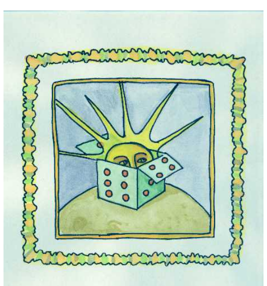
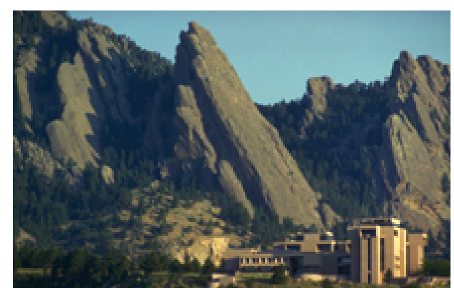
Combining climate model experiments

Claudia Tebaldi, Doug Nychka
National Center for Atmospheric Research Richard Smith
UNC-Chapel Hill

- **Climate change and climate models**
- **Inference for a single region**
- **Two-way effects for several regions.**

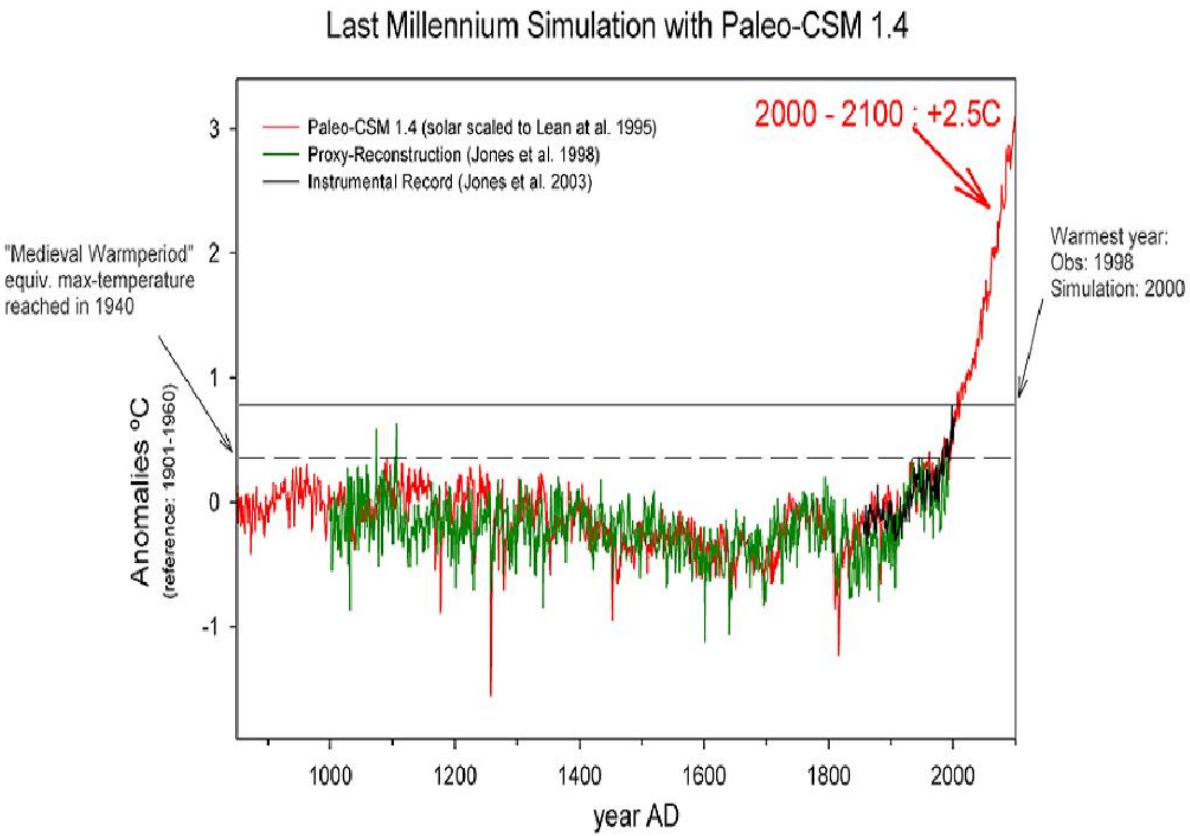
How does a changed (warmer) climate effect human health?



