Using Statistics to assess climate uncertainty

Paul Northrop,

University College London

paul@stats.ucl.ac.uk

This is joint work with Richard Chandler.

Overview

- Climate uncertainty and design of experiments
- Examples : 21st century
  1. global temperature
  2. regional temperature
  3. regional precipitation
- Implications for the design of climate model experiments

Design of climate model experiments

- Prediction from climate models is time-consuming
- We want to make best use of this time/effort
- Experimental design: how best to organise an experiment in order to answer the question(s) of interest with sufficient precision
- Question: “Which sources of climate uncertainty are most important?”
  - IPCC AR4 data
  - Simple probability model
  - Implications for design of climate experiments

Some sources of uncertainty in climate predictions

- Climate model (GCM)
- SRES emission scenario (A1B, A2, B1, …)
- GCM run

The greater the variability in climate predictions over, say, GCMs, the more the choice of GCM matters.

IPCC AR4 data

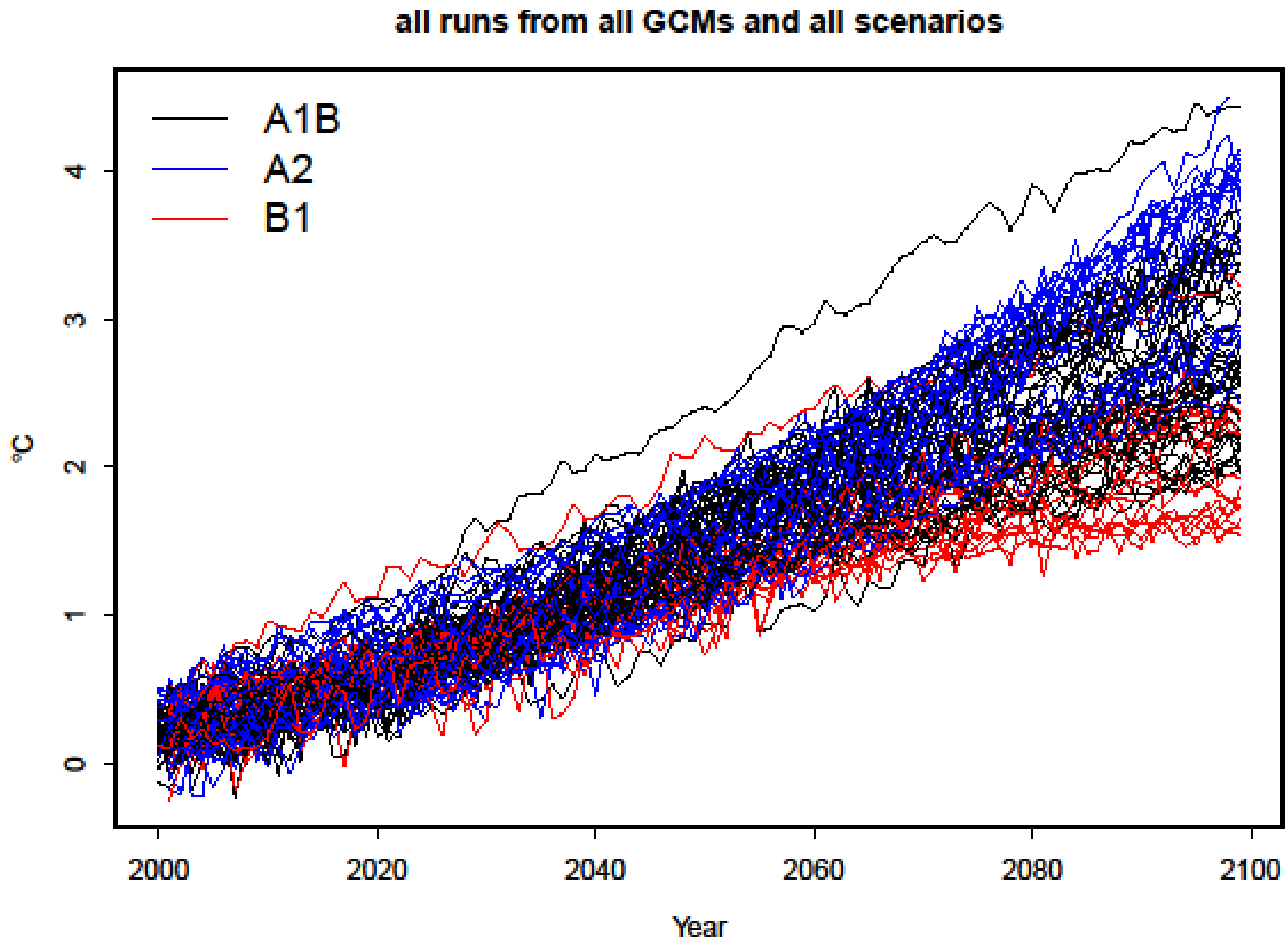
- WCRP CMIP3 Multi-Model Dataset Archive at PCMDI
- 24 GCMs
- 3 scenarios: A1B, A2, B1
- Some GCMs have multiple runs per scenario; some have none.
- Not a designed experiment
- There is scope to increase the usefulness of runs

Example 1 : 21st century global (surface air) temperature

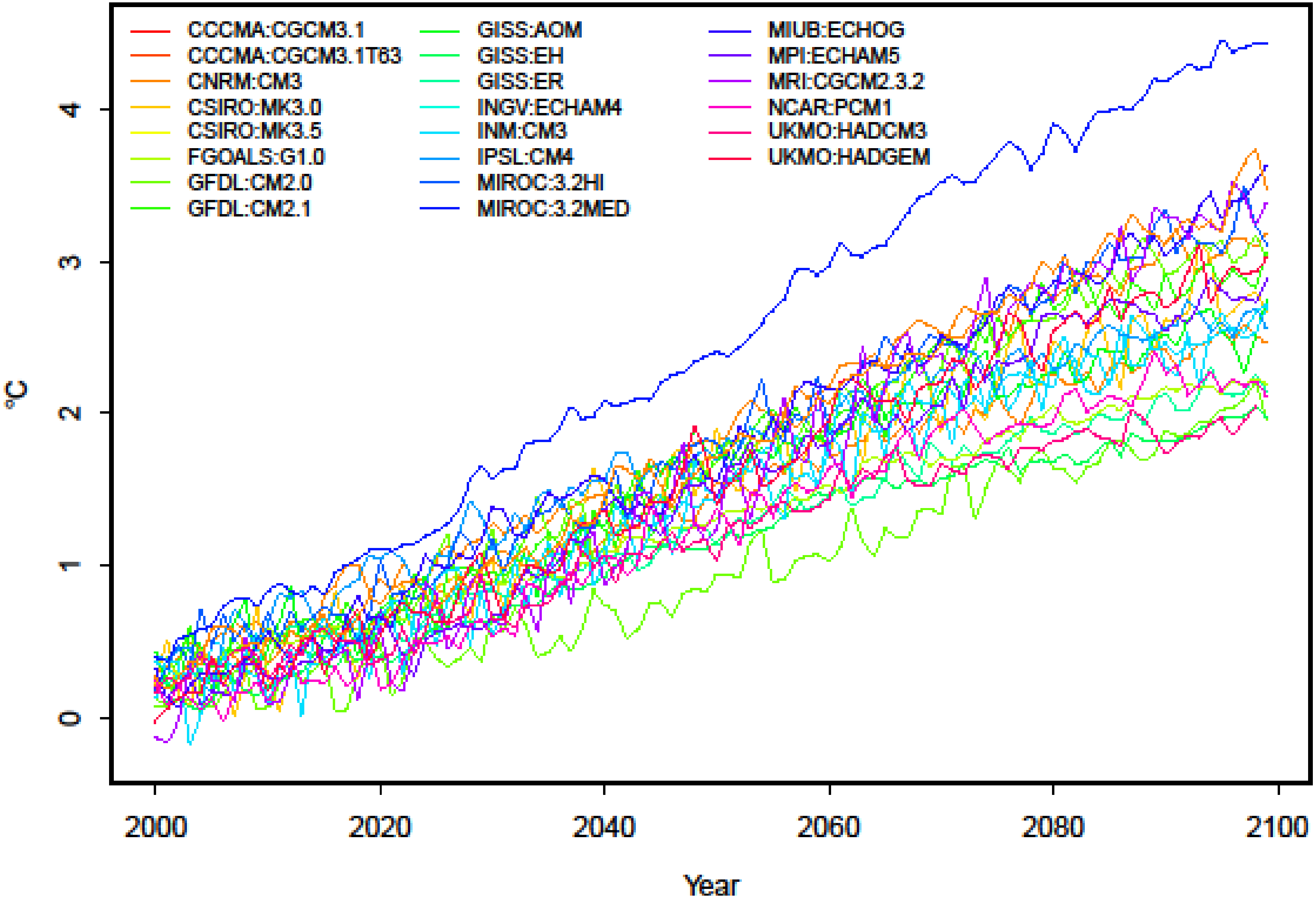
- We define an index of temperature change
- Baseline: mean temperature in 1980-1999
- 2 time horizons:
    1. Change in 2020-2049 mean from 1980-1999 mean
    2. Change in 2069-2098 mean from 1980-1999 mean
- Units are °C throughout

