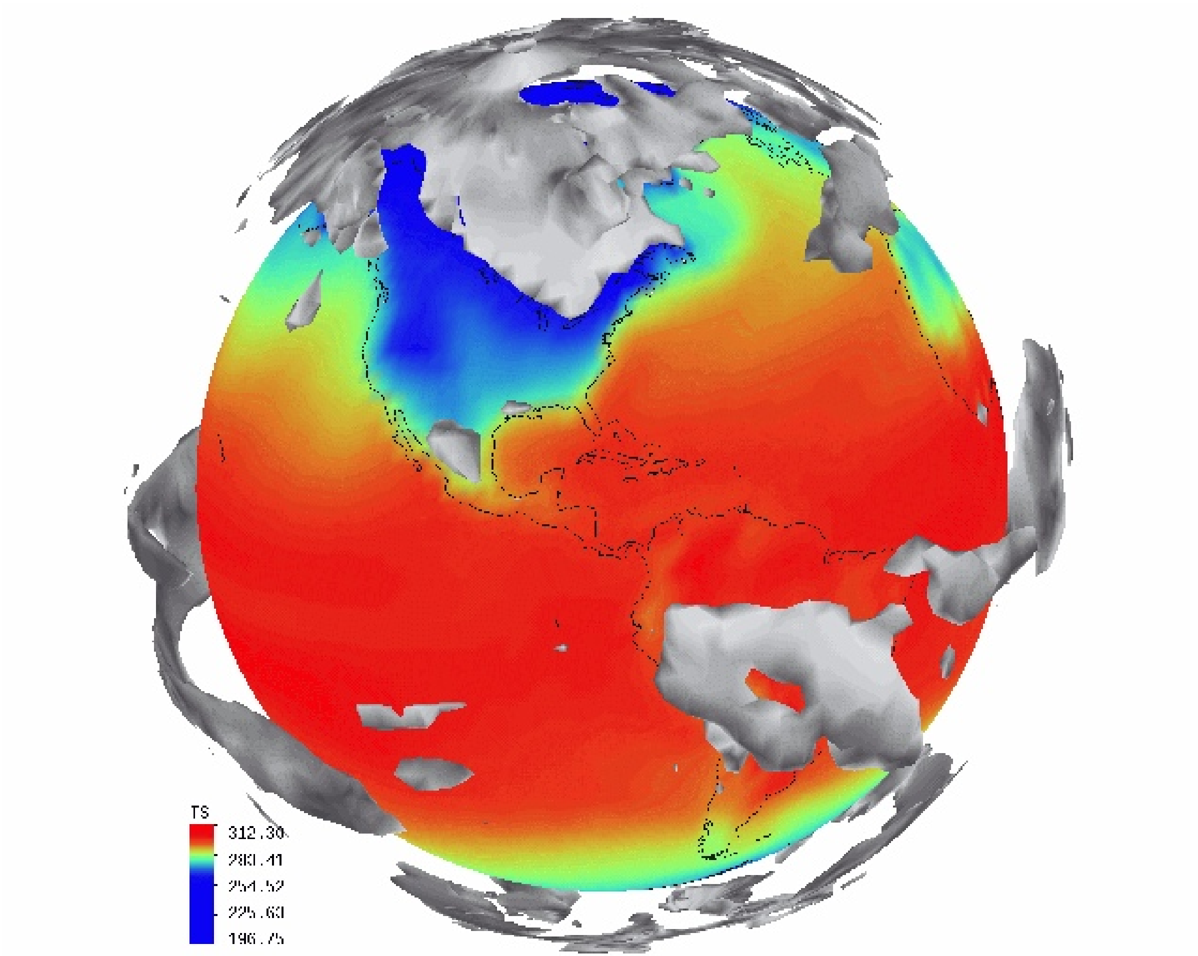
Claudia Tebaldi, Reinhard Furrer
Linda Mearns, Doug Nychka
National Center for Atmospheric Research
Richard Smith - UNC-Chapel Hill
Steve Sain - CU-Denver

- Statistical models for climate experiments
- Inference for a single region
- ANOVA models across regions
- Spatial models for temperature fields

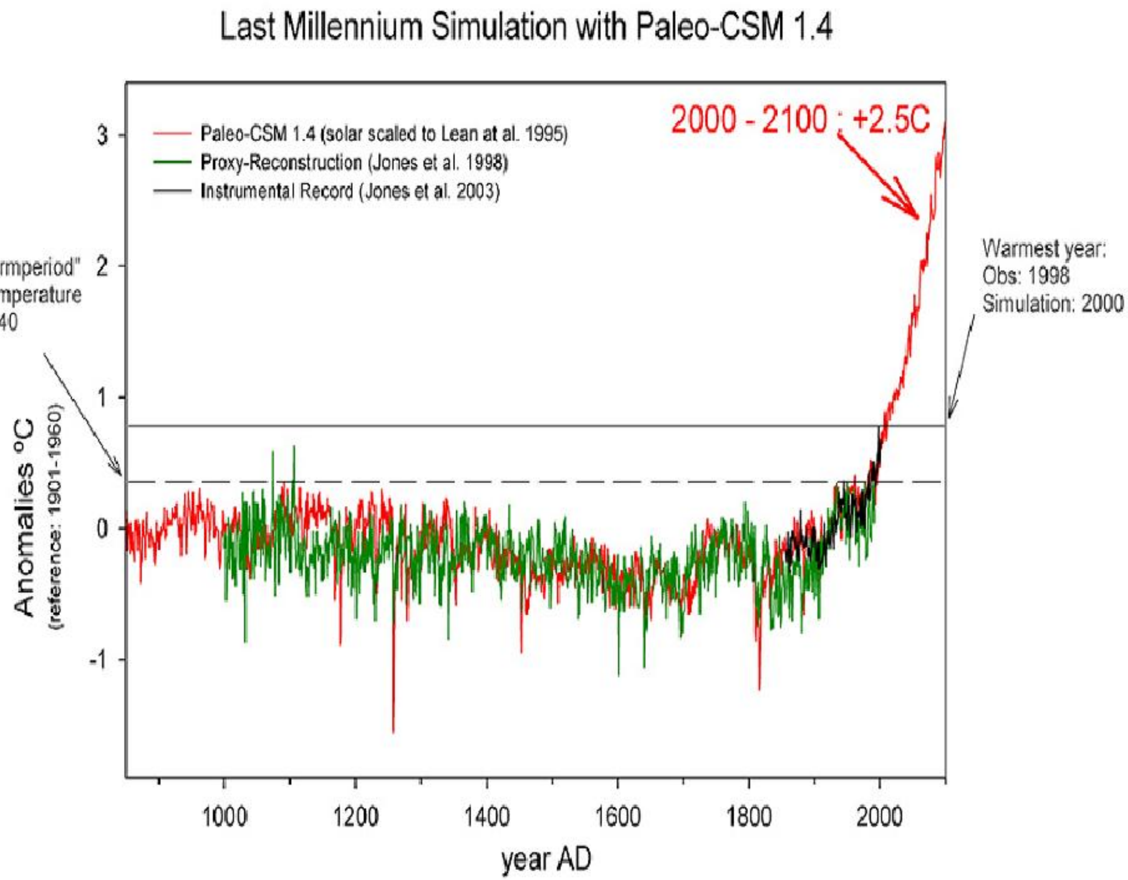


Climate: **What you expect ...**

Weather: **What you get.**



An Atmosphere-Ocean General Circulation Model (AOGCM)



Motivation

Based on model results, what will the climate be like in 2100?

