Partitioning uncertainty in climate predictions using data from undesigned climate experiments

Paul Northrop
Richard Chandler
Department of Statistical Science
University College London

RSS2012, Telford
6th September 2012

Human-induced climate change

Greenhouse gases (warming)
CO_2 (fossil fuels, deforestation)
methane (livestock), N_2O (fertilizers)
Industrial air pollution (cooling)
sulphates, soot

↓

Increase in global temperature
Arctic shrinkage, sea level rise
change in disease patterns, species extinctions

↓

Changes in climate (long-term weather)
change in rainfall patterns, extreme weather?
floods? drought?

UCL

Question: What will the climate be like over the next 100 years?
Answer: It will depend on

- what the world is like in the future - guesses (scenarios);
- how the atmosphere/oceans react - climate models.

IPCC Fourth Assessment Report (AR4) 2007

- IPCC : Intergovernmental panel on climate change.
- Provides an objective assessment of the latest literature on climate change.
- Provides climate predictions for 2000-2100 from many climate models under different socio-economic scenarios.

Socio-economic scenarios

(Educated) guesses at what the world will be like over the next 100 years.

B1: low emissions, clean and efficient technologies, global sustainability, population peaks in 2050.

A2: medium-high emissions, economic growth on regional scales, increasing population.

A1B: high emissions (a balance across all fuel sources), very rapid economic growth, market forces dominate, population peaks in 2050.

. . . and many others.

General Circulation Models (GCMs)

- mathematical models of atmosphere/oceans based on the laws of physics/chemistry;
- require large amounts of computing power and time;
- produce predictions of climate variables.

Comments

- Different scientists produce different GCMs.
- Predictions depend on scenario.
- Small changes in the input conditions can produce big changes in the predictions, i.e. **different runs of the model produce different predictions.**

IPCC AR4 data

GCM number	GCM name	A1B	A2	B1
1	bccr:bcm2:0	1	1	1
2	cccma:cgcm3:1	5	5	5
3	cccma:cgcm3:1:t63	1	0	1
4	cnrm:cm3	1	1	1
5	csiro:mk3:0	1	1	1
6	csiro:mk3:5	1	1	1
7	gfdl:cm2:0	1	1	1
8	gfdl:cm2:1	1	1	1
9	giss:aom	2	0	2
10	giss:model:e:h	3	0	0
11	giss:model:e:r	5	1	1
12	iap:fgoals1:0:g	3	0	3
13	ingv:echam4	1	1	0
14	inmcm3:0	1	1	1
15	ipsl:cm4	1	1	1
16	miroc3:2:hires	1	0	1
17	miroc3:2:medres	3	3	3
18	miub:echo:g	3	3	3
19	mpi:echam5	4	3	3
20	mri:cgcm2:3:2a	5	5	5
21	ncar:ccsm3:0	7	5	8
22	ncar:pcm1	4	4	4
23	ukmo:hadcm3	1	1	1
24	ukmo:hadgem1	1	1	0
total		57	40	48

No design. Not the best use of a lot of computational effort!

Sources of climate uncertainty

Question: Is variability in predicted climate variables due mainly to choice of

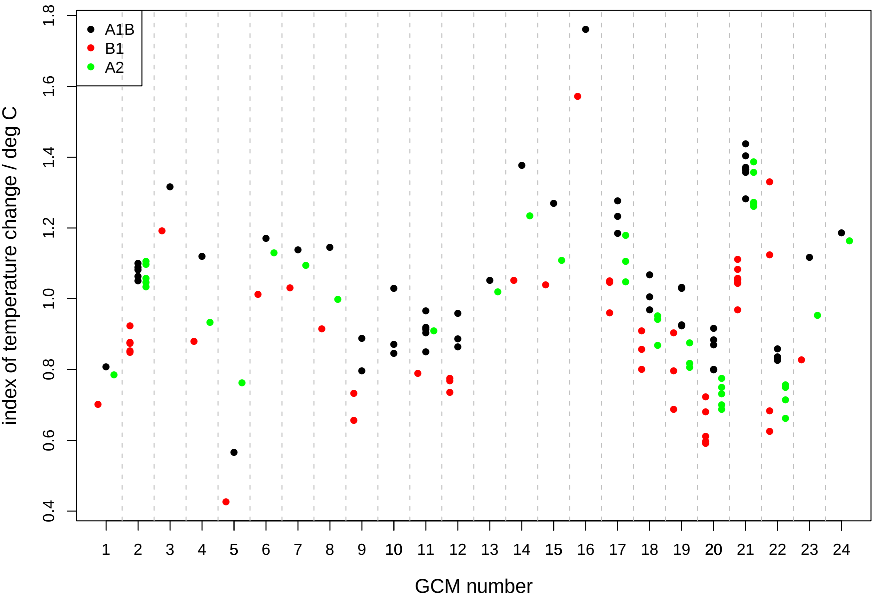
- **GCM,**
- socio-economic **scenario,** or
- GCM **run?**