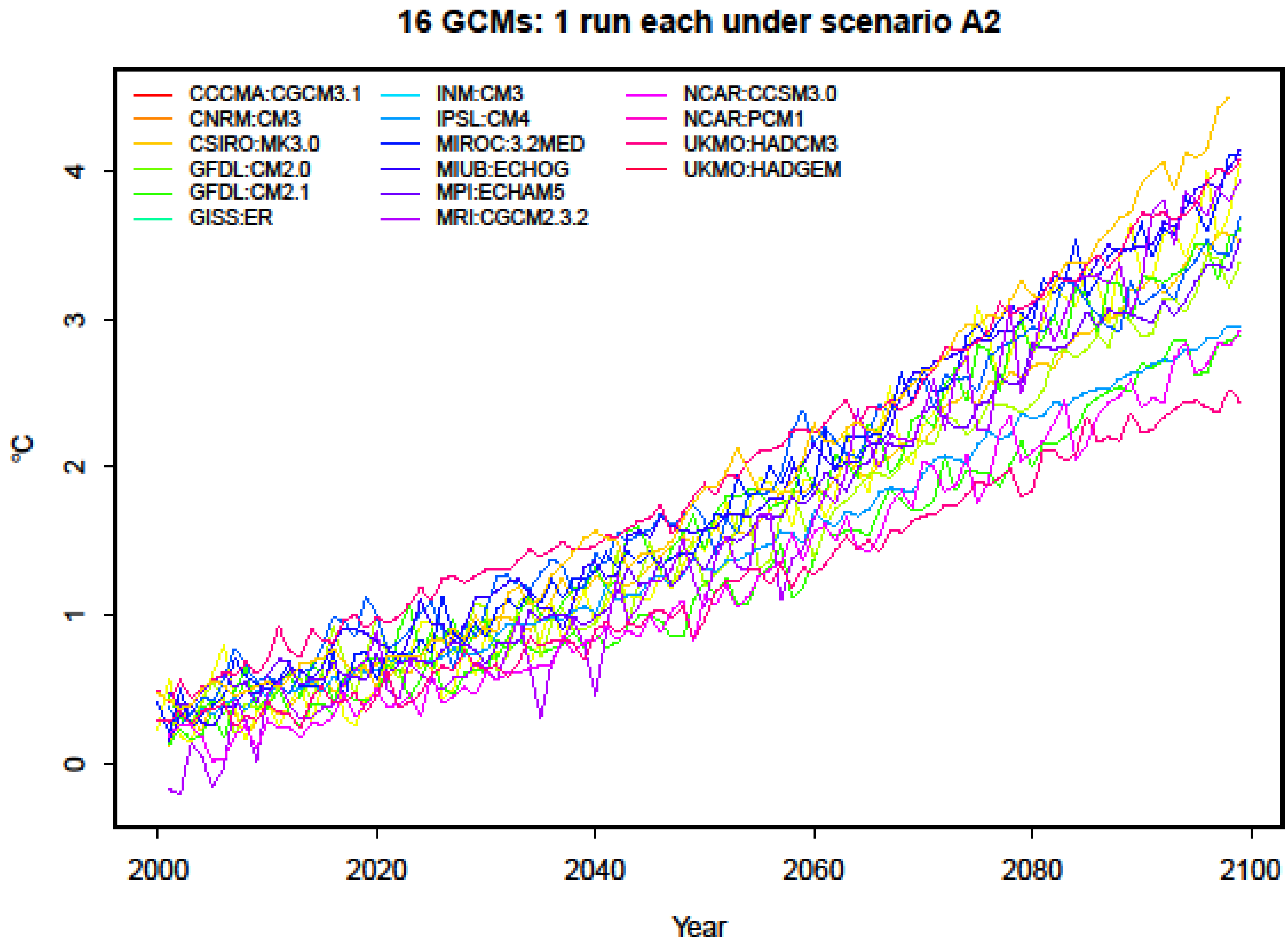
PCMDI data

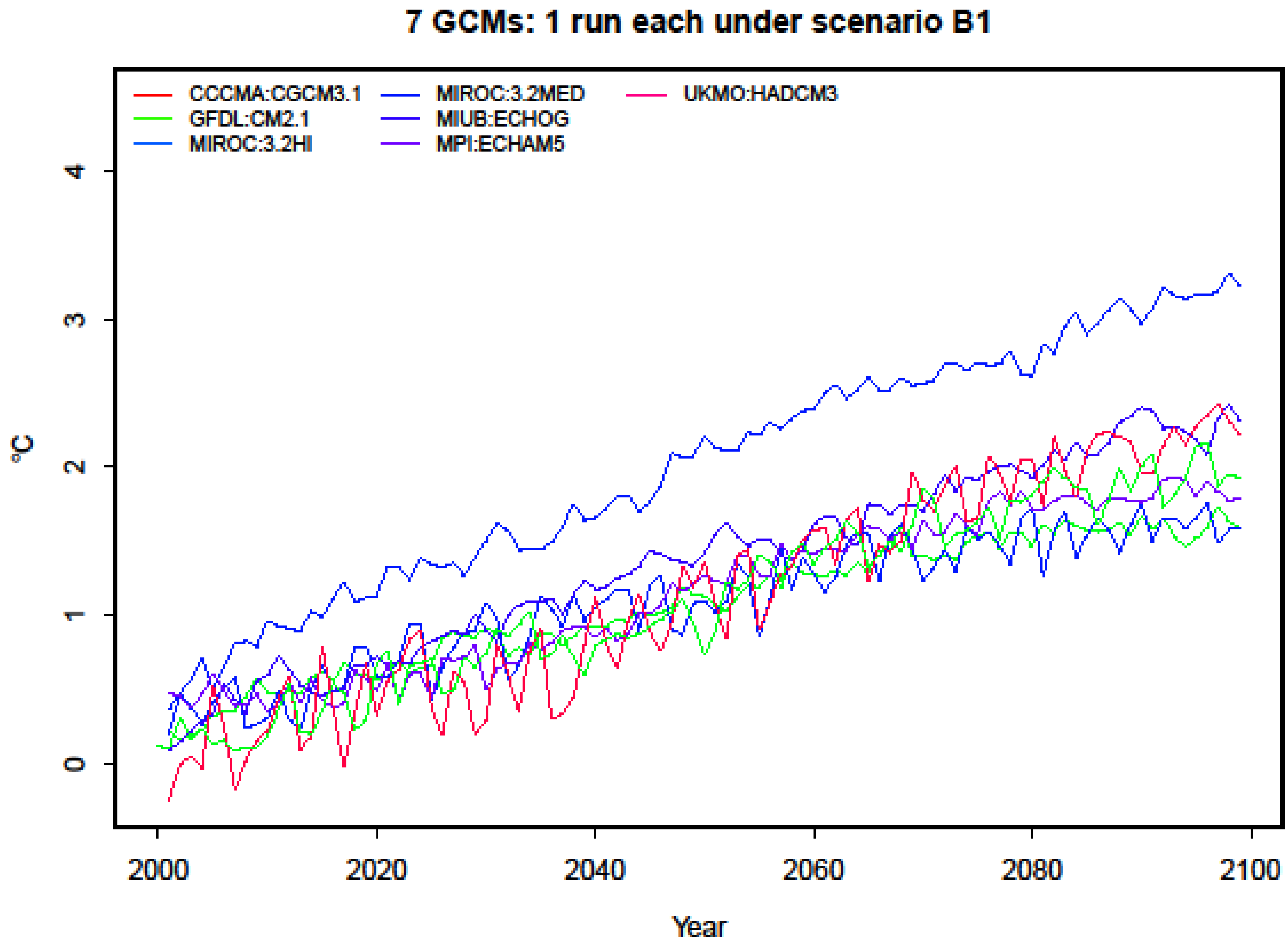
GCM	A1B	A2	B1
bccr:bcm2:0	1	1	1
cccma:cgcm3:1	5	5	5
cccma:cgcm3:1:t63	1	0	1
cnrm:cm3	1	1	1
csiro:mk3:0	1	1	1
csiro:mk3:5	1	1	1
gfdl:cm2:0	1	1	1
gfdl:cm2:1	1	1	1
giss:aom	2	0	2
giss:model:e:h	3	0	0
giss:model:e:r	5	1	1
iap:fgoals1:0:g	3	0	3
ingv:echam4	1	1	0
inmcm3:0	1	1	1
ipsl:cm4	1	1	1
miroc3:2:hires	1	0	1
miroc3:2:medres	3	3	3
miub:echo:g	3	3	3
mpi:echam5	4	3	3
mri:cgcm2:3:2a	5	5	5
ncar:ccsm3:0	7	5	8
ncar:pcm1	4	4	4
ukmo:hadcm3	1	1	1
ukmo:hadgem1	1	1	0
	57	40	48


