Ensemble of Climate Models

Claudia Tebaldi
Climate Central
and Department of Statistics, UBC

Reto Knutti, Reinhard Furrer, Richard Smith, Bruno Sanso'

Outline

Multi-model ensembles (MMEs) - a description at face value

- What they are and what they can provide
- Where to find them
- How they are used (in particular in IPCC-AR4)

Multi-model ensembles and future climate change projections

- What are the main uncertainties in future projections, even given a specific emission scenario
- What MMEs can characterize and what they cannot
- Main issues, challenges and value of multi-model analysis

An example of analysis, summary and way forward

http://www-pcmdi.llnl.gov/ipcc/about_ipcc.php

This unprecedented collection of recent model output is officially known as the "WCRP CMIP3 multi-model dataset." It is meant to serve IPCC's Working Group 1, which focuses on the physical climate system -- atmosphere, land surface, ocean and sea ice -- and the choice of variables archived at the PCMDI reflects this focus. A more comprehensive set of output for a given model may be available from the modeling center that produced it.

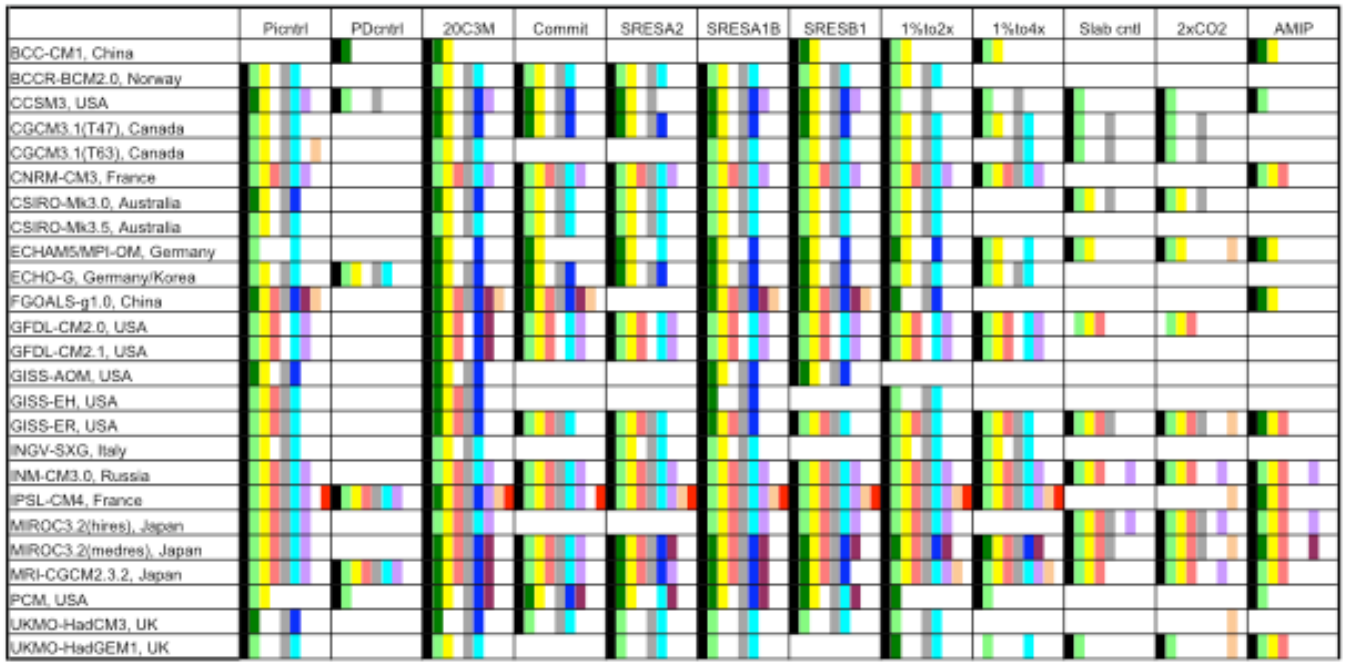
With the consent of participating climate modelling groups, the WGCM has declared the CMIP3 multi-model dataset open and free for non-commercial purposes. After registering and agreeing to the "terms of use," anyone can now obtain model output via the ESG data portal, ftp, or the OPeNDAP server.

As of January 2007, over 35 terabytes of data were in the archive and over 337 terabytes of data had been downloaded among the more than 1200 registered users. Over 250 journal articles, based at least in part on the dataset, have been published or have been accepted for peer-reviewed publication.

Data Availability Summary (as of 16 July 2007)

shaded area indicates that at least some but not necessarily all fields are available for data type indicated

- time-independent land surface
- >1 1 monthly-mean atmosphere
- daily-mean atmosphere
- 3-hourly atmosphere
- time-independent ocean
- >1 1 monthly-mean ocean
- >1 1 Extreme Indices
- Forcing
- ISCCP Simulator



These are General Circulation Models

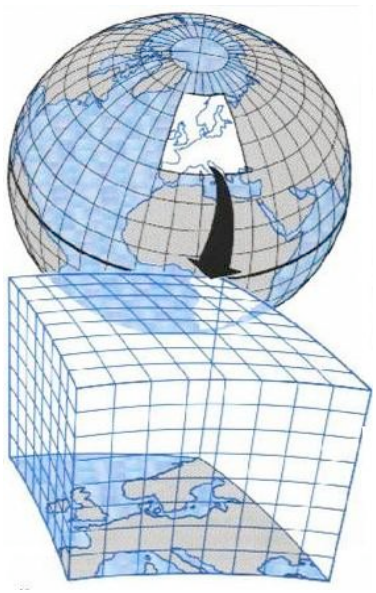
They represent fully coupled ocean-atmosphere (and land and ice) processes.

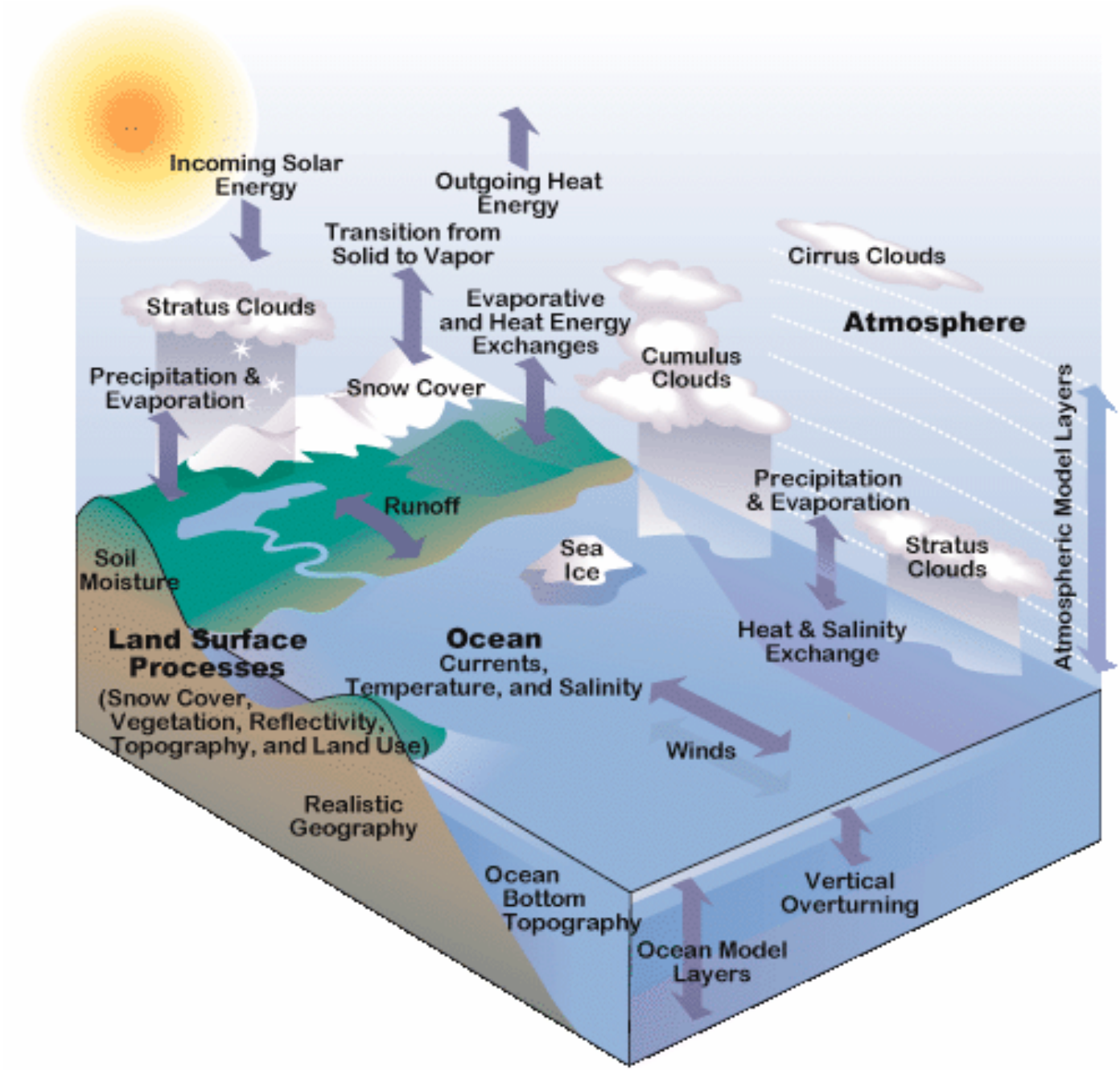
The most recent ensemble has a median horizontal resolution ~250Km.