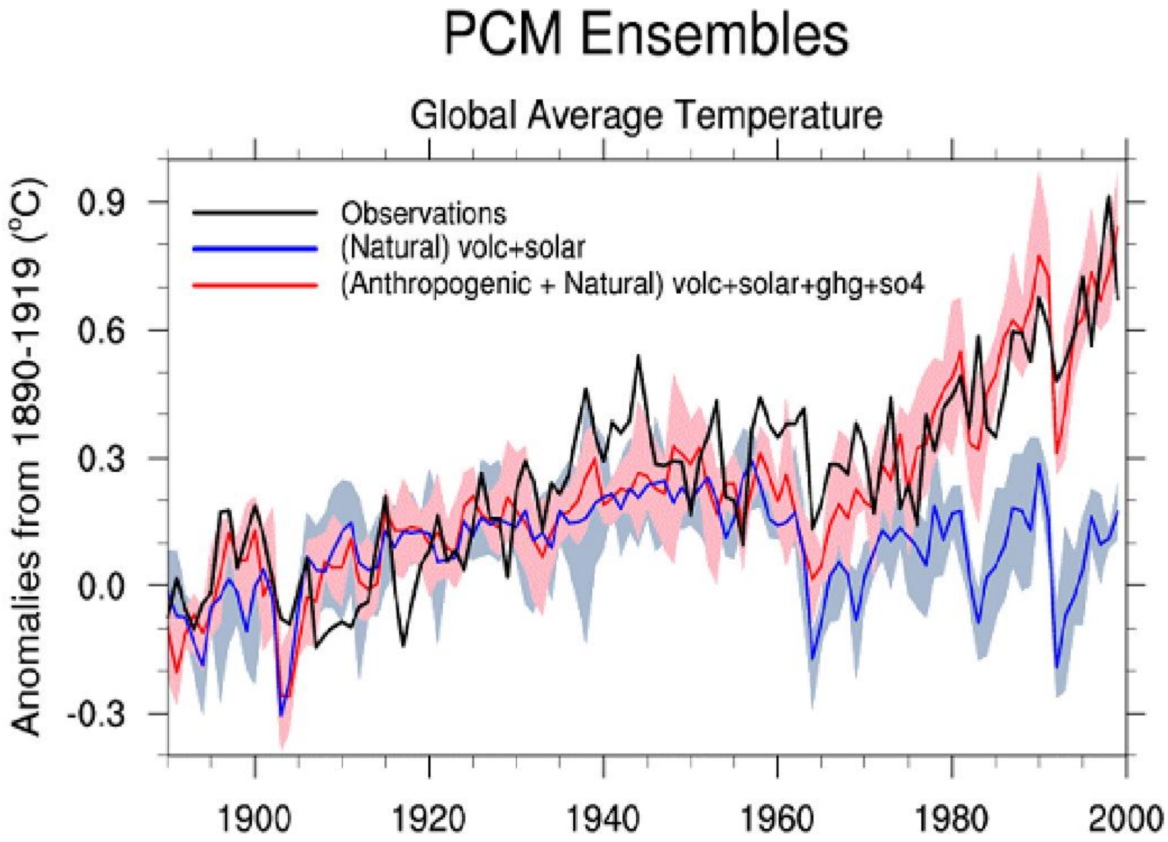


Climate: **What you expect ...** *Weather:* **What you get.**

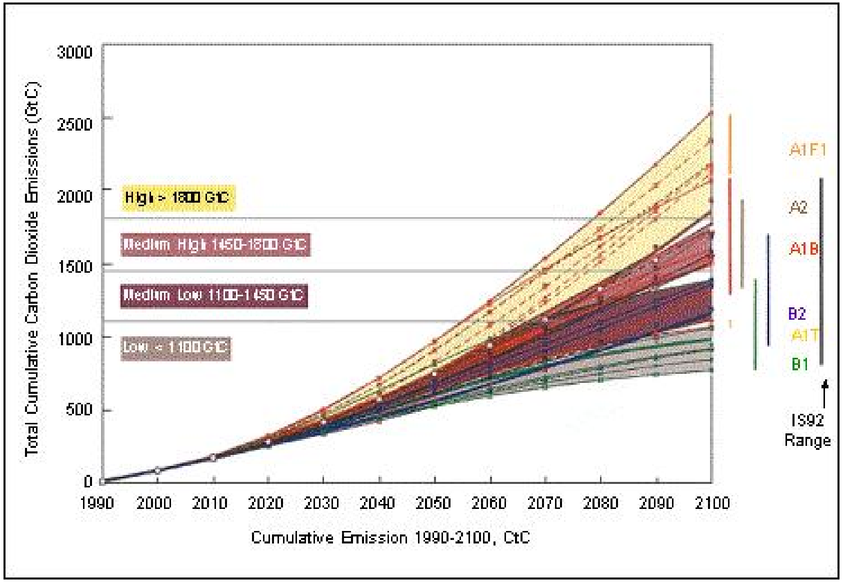
Recent warming appears unusual from the proxy records of global temperatures