...or a mixture of these?
...does it matter how far into the future we want to look?
...does the climate variable matter?
...does the region of the world matter?

Simple measures of climate change, e.g. global temperature

1. average change in **2020–2049** (relative to 1980–1999)
2. average change in **2069–2098** (relative to 1980–1999)

Global temperature change 2020–2049



A 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	$\alpha_i \overset{iid}{\sim} N(0, \sigma_G)$
β_j	adjustment for scenario j	$\beta_j \overset{iid}{\sim} N(0, \sigma_S)$
γ_{ij}	scenario-specific adjustment for GCM i	$\gamma_{ij} \overset{iid}{\sim} N(0, \sigma_{GS})$
ϵ_{ijk}	residual effect of variability over runs	$\epsilon_{ijk} \overset{iid}{\sim} N(0, \sigma_R)$

- $\alpha_i \overset{iid}{\sim} N(0, \sigma_G)$ means $\alpha_1, \alpha_2, \ldots$ are independent and normally distributed with mean 0 and st. dev. σ_G.
- Imagine a population of GCMs,each producing a separate effect on Y_{ijk}.
- We assume that all random variables are independent.

Interpretation of parameters

μ	overall temperature change (headline value in newspaper)
σ_G	variability over GCMs
σ_S	variability over scenarios
σ_{GS}	variability of scenario-specific adjustment for GCM
σ_R	variability over runs

- Large value of $\sigma \Rightarrow$ variable makes a big difference to predictions of global temperature:
  e.g. if σ_G is large then the choice of GCM really matters.
- Large variability $\Rightarrow$ large uncertainty.

Statistical inference

Issues

1. Scenario has only 3 levels. Little 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 with weakly-informative priors.

Weakly-informative priors

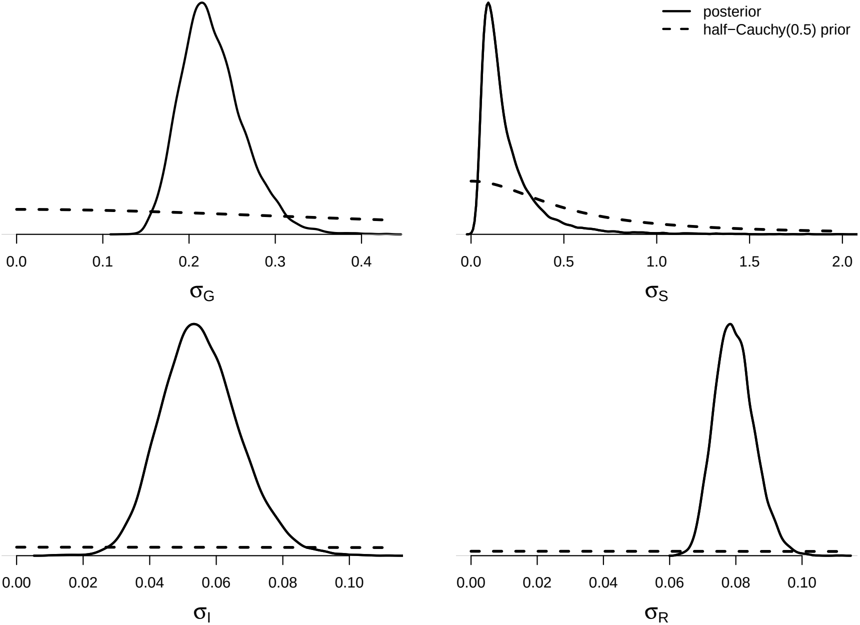
Use a prior distribution that is

- weakly-informative for σ_S, and
- uninformative for $\sigma_G, \sigma_{GS}, \sigma_R$ (and μ).