A typical experiment consist of a handful of runs of 250 years' worth of simulated climate (output saved as 6-hourly, daily, monthly…).

~2 months' worth of (super) computer time for a single run.

Often each experiment consists of an "ensemble" of runs (different initial conditions).




Incoming Solar Energy
Outgoing Heat Energy
Transition from Solid to Vapor
Cirrus Clouds
Stratus Clouds
Evaporative and Heat Energy Exchanges
Atmosphere
Cumulus Clouds
Precipitation & Evaporation
Snow Cover
Atmospheric Model Layers
Precipitation & Evaporation
Runoff
Stratus Clouds
Sea Ice
Soil Moisture
Land Surface Processes
(Snow Cover, Vegetation, Reflectivity, Topography, and Land Use)
Ocean
Currents, Temperature, and Salinity
Heat & Salinity Exchange
Winds
Realistic Geography
Ocean Bottom Topography
Vertical Overturning
Ocean Model Layers

The way each model discretizes its domain, the processes that it chooses to represent explicitly, those that it needs to parameterize (and the value for those parameters*) form its structural identity.

There is no wrong or right way to do it. No model is *the true model*.

Multi-model ensembles get at this structural uncertainty.

Other uncertainties are:

- -parameter uncertainty
- -initial conditions uncertainty
- -boundary uncertainty (SRES scenarios)

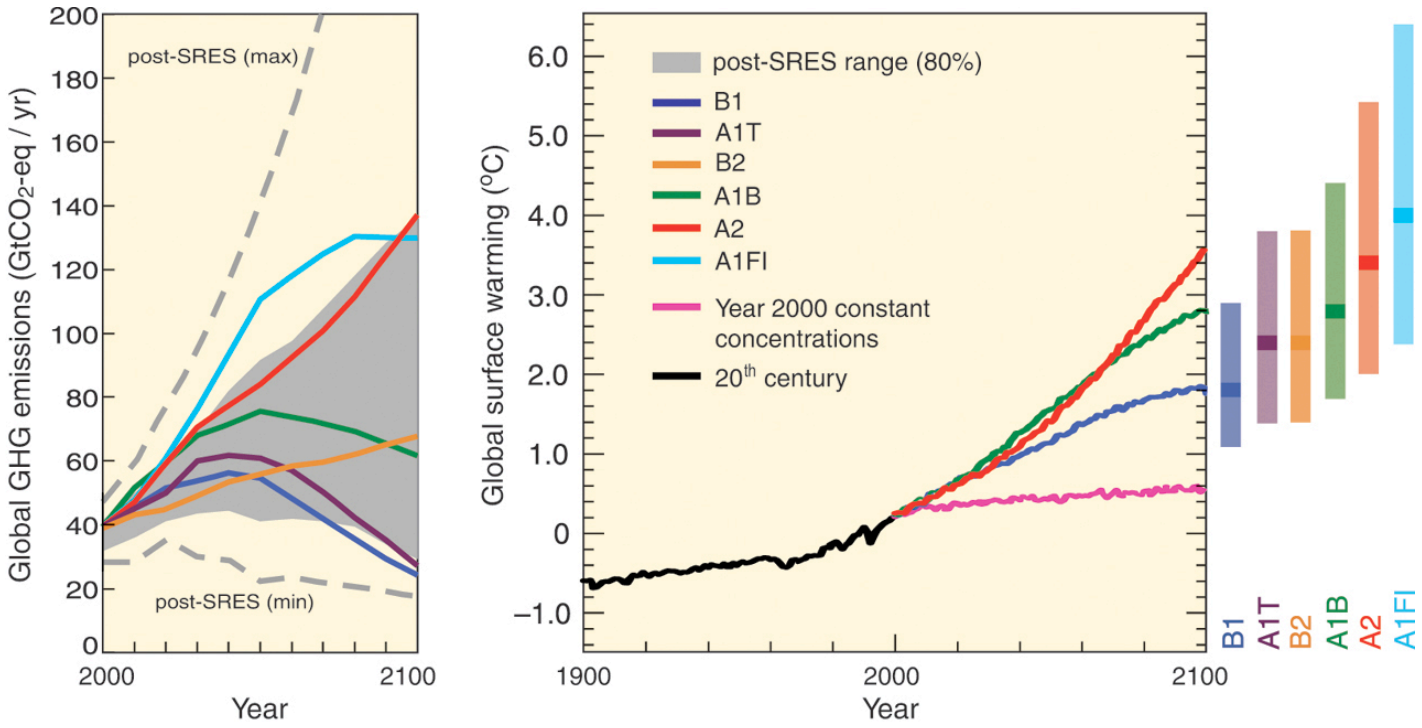
Perturbed physics ensembles deal with parameter uncertainty within a single model.

The most recent multi-model ensemble is made of about 20 state-of-the-art GCMs

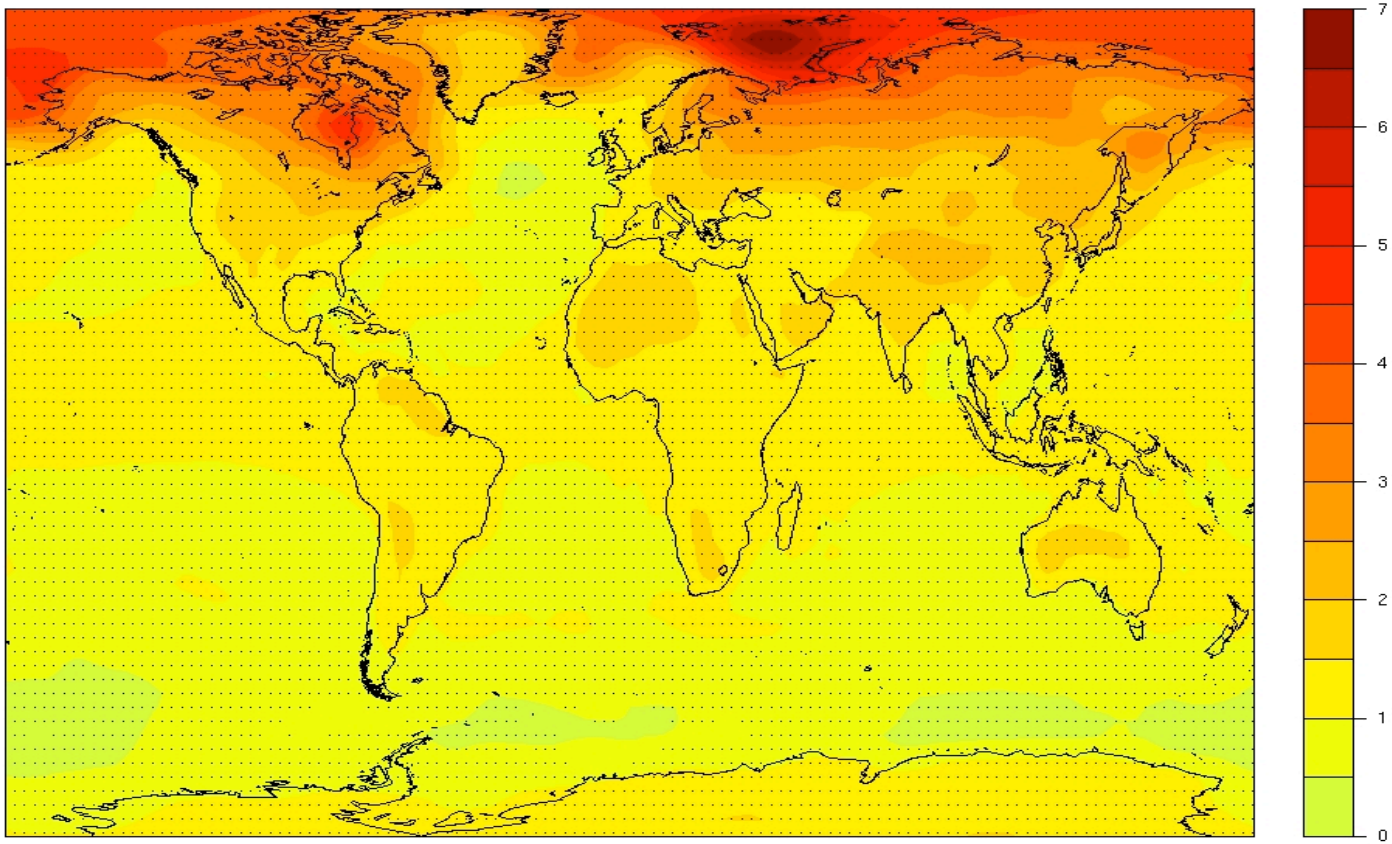
When it comes to future projections, each model gives its own projection.