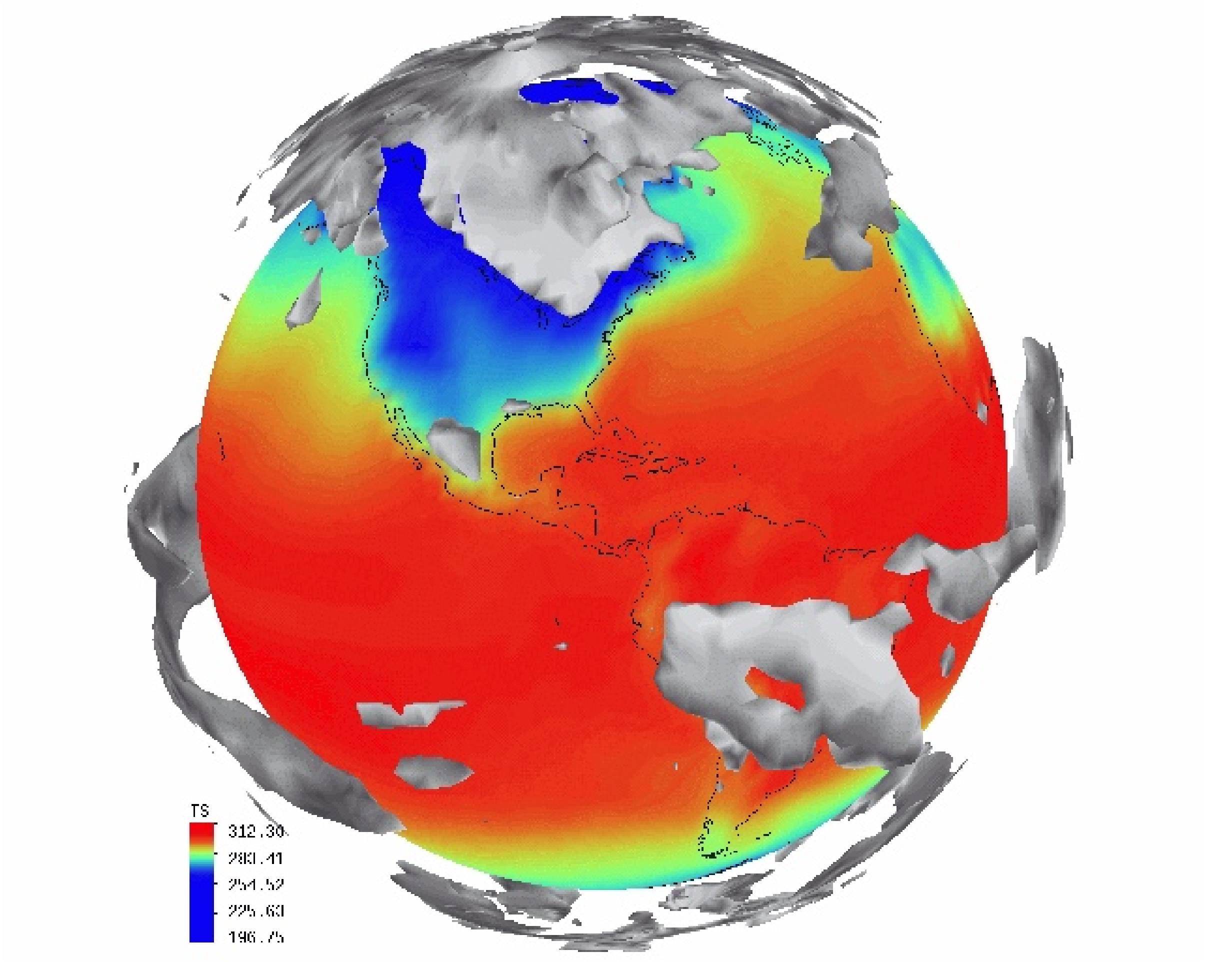
Evidence for attribution due to human activities



Scenarios for emissions in the future



A snapshot of a climate model



How do they do it?

Modeling the atmosphere

The physical equations to describe atmospheric motion are derived from fluid mechanics and thermodynamics.

The complete state depends on:

- 3-d wind field, $\boldsymbol{v}$
- pressure p
- temperature T
- heating by radiation $Q_{\mathbf{r}ad}$, condensation $Q_{\mathbf{c}on}$
- evaporation E and condensation C from clouds
- D_H D_M and D_q are diffusion terms.

Climate System Model (CSM)

General Circulation Model (GCM): **A deterministic numerical model that describes the circulation of the atmosphere by solving the primitive equations in a discretized form.**

- **Conceptually based on grid boxes (for the NCAR climate system model: there are $128 \times 64 \times 17 \approx 141K$) and the state of the atmosphere is the average quantities for each box ($\approx 1M$ real numbers).**
- **Each grid cell is large (for NCAR CSM/PCM $\approx$ 170 $\times$ 170 miles) and so important processes that affect large scale flow are not resolved by the grid.**
- **GCM must be stepped on the order of minutes, even for a 200+ year numerical experiment! To halve the horizontal resolution the amount of computing goes up by $2^4 = 16$.**

Climate System Model

A GCM coupled to other models for the ocean, ice , land, chemistry, etc. to model the entire climate system. Coupling these components without overt flux adjustments is an important feature.

Called an AOGCM because air/ocean coupling is the most important.

- **Several models years can be simulated per *day* on a large computer. Full numerical experiments are limited and expensive.**
- **The NCAR model takes 50+ people to maintain and develop**
- **Completely determinisitic!**

Motivation

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

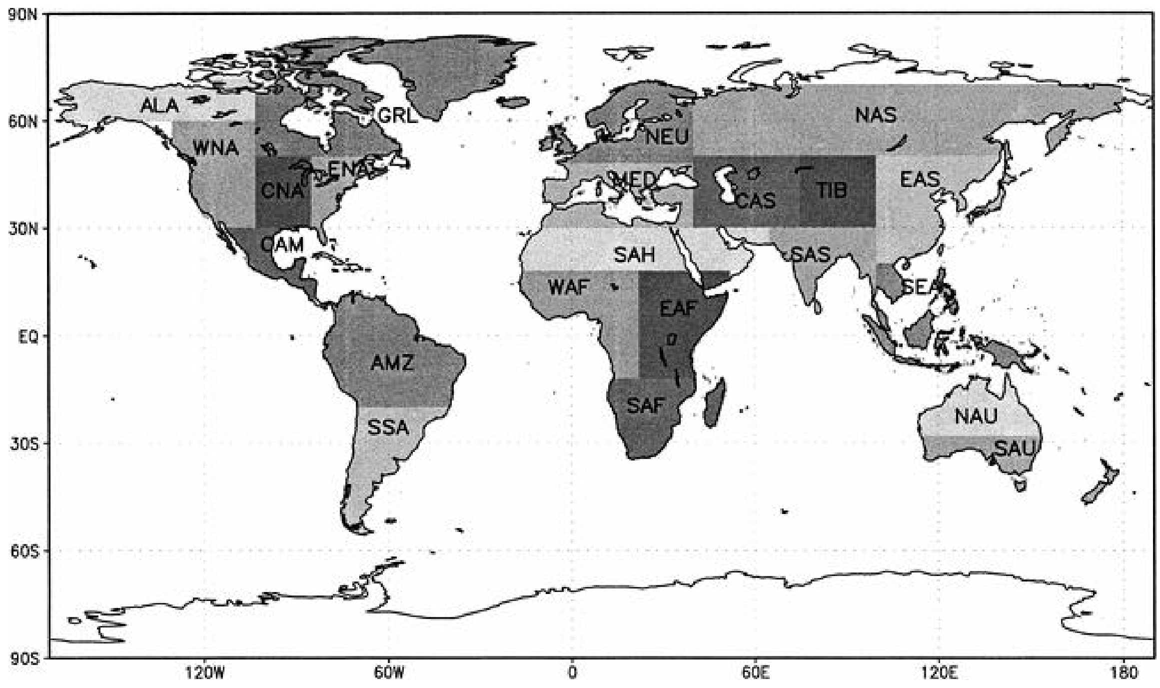
- **Reconciling different projections - no model is the true model!**
- **Offering stake-holders and policy-makers a probabilistic forecast.**
- **Substituting formal probabilistic assumptions for heuristic criteria, and testing sensitivity of the results to them.**

