Statistics for Climate Model Experiments
Part 1: The Nature of the Beasts

Claudia Tebaldi

Climate Central

currently visiting

the Institute for Mathematics Applied to the Geosciences

NCAR

Boulder

ctebaldi@climatecentral.org

The international Graduate Summer School on Statistics and Climate Modeling

Framework

There is an inescapable dose of uncertainty in climate model projections, and that is where statisticians like to (should!) enter the game.

There are complex issues characterizing model output that hamper statistical modeling.

Nonetheless, formal statistical handling of model output is a step forward with respect to heuristic approaches in order to quantify uncertainties, distinguish their source and highlight promising and necessary paths of further research (not only for statisticians, but as a feedback to climate and impact research).

Climate Change Projections and their Uncertainty

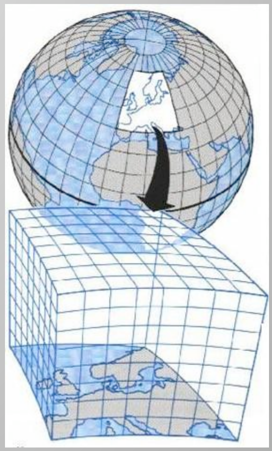
Sources:

Natural variability of climate, e.g., decadal oscillations.

Alternative economic/technological pathways and ensuing rates of greenhouse gases' emission.

Modeling uncertainties, mainly due to the need of approximating sub-scale processes that are not explicitly resolved.

Ensembles of models let us address these uncertainties.



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.

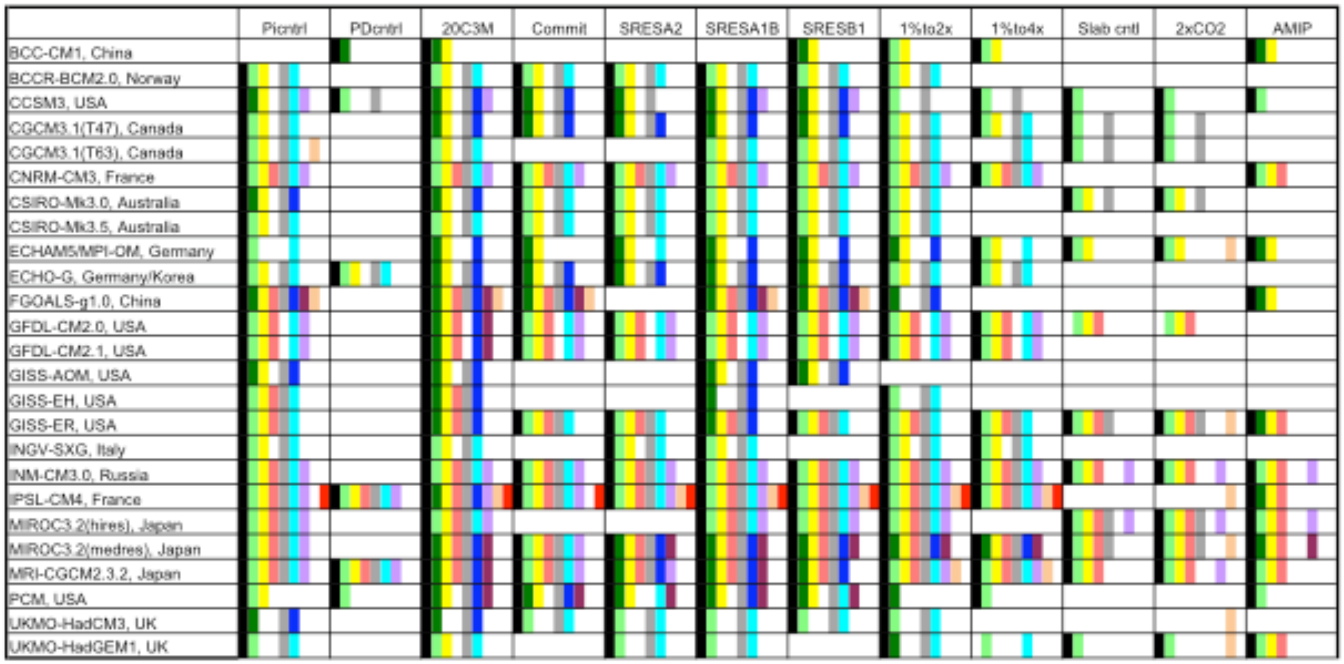
With the consent of participating climate modelling groups, the WGCM has declared the CMIP3 multi-model dataset open and free for non-commercial purposes. After registering and agreeing to the "terms of use," anyone can now obtain model output via the ESG data portal, ftp, or the OPeNDAP server.

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.

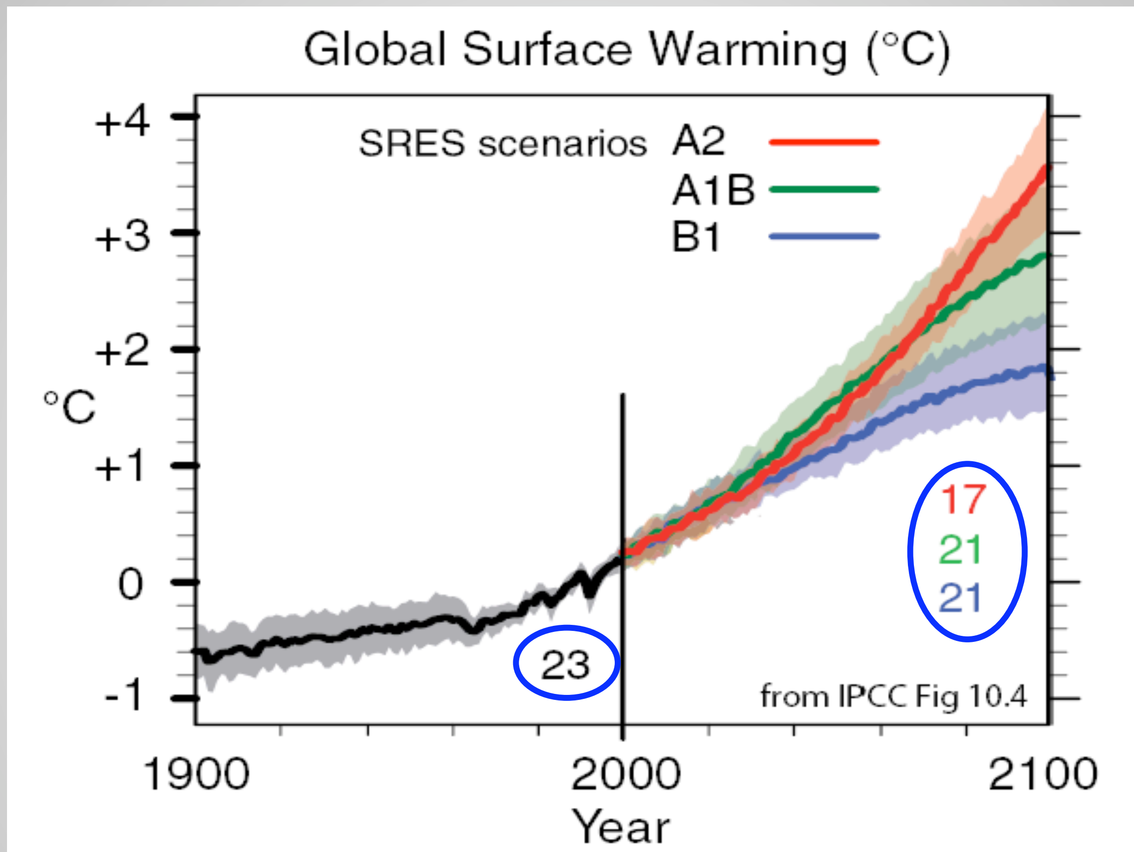
Data Availability Summary (as of 16 July 2007)