Sometime only in absolute value, sometime in sign too!

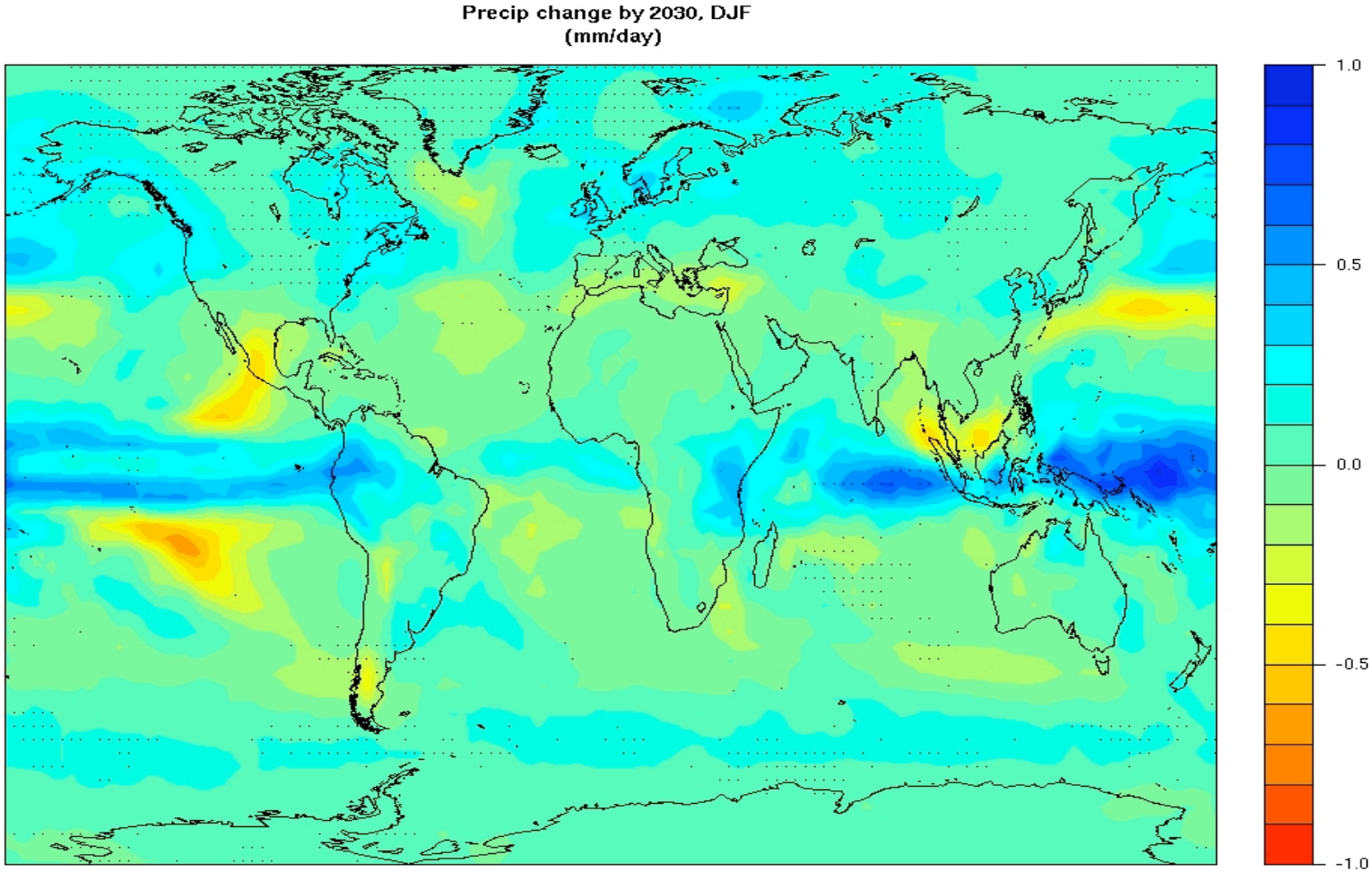
Agreement becomes harder to get the finer the spatial and temporal scale.


post-SRES (max)
post-SRES (min)
Global GHG emissions (GtCO2-eq / yr)
Year
post-SRES range (80%)
B1
A1T
B2
A1B
A2
A1FI
Year 2000 constant concentrations
20th century
Global surface warming (°C)
Year

Temperature change by 2030, DJF
(degrees C)



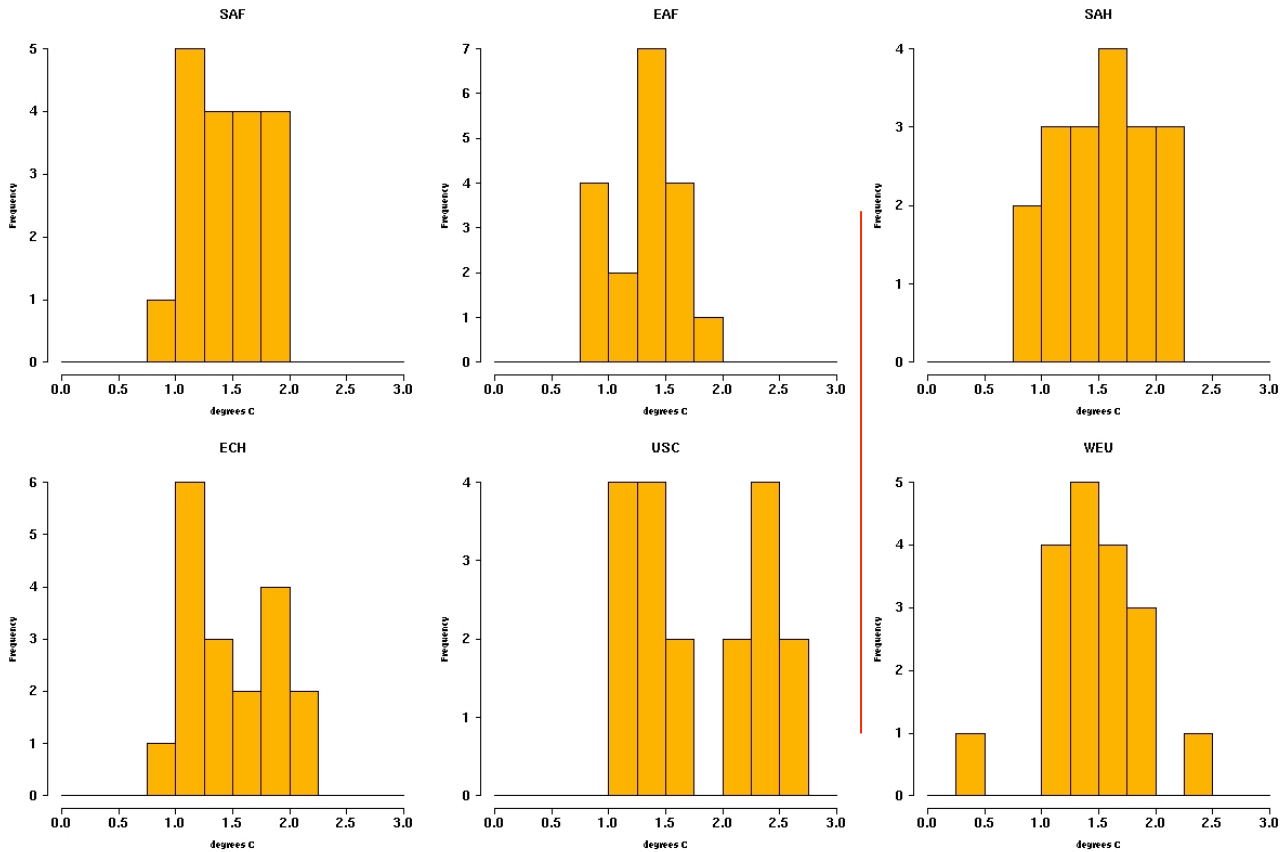
Stippling means 90% of models or more agree in the sign of change