Impacts of climate change include: Extremes in summer temperatures, Possible degradation in air quality, Changes in the domain of vector-borne diseases. All of these have implications for human health.

The Data

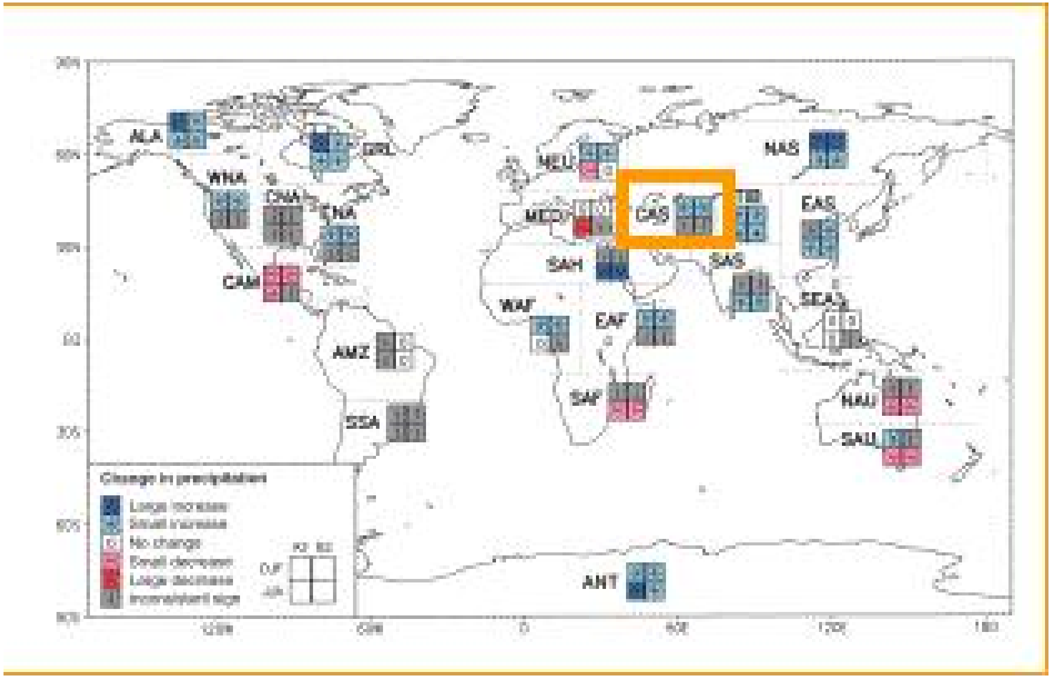
- 9 AOGCMs;
- 22 Regions;
- 2 Seasons;
- Simulated Temperature values in 30-years averages (X, 1961-1990; Y, 2071-2100 (A2));
- Observed Temperature average, X_0, for 1961-1990. (Allows for an estimate of model bias for current climate.)

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

- *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 take into account:
  model performance (**BIAS**)
  and
  model agreement (**CONVERGENCE**).

Reliability Ensemble Average (cont'd)

Given the single AOGCM responses:

$$\{\Delta T_i\}_{i=1,\ldots,9}$$

The summary is given by a weighted average:

$$\widehat{\Delta T} = \sum_i \frac{R_i \Delta T_i}{\sum_i R_i}$$

where the weights are iteratively recomputed, since they include $\widehat{\Delta T}$ itself, the target of the estimation:

$$R_i = K_{nat.\ var.} \cdot \left(\frac{1}{|T_0 - T_i|} \cdot \frac{1}{|\widehat{\Delta T} - \Delta T_i|} \right)^p$$

Incidentally: This is robust estimation!

The (implicit) loss function minimized is:

$$\sum_i w_i |\Delta T_i - \boldsymbol{\delta}|^{2-p}$$

If $\boldsymbol{p} = 1$, $\hat{\boldsymbol{\delta}}$ is the (weighted) median of the 9 AOGCM responses.

A Bayesian model for future climate outcomes

For one region:

Model i produces a current temperature reconstruction

$$X_i \sim N[\mu, \lambda_i^{-1}]$$

and a future temperature projection

$$Y_i \sim N[\nu, (\theta\lambda_i)^{-1}]$$

The observed current temperature is

$$X_0 \sim N[\mu, \lambda_0^{-1}]$$

True current temperature μ, true future temperature ν, AOGCM's precision λ_i, "inflation/deflation" of precision future θ

A Bayesian model for future climate outcomes (cont'd)

The ith model has some unknown precision λ_i

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

Prior distribution is

$$\lambda_i \sim \Gamma(a, b)$$

with $a = b = .001$

Very weak prior assumption – nevertheless proper posteriors result.