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

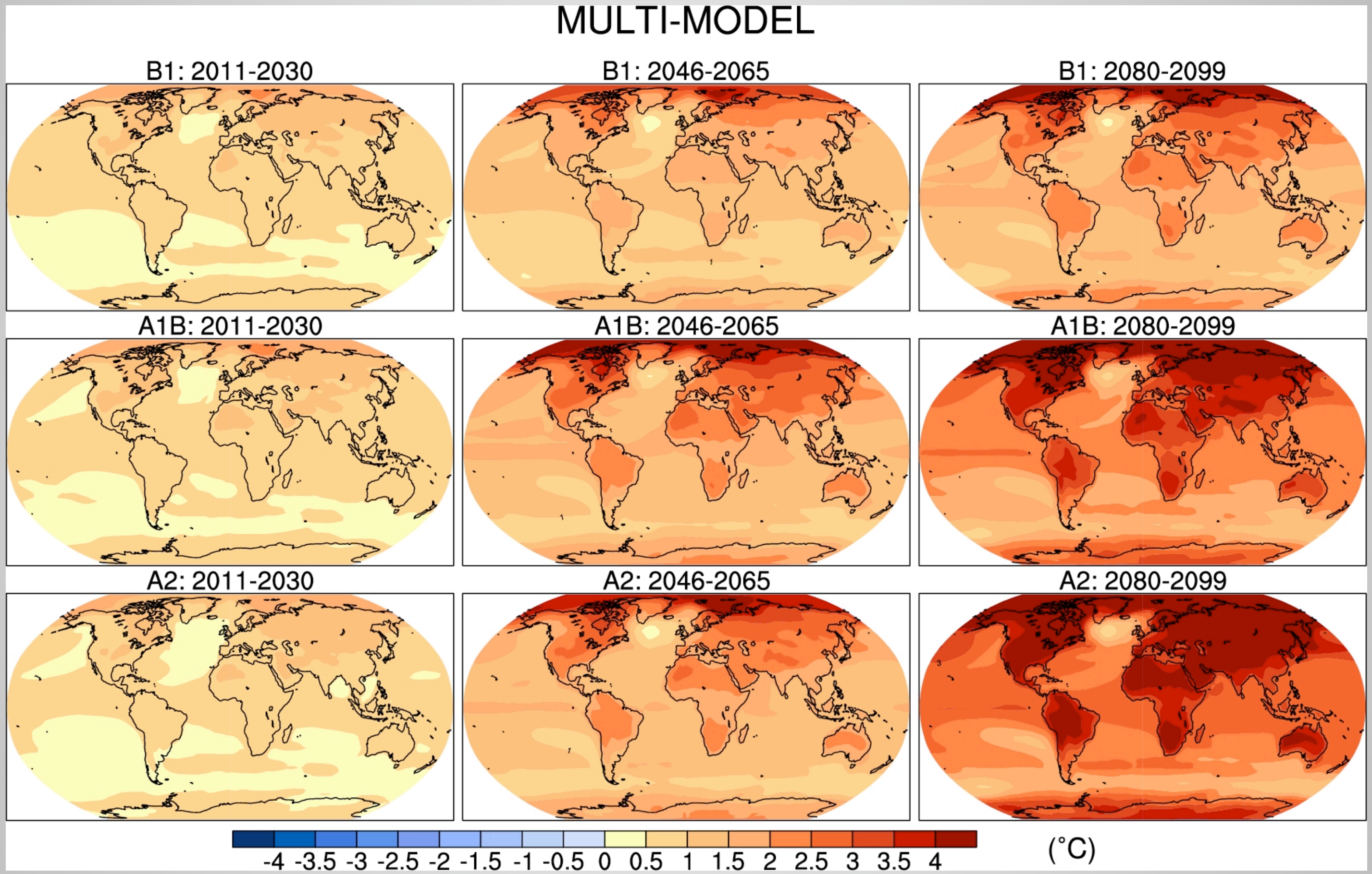


Model uncertainties



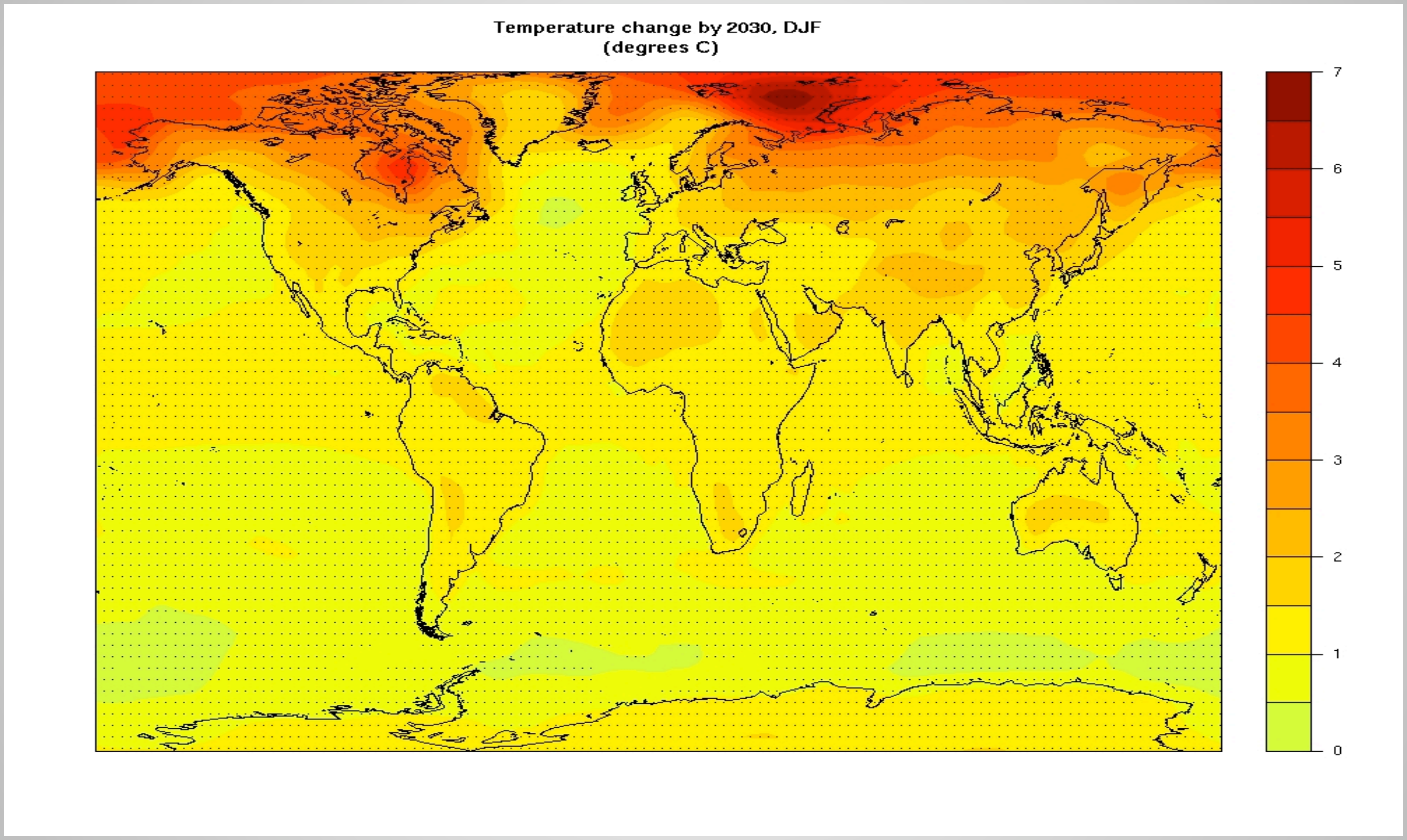
Times series of global average temperature change under different greenhouse gas emission scenarios (different colors). Thick line is ensemble mean, shading is ONE standard deviation of ensemble.

Scenario uncertainties



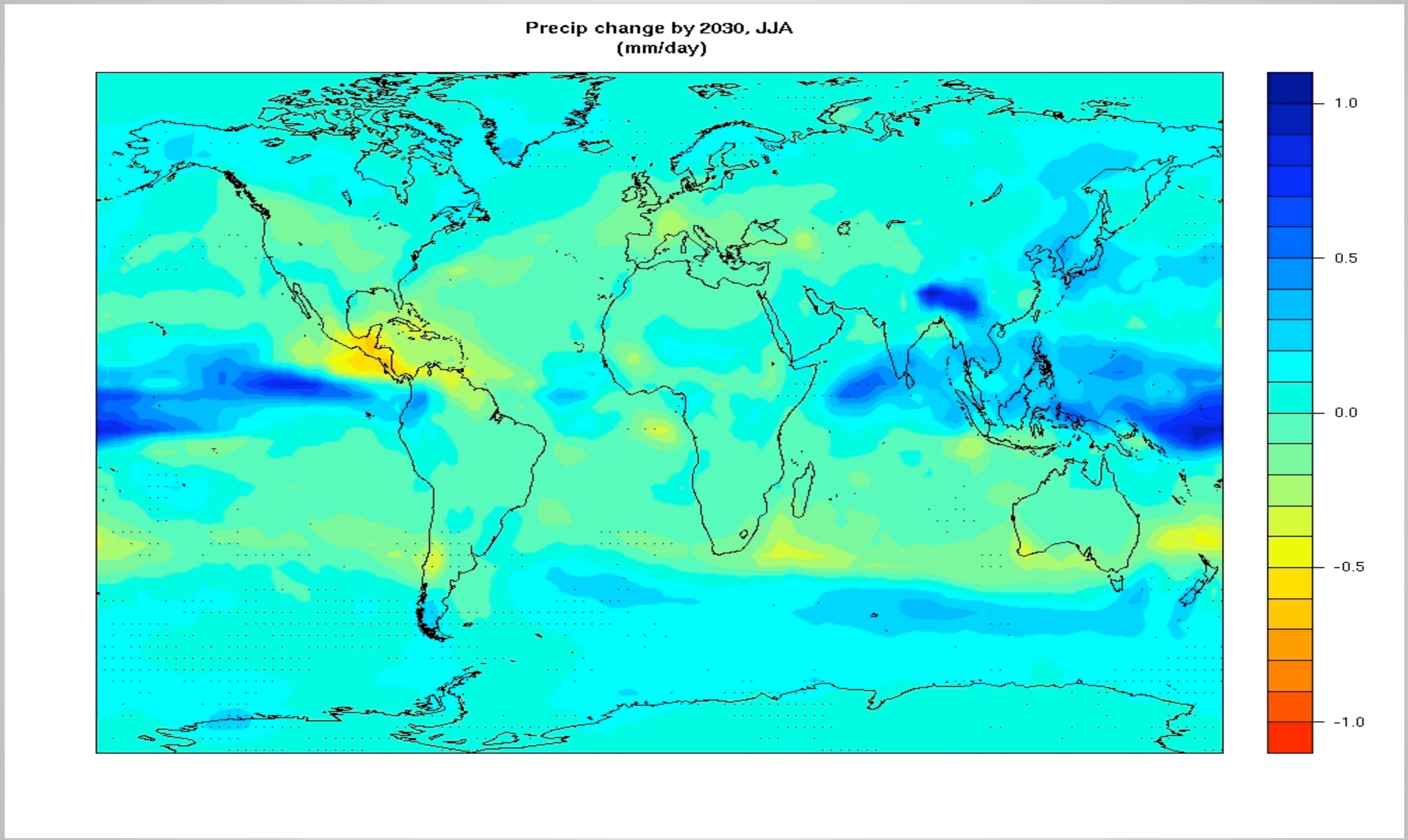
Scenarios don't matter that much until about 2050.

But what is behind ensemble averages?