Stippling means 90% or more of models agree in the sign of change

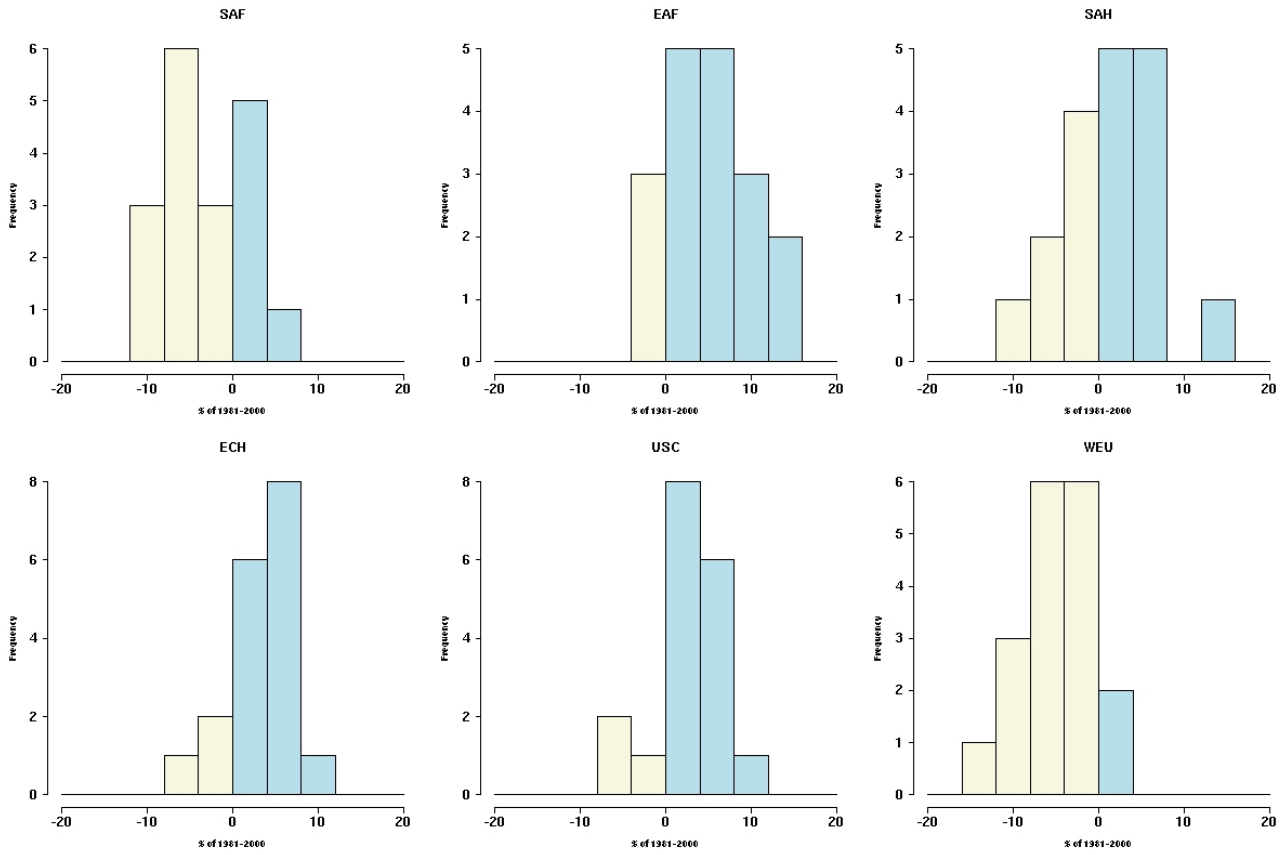
What is behind agreement/disagreement:
A distribution of projected changes

Looking at regional averages of temperature change



What is behind agreement/disagreement:
A distribution of projected changes

Looking at regional averages of % precipitation change



IPCC-AR4 WG1 Chapter 10 and 11, Global and Regional Future Projections:

Most of the projected changes are shown as multi-model means.
Inter-model standard deviation/range is often used as a measure of the uncertainty

What are we looking at?
What are the uncertainties characterized?
What are those left out?

Why means? Why standard deviations?

The problem of synthesizing multiple models projections has a Bayesian flavor:

We start with a prior distribution (of models) which determines their initial weight.

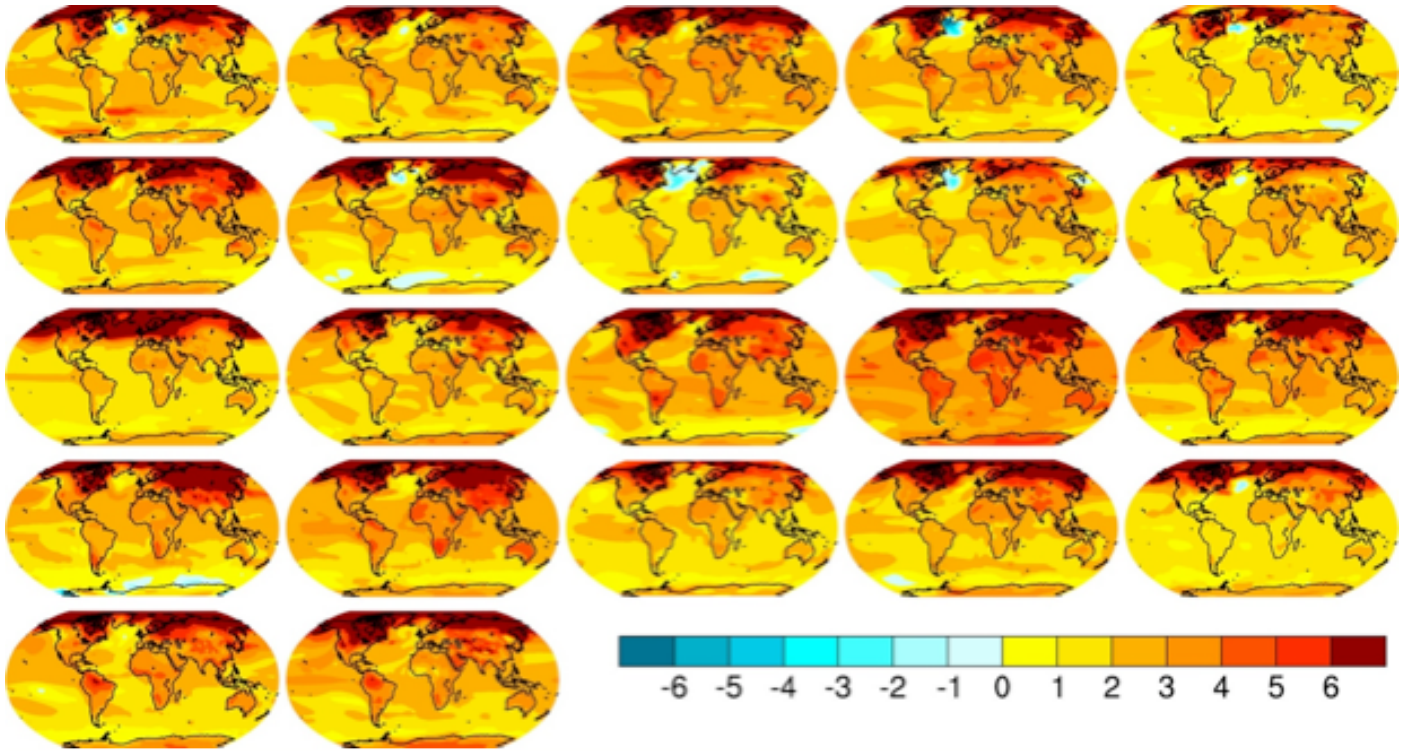
We observe output and we compare the real quantities of interest building a likelihood.

Combining the two we get at a posterior/final distribution.

At this point in time, for most questions, the likelihood is not robustly constraining the final results, and the prior remains a crucial ingredient.