Impacts of climate change: Extremes in temperatures, Possible degradation in air quality, Changes in the domain of vector-borne diseases.

- **Reconciling different projections - no model is the true model!**
- **Offering stake-holders and policy-makers a probabilistic forecast.**
- **Substituting formal probabilistic assumptions for heuristic criteria, and testing sensitivity of the results to them.**

The main points

Likelihood: **Formulate a statistical description of model bias and variability where each model is a "sample" from a superpopulation of AOGCMs.**

Prior: **Include any prior knowledge on the model biases.**

Posterior **Using Bayes compute the distribution of possible climate change** ***given*** **the model experiments.**

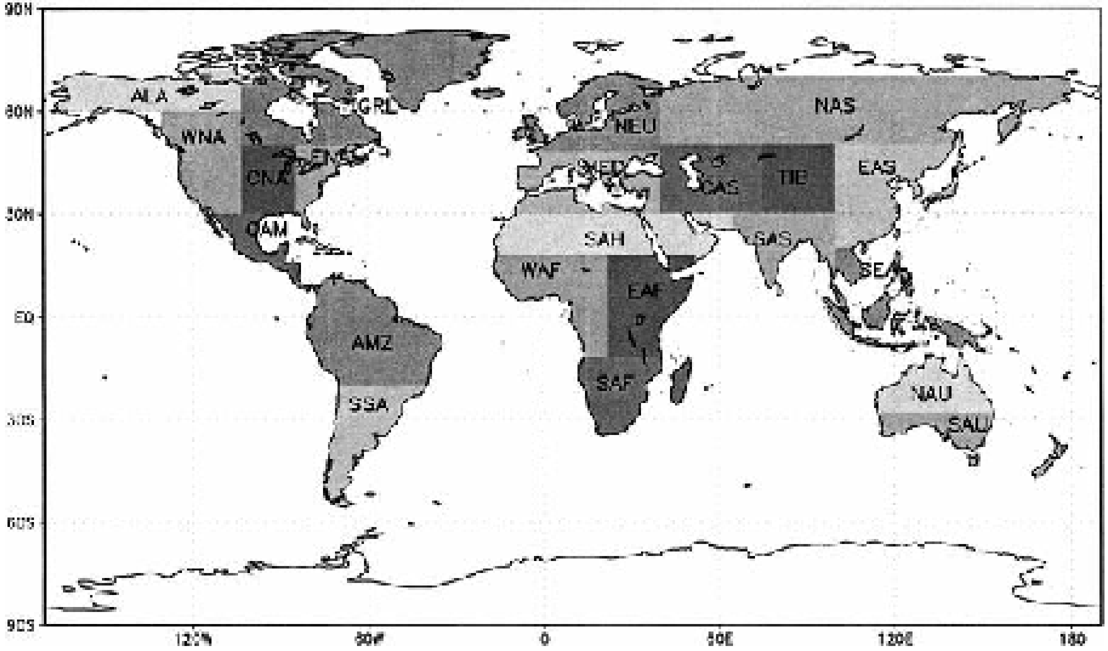
$$Likelihood \times Prior \rightarrow Posterior$$

A test suite of regional AOGCM experiments

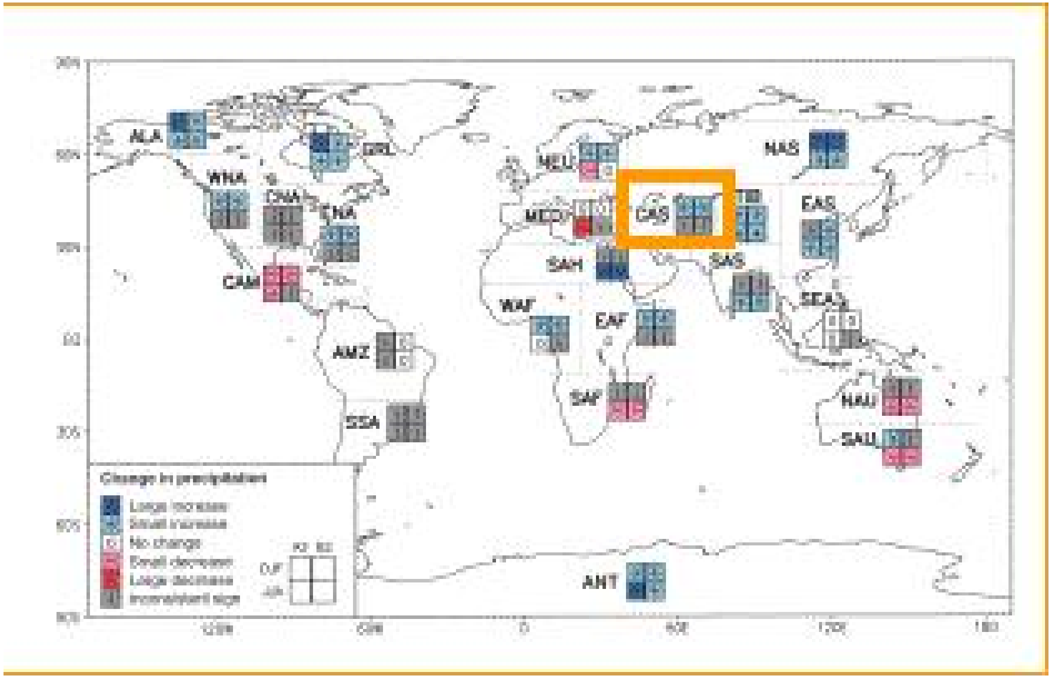
- **9 AOGCMs;**
- **22 Regions;**
- **2 Seasons;**
- **Simulated Temperature values in 30-years averages (X, 1961-1990; Y, 2071-2100 (A2));**
- **Observed Temperature average, X_0, for 1961-1990. (Allows for an estimate of model bias for current climate.)**

The data are the X's and Y's.

Regions


90N
60N
30N
EQ
30S
60S
90S
120W
60W
0
60E
120E
180
ALA
GRL
WNA
CNA
ENA
CAM
AMZ
SSA
NEU
MED
SAH
WAF
EAF
SAF
NAS
CAS
TIB
EAS
SAS
SEA
NAU
SAU

State-of-the art inference for the last IPCC report



Some background: Reliability Ensemble Average (REA)

- *Journal of Climate*, May 2002:Calculation of Average, Uncertainty Range and Reliability of Regional Climate Change from AOGCM Simulations....., by Giorgi and Mearns.
- Combine regional climate results , based on a **WEIGHTED AVERAGE**.
- Weights are implicit but quantify:
  BIAS: model performance for present climate
  and
  CONVERGENCE: model agreement for future projections.