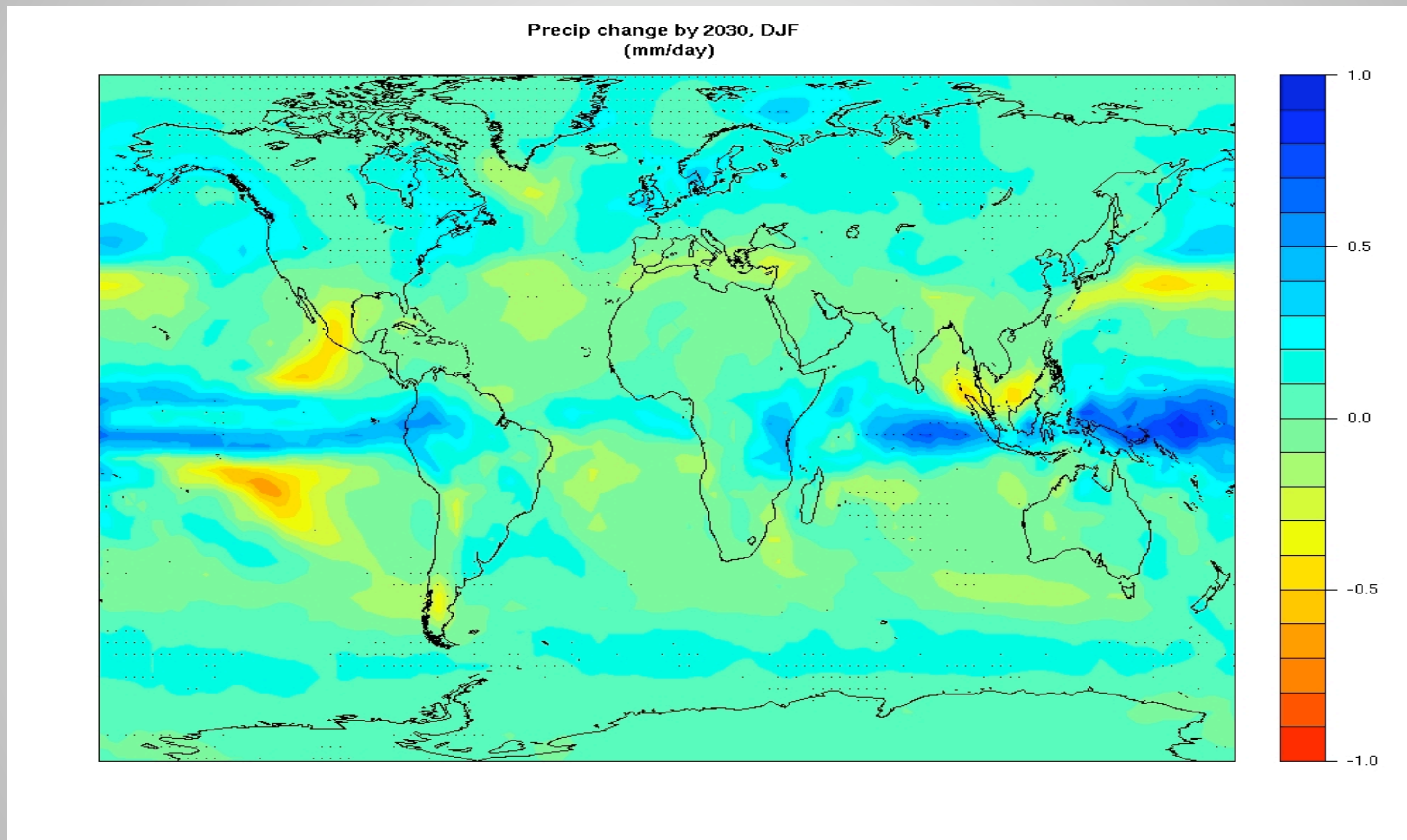


Stippling means 90% of models agree in sign

For precipitation projections agreement is much harder to find

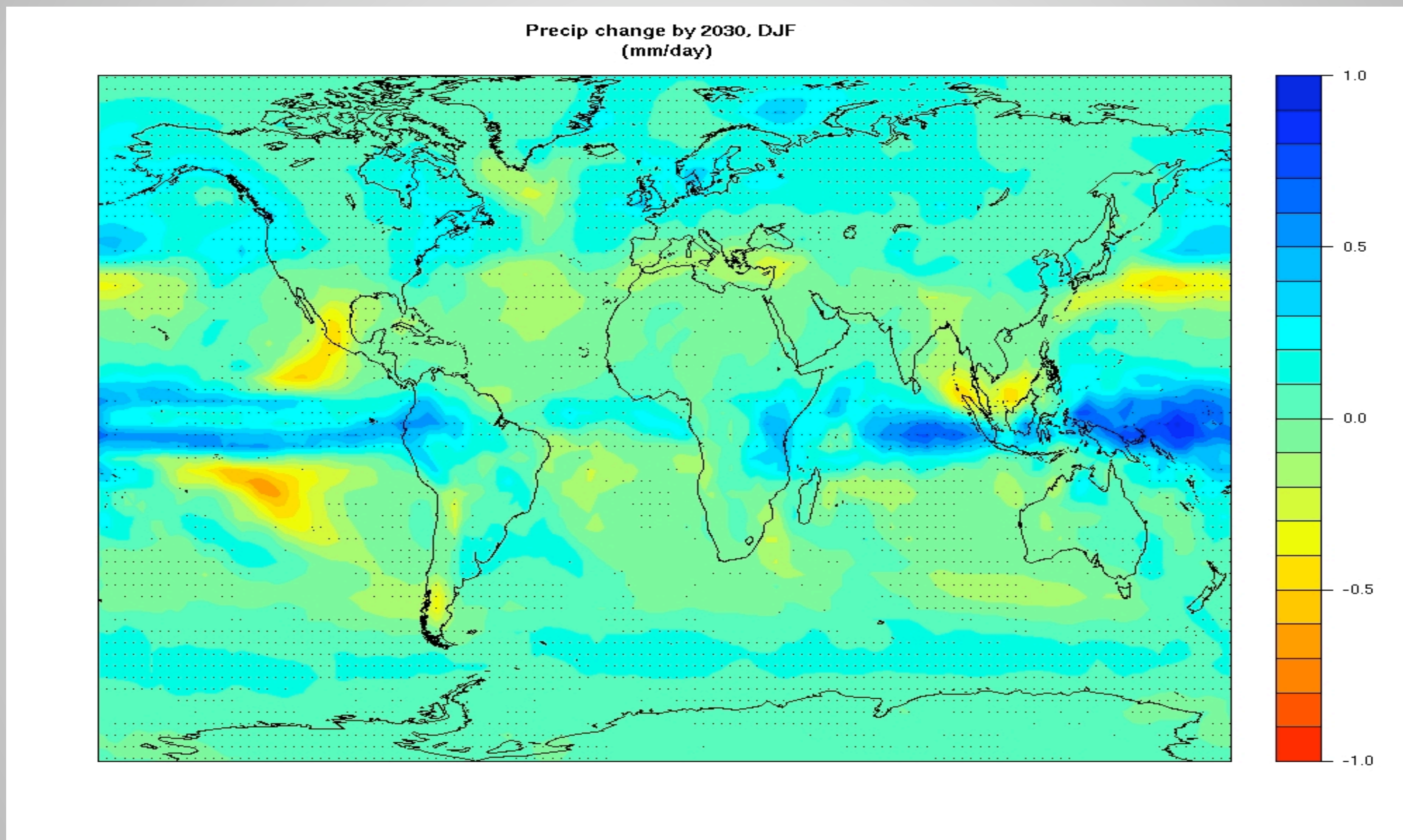


Stippling means 90% of models agree in sign



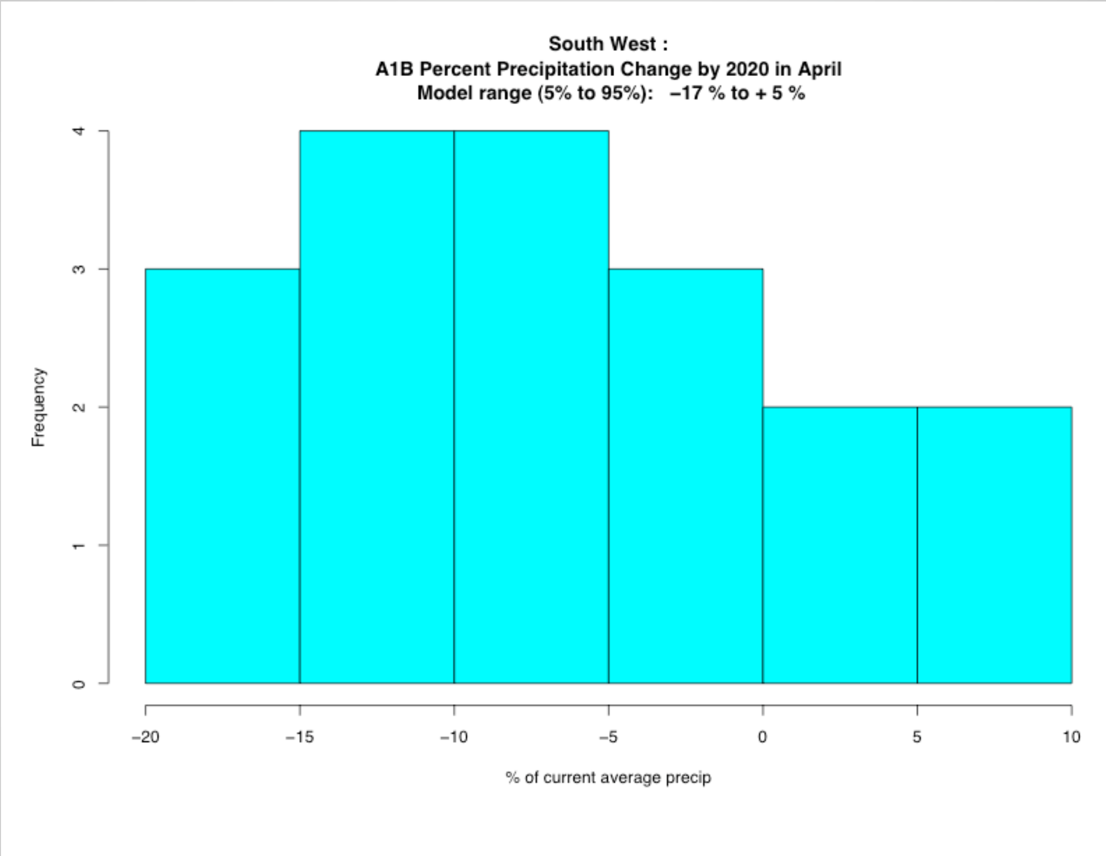
Stippling means 90% of models agree in sign

