Exploiting strength, discounting weakness: combining information from multiple climate simulators

Richard E. Chandler[§] Bryson Bates[†]

[§]Department of Statistical Science, University College London
(richard@stats.ucl.ac.uk)

[†]CSIRO Climate Adaptation Flagship
(bryson.bates@csiro.au)

Greenhouse 2011, Cairns, April 2011

Motivation

- Climate change impacts studies / adaptation strategies informed by projections of future climate
- Projections vary between climate simulators $\Rightarrow$ choice of simulator (particularly GCM) often represents significant source of uncertainty
- Users advised to consider information from “ensemble” of simulators — multimodel ensemble (MME)
- Questions:
    - How to use / combine information from multiple simulators?
    - How to provide decision-relevant uncertainty assessment?

Rationale for approach taken here

- Aim for generic conceptual representation of MME structure
- Representation should be transparent and intuitive
- Want framework that is computationally tractable
- Want to learn how to combine information (avoid heuristics)
- Want ability to make probability statements about reality

A new(-ish) conceptual framework

Reference:

Leith & Chandler, *J. Roy. Statist. Soc. Series C*, 2010

- Motivation: projections from climate simulators aim to reproduce statistical properties of actual climate

A new(-ish) conceptual framework

Reference:

Leith & Chandler, *J. Roy. Statist. Soc. Series C*, 2010

- Motivation: projections from climate simulators aim to reproduce statistical properties of actual climate
- Interpretation: outputs from different simulators should have similar structure, hence can be described using same form of statistical model — but with different parameters

A new(-ish) conceptual framework

Reference:

Leith & Chandler, *J. Roy. Statist. Soc. Series C*, 2010

- Motivation: projections from climate simulators aim to reproduce statistical properties of actual climate
- Interpretation: outputs from different simulators should have similar structure, hence can be described using same form of statistical model — but with different parameters
- Statistical model is emulator for system and simulators

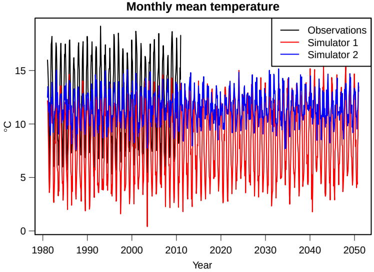
A new(-ish) conceptual framework

Reference:

Leith & Chandler, *J. Roy. Statist. Soc. Series C*, 2010

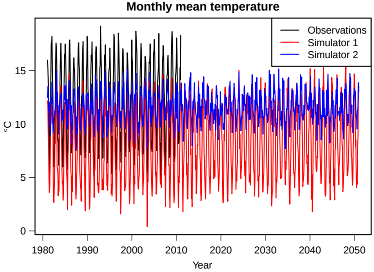
- Motivation: projections from climate simulators aim to reproduce statistical properties of actual climate
- Interpretation: outputs from different simulators should have similar structure, hence can be described using same form of statistical model — but with different parameters
- Statistical model is emulator for system and simulators
- Emulator parameters are descriptors of system / simulator outputs

Emulation in a toy example



- Two simulators, monthly mean temperature
- Simulator 1: reasonable mean temp, hopeless seasonality
- Simulator 2: vice versa

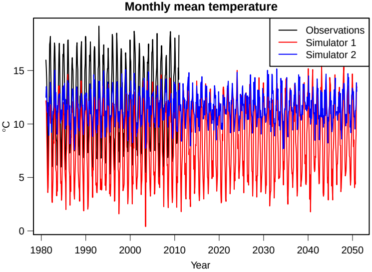
Emulation in a toy example



- Two simulators, monthly mean temperature
- Simulator 1: reasonable mean temp, hopeless seasonality
- Simulator 2: vice versa

- Possible emulator: $Y_t = \beta_0 + \beta_1 \cos \frac{2\pi t}{365} + \beta_2 \sin \frac{2\pi t}{365} + \beta_3 t + \varepsilon_t$, $\varepsilon_t \sim N(0, \sigma^2)$.