A Bayesian model for future climate outcomes (cont'd)

Priors for μ, ν and θ are:

$$\mu \sim U(-\infty, +\infty)$$

$$\nu \sim U(-\infty, +\infty)$$

$$\theta \sim \Gamma(c, d)$$

with $c = d = .001$

As non-committed as we can be.

Perhaps expert knowledge could be included.

Gibbs sampler

- **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 verified by standard diagnostic tools.**

Conditional distributions for present and future temperature

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

Define

$$\widetilde{\mu} = (\textstyle\sum_{i=0}^{9} \lambda_i X_i)/(\sum_{i=0}^{9} \lambda_i)$$

and

$$\widetilde{\nu} = (\textstyle\sum_{i=1}^{9} \lambda_i Y_i)/(\sum_{i=1}^{9} \lambda_i)$$

Then, posteriors for present and future true temperatures:

$$\begin{aligned} \mu|\ldots &\sim N[\widetilde{\mu}, (\textstyle\sum_{i=0}^{9} \lambda_i)^{-1}] \\ \nu|\ldots &\sim N[\widetilde{\nu}, (\theta \textstyle\sum_{i=1}^{9} \lambda_i)^{-1}] \end{aligned}$$

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

The posterior for λ_i is $\Gamma[a+1, b+\frac{1}{2}((X_i-\widetilde{\mu})^2+\theta(Y_i-\widetilde{\nu})^2)]$