Even when lowering the bar…



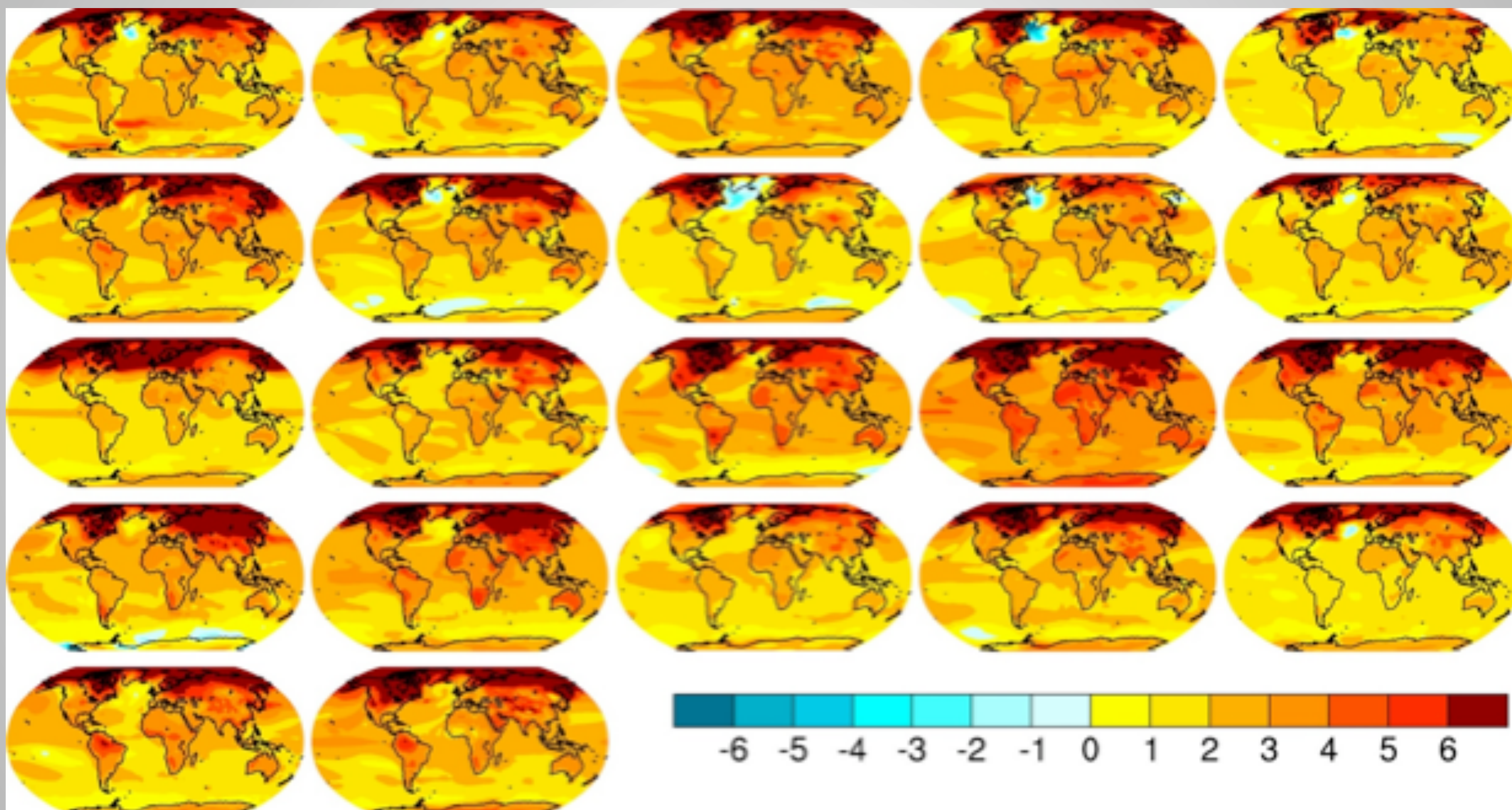
Stippling means 2/3 of models agree in sign

What does this mean for a given region?



Histogram of 18 models' projections of percent precipitation change for the Southwest of the United States

So different models give different projections.
How do we synthesize them?
What kind of sample is a multi-model ensemble?



Before getting into statistical modeling of this sample, let's try to find answers to some pretty general questions:

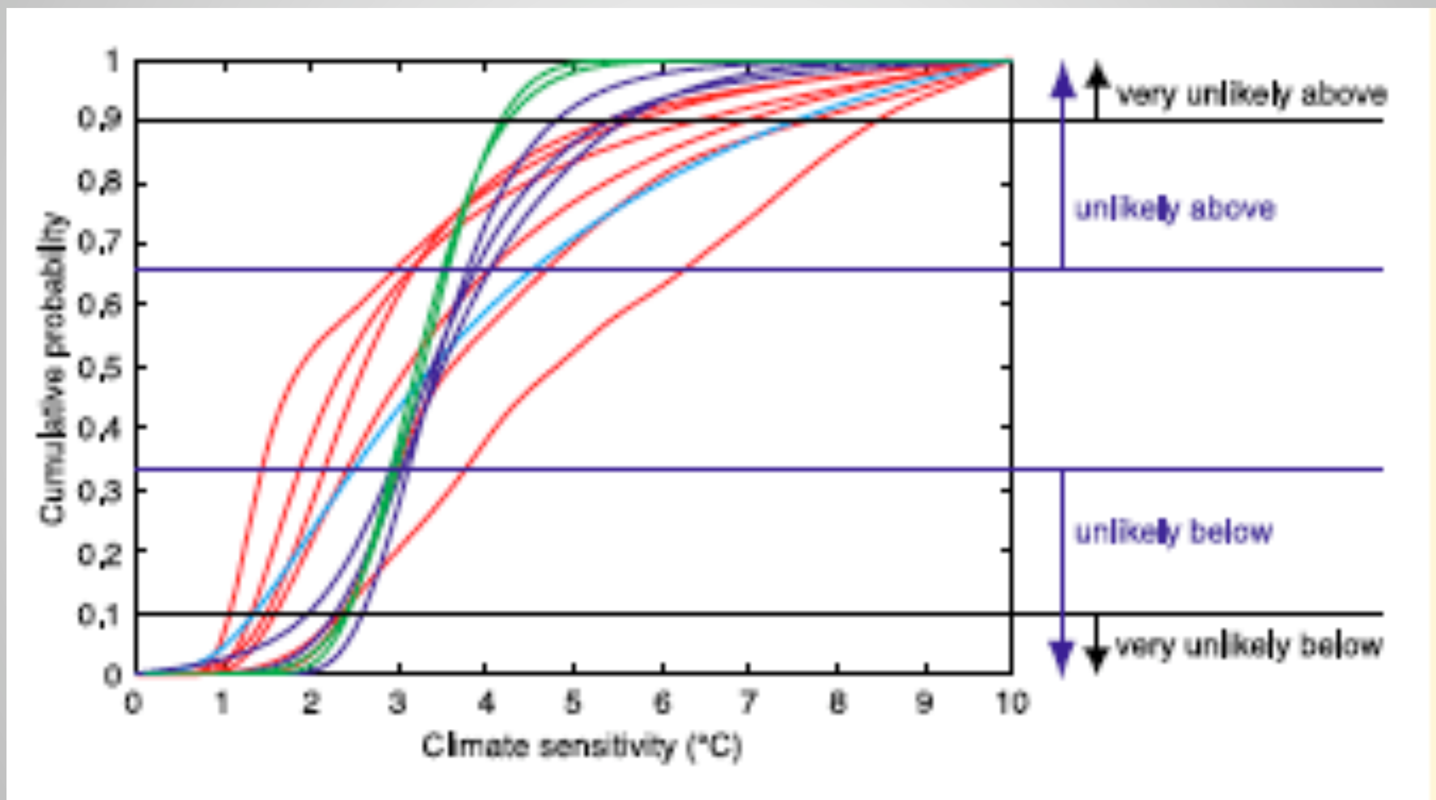
✓What is underneath (or around) an ensemble mean?

•Are we seeing the whole picture? Are there dimensions of the uncertainty we are not exploring through these ensembles?

•Are we justified in calculating simple means and ranges?

•Should we rather bring model performance to bear?

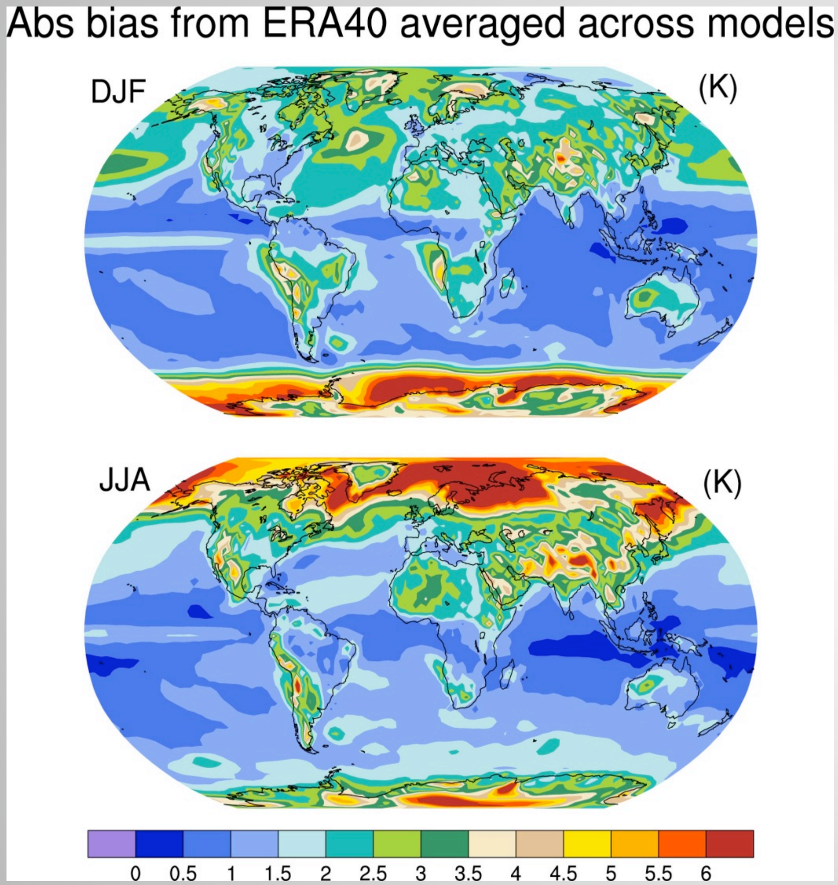
Models do not sample a wide range of uncertainty: they are not a random sample, nor are they a systematic sample