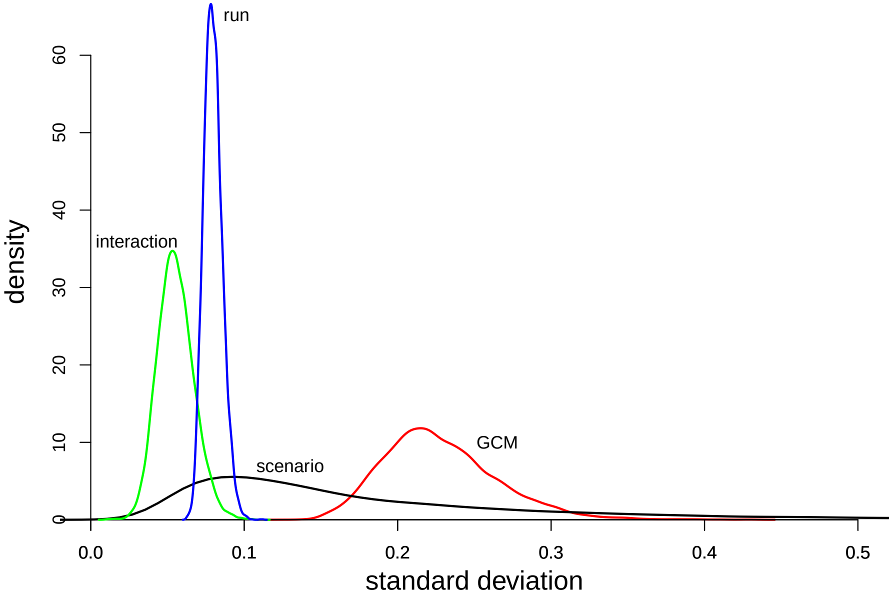
Idea

- downweight unrealistic values of σ_S - because the data aren't informative enough to discount these values;
- otherwise let the data speak for themselves;
- Gelman (2006) argues against improper uniform priors and the inverse-gamma family (improper posteriors can result and/or undue sensitivity to parameters of priors)
- ... and for a half-Cauchy prior ...