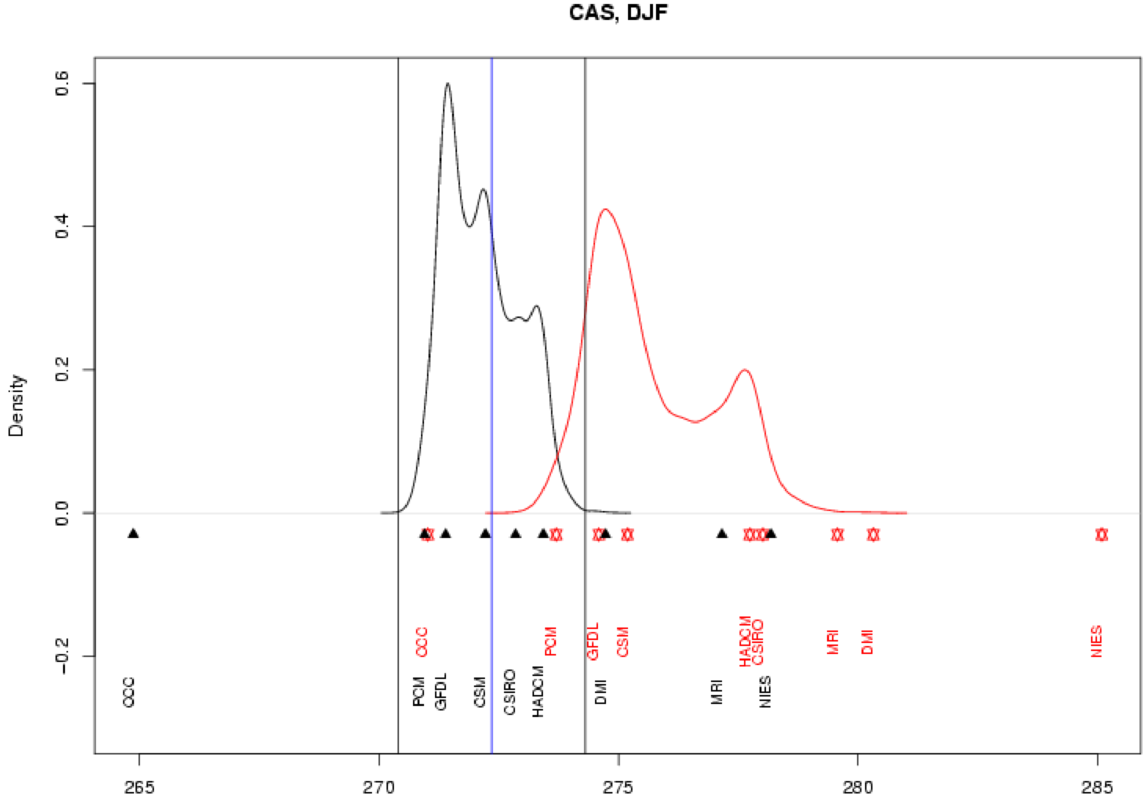
The posterior mean for λ_i is

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

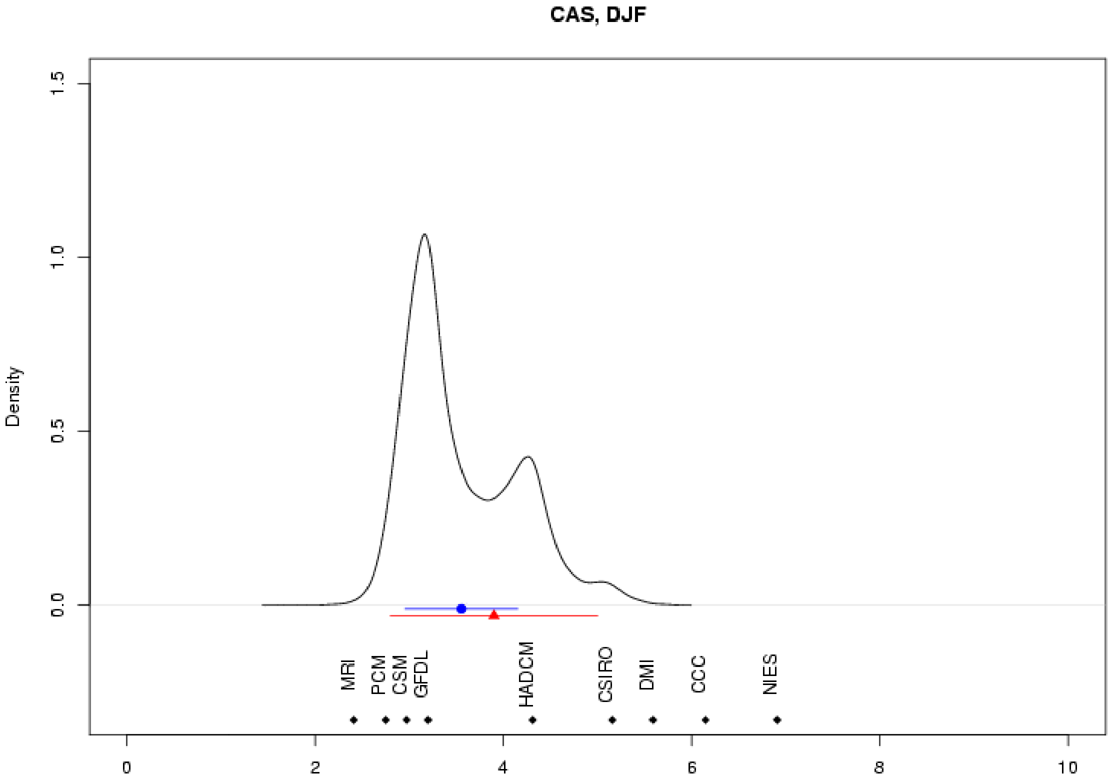
Large only if both $|X_i-\widetilde{\mu}|$ ("bias")
and $|Y_i-\widetilde{\nu}|$ ("convergence") are small

The "bias" becomes exactly $|X_i-X_0|$
if $\lambda_0\rightarrow\infty$ in which case $\widetilde{\mu}\rightarrow X_0$

A tour of Central Asia: posteriors for μ and ν

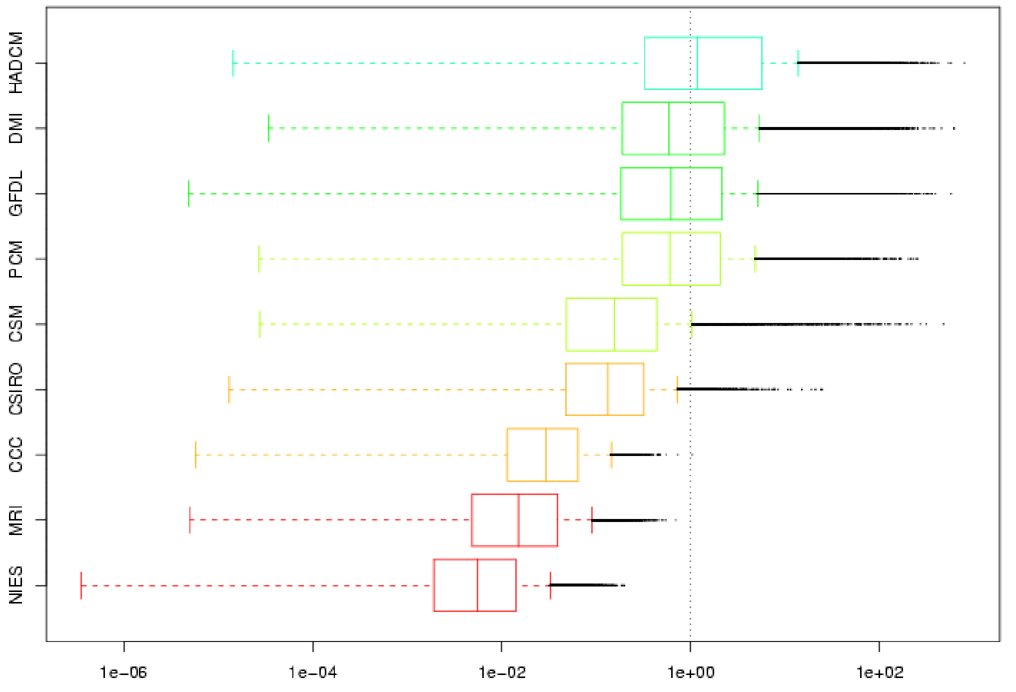


Posterior for climate change $\Delta T = \nu - \mu$



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

A tour of Central Asia
Model precision λ_i



	NIES	MRI	CCC	CSIRO	CSM	PCM	GFDL	DMI	HADCM
$\tilde{\lambda}_i / \sum_i \tilde{\lambda}_i \times 100$	0.04	0.12	0.18	1.09	5.00	12.73	19.95	23.08	37.81

Model precision as weight

Notice the amount of additional information when going from a table to a picture of distributions.

Clear ranking of models, but substantial spread and uncertainty (overlapping of the distributions).