A Bayesian model: models projections and observations

Linear random effects model for a region:

For model $\boldsymbol{i}$

current temperature

$$\boldsymbol{X}_i = \mu + b_i + u_i$$

future projection

$$\boldsymbol{Y}_i = \nu + b'_i + v_i$$

observed temperature

$$\boldsymbol{X}_0 = \mu + e$$

True current temperature $\boldsymbol{\mu}$,
"true" future temperature $\boldsymbol{\nu}$

Model projection = true climate + model bias + noise

Key Assumption:

$$\boldsymbol{X}_i = \mu + b_i + u_i$$

$$\boldsymbol{Y}_i = \nu + b_i' + v_i$$

$E[b_i] = E[b_i'] = 0$

AOGCM's biases are treated as a random effect with zero mean.

The noise is due to the internal variability of the model (weather).

An identifiable model

The bias and internal variability are not identifiable with only one experiment per model.

Combine the model variability and the bias random effect into one variance term:

$$\boldsymbol{X}_i = \mu + e_i$$

$$\boldsymbol{Y}_i = \nu + \epsilon_i$$

$$\boldsymbol{X}_0 = \mu + e$$

The random components are mean zero, Gaussian

$VAR(e_i) = \lambda_i$ **and** $VAR(\epsilon_i) = \theta\lambda_i$

The goal

$$\boldsymbol{X}_i = \mu + e_i$$

$$\boldsymbol{Y}_i = \nu + \epsilon_i$$

$$\boldsymbol{X}_0 = \mu + e$$

The posterior for $(\nu - \mu)$ represents the uncertainty in the change in climate

Why not use the sample mean?

- **Not weighted by how well a model's matches current climate**

Why not use the sample mean?

- Not weighted by how well a model's matches current climate
- Not robust if a particular model projection disagrees with the others.

Why not use the sample mean?

- Not weighted by how well a model's matches current climate
- Not robust if a particular model projection disagrees with the others.
- What is the uncertainty?

Why not use the sample mean?

- Not weighted by how well a model's matches current climate
- Not robust if a particular model projection disagrees with the others.
- What is the uncertainty?

A Bayesian model continued: The priors

λ_i: Precision of the ith model

Bias of the ith model and Convergence of the ith model within the ensemble give information on λ_i

A Bayesian model continued: The priors

$\boldsymbol{\lambda_i}$: **Precision of the ith model**

Bias of the ith model and Convergence of the ith model within the ensemble give information on $\boldsymbol{\lambda_i}$

Prior distribution for $\boldsymbol{\lambda_i}$ *is*

$$\lambda_i \sim \text{Gamma}(.001, .001)$$

a very weak prior assumption.

More Priors

Priors for μ, ν and θ are:

$$\mu \sim \text{Uniform}(-\infty, +\infty)$$

$$\nu \sim \text{Uniform}(-\infty, +\infty)$$

$$\theta \sim \text{Gamma}(.001, .001)$$

As non-committed as we can be!

More Priors

Priors for μ, ν and θ are:

$$\mu \sim \text{Uniform}(-\infty, +\infty)$$

$$\nu \sim \text{Uniform}(-\infty, +\infty)$$

$$\theta \sim \text{Gamma}(.001, .001)$$

As non-committed as we can be!

Perhaps expert knowledge could be included ...

Markov Chain Monte Carlo

Statistical computation **The posterior does not have a simple form. As an alternative one generates a zillion samples from the posterior and makes a histogram of the density.**

- **Simple Gibbs sampler – all full conditionals are either gammas or Gaussians.**

Markov Chain Monte Carlo

Statistical computation **The posterior does not have a simple form. As an alternative one generates a zillion samples from the posterior and makes a histogram of the density.**

- **Simple Gibbs sampler – all full conditionals are either gammas or Gaussians.**
- **Conclusions based on a total of $50,000$ values for each parameter, representing a sample from its posterior distribution.**

Markov Chain Monte Carlo

Statistical computation **The posterior does not have a simple form. As an alternative one generates a zillion samples from the posterior and makes a histogram of the density.**

- **Simple Gibbs sampler – all full conditionals are either gammas or Gaussians.**
- **Conclusions based on a total of $50,000$ values for each parameter, representing a sample from its posterior distribution.**
- **Convergence of algorithm verified by standard diagnostic tools.**

Markov Chain Monte Carlo

Statistical computation **The posterior does not have a simple form. As an alternative one generates a zillion samples from the posterior and makes a histogram of the density.**

- **Simple Gibbs sampler – all full conditionals are either gammas or Gaussians.**
- **Conclusions based on a total of $50,000$ values for each parameter, representing a sample from its posterior distribution.**
- **Convergence of algorithm verified by standard diagnostic tools.**
- **You can do this at home, complete R source code is posted:** `www.cgd.ucar.edu/~nychka/man.html`

Conditional distributions for present and future temperature

Assume $\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \ldots, \boldsymbol{\lambda}_9$ known.

Posteriors for present and future true temperatures are centered around

$$\widetilde{\boldsymbol{\mu}} = (\Sigma_{i=0}^{9} \boldsymbol{\lambda}_i \boldsymbol{X}_i)/(\Sigma_{i=0}^{9} \boldsymbol{\lambda}_i)$$

and

$$\widetilde{\boldsymbol{\nu}} = (\Sigma_{i=1}^{9} \boldsymbol{\lambda}_i \boldsymbol{Y}_i)/(\Sigma_{i=1}^{9} \boldsymbol{\lambda}_i)$$

Conditional distributions for present and future temperature

Assume $\boldsymbol{\lambda_1, \lambda_2, \ldots, \lambda_9}$ known.

Posteriors for present and future true temperatures are centered around

$$\widetilde{\boldsymbol{\mu}} = (\Sigma_{i=0}^{9} \boldsymbol{\lambda}_i \boldsymbol{X}_i)/(\Sigma_{i=0}^{9} \boldsymbol{\lambda}_i)$$

and

$$\widetilde{\boldsymbol{\nu}} = (\Sigma_{i=1}^{9} \boldsymbol{\lambda}_i \boldsymbol{Y}_i)/(\Sigma_{i=1}^{9} \boldsymbol{\lambda}_i)$$

A weighted average!

But $\boldsymbol{\lambda_i}$ is unknown, so...back to bias and convergence!

The posterior mean for $\boldsymbol{\lambda_i}$ is

$$\frac{a+1}{b+\frac{1}{2}((X_i-\widetilde{\mu})^2+\theta(Y_i-\widetilde{\nu})^2)}$$

Precision is large only if both $|X_i - \widetilde{\mu}|$ (BIAS)

But λ_i is unknown, so...back to bias and convergence!

