Quantifying sources of uncertainty in projections of future climate

Paul Northrop and Richard Chandler
Department of Statistical Science, University College London
p.northrop@ucl.ac.uk, r.chandler@ucl.ac.uk

UCL

Abstract

Projections of future climate are often based on General Circulation Models (GCMs), deterministic models of the earth's atmosphere and oceans. Projections can vary greatly between GCMs, the greenhouse gas emissions scenario envisaged for the future, and model simulations based on different initial conditions. See Chandler et al. (2010) for background information. We use a Bayesian analysis of a two-way random effects ANOVA model to assess which of these three sources of uncertainty are of greatest importance, and the extent to which this depends on climate variable, region of the world and time horizon. Given that projections are only available under 3 scenarios a weakly-informative half-Cauchy prior distribution is used to downweight the posterior probability of physically implausible values.

Sources of climate uncertainty

Is variability in a projected climate variable due mainly to choice of

- General Circulaton Model **GCM** (climate simulator),
- future greenhouse gas emissions **scenario,** or
- GCM **run** (simulation number)?

…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?

21st century climate projections

We use data from the World Climate Research Programme's (WCRP's) Coupled Model Intercomparison Project phase 3 (CMIP3) multi-model dataset (Meehl et al., 2007).

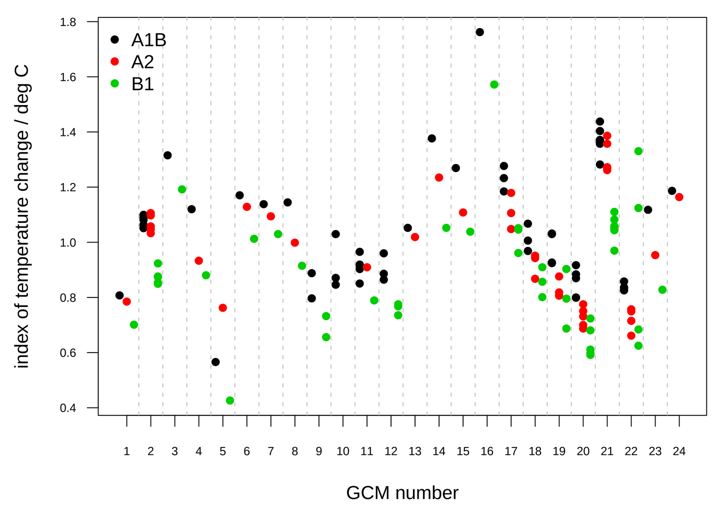
		scenario		
GCM number	GCM name	A1B	A2	B1
1	bccr:bcm2:0	1	1	1
2	cccma:cgcm3:1	5	5	5
3	cccma:cgcm3:1:t63	1	0	1
4	cnrm:cm3	1	1	1
5	csiro:mk3:0	1	1	1
6	csiro:mk3:5	1	1	1
7	gfdl:cm2:0	1	1	1
8	gfdl:cm2:1	1	1	1
9	giss:aom	2	0	2
10	giss:model:e:h	3	0	0
11	giss:model:e:r	5	1	1
12	iap:fgoals1:0:g	3	0	3
13	ingv:echam4	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

Numbers of runs for each combination of 24 GCMs and three socio-economic scenarios (A1B, A2, B1) for the climate experiments in the CMIP3 archive.

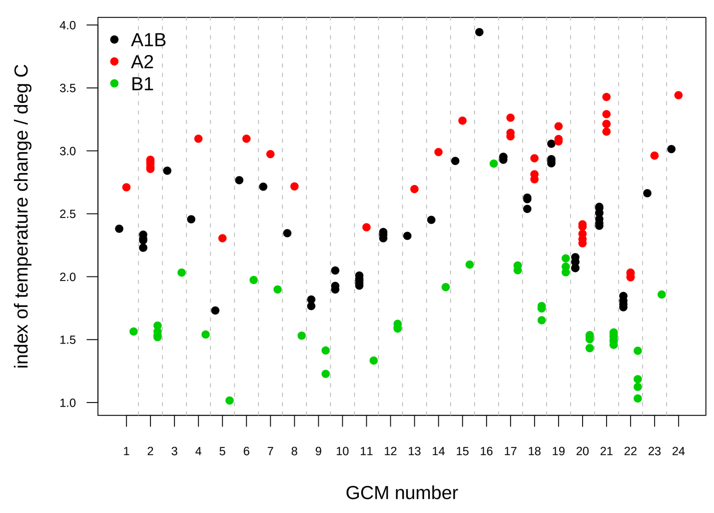
- No design ("ensemble of opportunity");
- Lack of balance.
- Zero cells.
- Scenario has only 3 levels.
- Each run takes approx. 1 month.

Simple indices of climate change

1. **2020-2049** mean − 1980-1999 mean.
2. **2069-2098** mean − 1980-1999 mean.