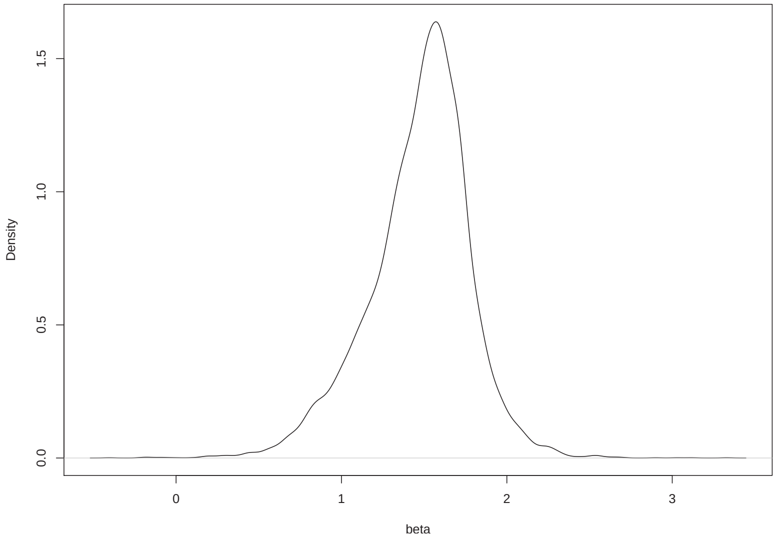
Extensions

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
$$\boldsymbol{X}_i \sim N[\boldsymbol{\mu}, (\boldsymbol{\lambda}_i)^{-1}]$$
and
$$\boldsymbol{Y}_i \sim N[\boldsymbol{\nu} + \boldsymbol{\beta}(\boldsymbol{X}_i - \boldsymbol{\mu}), (\boldsymbol{\theta}\boldsymbol{\lambda}_i)^{-1}]$$
2. Assume heavy-tailed distributions instead of gaussians for $\boldsymbol{X}_i$ and $\boldsymbol{Y}_i$

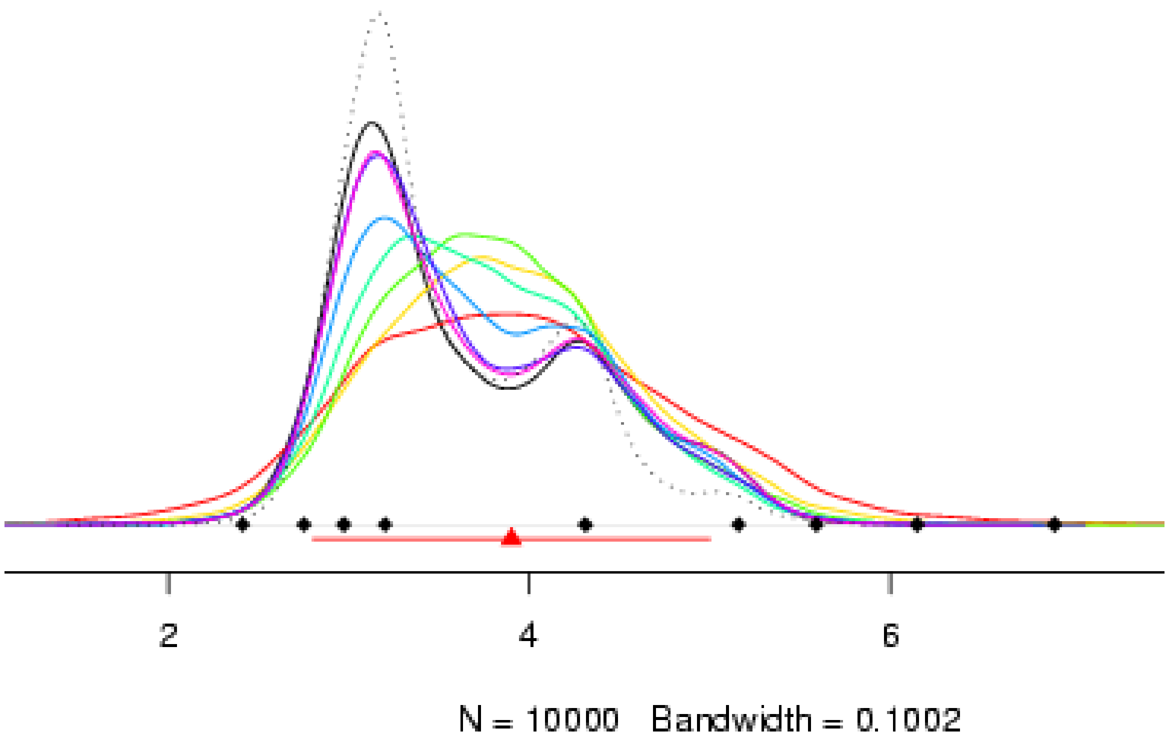
A tour of Central Asia
Regression coefficient between future and present climate β



Different from 0!

A tour of Central Asia
Climate change under different statistical assumptions

Results varying across T-family.



A multivariate version

Are the temperatures of an AOGCM in different regions correlated?

Double indexing...

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

Then:

$$
\begin{aligned}
X_{0j} &\sim N[\mu_j, \lambda_{0j}^{-1}], \\
X_{ij} &\sim N[\mu_j + \alpha_i, (\phi_j \lambda_i)^{-1}], \\
Y_{ij} &\sim N[\nu_j + \alpha_i{}' + \beta_x (X_{ij} - \mu_j - \alpha_i), (\theta_j \lambda_i)^{-1}], \\
\alpha_i{}' &\sim N[\beta_\alpha \alpha_i, (\psi_i)^{-1}].
\end{aligned}
$$

Main features

- **Still region specific μ_j and ν_j.**
- **The additive effects α_i and α'_i, common to all regions for a given model, introduce correlation.**
- **$\boldsymbol{\beta}_{\boldsymbol{\alpha}}$ and $\boldsymbol{\beta}_{\boldsymbol{x}}$ introduce correlation between regions as well, in addition to allowing for correlation between future and current responses.**

Main features (cont'd)

- Systematic variations of precision with regions, but retaining a "model precision" component: the precision is a product of two factors.