The posterior mean for λ_i is

$$\frac{a+1}{b+\frac{1}{2}((X_i-\widetilde{\mu})^2+\theta(Y_i-\widetilde{\nu})^2)}$$

Precision is large only if both $|X_i - \widetilde{\mu}|$ (BIAS)

and $|Y_i - \widetilde{\nu}|$ (CONVERGENCE) are small

But λ_i is unknown, so...back to bias and convergence!

The posterior mean for λ_i is

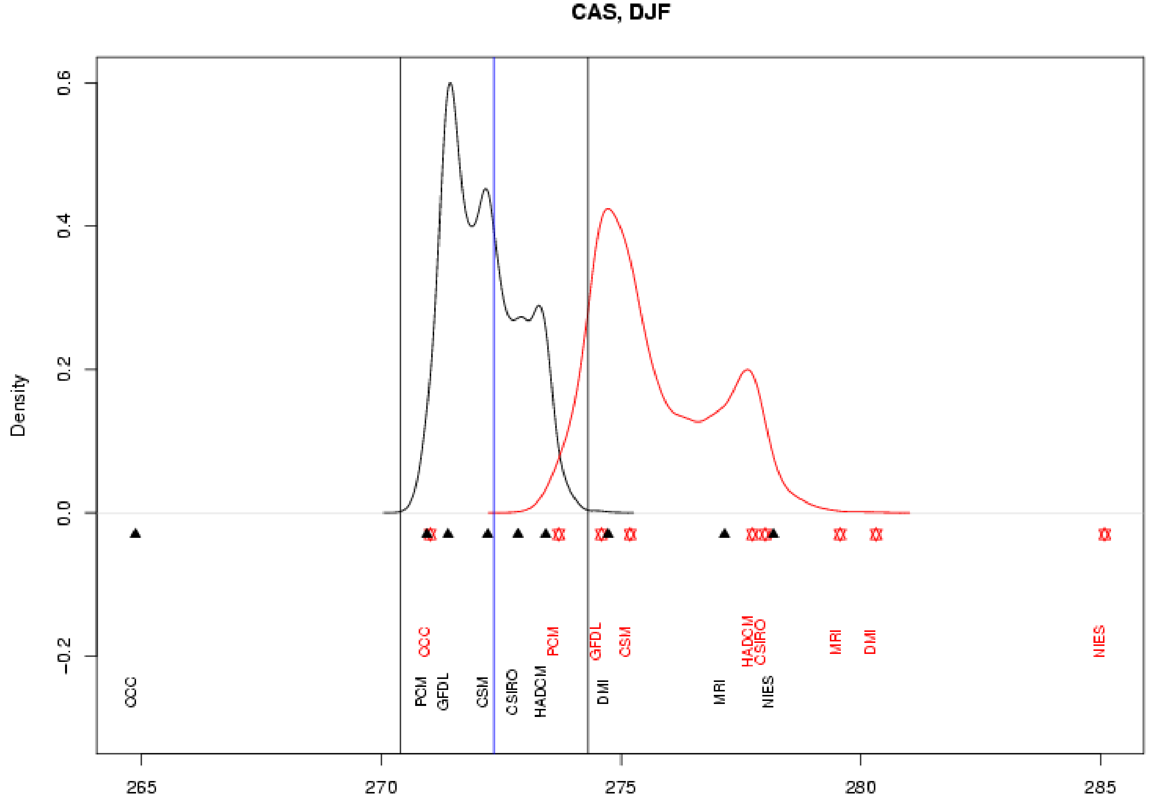
$$\frac{a+1}{b+\frac{1}{2}((X_i-\widetilde{\mu})^2+\theta(Y_i-\widetilde{\nu})^2)}$$

Precision is large only if both $|X_i - \widetilde{\mu}|$ (BIAS)

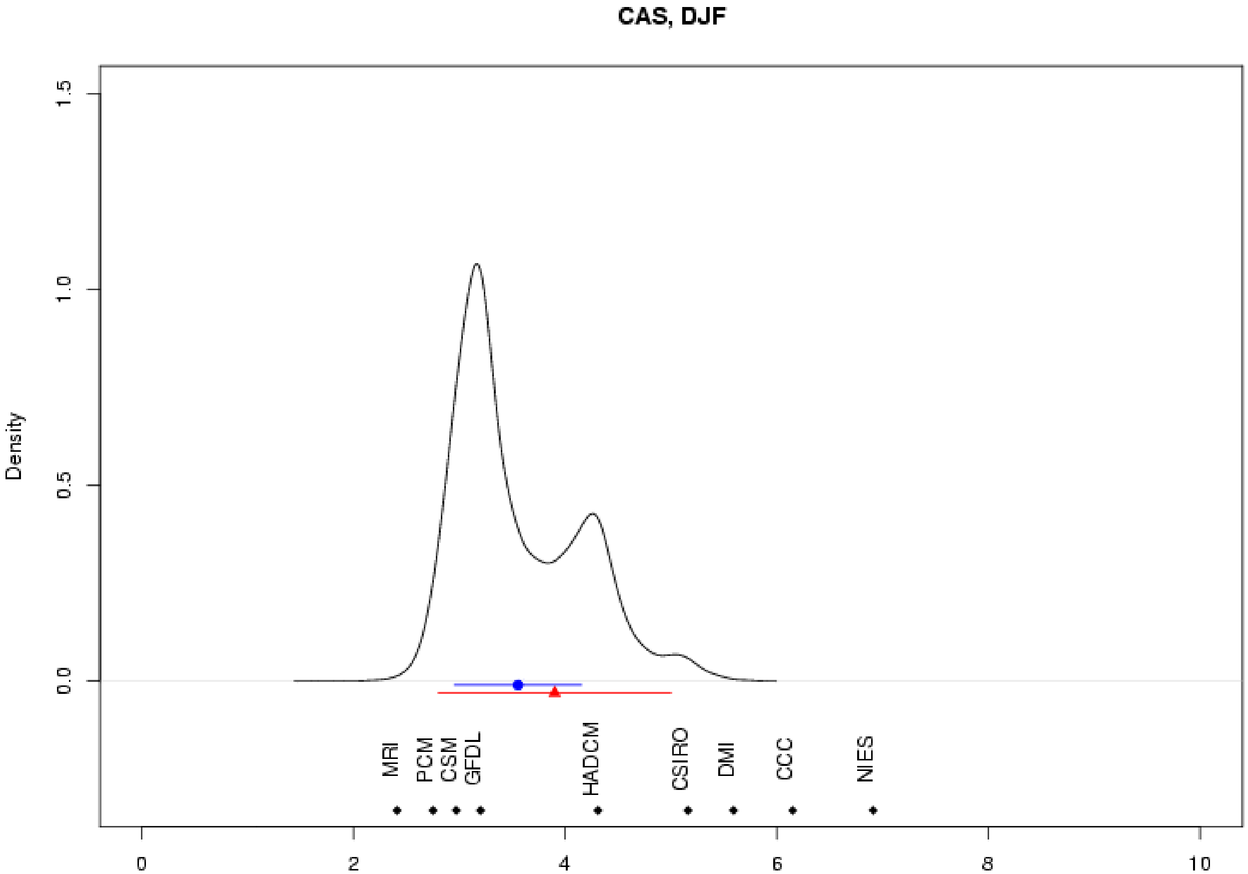
and $|Y_i - \widetilde{\nu}|$ (CONVERGENCE) are small

The BIAS is just $|X_i - X_0|$
if the observations are assumed to have no error.