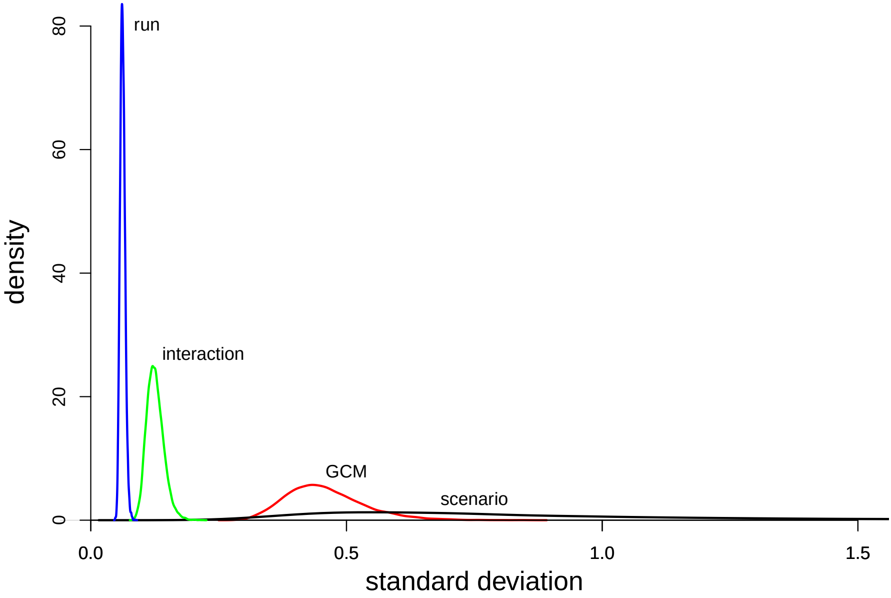
Global temperature 2020–2049



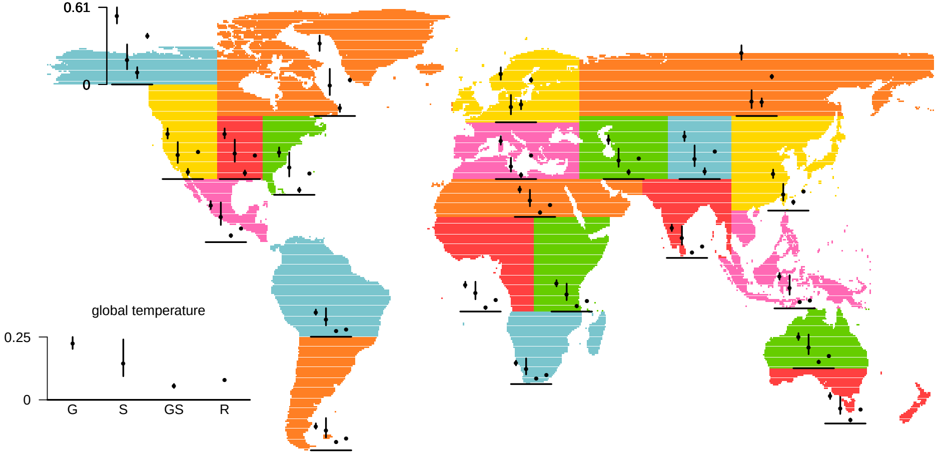
Global temp. 2020–2049: posterior distns



Global temp. 2069–2098: posterior distns

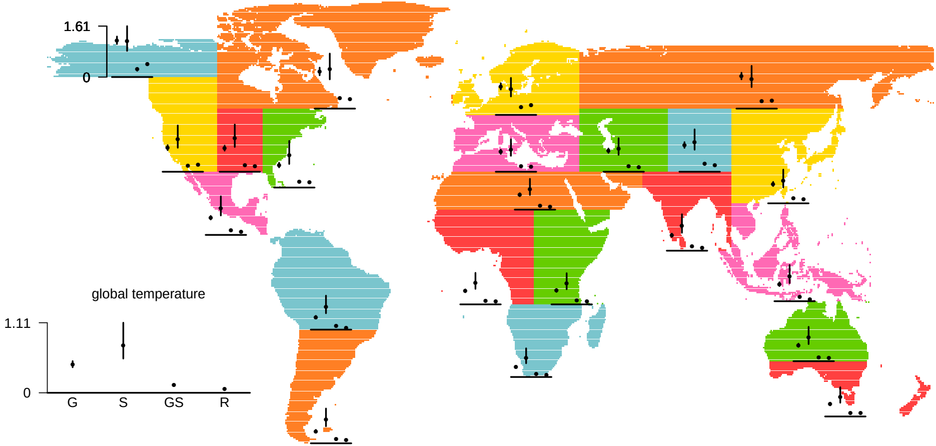


Regional temperature: 2020–2049



median (•) and line between quartiles

Regional temperature: 2069–2098



median (•) and line between quartiles

Summary for temperature

2020–2049

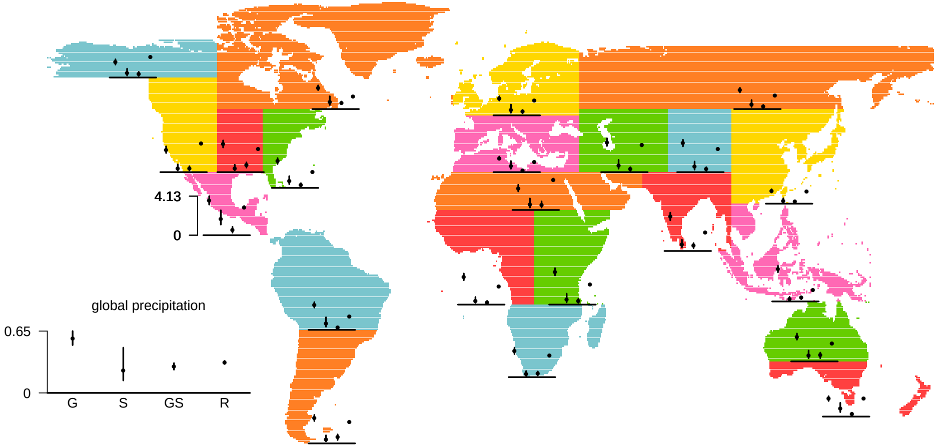
- Global : variability over GCMs $>$ scenario $>$ runs
- Regional: runs matters more than scenario in some areas, e.g. In the north