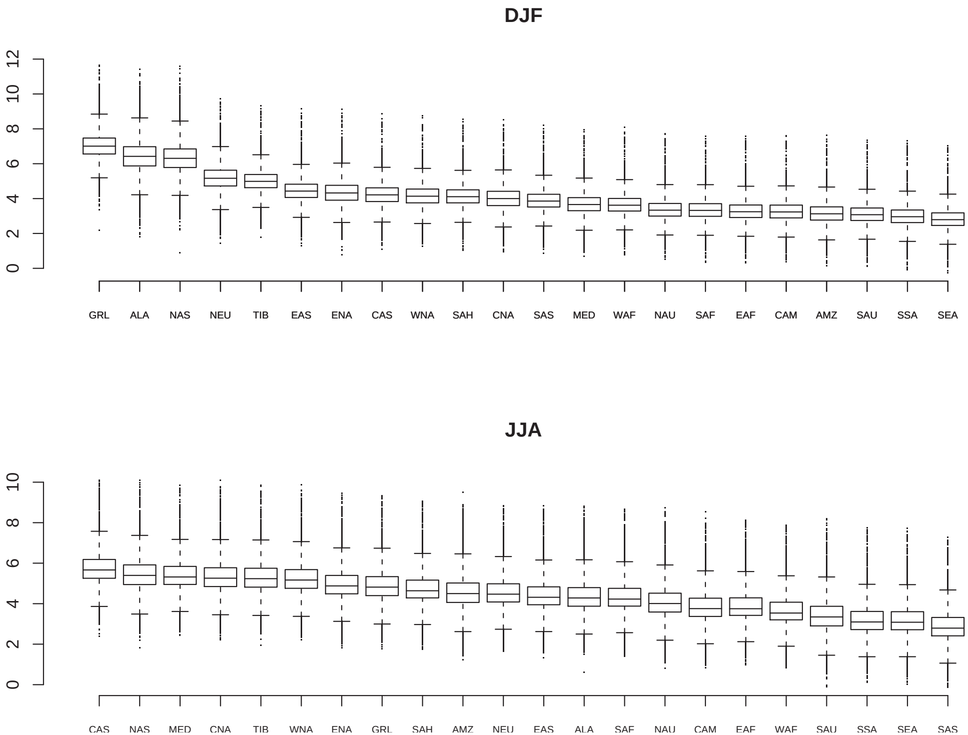
  $\boldsymbol{\lambda_i}$ model-specific

  $\boldsymbol{\theta_j, \phi_j}$ region-specific

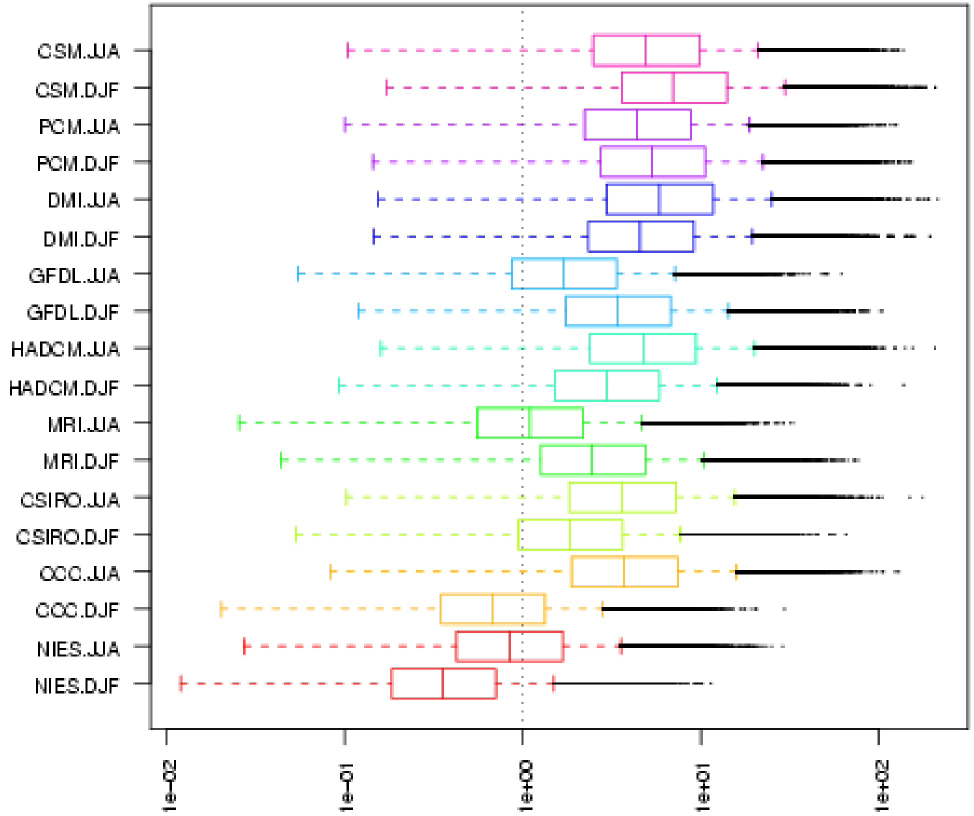
  - We borrow strength from all the regional responses in estimating $\boldsymbol{\lambda_i}$'s;
  - We gather information from all the models in the posterior distribution of $\boldsymbol{\theta_j, \phi_j}$'s.

- Two different region-specific factors, $\boldsymbol{\theta_j}$ and $\boldsymbol{\phi_j}$ in the precisions of present and future temperatures' distributions: the "quality" of the regional climate simulation may vary between the two simulation periods.

Climate change
The big picture



Model-specific precision factors λ_i



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 chnaged climate.**
- **We can perform sensitivity analysis to prior assumptions.**