A tour of Central Asia: posteriors for μ and ν

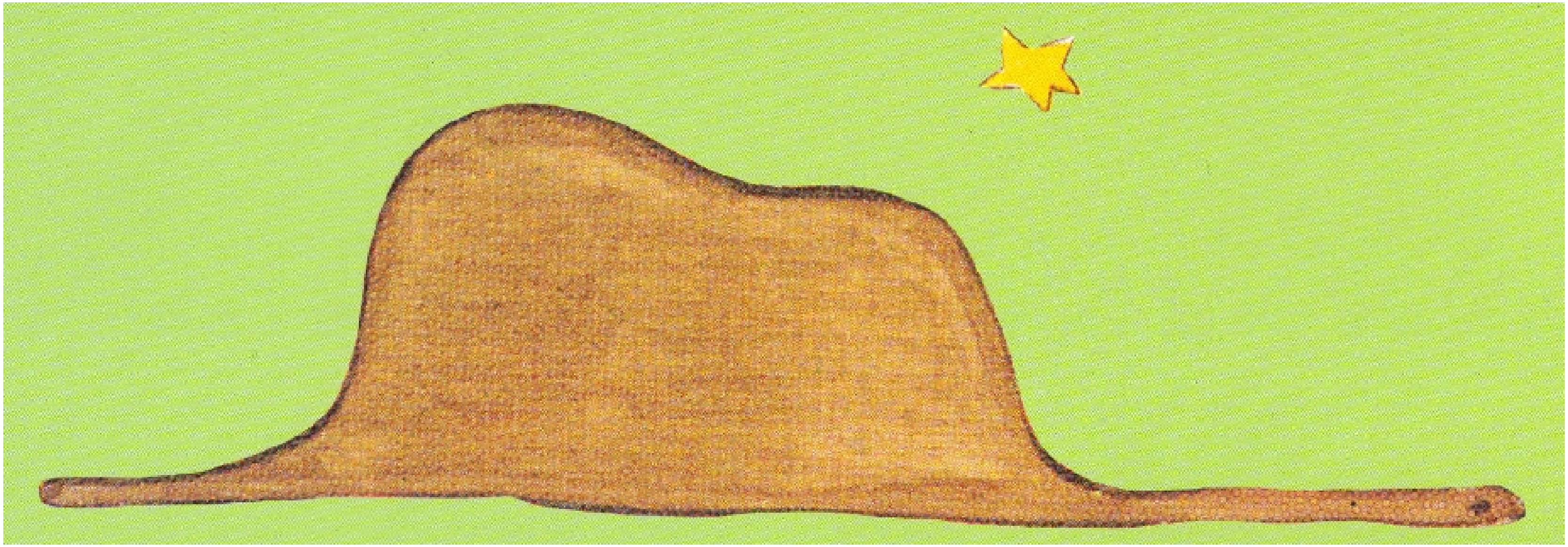


Posterior for climate change $\Delta T = \nu - \mu$



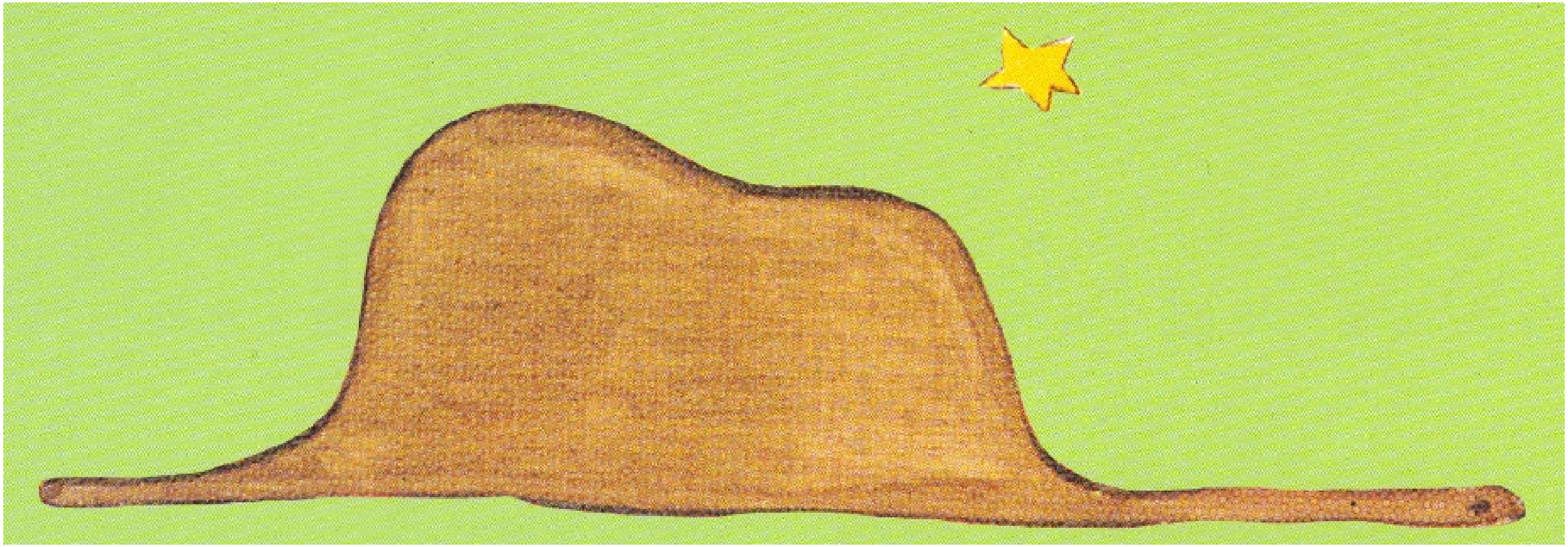
NIES	MRI	CCC	CSIRO	CSM	PCM	GFDL	DMI	HADCM
5.83	4.81	-7.48	0.50	-0.13	-1.40	-0.96	2.38	1.08

What is in a distribution?



The “fedora hat” shape would not be well explained by the mean and standard deviation.

What is in a distribution?