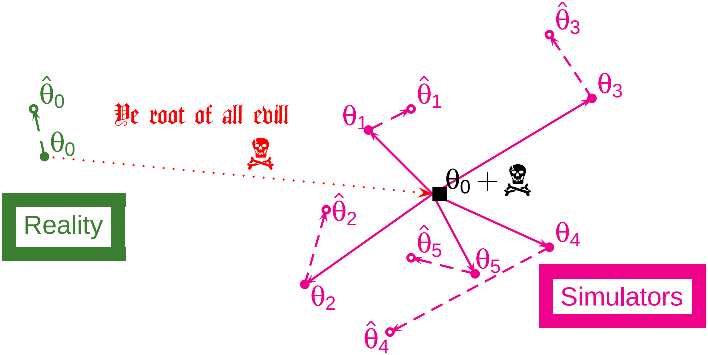
Emulation in a toy example



- Two simulators, monthly mean temperature
- Simulator 1: reasonable mean temp, hopeless seasonality
- Simulator 2: vice versa

- Possible emulator: $Y_t = \beta_0 + \beta_1 \cos \frac{2\pi t}{365} + \beta_2 \sin \frac{2\pi t}{365} + \beta_3 t + \varepsilon_t$, $\varepsilon_t \sim N(0, \sigma^2)$.
- Descriptor vector: $\theta = (\beta_0, \beta_1, \beta_2, \beta_3, \sigma^2)$
- Can use data from each source to estimate corresponding θ e.g. using maximum likelihood — estimator is $\hat{\theta}$, say

Suggested schematic of problem structure



Given m simulators and one set of climate observations, want to use MLEs $\hat{\theta}_0, \hat{\theta}_1, \ldots, \hat{\theta}_m$ to learn about θ_0

Learning about θ_0

- Convenient to take Bayesian approach initially: want posterior $\pi\left(\theta_0 | \hat{\theta}_0, \hat{\theta}_1, \ldots, \hat{\theta}_m\right)$
- **NB** optimal (minimum MSE) estimate of θ_0 is posterior mean: $\mathrm{E}\left(\theta_0 | \hat{\theta}_0, \hat{\theta}_1, \ldots, \hat{\theta}_m\right) = \tau$ say.

Learning about θ_0

- Convenient to take Bayesian approach initially: want posterior $\pi\left(\theta_0|\hat{\theta}_0, \hat{\theta}_1, \ldots, \hat{\theta}_m\right)$
- **NB** optimal (minimum MSE) estimate of θ_0 is posterior mean: $E\left(\theta_0|\hat{\theta}_0, \hat{\theta}_1, \ldots, \hat{\theta}_m\right) = \tau$ say.

Result

Posterior = Prior × Likelihood for observations × Likelihood for simulator outputs

The Gaussian case

- Simplest case: everything Gaussian:

$$\hat{\theta}_i \sim \mathrm{MVN}(\theta_i, \mathbf{J}_i) \quad i = 0, \ldots, m$$

$$\theta_i = \theta_0 + \delta_i \ , \ \delta_i \sim \mathrm{MVN}\left(\text{☠}, \mathbf{C}_i\right) \qquad i = 1, \ldots, m$$

$$\theta_0 \sim \mathrm{MVN}(\mu_0, \Sigma_0) \ , \ \text{☠} \sim \mathrm{MVN}(\mathbf{0}, \Lambda)$$

$$\{\delta_i\} \text{ mutually independent given } \text{☠}$$

The Gaussian case

- Simplest case: everything Gaussian:

$$\hat{\theta}_i \sim \mathrm{MVN}(\theta_i, \mathbf{J}_i) \quad i = 0, \ldots, m$$

$$\theta_i = \theta_0 + \delta_i \;,\; \delta_i \sim \mathrm{MVN}\left(\skull, \mathbf{C}_i\right) \quad i = 1, \ldots, m$$

$$\theta_0 \sim \mathrm{MVN}(\mu_0, \Sigma_0) \;,\; \skull \sim \mathrm{MVN}(\mathbf{0}, \Lambda)$$

$$\{\delta_i\} \text{ mutually independent given } \skull$$

- $\{\mathbf{J}_i\}$ measures internal variability from each data source
- $\mathbf{C}_i$ measures (lack of) consensus for simulator i
- Λ measures collective tendency of simulators to deviate from reality

Comments on simplification

- Gaussian assumptions not critical — subsequent calculations lead to optimal linear combination of information even in non-Gaussian settings
- Mutual independence of $\{\delta_i\}$: probably unrealistic (simulators come in “families”) but easily handled within general framework

Gaussian case: the posterior

- Posterior turns out to be $\mathrm{MVN}(\tau, \mathbf{S})$, where:

$$\mathbf{S}^{-1} = \Sigma_0^{-1} + \mathbf{J}_0^{-1} + \sum_{i=1}^{m} \mathbf{W}_i^{-1} \; ;$$

$$\tau = \mathbf{S}\left[\Sigma_0^{-1}\mu_0 + \mathbf{J}_0^{-1}\hat{\theta}_0 + \sum_{i=1}^{m} \mathbf{W}_i^{-1}\hat{\theta}_i\right] \; ;$$

$$\mathbf{W}_i = \mathbf{D}_i\left(\mathbf{I} + \sum_{k=1}^{m} \mathbf{D}_k^{-1}\Lambda\right) \; ; \text{ and } \quad \mathbf{D}_i = \mathbf{J}_i + \mathbf{C}_i$$

- Colour coding: prior, observations, simulators.