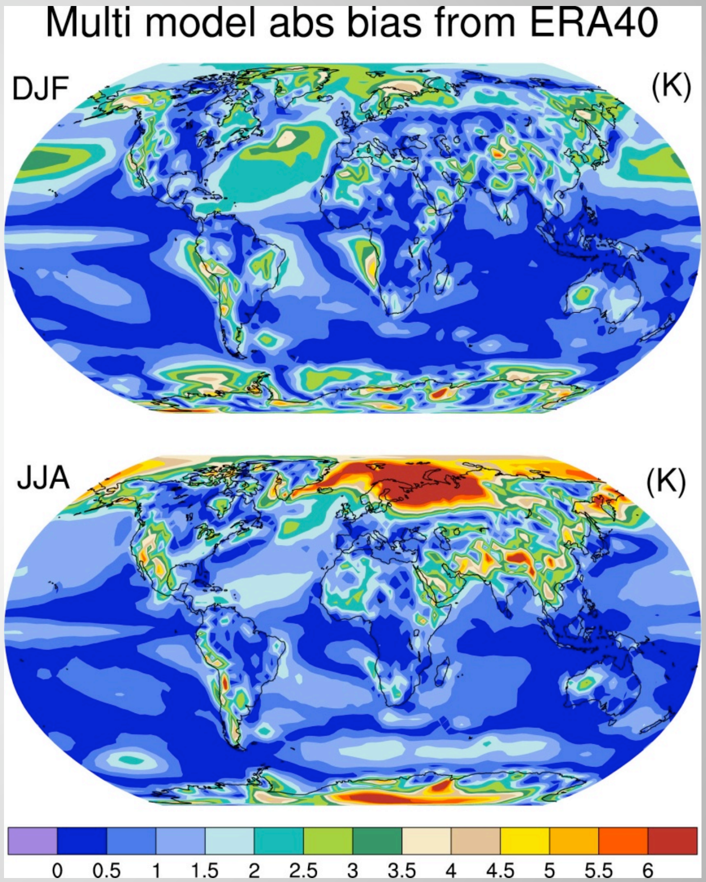


Climate sensitivity of CMIP3 models is green curve

IPCC-AR4 wg1 Ch10

However, errors tend to cancel
i.e., averaging is good

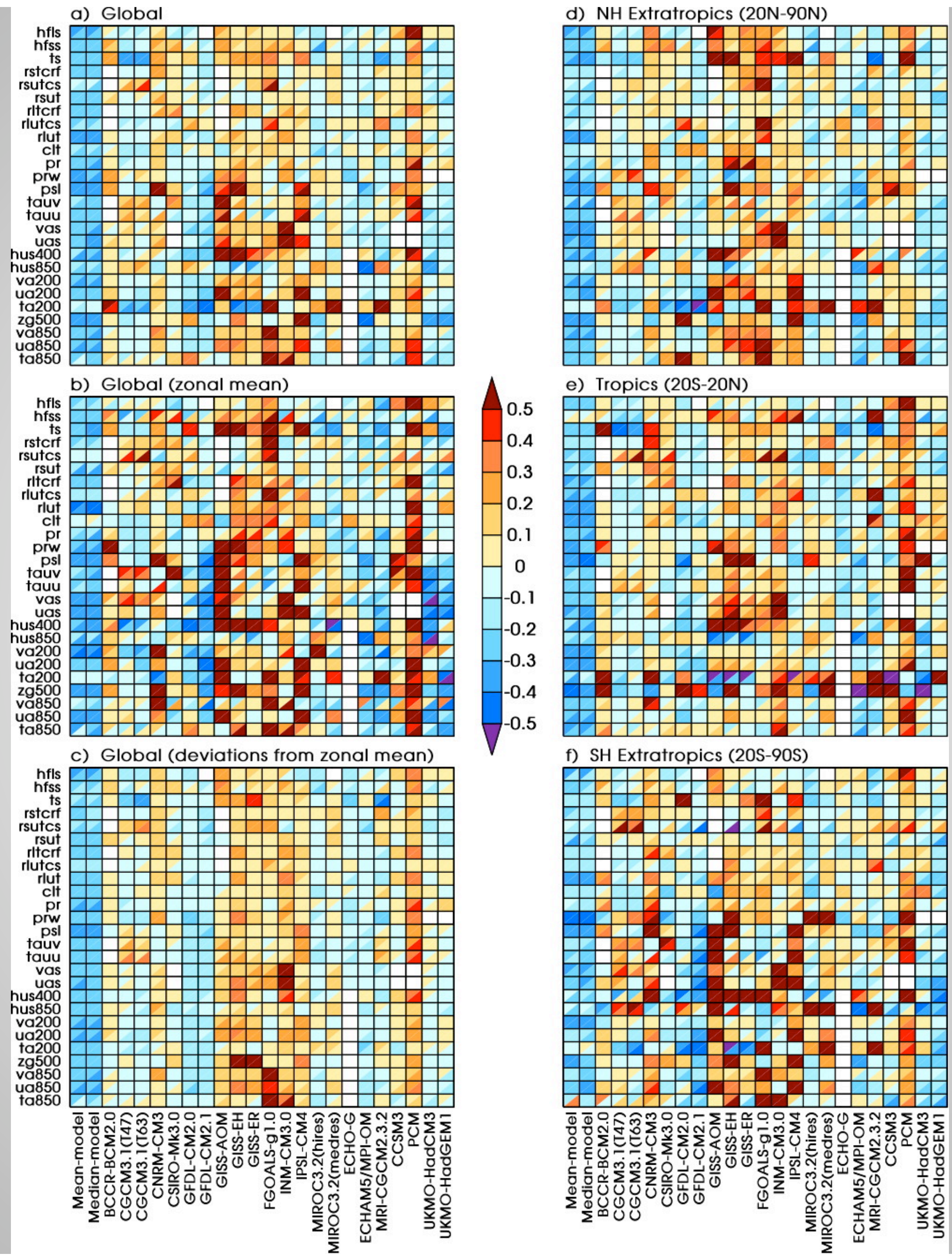




Comparison of average bias across models or bias of the ensemble average.

Dark blue is good. Red is bad.

Different models have different strengths and weaknesses



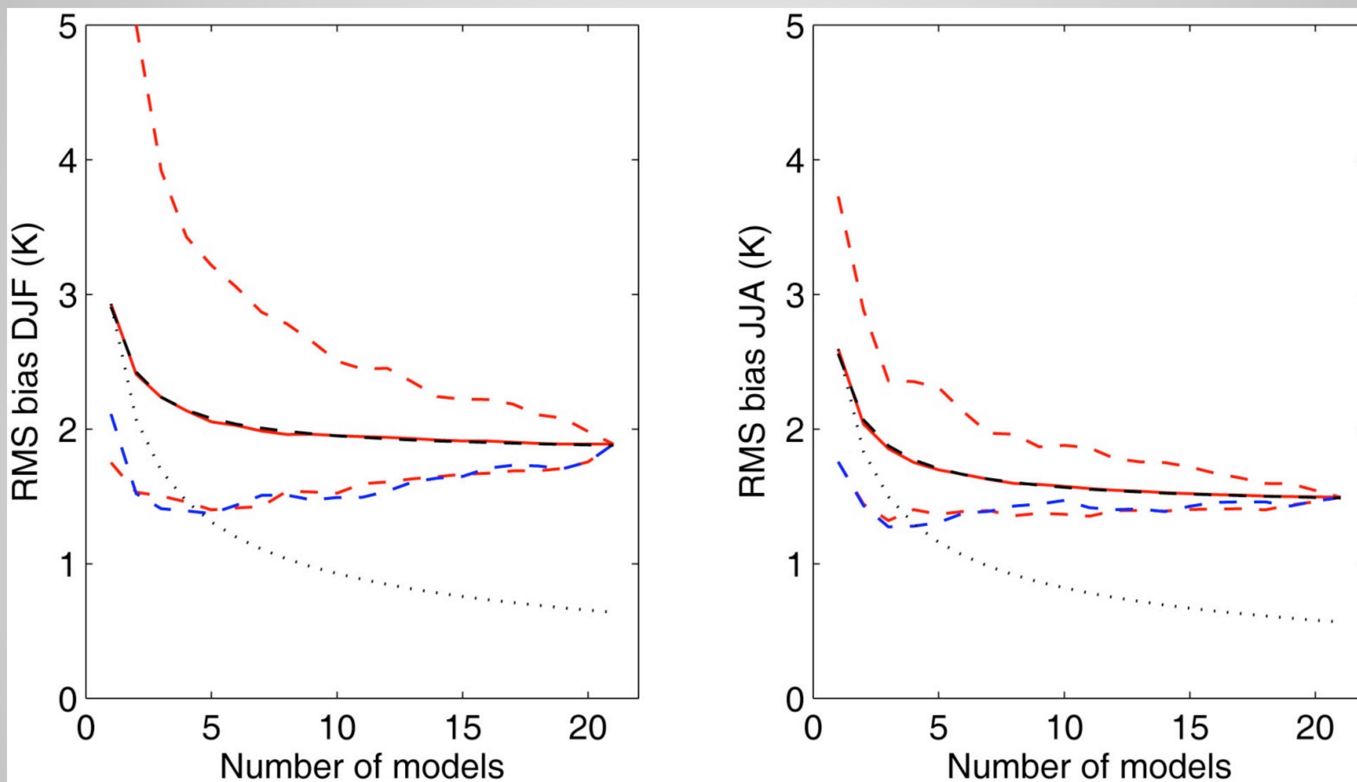
Metrics of performance for different models (columns), across different diagnostics (rows).

Blue is good, Red is bad.

First two columns are ensemble mean and median.