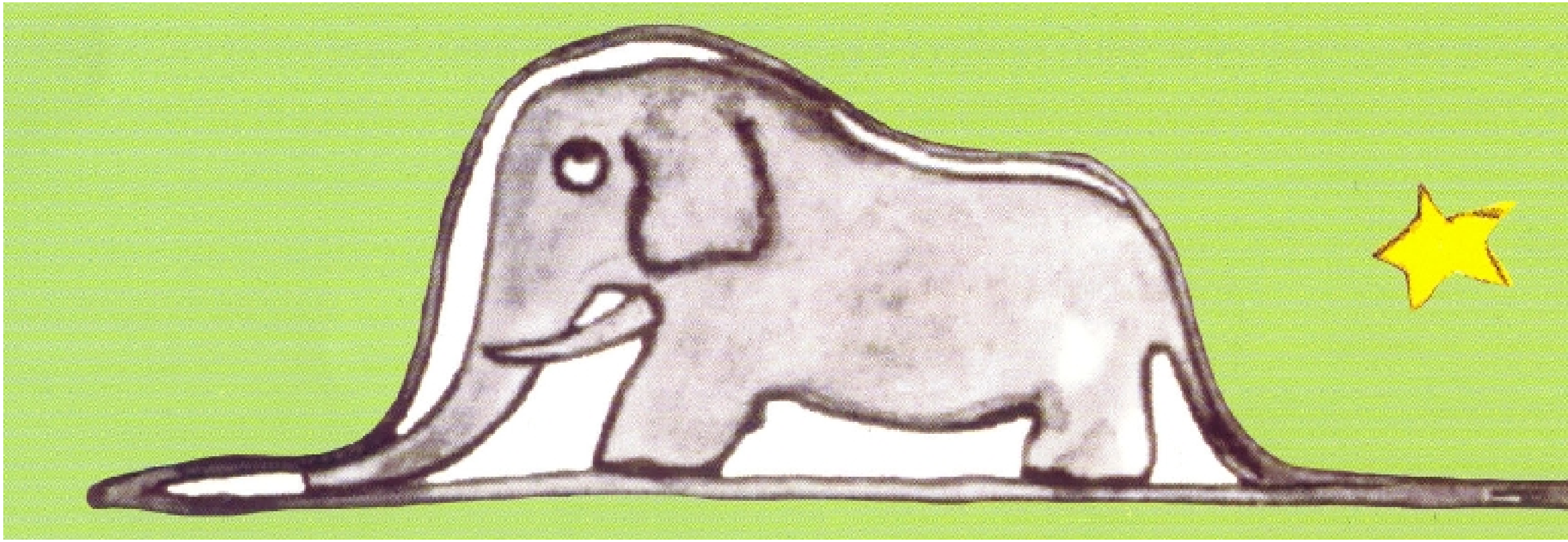


The “fedora hat” shape would not be well explained by the mean and standard deviation.

One explanation is that this “random sample” of models is actually is not as representative of the full range of physics.

A more biological explanation of the shape

A more biological explanation of the shape



A (large) snake who has swallowed an elephant.

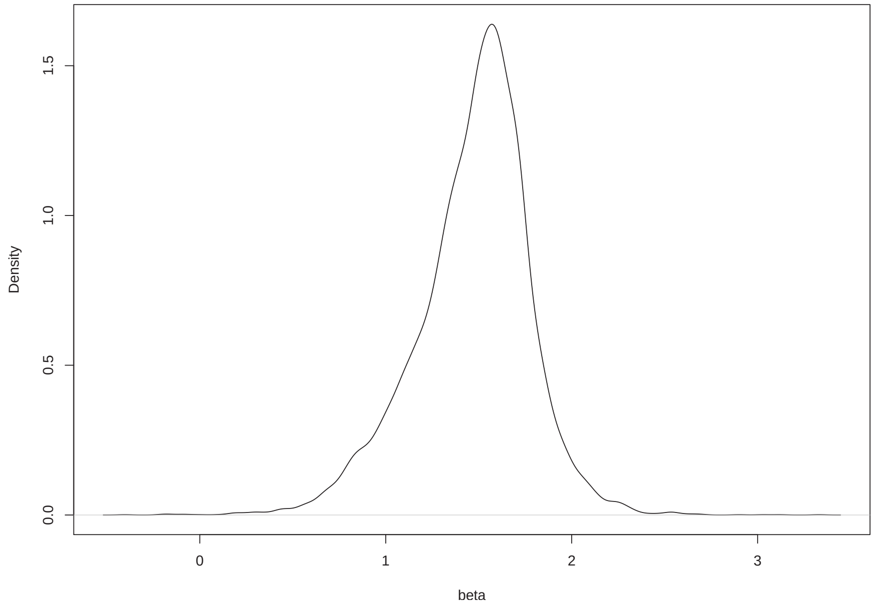
Model Extensions

1. Is $\boldsymbol{Y}_i$ (cor)related with $\boldsymbol{X}_i$?
2. Do we have real outliers among $\boldsymbol{X}_i$ and $\boldsymbol{Y}_i$?

Easily modeled:

1. Assume
$$\boldsymbol{X}_i = \mu + e_i$$
and
$$\boldsymbol{Y}_i = \nu + \beta(X_i - \mu) + \epsilon_i$$
2. Assume heavy-tailed distributions instead of Gaussians for $\boldsymbol{X}_i$ and $\boldsymbol{Y}_i$

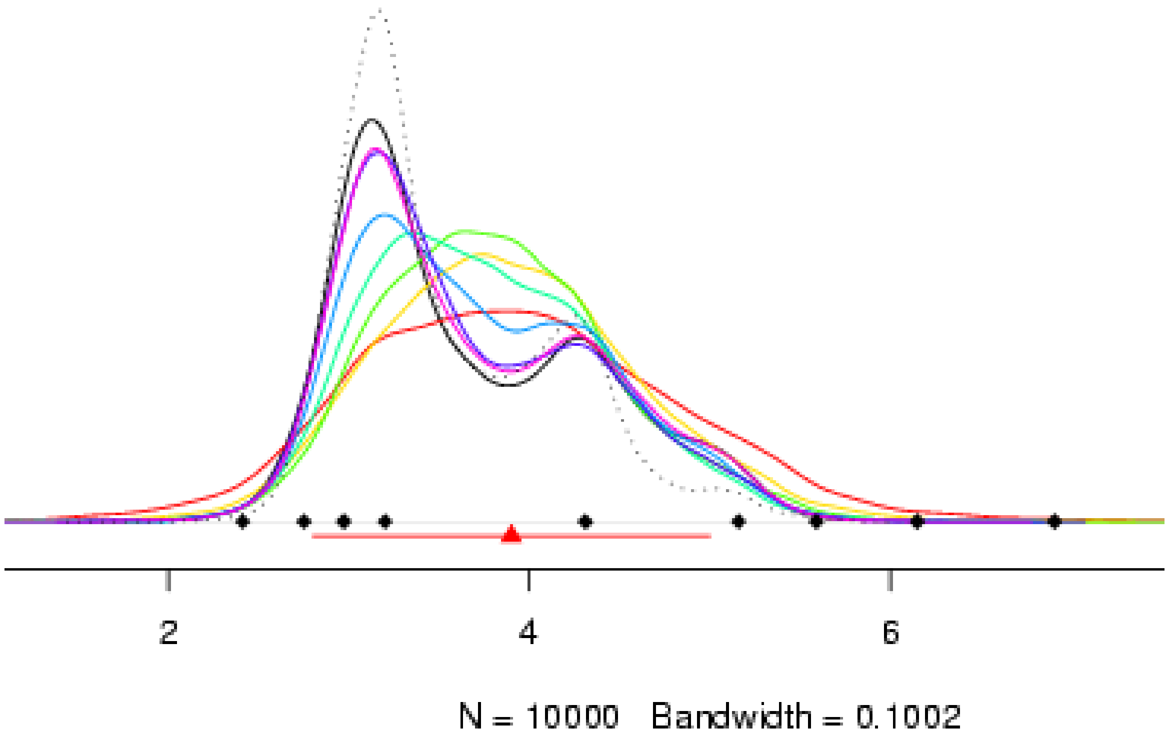
A tour of Central Asia
Regression coefficient between future and present climate β