all runs from all GCMs and all scenarios
A1B
A2
B1
4
3
2
1
0
°C
2000
2020
2040
2060
2080
2100
Year

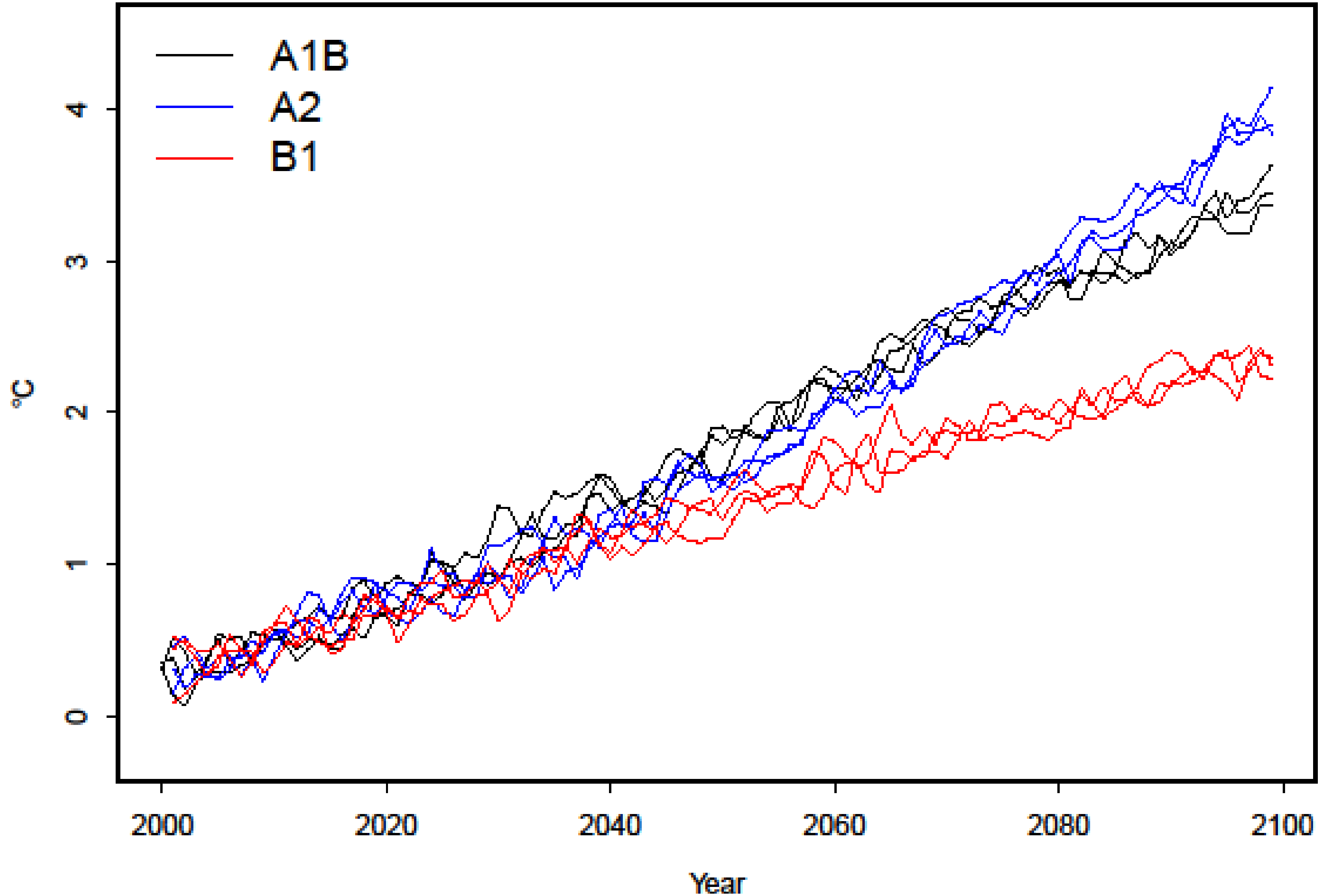
22 GCMs: 1 run each under scenario A1B




16 GCMs: 1 run each under scenario A2
CCCMA:CGCM3.1
CNRM:CM3
CSIRO:MK3.0
GFDL:CM2.0
GFDL:CM2.1
GISS:ER
INM:CM3
IPSL:CM4
MIROC:3.2MED
MIUB:ECHOG
MPI:ECHAM5
MRI:CGCM2.3.2
NCAR:CCSM3.0
NCAR:PCM1
UKMO:HADCM3
UKMO:HADGEM
°C
0
1
2
3
4
2000
2020
2040
2060
2080
2100
Year


7 GCMs: 1 run each under scenario B1
CCCMA:CGCM3.1
GFDL:CM2.1
MIROC:3.2HI
MIROC:3.2MED
MIUB:ECHOG
MPI:ECHAM5
UKMO:HADCM3
°C
0
1
2
3
4
2000
2020
2040
2060
2080
2100
Year

GCM MIROC:3.2MED, 3 runs from each scenario



2-way random effects ANOVA

Y_{ijk} = measure of change for **GCM** i, **scenario** j and **run** k.

$$Y_{ijk} \;=\; \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ijk},$$

μ	overall mean change	
α_i	adjustment for GCM i	$\alpha_i \overset{iid}{\sim} N\left(0, \sigma_G^2\right)$
β_j	adjustment for scenario j	$\beta_j \overset{iid}{\sim} N\left(0, \sigma_S^2\right)$
γ_{ij}	scenario-specific adjustment for GCM i	$\gamma_{ij} \overset{iid}{\sim} N\left(0, \sigma_I^2\right)$
ϵ_{ijk}	residual effect of variability over runs	$\epsilon_{ijk} \overset{iid}{\sim} N\left(0, \sigma_R^2\right)$