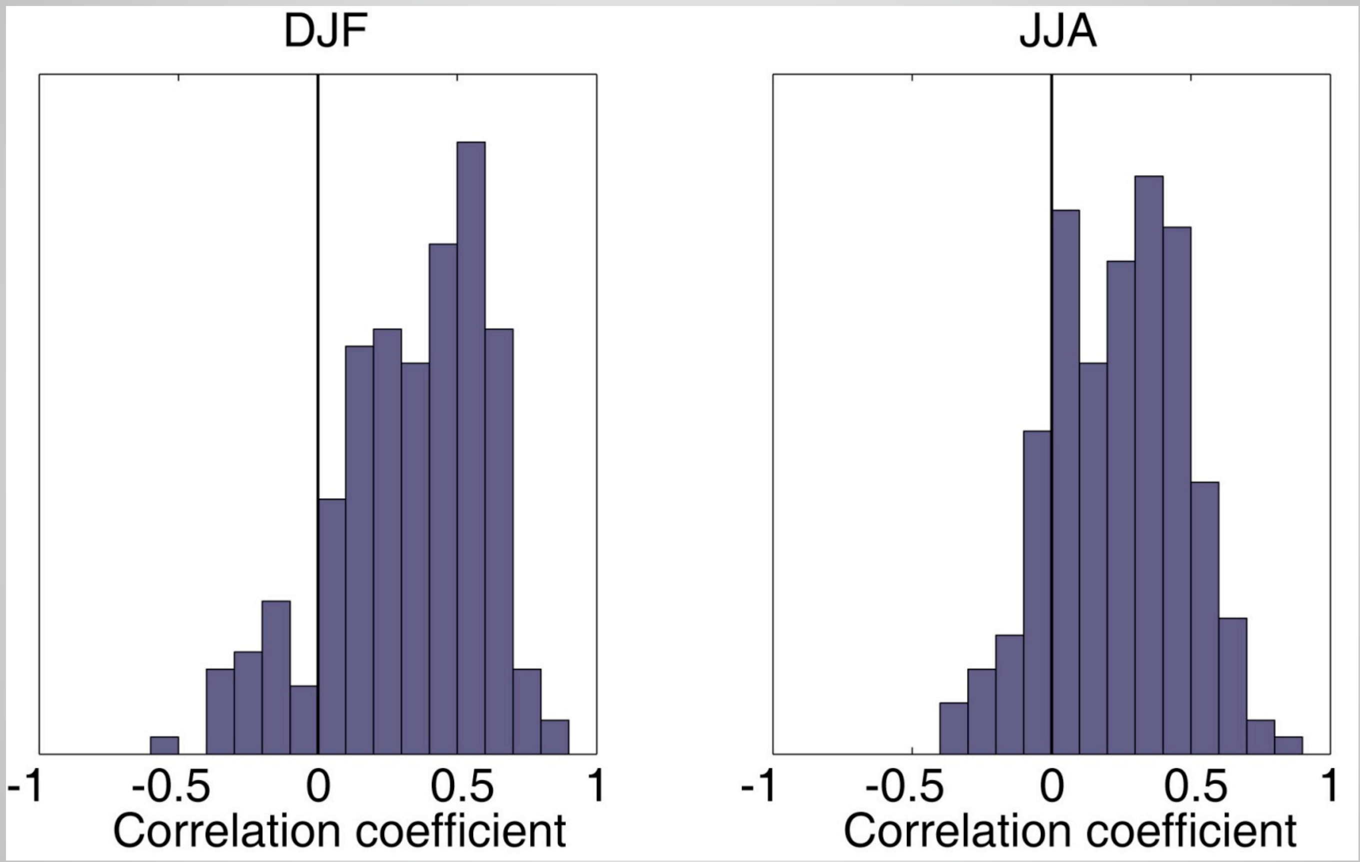
Gleckler et al., (2008) JGR-Atmos

Averaging, though, has decreasing marginal value



Error goes down as SQRT(N) with Number of Models, but does not go to zero

Fields of errors from different models are correlated

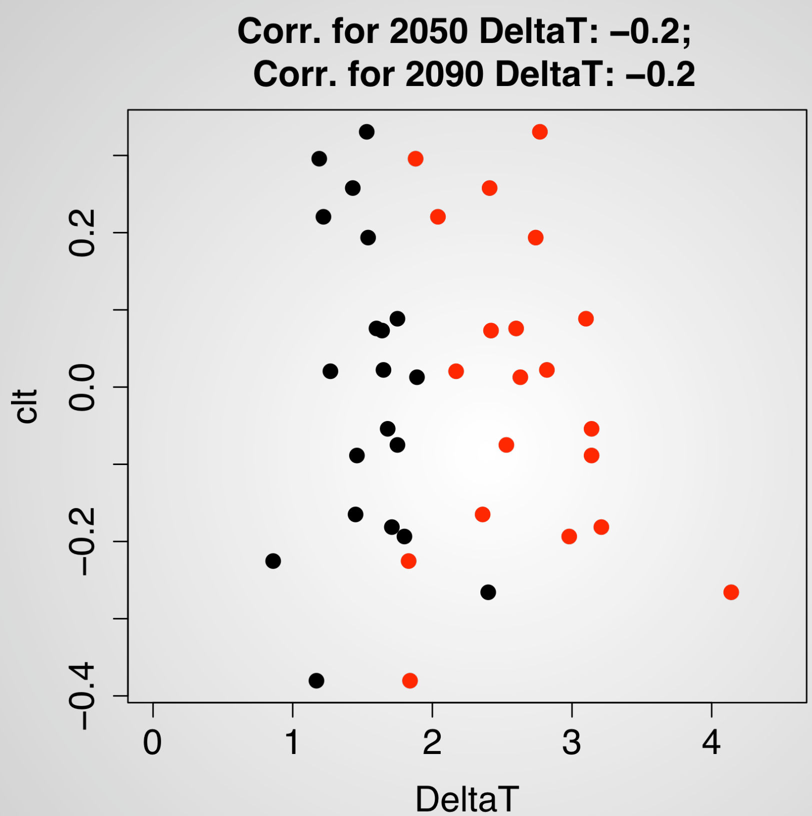


Correlation coefficients between pairs of bias fields across all possible pairs of models

So, it seems as if not all models are created equal. Combining them rather than taking a single one seems to be a good idea.

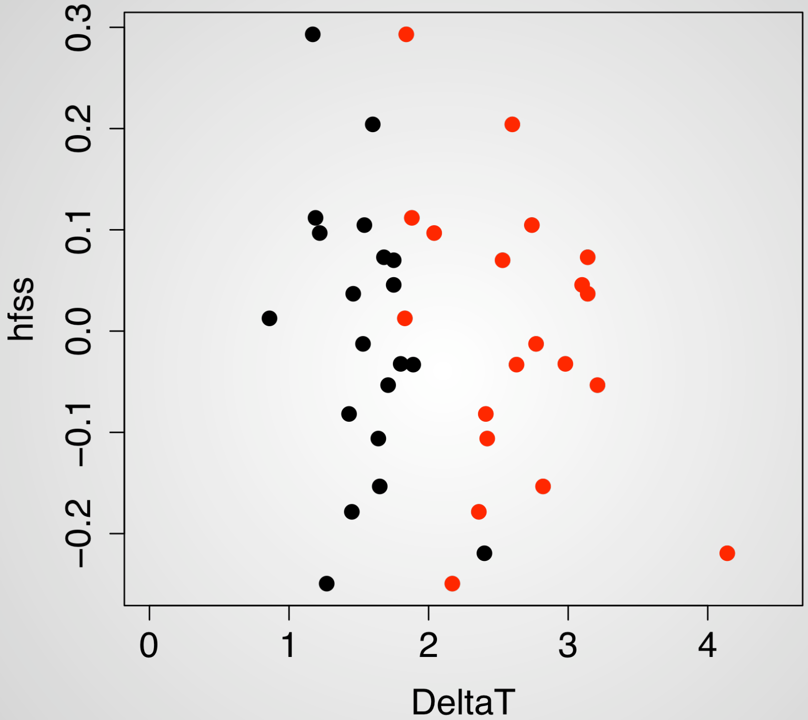
Using model performance in replicating observed climate to weigh a model more or less seems like a good idea too, except….

Performance does not seem to be correlated to projections of change!

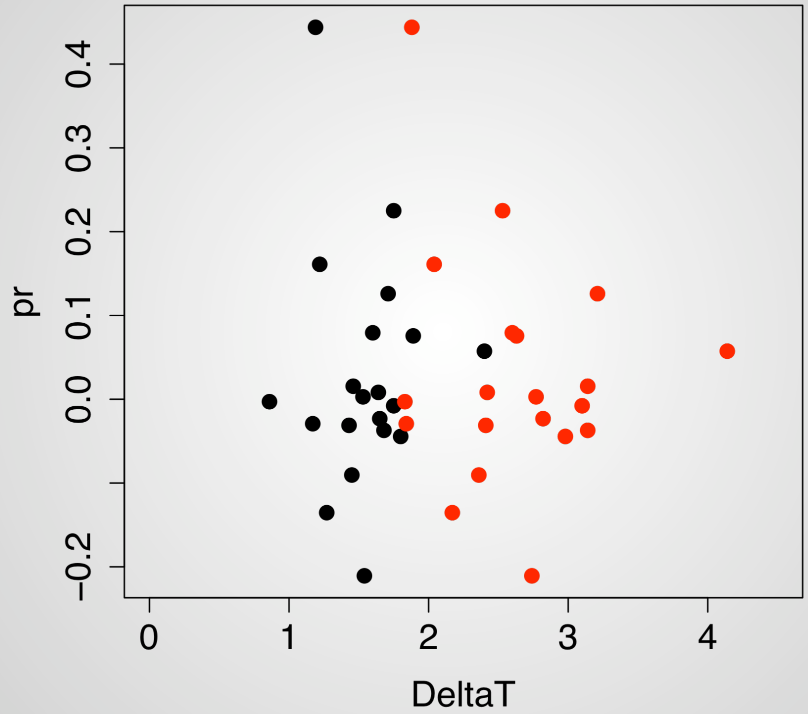


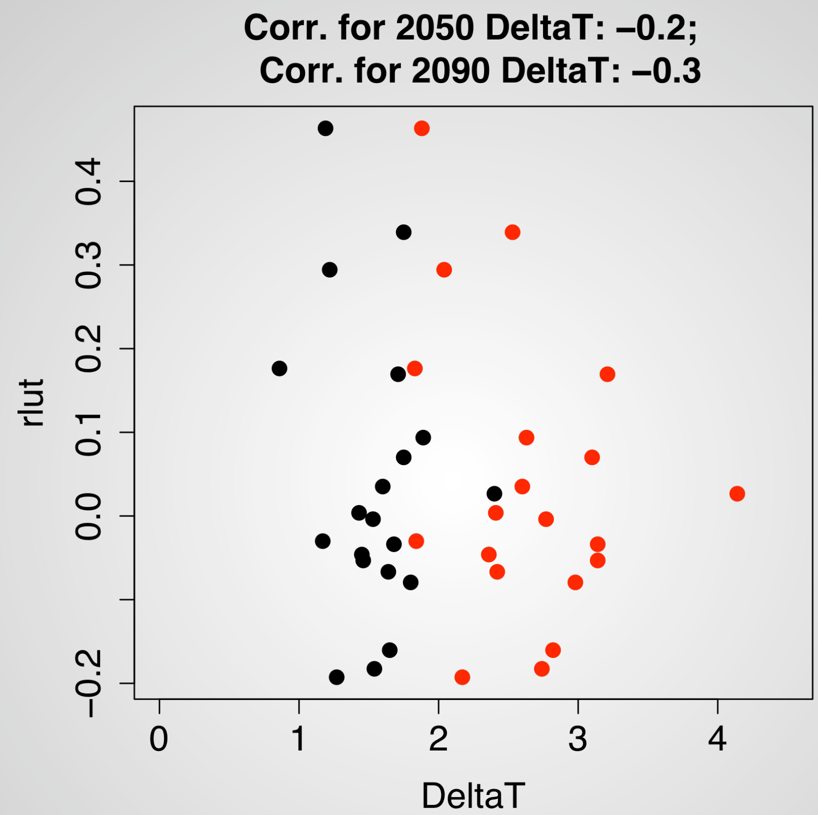
Model performance (one of the diagnostics from previous page) vs. model projected temperature change at 2050 or 2100 (or their difference). Low correlation!