Making sense of the formula

- Posterior mean is $\tau = \mathbf{S}\left[\Sigma_0^{-1}\mu_0 + \mathbf{J}_0^{-1}\hat{\theta}_0 + \sum_{i=1}^{m}\mathbf{W}_i^{-1}\hat{\theta}_i\right]$ with $\mathbf{S}^{-1} = \Sigma_0^{-1} + \mathbf{J}_0^{-1} + \sum_{i=1}^{m}\mathbf{W}_i^{-1}$

Making sense of the formula

- Posterior mean is $\tau = \mathbf{S}\left[\Sigma_0^{-1}\mu_0 + \mathbf{J}_0^{-1}\hat{\theta}_0 + \sum_{i=1}^{m}\mathbf{W}_i^{-1}\hat{\theta}_i\right]$ with $\mathbf{S}^{-1} = \Sigma_0^{-1} + \mathbf{J}_0^{-1} + \sum_{i=1}^{m}\mathbf{W}_i^{-1}$
- Simple example:
  - Suppose θ has one component $\Rightarrow$ all quantities are scalars
  - Suppose $m = 2$, $\Sigma_0^{-1} = 0.1$, $\mathbf{J}_0^{-1} = 10$, $\mathbf{W}_1^{-1} = 6$, $\mathbf{W}_2^{-1} = 7$
  - Then $\tau = \left(0.1\mu_0 + 10\hat{\theta}_0 + 6\hat{\theta}_1 + 7\hat{\theta}_2\right) \div (0.1 + 10 + 6 + 7)$ i.e. weighted average of μ_0, $\hat{\theta}_0$, $\hat{\theta}_1$ & $\hat{\theta}_2$

Making sense of the formula

- Posterior mean is $\tau = \mathbf{S}\left[\Sigma_0^{-1}\mu_0 + \mathbf{J}_0^{-1}\hat{\theta}_0 + \sum_{i=1}^{m}\mathbf{W}_i^{-1}\hat{\theta}_i\right]$ with $\mathbf{S}^{-1} = \Sigma_0^{-1} + \mathbf{J}_0^{-1} + \sum_{i=1}^{m}\mathbf{W}_i^{-1}$
- Simple example:
  - Suppose θ has one component $\Rightarrow$ all quantities are scalars
  - Suppose $m = 2$, $\Sigma_0^{-1} = 0.1$, $\mathbf{J}_0^{-1} = 10$, $\mathbf{W}_1^{-1} = 6$, $\mathbf{W}_2^{-1} = 7$
  - Then $\tau = \left(0.1\mu_0 + 10\hat{\theta}_0 + 6\hat{\theta}_1 + 7\hat{\theta}_2\right) \div (0.1 + 10 + 6 + 7)$ i.e. weighted average of μ_0, $\hat{\theta}_0$, $\hat{\theta}_1$ & $\hat{\theta}_2$
- For vector θ, τ is 'matrix-weighted average' of prior mean, estimate from actual climate and estimates from simulators

Some consequences: simulator weighting

- Matrix-valued weights mean different weightings for different components of θ: simulators contribute where informative, not elsewhere ('exploit strength, discount weakness')

Some consequences: simulator weighting

- Matrix-valued weights mean different weightings for different components of θ: simulators contribute where informative, not elsewhere ('exploit strength, discount weakness')
- Weights emerge indirectly from judgments about / estimates of internal variability ($\{\mathbf{J}_i\}$), consensus ($\{\mathbf{C}_i\}$) and ☠ (Λ)

Some consequences: simulator weighting

- Matrix-valued weights mean different weightings for different components of θ: simulators contribute where informative, not elsewhere ('exploit strength, discount weakness')
- Weights emerge indirectly from judgments about / estimates of internal variability ($\{\mathbf{J}_i\}$), consensus ($\{\mathbf{C}_i\}$) and ☠ (Λ)
- **NB** weights are related to *expected* rather than *observed* variation (surprising?)
  - If $\{\mathbf{D}_i\}$ are all equal, weights are the same for all simulators

Value of information

- Posterior precision is $\mathbf{S}^{-1} = \Sigma_0^{-1} + \mathbf{J}_0^{-1} + \left[\Lambda + \left(\sum_{k=1}^{m} \mathbf{D}_k^{-1}\right)^{-1}\right]^{-1}$