In particular, we are interested in the relative magnitudes of the variance components

$$\sigma_G^2,\ \sigma_S^2 \text{ and } \sigma_R^2$$

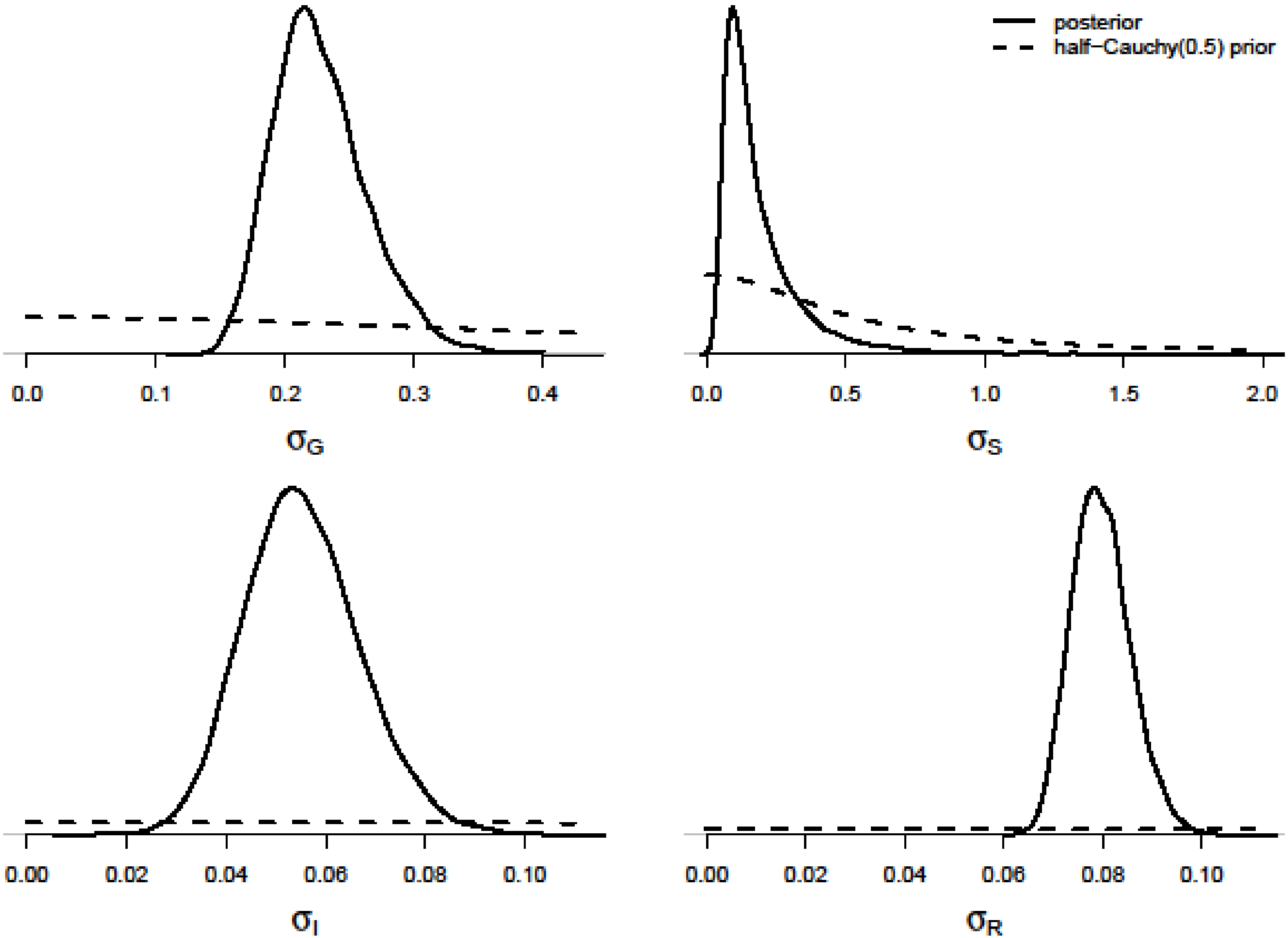
Statistical inference

- Issues
  1. Scenario has only 3 levels. Lack of information about variability over scenarios (σ_S).
  2. Lack of balance.
  3. No runs for some GCM-scenario combinations
- REML (cf. posterior mode). Gilmour & Goos (2009) argue against REML : σ_S tends to be underestimated.
- Bayesian inference (cf. posterior mean or median) with weakly-informative priors

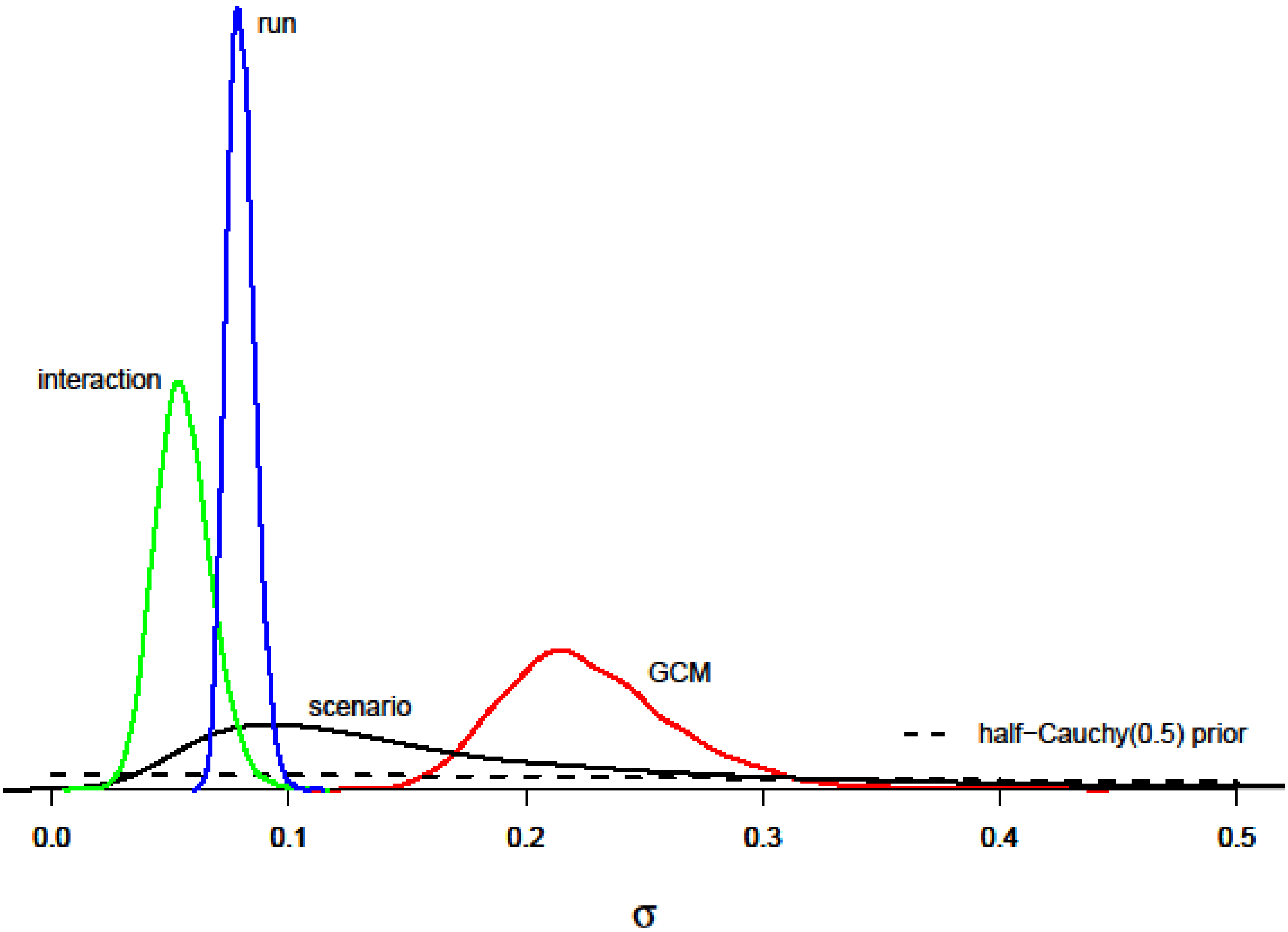
Weakly-informative priors for variance parameters