2069–2098

- Scenario matters more as we move through the 21st century (obviously!)
- Scenario is at least as important as GCM in most regions

Regional rainfall: 2020–2049

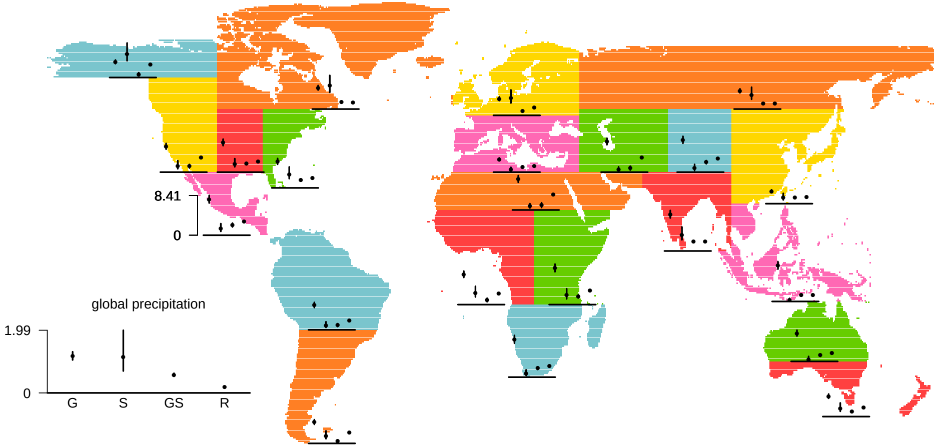
Percentage change in mean from 1980–1999



median (•) and line between quartiles

Regional rainfall: 2069–2098

Percentage change in mean from 1980–1999



median (•) and line between quartiles

Summary for rainfall

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

Concluding remarks

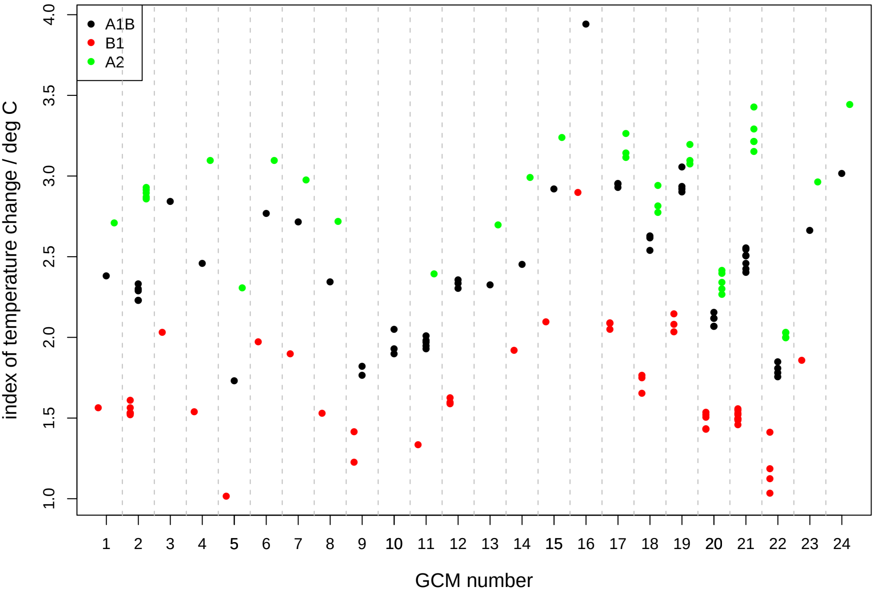
1. Climate uncertainty depends on
   - the **variable** of interest;
   - the **region** of the world;
   - the **time horizon**.
2. Simple statistical models are useful.
3. Scope to improve design of climate experiments.

References

- Gelman, A. (2006) Bayesian Analysis, 1(3), 515–533
- Gilmour, S.G. and Goos, P. (2009) Appl. Statist. 58(4), 467–484
- 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 your attention

Global temperature change 2069–2098



Experimental design (2-way RE ANOVA)

- Review of design for variance components estimation: Khuri, A.I. (2000) Inst. Statist. Rev., 68(3),311–322
- 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 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.