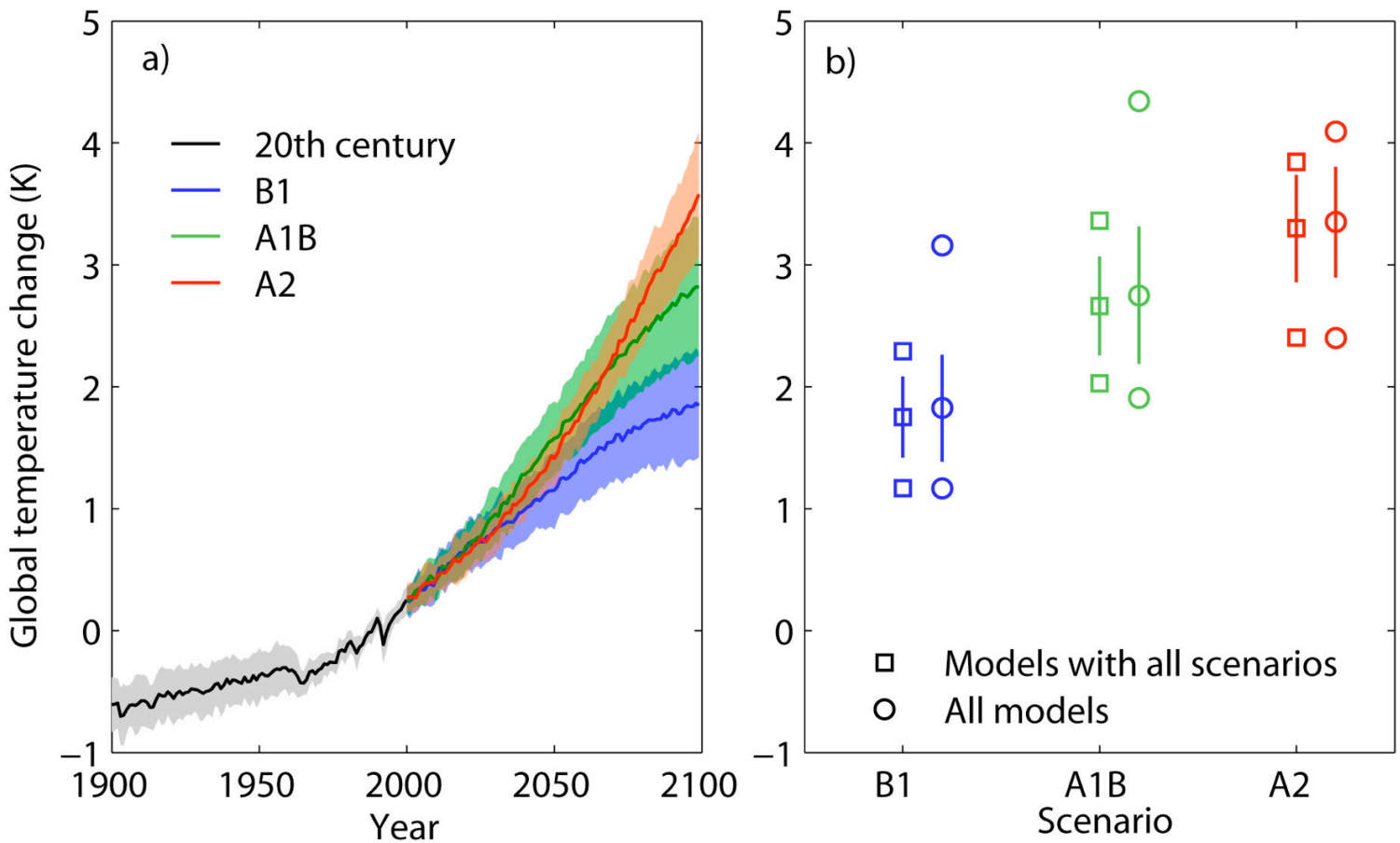
So, what is the prior we are dealing with when we look at multiple models, and what kind of sample are these 20-something models we use all the time?

What kind of sample is the CMIP3 ensemble?

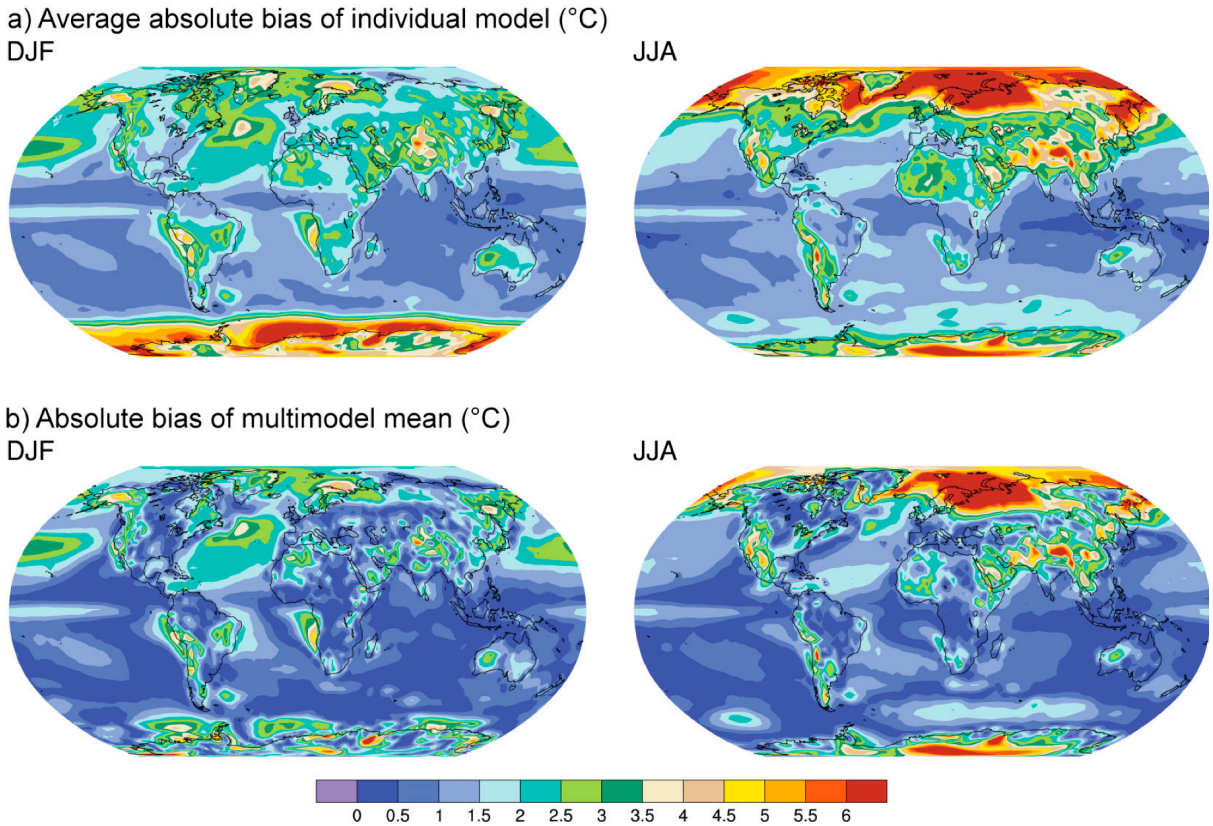


✓A collection of best guesses
✓Likely mutually dependent (models share components)
✓Not random, but not systematic either