- ... let the data speak for themselves, but downweight unrealistic possibilities
- The data provide little information about σ_S
- Gelman (2006) argues against improper uniform priors and the inverse-gamma family.
- ... and for a half-Cauchy prior
- Idea: we choose the prior so that
  1. It is weakly-informative for σ_S
  2. It is non-informative for σ_G, σ_I and σ_R

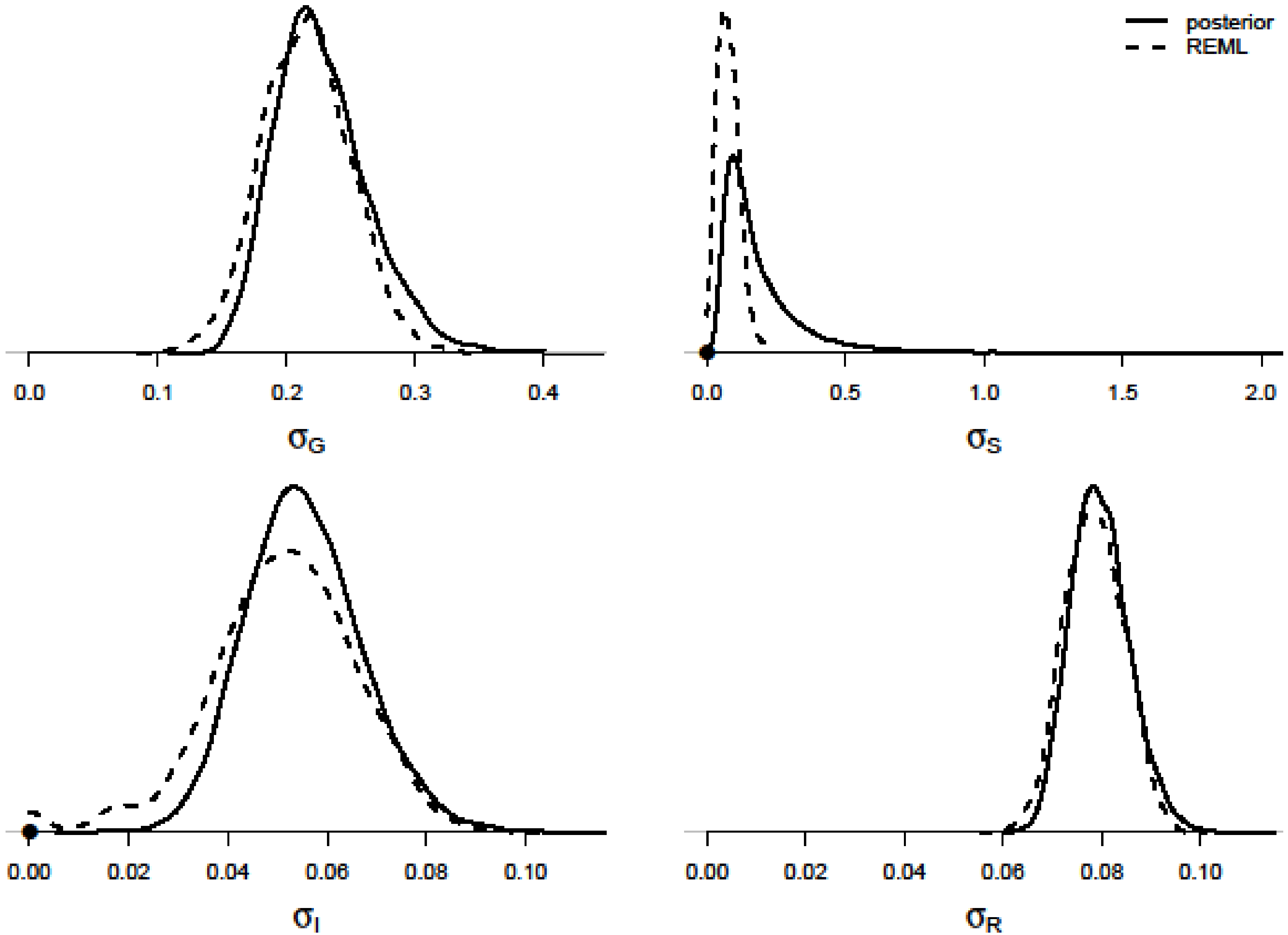
Global temperature 2020-2049



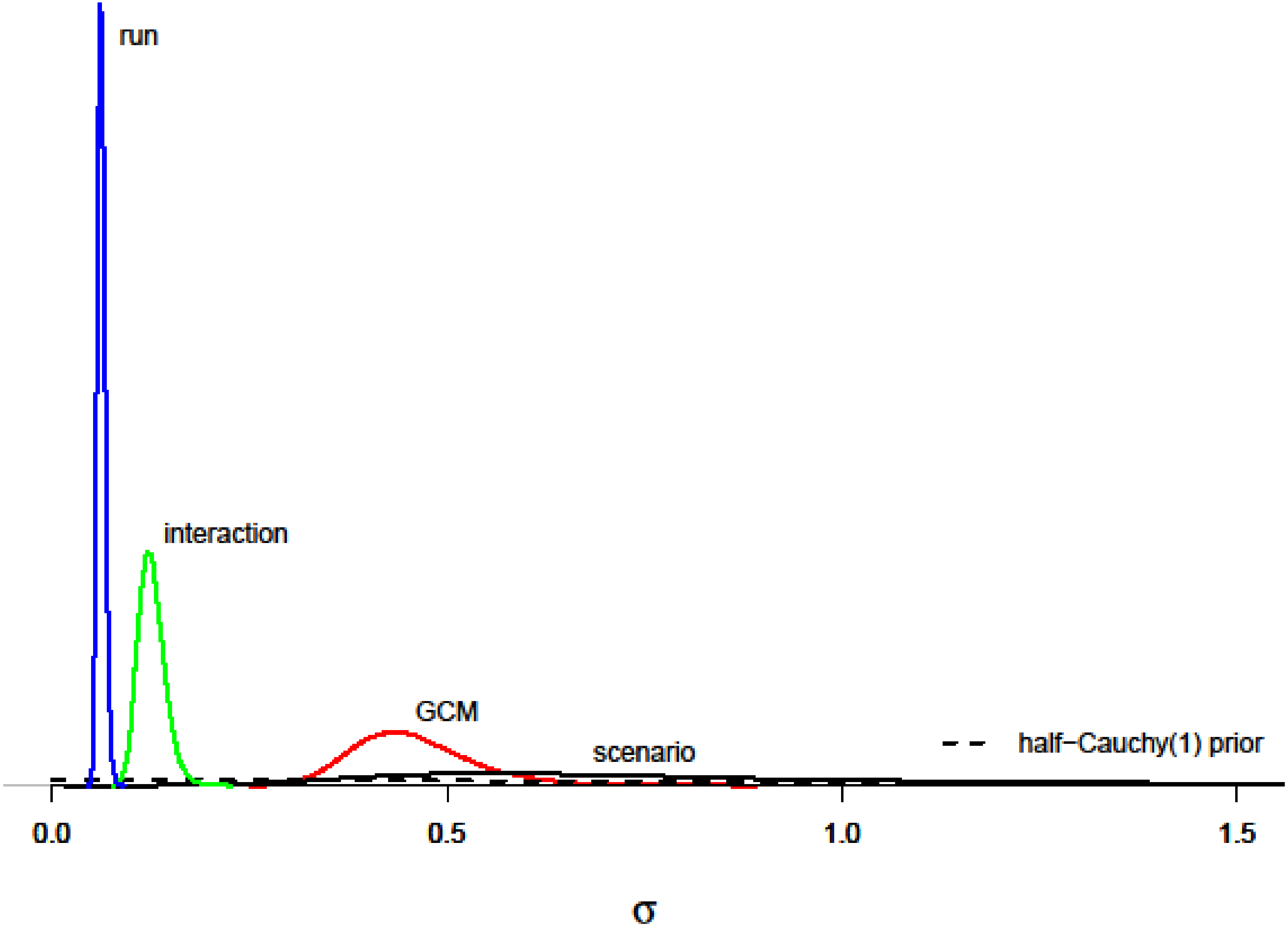
Global temperature 2020-2049



Global temperature 2020-2049

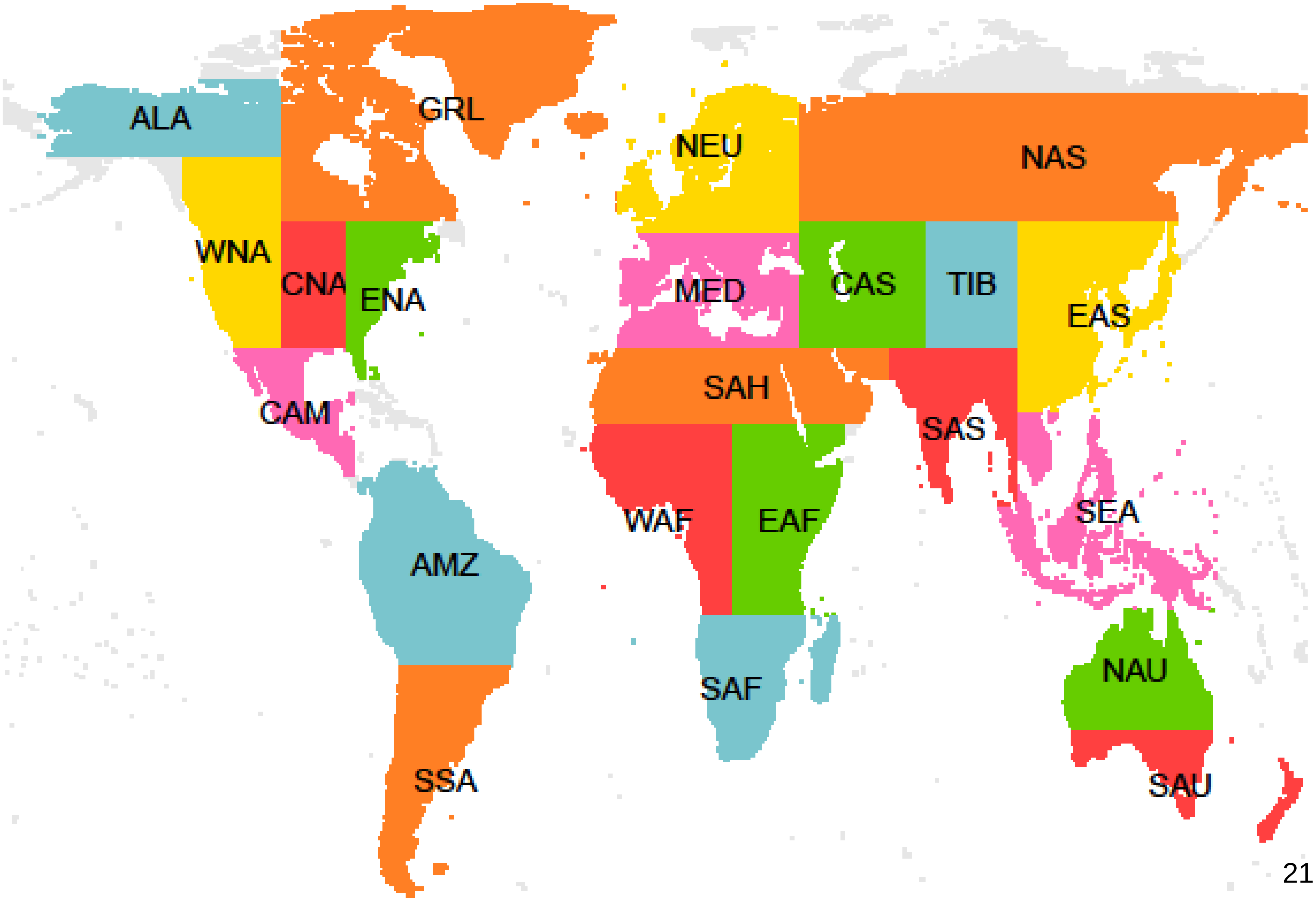