Corr. for 2050 DeltaT: –0.3;
Corr. for 2090 DeltaT: –0.3

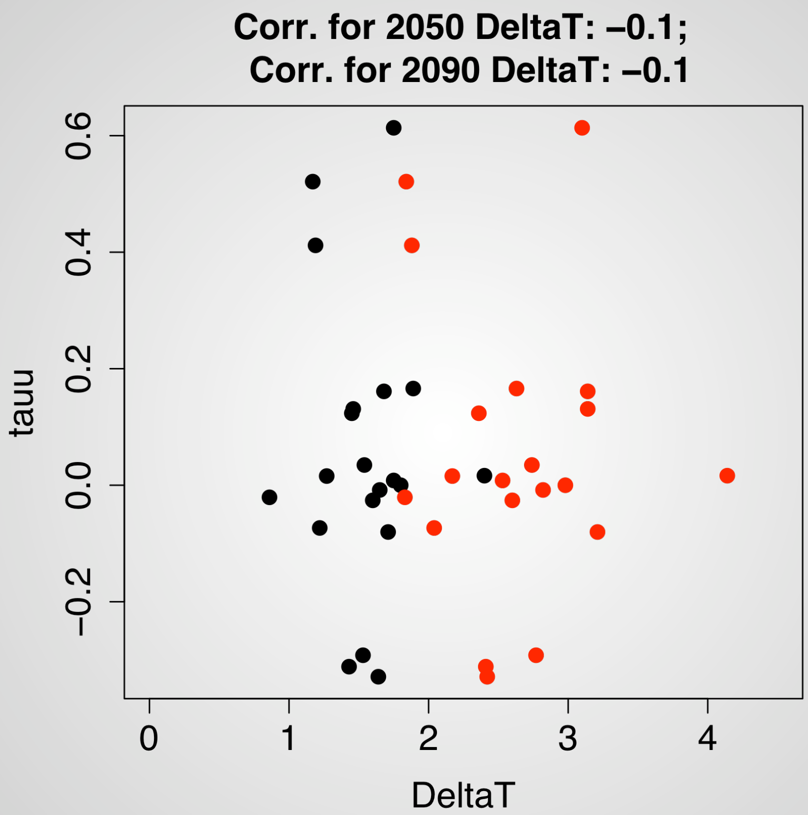


Corr. for 2050 DeltaT: 0;
Corr. for 2090 DeltaT: −0.2

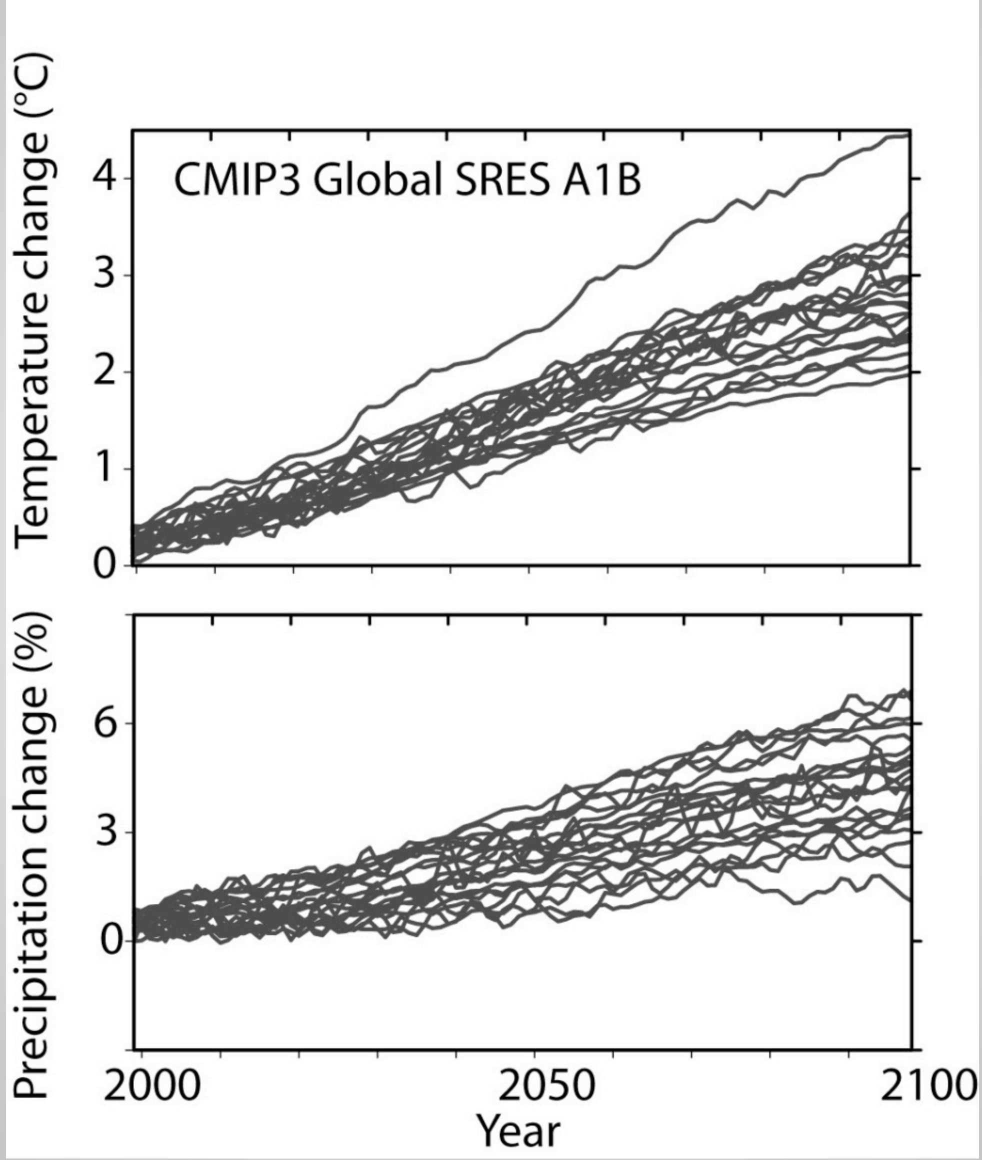


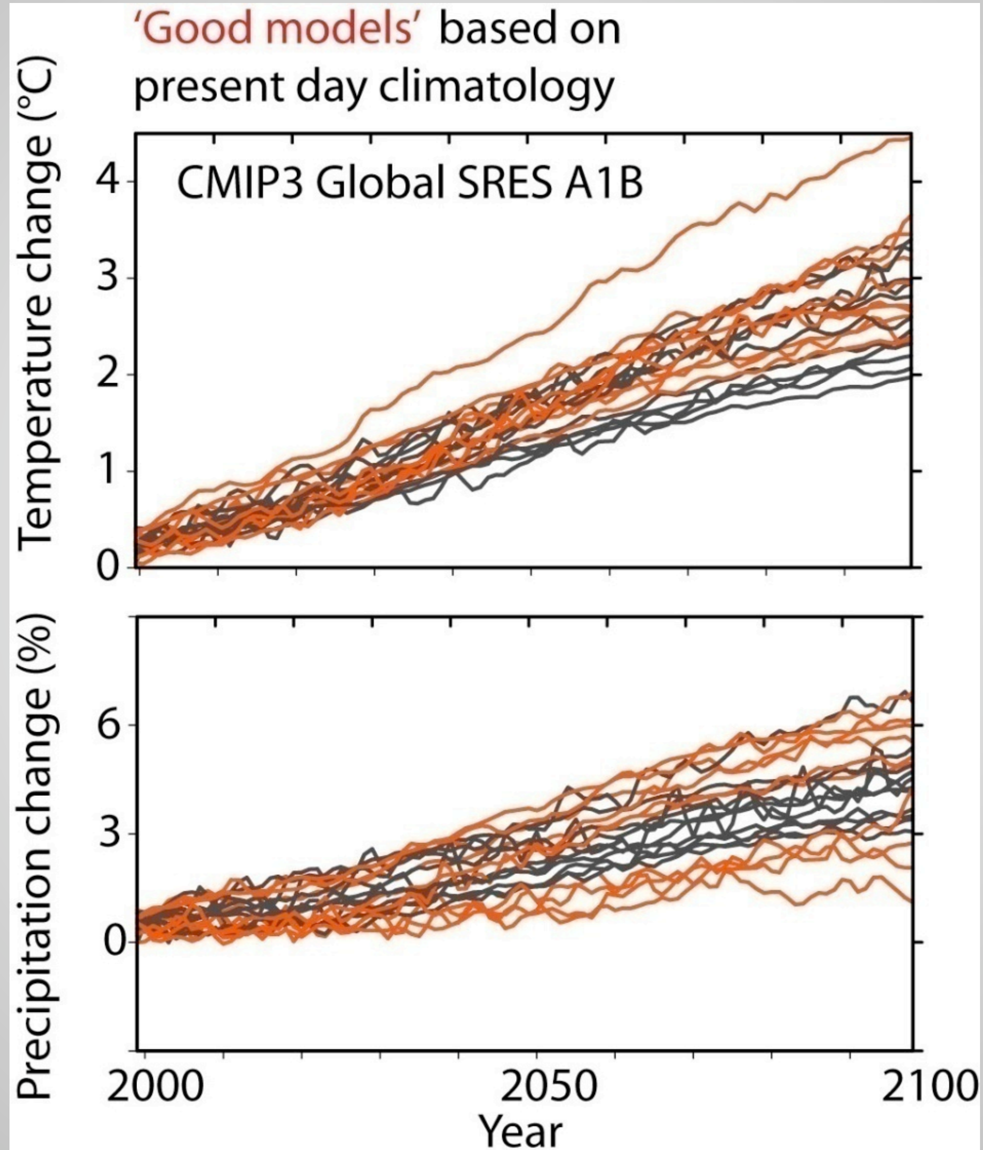

Corr. for 2050 DeltaT: –0.2;
Corr. for 2090 DeltaT: –0.3
rlut
−0.2
0.0
0.1
0.2
0.3
0.4
0
1
2
3
4
DeltaT