Global temperature change: 2020–2049



Global temperature change: 2068–2098

Related work

Yip et al. (2011) create balance, by using data only from the 7 GCMs that have multiple for each scenario, followed by a classical ANOVA decomposition of variability. We seek to avoid discarding data using a Bayesian analysis of a two-way random effects ANOVA model.

A two-way random effects ANOVA

Let 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	$\alpha_i \overset{\text{i.i.d.}}{\sim} N(0, \sigma_G^2)$
β_j	adjustment for scenario j	$\beta_j \overset{\text{i.i.d.}}{\sim} N(0, \sigma_S^2)$
γ_{ij}	scenario-specific adjustment for GCM i	$\gamma_{ij} \overset{\text{i.i.d.}}{\sim} N(0, \sigma_{GS}^2)$
ϵ_{ijk}	residual effect of variability over runs	$\epsilon_{ijk} \overset{\text{i.i.d.}}{\sim} N(0, \sigma_R^2)$

We assume that all random variables are independent.

View the GCMs and scenarios as random samples from notional super-populations of GCMs and scenarios.

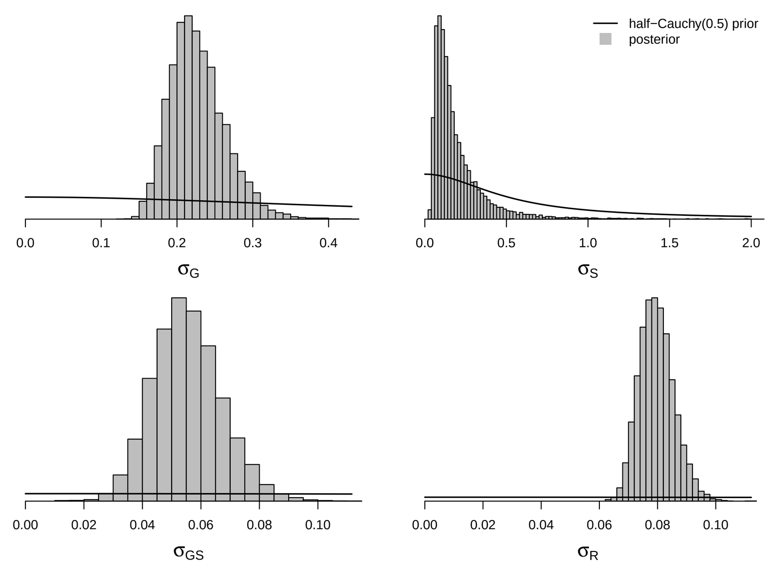
Inference

- Little information about variability over scenarios σ_S (only 3 levels). Data highly unbalanced.
- REML tends to underestimate σ_S; $\widehat{\sigma}_S = 0$ is common (Gilmour and Goos, 2009).
- Bayesian inference with a weakly-informative prior for σ_S (Gelman, 2006):
  - $N(0, 10^6)$ prior for μ; independent half-Cauchy(A) priors for the super-population SDs:

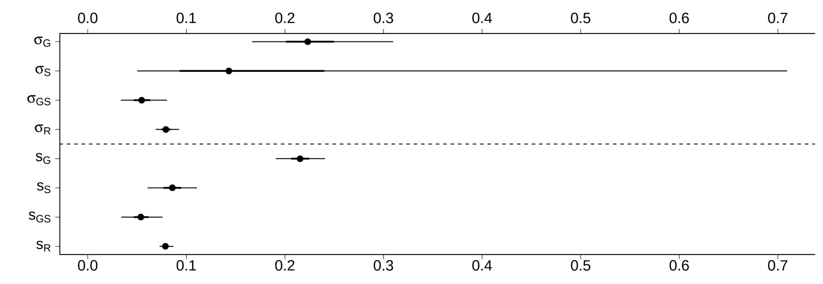
$$\pi(\sigma) = \frac{2}{\pi A}\left(1 + \frac{\sigma^2}{A^2}\right)^{-1}, \qquad \sigma > 0;$$

  - A chosen to downweight unrealistic values of σ_S, e.g. for 2020–2049 consider $4\sigma_S = 10°C$ to be very unlikely.
  - Gelman (2006) argues against an improper uniform prior for σ_S (posterior unrealistically broad) and inverse-gamma(ϵ, ϵ) (posterior sensitive to ϵ).
  - Also look at finite-population SDs s_G, s_S, s_{GS} and s_R, e.g. $s_G^2 = (1/23)\sum_{i=1}^{24}(\alpha_i - \bar{\alpha})^2$.
  - Use R package `arm` (Gelman et al., 2010) (calls winBUGS) to perform MCMC.