Global temperature 2069-2098

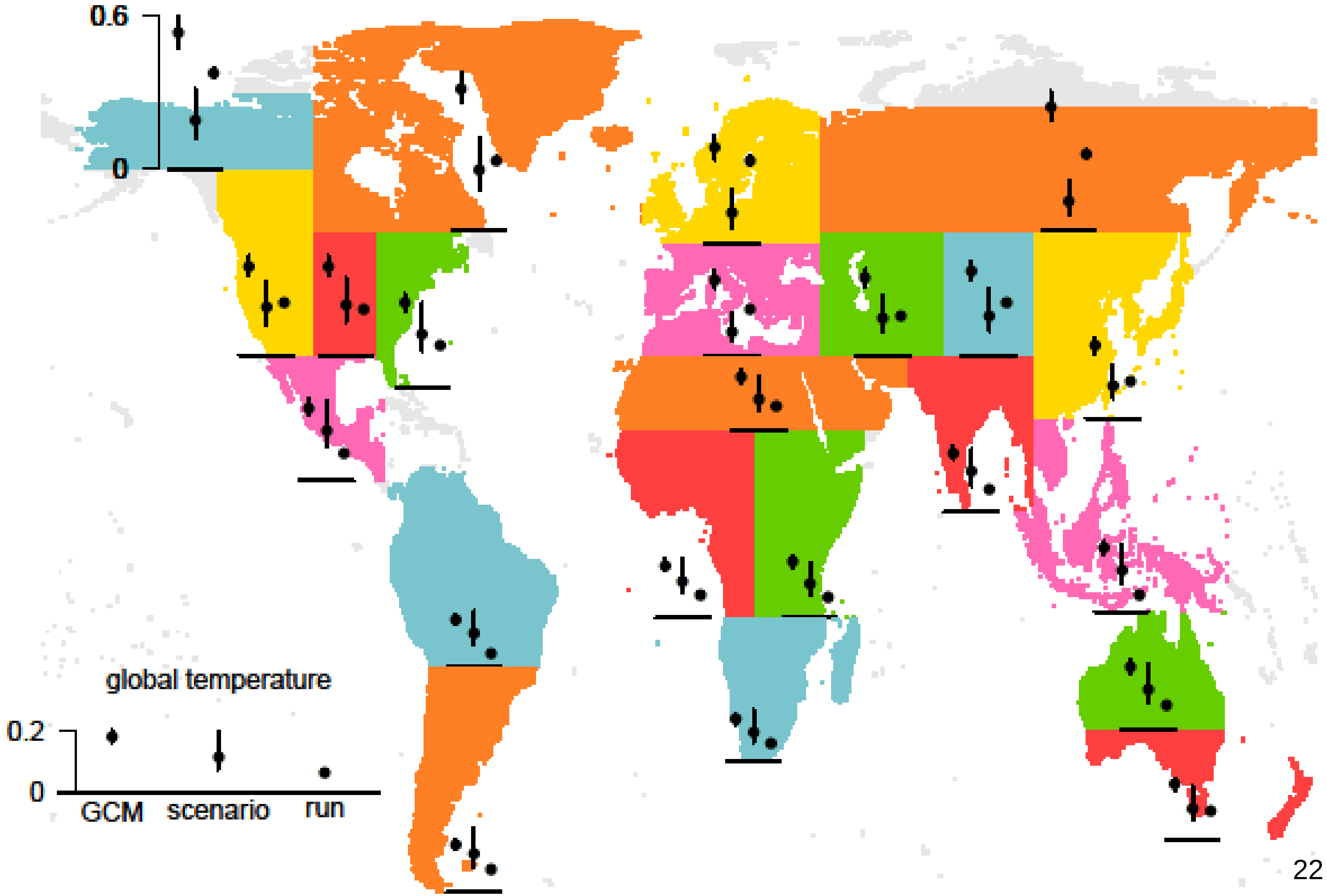


Example 2 : regional temperature

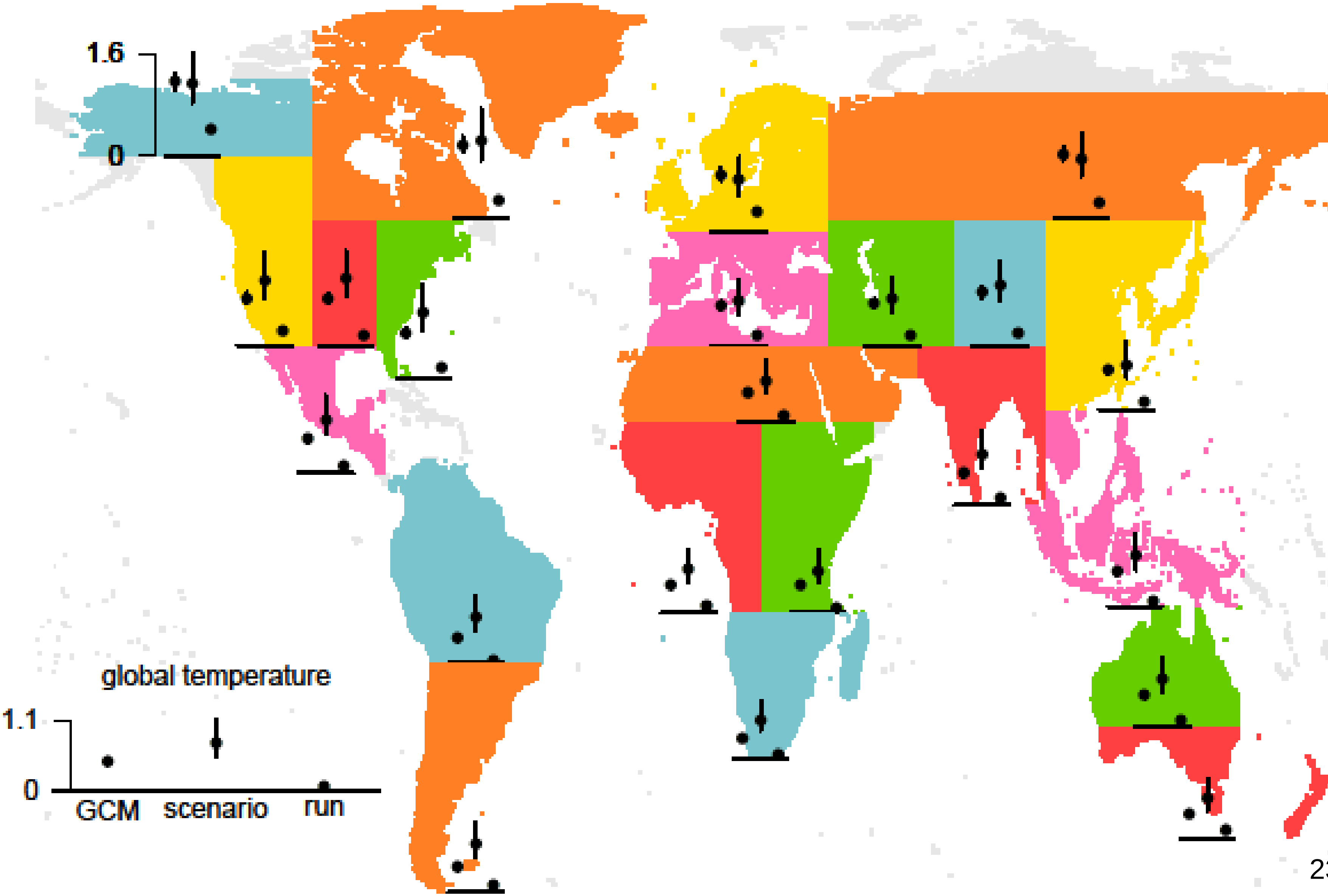
- 22 land regions (IPCC Data Distribution Centre)
- We repeat the Bayesian analysis with weakly-informative priors within each region