A simple example and a simple fix

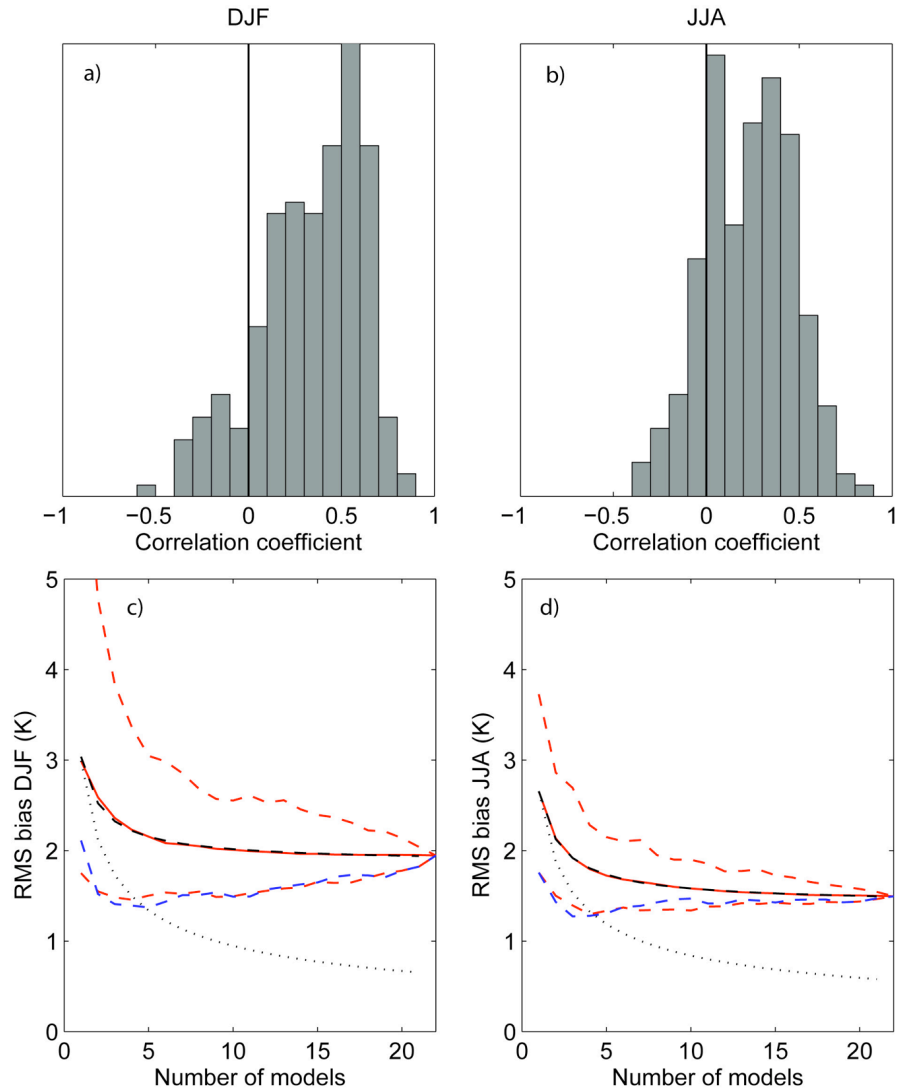


The model mean is better than a single model



But there exist areas where error persists

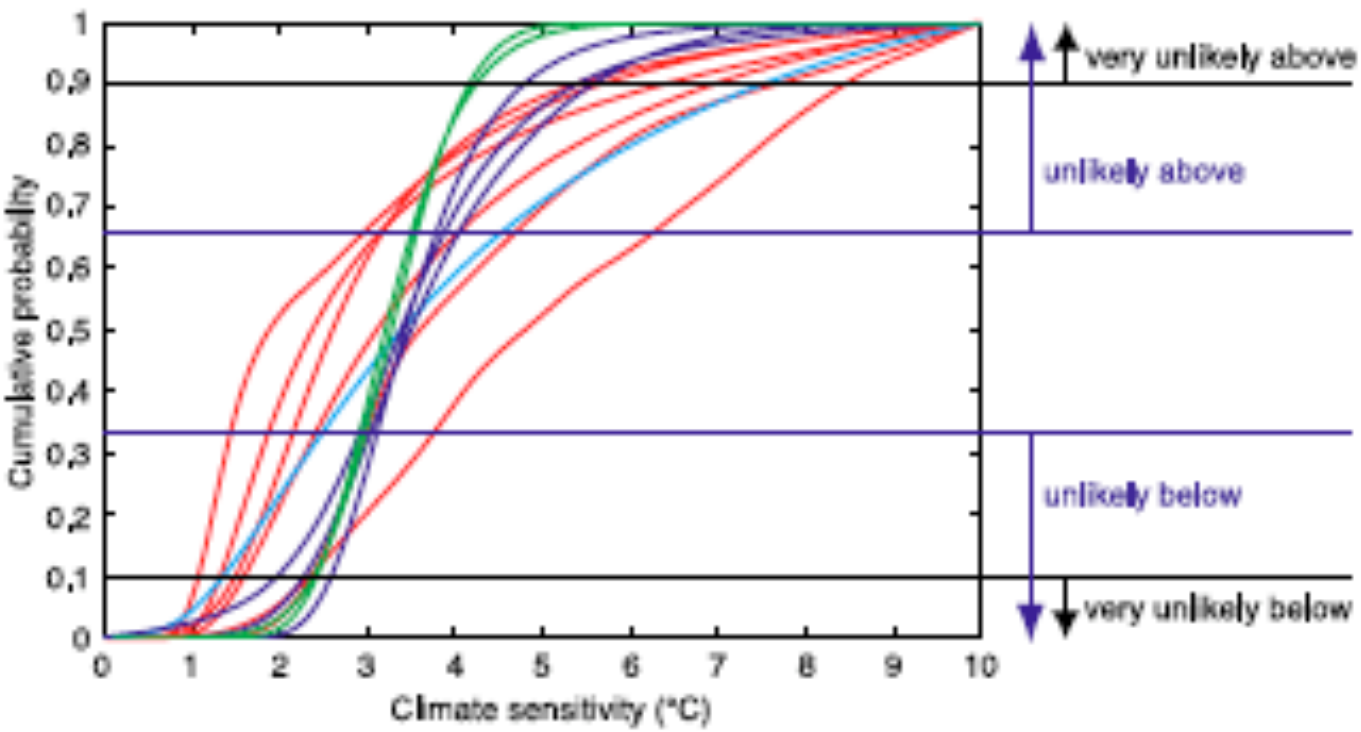
Model errors are not independent



Histograms of correlation coefficients of all possible pairs of model bias (simulated minus observed fields of mean temperature 1980-2000).

Behavior of RMS error with increasing number of models averaged.

They are not designed to sample a wide range of uncertainty



IPCC-AR4 wg1 Ch10

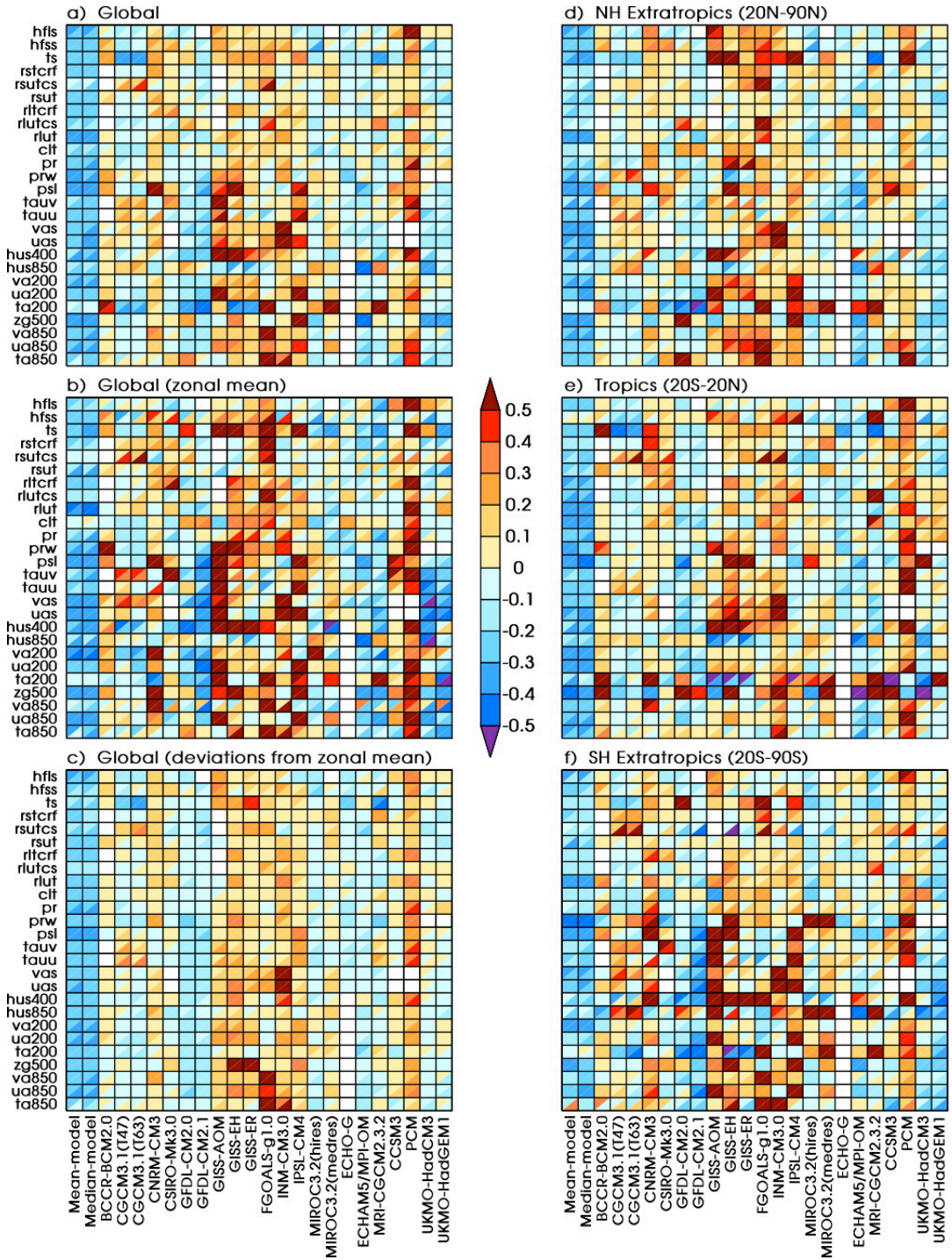
Climate sensitivity of CMIP3 models is green curve