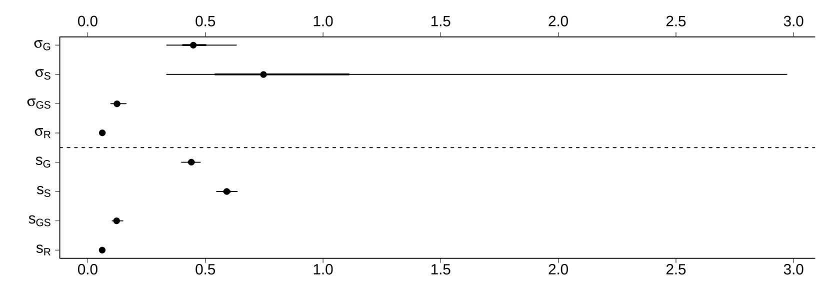
Global temperature change



2020–2049. Prior densities and posterior samples.



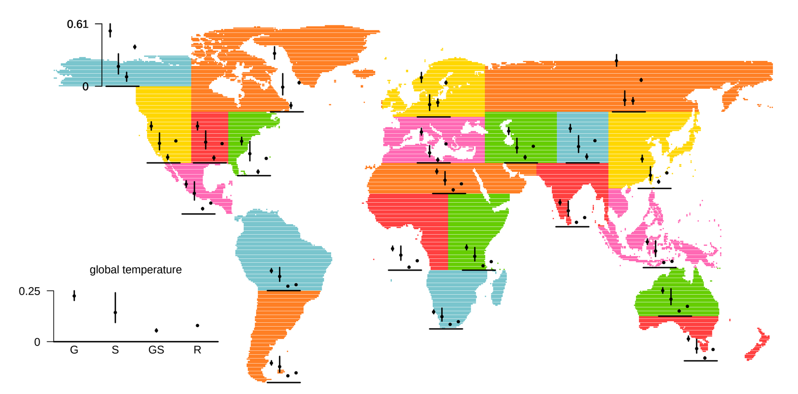
2020–2049. A=0.5. Posterior medians, 50% and 95% CIs



2069–2098. A=1. Posterior medians, 50% and 95% CIs

- Choice of GCM matters more than choice of GCM run;
- Choice of scenario matters more later in the century;
- Posterior median of σ_S increases (slowly) with A (and posterior becomes increasingly heavy-tailed). Other posteriors are insensitive to A.

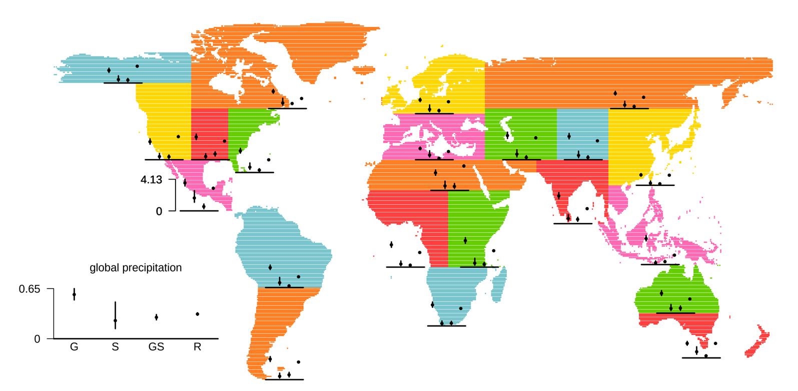
Regional temperature change 2020–2049



Posterior medians and 50% CIs of super-population SDs.

- Choice of GCM matter most;
- Choice of GCM run matters more than scenario in some (northern) regions;
- Large uncertainty about variability σ_S of scenario super-population;
- (As expected) scenario matters more for the late-century index (plot not shown).

Regional precipitation change (% change from 1980–1999 mean) 2020–2049



Posterior medians and 50% CIs of super-population SDs.

- Choice of GCM run matters more than for temperature. …and more than choice of GCM in Alaska;
- In many regions choice of scenario is relatively unimportant (even in late-century plot, not shown).

Remarks

1. Climate uncertainty depends on: **climate variable**; **region**; **time horizon**.
2. Scope to improve design of climate experiments.
3. Bayesian analysis copes with non-orthogonal design and a factor with a small number of levels.

References

Chandler, R., J. Rougier, and M. Collins (2010). Climate change: making certain what the uncertainties are. *Significance 7*(1), 9–12.

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis 1*(3), 515–533.

Gelman, A., Y.-S. Su, M. Yajima, J. Hill, M. G. Pittau, J. Kerman, and T. Zheng (2010). *arm: Data Analysis Using Regression and Multilevel/Hierarchical Models.* R package version 1.3-02.

Gilmour, S. G. and P. Goos (2009). Analysis of data from non-orthogonal multistratum designs in industrial experiments. *Appl. Statist. 58*(4), 567–484.

Meehl, G. A., C. Covey, T. Delworth, M. Latif, B. McAvaney, J. F. B. Mitchell, R. J. Stouffer, and K. E. Taylor (2007). The WCRP CMIP3 multi-model dataset: A new era in climate change research. *Bulletin of the American Meteorological Society 88*, 1383–1394.

Yip, S., C. A. T. Ferro, D. B. Stephenson, and E. Hawkins (2011). A simple, coherent framework for partitioning uncertainty in climate predictions. *Journal of Climate 24*(17), 4634–4643.