ALA
GRL
NEU
NAS
WNA
CNA
ENA
MED
CAS
TIB
EAS
CAM
SAH
SAS
WAF
EAF
SEA
AMZ
NAU
SAF
SSA
SAU

Temperature 2020−2049 : posterior quartiles of random effect s.d.s



Temperature 2069–2098 : posterior quartiles of random effect s.d.s



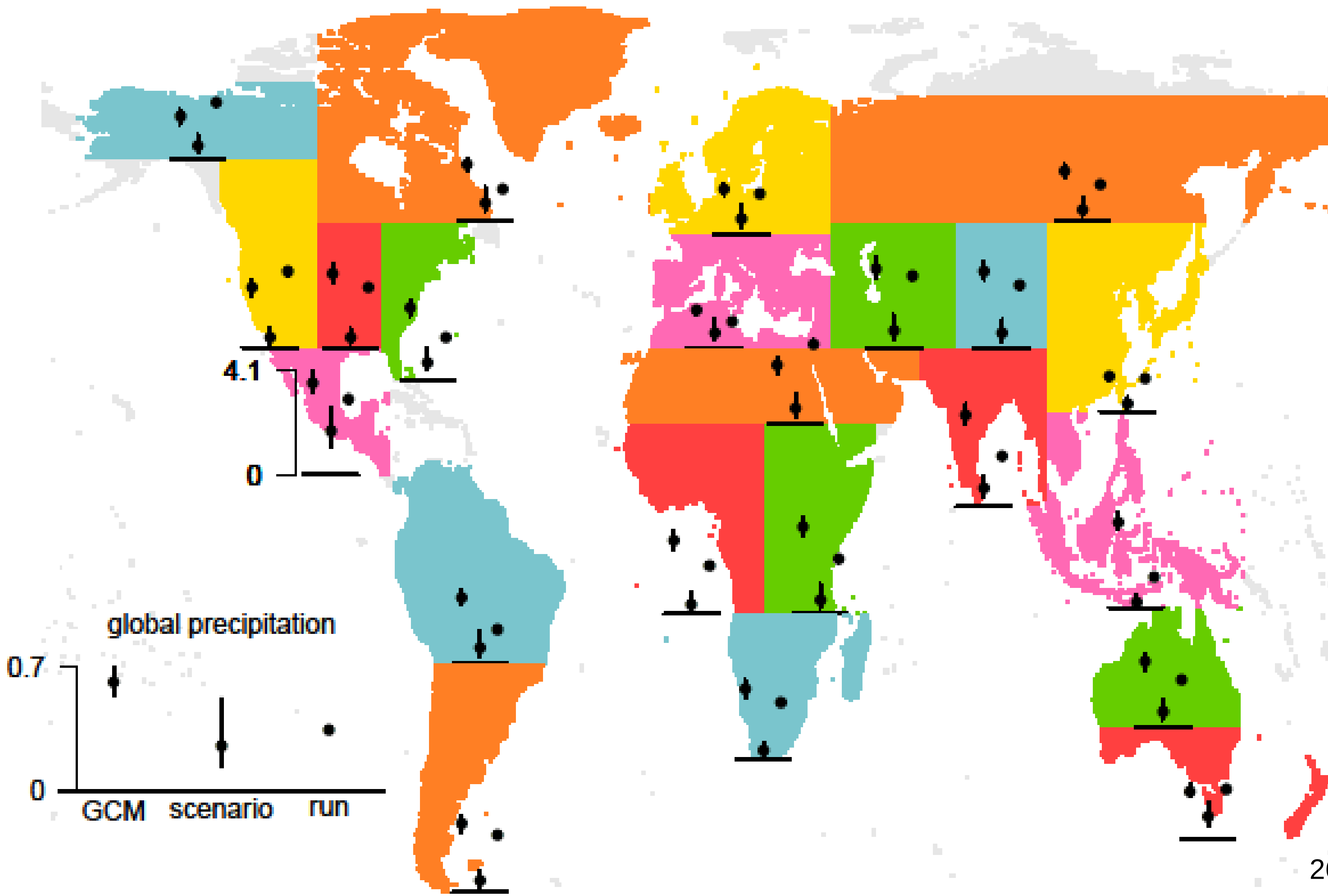
Summary for temperature

- 2020-2049
  - Global : variability over GCMs > scenario > runs
  - Regional: runs matters more than scenario in some areas, e.g. In the north
- 2069-2098
  - Scenario matters more as we move through the 21st century (obviously!)
  - Scenario is at least as important as GCM in most regions

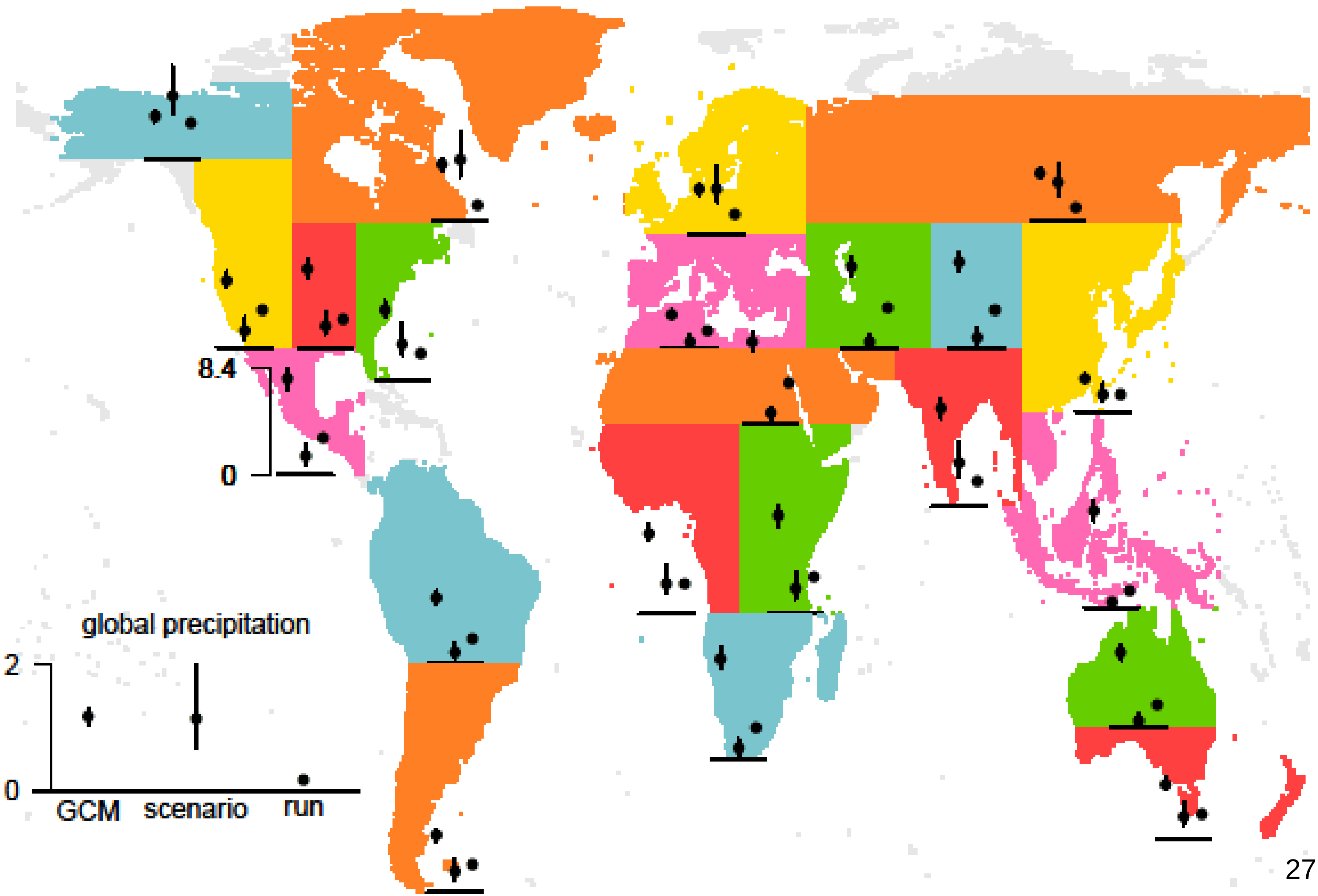
Example 3: regional precipitation

- Variable: precipitation flux, converted to mm/day
- Same idea as regional temperature
- Index of change is the % change in mean from the baseline period of 1980-1999

Precipitation 2020−2049 : posterior quartiles of random effect s.d.s



Precipitation 2069−2098 : posterior quartiles of random effect s.d.s



Summary for precipitation

- 2020-2049
  - Global : variability over GCMs largest, but relatively high variability over different runs from the same GCM
  - Regional : a similar picture. In some areas (e.g. Alaska) var. over runs > var. over GCMs
- 2069-2098
  - Global : choice of scenario becoming more important as century progresses
  - In many regions scenario is relatively unimportant

Summary

- The relative importance of GCM, scenario and run depends on
  1. climate variable
  2. region
  3. time horizon
  4. season ?
- Implications for the efficient design of future climate experiments ...

Experimental design (general idea)

- Probability model with parameters, e.g. $\sigma_G, \sigma_S, \sigma_R$
- Frame questions in terms of parameters
- Fixed resources : how do we collect data to estimate parameter(s) of interest with greatest precision, e.g. to minimize $\mathrm{var}(\hat{\sigma}_G)$?
- Unlimited resources: what data are required to estimate parameter(s) with desired precision ?
- Use of stochastic simulation

Experimental design (in current context)

- Review of design for variance components estimation by Khuri (2000)
- Optimal design depends on $\sigma_G, \sigma_S, \sigma_R$ (prior information.; adaptive designs?)

1. Fixed number of GCMs and scenarios
   - Balanced design is optimal

2. Can choose numbers of GCMs and scenarios
   - If σ_R is dominant balanced design is optimal
   - If not, there are more efficient unbalanced designs
   - If $\sigma_S >> \sigma_R$ we need large number of scenarios and small number of runs per scenario

Concluding remarks

- Strategic planning of climate experiments can increase increase the cost-effectiveness and usefulness of climate model runs
- (Relatively) simple probability models can inform design
- Better representation of future conditions needed
- `Representative Concentration Pathways' : "designed to span a wide range of outcomes" (Chandler *et al*. (2010)

References

- Chandler *et al*. (2010) *Significance*, 7(1), 9-12
- Gelman, A. (2006) *Bayesian Analysis*, 1(3), 515-533
- Gilmour, S.G. and Goos, P. (2009) *Appl. Statist.* 58(4), 467-484
- Khuri, A.I. (2000) *Inst. Statist. Rev.*, 68(3),311-322
- Software. R : http://www.R-project.org
- WCRP CMIP3 Multi-Model Dataset Archive at PCMDI https://esgcet.llnl.gov:8443/home/publicHomePage.do
- KNMI Climate Explorer http://climexp.knmi.nl/

Thank you for listening