Ensemble means and ensemble ranges are easy
to interpret, but they are not justified by the nature of the sample

The models have systematic and common errors

Likely, the uncertainty is larger than
what is represented by the ensemble of best guesses

Different models have different strengths and weaknesses

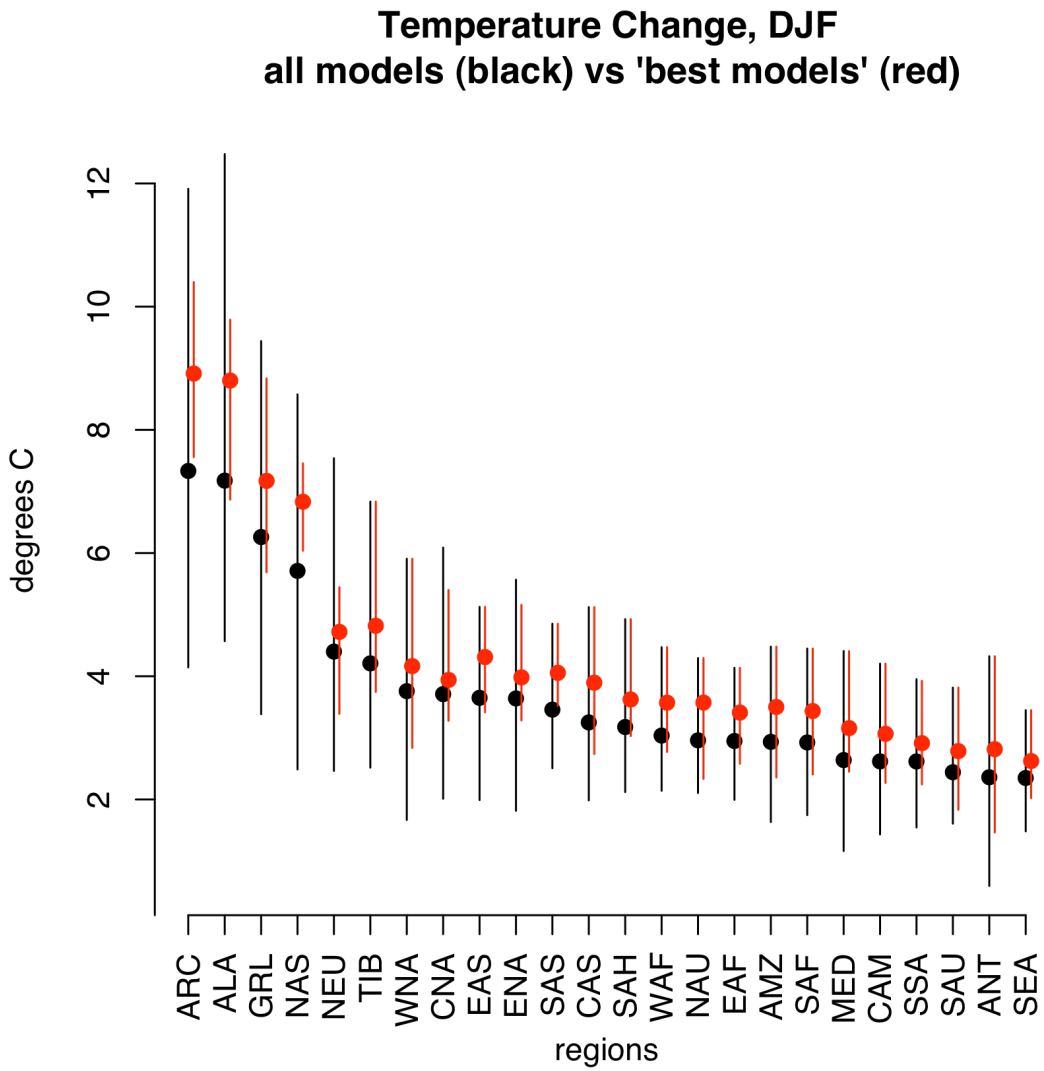


Metrics of performance for different models (columns), across different diagnostics (rows).
Blue is good, Red is bad.
First two columns are ensemble mean and median.

Gleckler et al., (2008) JGR-Atmos

Not all models are created equal

Should we use model performance in replicating observed climate to "weigh" a model more or less?


Temperature Change, DJF
all models (black) vs 'best models' (red)
degrees C
2
4
6
8
10
12
ARC
ALA
GRL
NAS
NEU
TIB
WNA
CNA
EAS
ENA
SAS
CAS
SAH
WAF
NAU
EAF
AMZ
SAF
MED
CAM
SSA
SAU
ANT
SEA
regions

Why statistical modeling of climate model output?

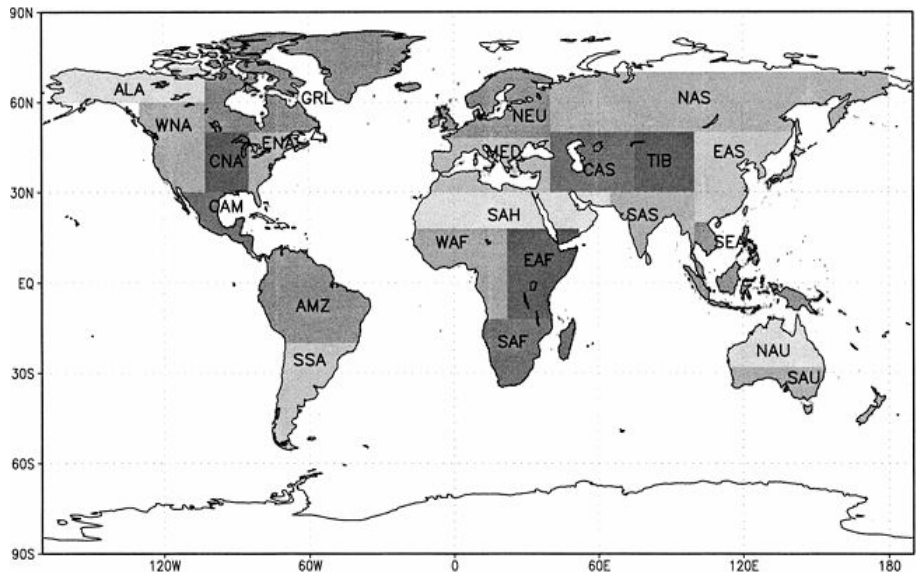
Simple means and standard deviations are not supported by the nature of the data.

Through statistical modeling of GCM data and observations we can combine information in ways that bring validation and performance to bear on the final optimal estimates of what we are after:

Usually, estimates of temperature or precipitation change for specific regions/seasons/time periods. Usually average quantities, since by definition combining models produces a "smoothed" version of the quantity we are after.

An example of how we combine GCM output/observations to produce probabilistic projections of Temperature and Precipitation change

The analysis is performed over regional and decadal averages, for a given season (e.g., DJF) and a specified emission scenario.



Data

Observed decadal means of temperature and precipitation

$$O_t = \begin{pmatrix} O_t^T \\ O_t^P \end{pmatrix}$$

$$t = 1950, 1960, \ldots 2000$$

Modeled decadal means of temperature and precipitation

$$X_{jt} = \begin{pmatrix} X_{jt}^T \\ X_{jt}^P \end{pmatrix}$$

$$t = 1950, 1960, \ldots, 2000, 2010, \ldots, 2100$$

$$j = 1, \ldots, M$$

Assumptions behind the statistical model, in words

•the true climate signal time series (both temperature and precipitation) is a piecewise linear trend with an elbow at 2000, to account for the possibility that future trends will be different from current trends;

•superimposed to the piecewise linear trends is a bivariate gaussian noise with a full covariance matrix, which introduces correlation between temperature and precipitation;

•observed decadal averages provide a good estimate of the current series and their correlation, and of their uncertainty;

• GCMs may have systematic additive bias, assumed constant along the length of the simulation;

•after the bias in each GCM simulation is identified there remains variability around the true climate signal that is model specific.