Different from 0!

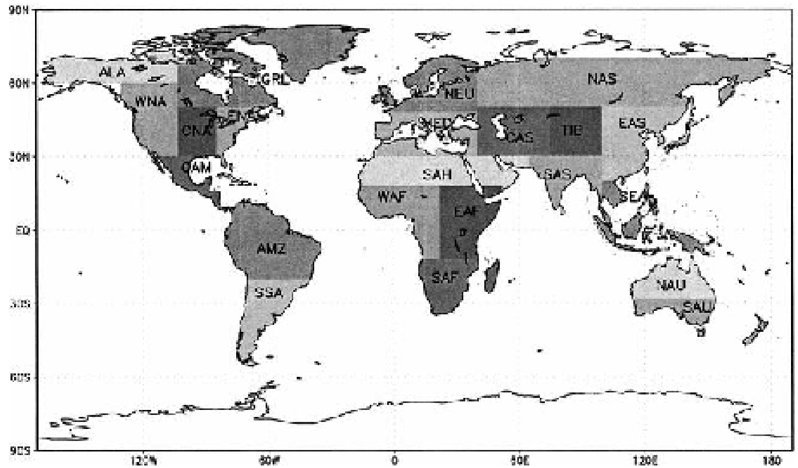
A tour of Central Asia
Different assumptions for error distribution.

Results varying across T-family.



A multivariate version

How can model results be combined across regions.



Index across models and regions

i indexes AOGCMs (9), j indexes regions (22)

Then:

$$X_{0j} = \mu_j + e_{0j}$$

Present temperatures

$$X_{ij} = \mu_j + b_i + e_{ij}$$

Future temperature projections across regions

$$Y_{ij} = \nu_j + b'_i + \beta_x(X_{ij} - \mu_j - b_i) + \epsilon_{ij}$$

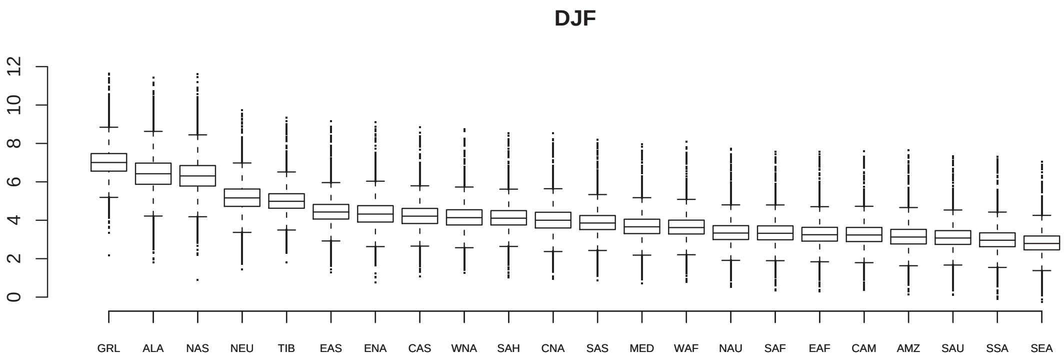
Linking biases

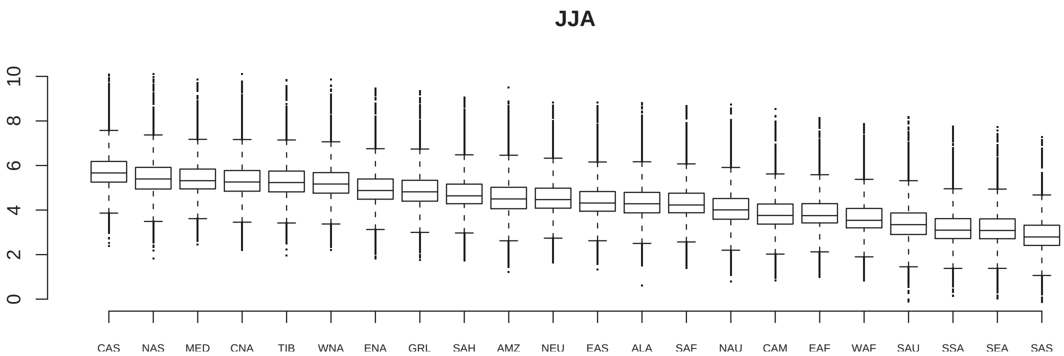
$$b'_i = \beta_b b_i + \omega_i$$

Main features

- Still region specific μ_j and ν_j.
- The additive effects b_i and b_i', common to all regions for a given model, introduce correlation.
- $\boldsymbol{\beta}_b$ and $\boldsymbol{\beta}_x$ introduce correlation between regions as well, in addition to allowing for correlation between future and current responses.