and so on and so on…



Another way to look at the same issue

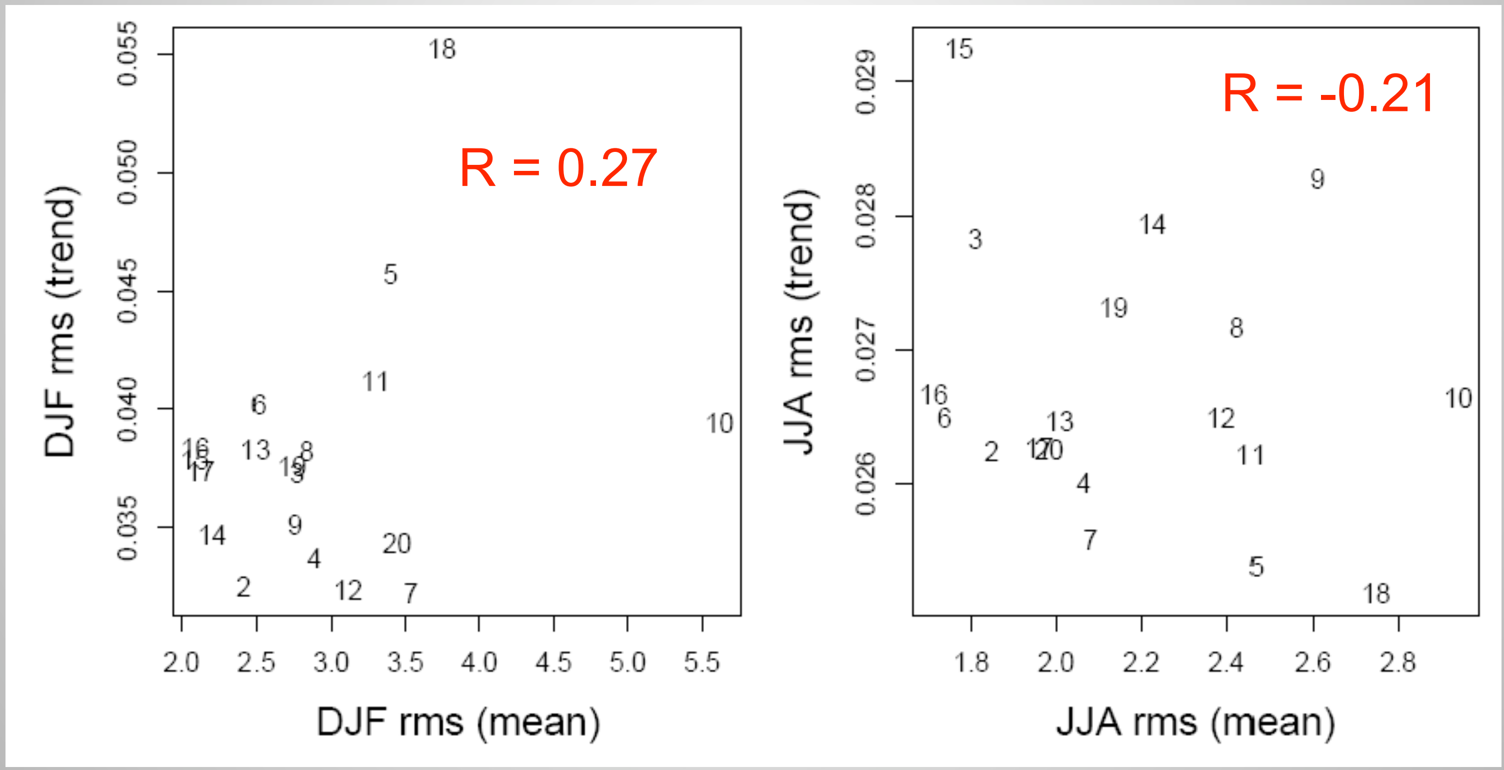



'Good models' based on
present day climatology
CMIP3 Global SRES A1B
Temperature change (°C)
0
1
2
3
4
Precipitation change (%)
0
3
6
2000
2050
2100
Year

And another….

Is model performance on the mean state indicative of the ability to simulate (future) trends?

Ability to simulate observed pattern of warming trend

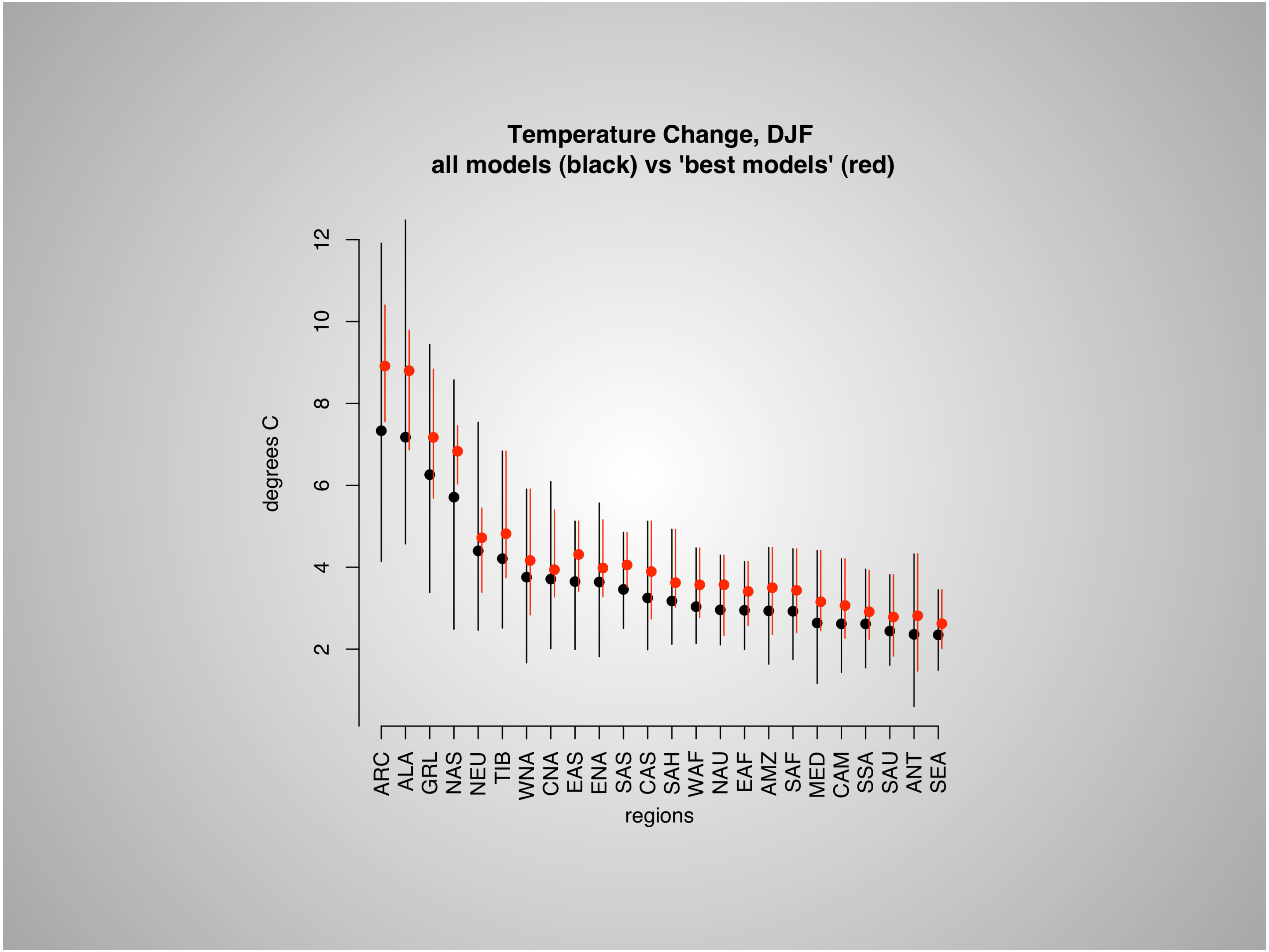


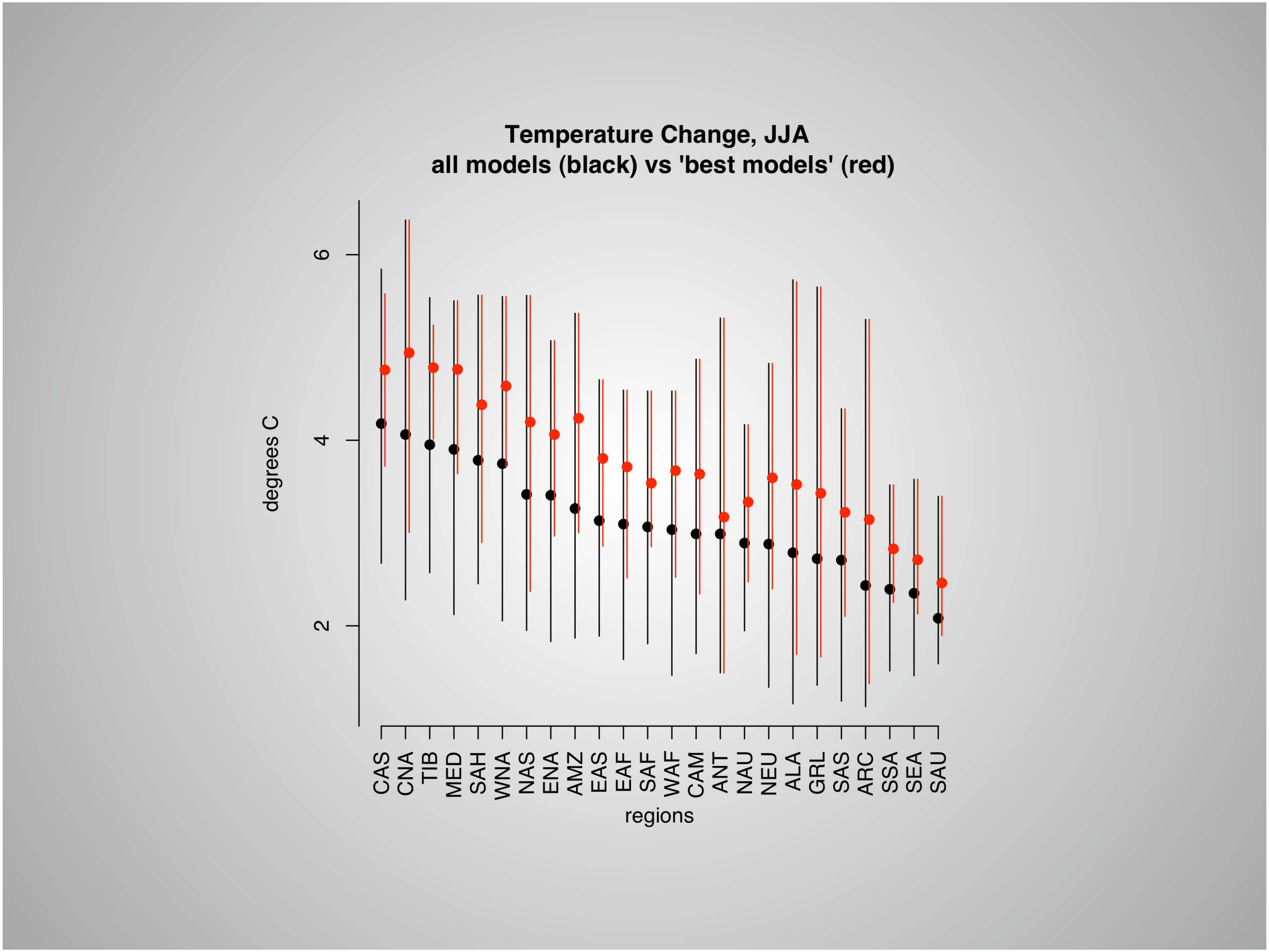
Ability to simulate observed pattern of mean climate

What if we apply a really arbitrary criterion?

An informal survey among climate scientists comes up with the same 5 climate models as “the best”.

Does it matter?


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



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

Looks like some complex definition of model performance actually constrains model projections.

Conclusions (Part 1)

The nature of climate model ensembles makes **simple means and standard deviations unsustainable** as summary statistics, not to mention as the basis for inference. Simple summary statistics may be easy to interpret but hide assumptions (IID) that undermine their use.

Rigorous statistical approaches, accounting for biases and dependence among climate models are a compelling alternative, even if you don’t make your living as a statistician.

The nature of climate simulations, however, does not show an obvious path to modeling multi-model ensembles and bringing their performance to bear.