Likelihood Model

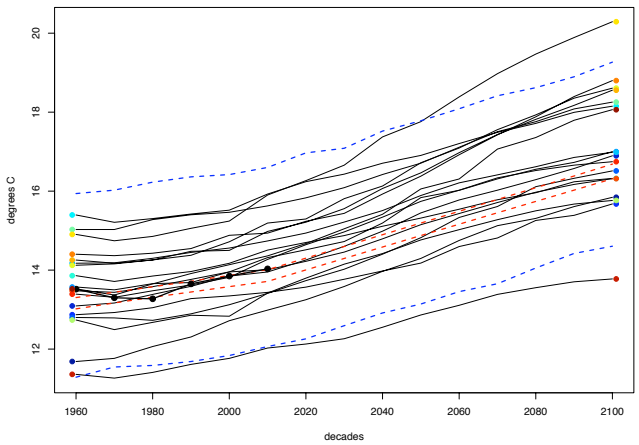
$$O_t^T \sim N\left[\mu_t^T;\eta^T\right]$$
$$O_t^P \sim N\left[\mu_t^P + \beta_{xo}(O_t^T - \mu_t^T);\eta^P\right]$$
$$X_{jt}^T \sim N\left[\mu_t^T + d_j^T;\phi_j^T\right]$$
$$X_{jt}^P \sim N\left[\mu_t^P + \beta_{xj}(X_{jt}^T - \mu_t^T - d_j^T) + d_j^P;\phi_j^P\right]$$

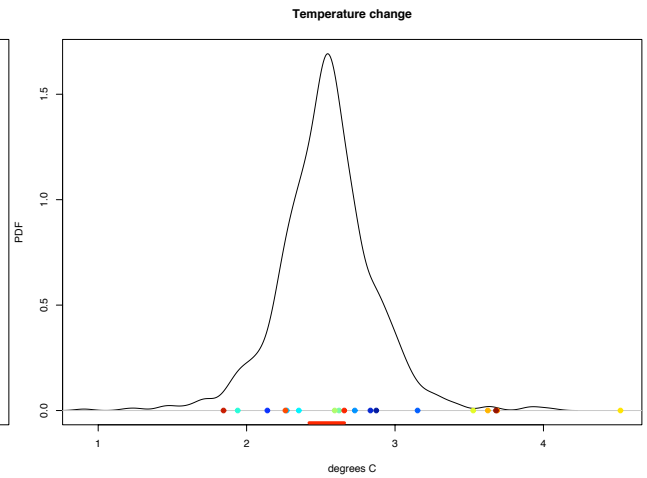
$$\begin{pmatrix}\mu_t^T\\ \mu_t^P\end{pmatrix} = \begin{pmatrix}\alpha^T + \beta^T t + \gamma^T (t - T_0)\chi_t\\ \alpha^P + \beta^P t + \gamma^P (t - T_0)\chi_t\end{pmatrix}$$

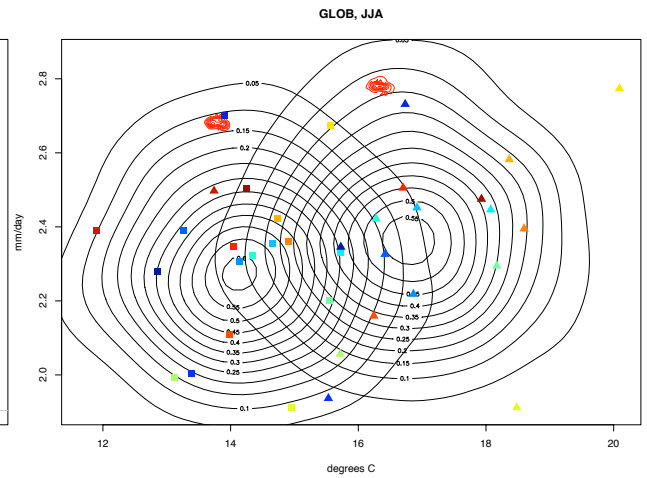
All unknown quantities

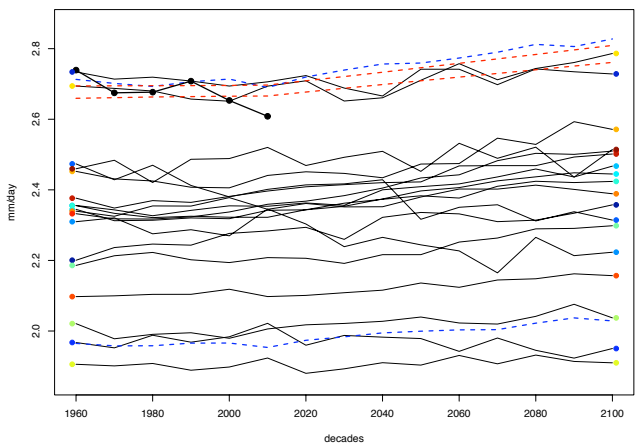
$$\alpha^T, \alpha^P, \beta^T, \beta^P, \gamma^T, \gamma^P, d_j^T, d_j^P, \phi_j^T, \phi_j^P \beta_{xo}, \beta_{xj}$$

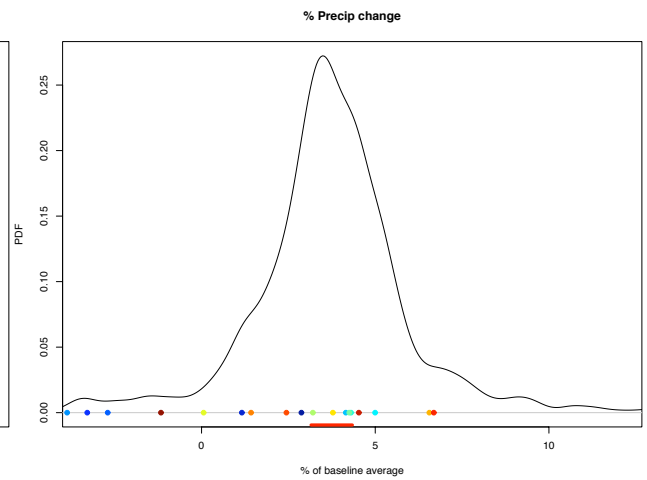
have uninformative prior distributions, and their joint posterior is estimated through MCMC.

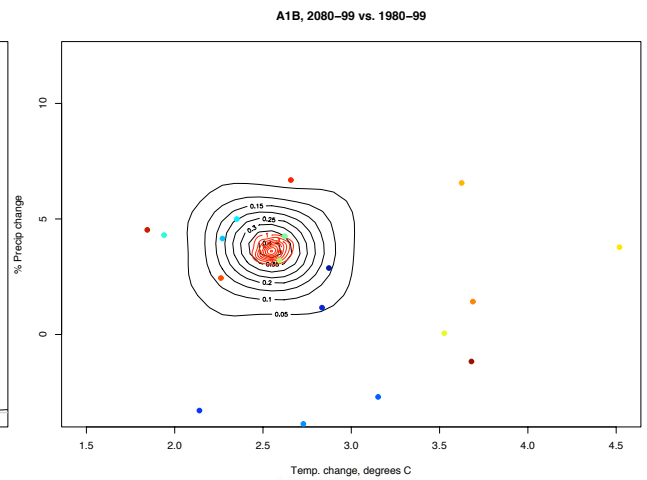

degrees C
decades


Temperature change
PDF
degrees C


GLOB, JJA
mm/day
degrees C


mm/day
decades


% Precip change
PDF
% of baseline average


A1B, 2080–99 vs. 1980–99
% Precip change
Temp. change, degrees C

This analysis

Can quantify systematic biases, models' reliability.

Can provide a probabilistic "best guess" for the climate signal as a weighted mean between observations and models' central tendency.

Can provide a predictive distribution for a "new model",

It is an informed synthesis of the data
You can accept it as such if you agree with its assumptions (likelihood of the data, priors on the unknown parameters)

What does it tell us about the real world?

Conclusions

Combining multi-model ensembles can help quantify uncertainties in future climate projections, exploring and comparing models' structural characteristics.

The non-systematic nature of the sampling, the in-breeding among the population of models the complexities of choosing and drawing inference from diagnostics, the impossibility of verification present unique challenges in the formulation of a rigorous statistical approach at quantifying these uncertainties.

Conclusions

Things are bound to become even more interesting, with the introduction of more complex GCMs and the increasing heterogeneity of these ensembles; ever increasing demands on computing resources and the consequent trade-offs of resolution vs scenarios vs ensemble size.

The way forward: coordinated experiments, multi-model plus single-model ensembles, process-based evaluation translated into metrics, representation of model dependencies, quantification of common biases.

All along, transparency and two way communication between scientists and users, stake holders, policy makers.