Regional climate change
The big picture





A spatial model for the full AOGCM output

$\boldsymbol{z}$ is a grid cell location.

$$X_0(\boldsymbol{z}) = \mu(\boldsymbol{z}) + u_0(\boldsymbol{z})$$

$$X_i(\boldsymbol{z}) = \mu(\boldsymbol{z}) + u_i(\boldsymbol{z})$$

$$Y_i(\boldsymbol{z}) = \mu(\boldsymbol{z}) + v_i(\boldsymbol{z})$$

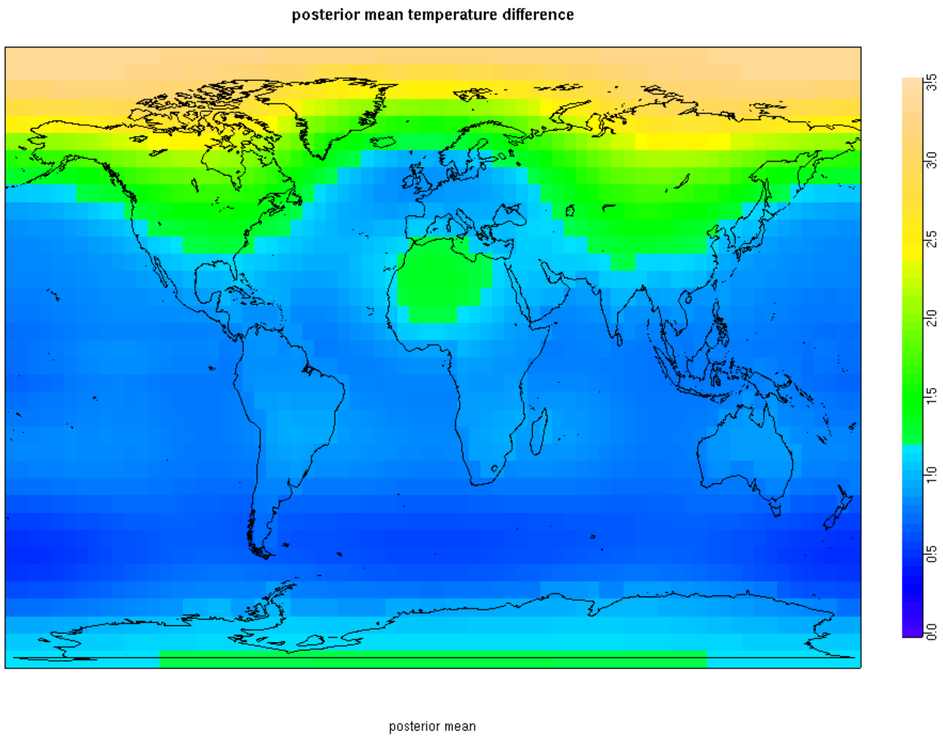
Here the climate change has the form

$$\mu(\boldsymbol{z}) = \sum \psi_k(\boldsymbol{z})\theta_k$$

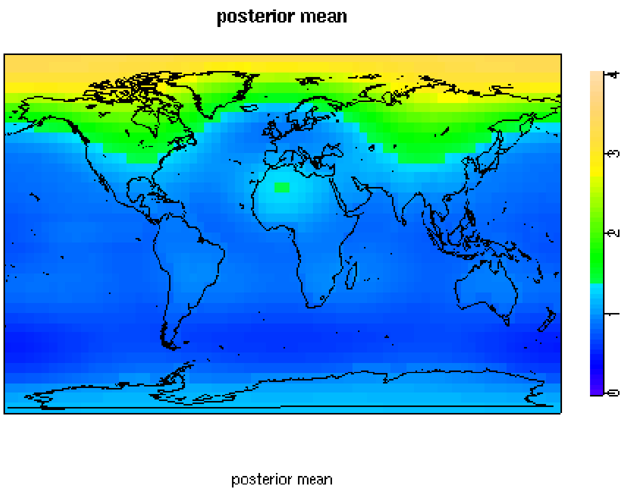
where $\{\psi_k\}$ are spherical harmonics. $u_i(\boldsymbol{z})$ and $v_i(\boldsymbol{z})$ are isotropic processes on the sphere.

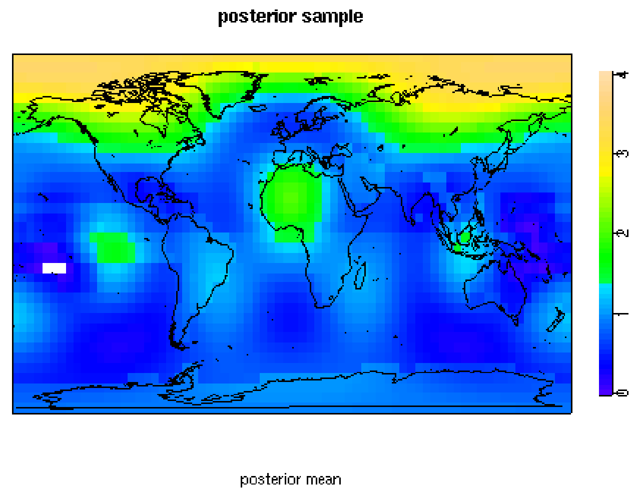
Preliminary results

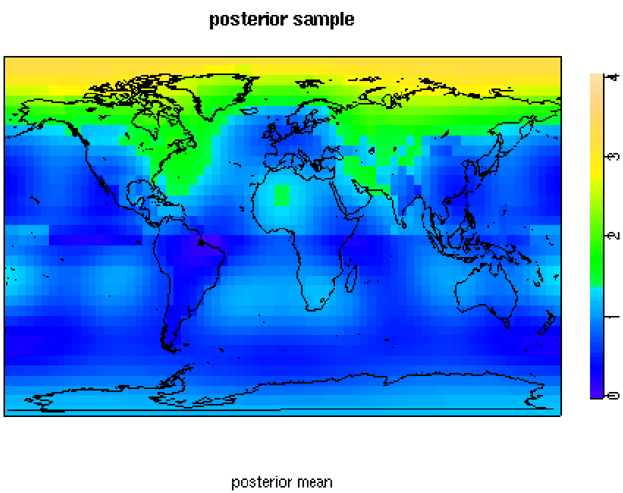
15 AOGCMS on a 5×5 grid taken from the MAGICC/SCENGEN package
Posterior mean field:

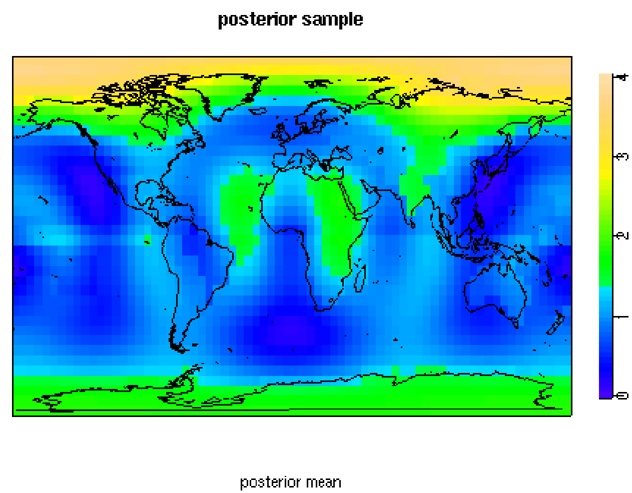


Some ensemble members









Conclusions

- **We have formalized the criteria of *bias* and *convergence* as a way of analyzing Multi-model ensembles.**
- **There is a hierarchy of models available. The assumptions for each are clearly stated. In particular the prior assumptions are vague, not constraining any of the parameters a priori.**
- **The posterior distributions from combining models can be used to propagate uncertainty into other models to assess the impacts of a changed climate.**
- **We can perform sensitivity analysis to prior assumptions.**