Value of information

- Posterior precision is $\mathbf{S}^{-1} = \Sigma_0^{-1} + \mathbf{J}_0^{-1} + \left[\Lambda + \left(\sum_{k=1}^{m} \mathbf{D}_k^{-1}\right)^{-1}\right]^{-1}$
- Increasing the number of (relevant) observations: as $\mathbf{J}_0 \to \mathbf{0}$, precision tends to ∞ & simulators / prior become irrelevant

Value of information

- Posterior precision is $\mathbf{S}^{-1} = \Sigma_0^{-1} + \mathbf{J}_0^{-1} + \left[\Lambda + \left(\sum_{k=1}^{m} \mathbf{D}_k^{-1}\right)^{-1}\right]^{-1}$
- Increasing the number of (relevant) observations: as $\mathbf{J}_0 \to \mathbf{0}$, precision tends to ∞ & simulators / prior become irrelevant
- Increasing the number of simulators: as $m \to \infty$, final contribution cannot increase beyond $\Lambda^{-1} \Rightarrow$ precision limited by ☠.

Value of information

- Posterior precision is $\mathbf{S}^{-1} = \Sigma_0^{-1} + \mathbf{J}_0^{-1} + \left[\Lambda + \left(\sum_{k=1}^{m} \mathbf{D}_k^{-1}\right)^{-1}\right]^{-1}$
- Increasing the number of (relevant) observations: as $\mathbf{J}_0 \to \mathbf{0}$, precision tends to ∞ & simulators / prior become irrelevant
- Increasing the number of simulators: as $m \to \infty$, final contribution cannot increase beyond Λ^{-1} $\Rightarrow$ precision limited by ☠.
- Increasing length of simulator runs or simulator consensus: again, final contribution cannot increase beyond Λ^{-1}

Implementation

- In practice, need to estimate $\{\mathbf{J}_i\}, \{\mathbf{C}_i\}, \Lambda$ — need Markov Chain Monte Carlo (MCMC) methods to handle estimation uncertainty
  - MCMC is computationally intensive ☹ . . .
  - . . . and often needs high levels of statistical literacy ☹

Implementation

- In practice, need to estimate $\{\mathbf{J}_i\}, \{\mathbf{C}_i\}, \Lambda$ — need Markov Chain Monte Carlo (MCMC) methods to handle estimation uncertainty
  - MCMC is computationally intensive ☹ . . .
  - . . . and often needs high levels of statistical literacy ☹
- Cheap alternative: plug simple estimates into earlier formulae
  - **NB** ignores estimation uncertainty — could be large
  - But probably better than much current practice
  - Estimates of $\{\mathbf{J}_i\}$: directly from likelihood-based fitting
  - Take $\mathbf{C}_i = \mathbf{C}$ for each simulator and estimate as sample covariance matrix of $\{\hat{\theta}_k : k = 1, \ldots, m\}$
  - Estimation of Λ depends on context, creativity needed . . .

Estimating Λ, one possibility

- Suppose descriptor is $\theta = (\psi'_{\text{historical}}\ \psi'_{\text{future}})'$ for some vector ψ

Estimating Λ, one possibility

- Suppose descriptor is $\theta = (\psi'_{\text{historical}}\ \psi'_{\text{future}})'$ for some vector ψ
- Can estimate $\skull_{\text{historic}}$ from difference between descriptor for observations and historical simulator consensus

Estimating Λ, one possibility

- Suppose descriptor is $\theta = (\psi'_{\text{historical}}\ \psi'_{\text{future}})'$ for some vector ψ
- Can estimate $☠_{\text{historic}}$ from difference between descriptor for observations and historical simulator consensus
- Then, since $\Lambda = \mathrm{E}\left(☠☠'\right)$, set $\hat{\Lambda}_{\text{historic}} = \hat{☠}_{\text{historic}}\hat{☠}'_{\text{historic}}$

Estimating Λ, one possibility

- Suppose descriptor is $\theta = (\psi'_{\text{historical}} \ \psi'_{\text{future}})'$ for some vector ψ
- Can estimate $☠_{\text{historic}}$ from difference between descriptor for observations and historical simulator consensus
- Then, since $\Lambda = \mathrm{E}\left(☠☠'\right)$, set $\hat{\Lambda}_{\text{historic}} = \hat{☠}_{\text{historic}}\hat{☠}'_{\text{historic}}$
- Judgements needed for remaining structure.

Estimating Λ, one possibility

- Suppose descriptor is $\theta = (\psi'_{\text{historical}}\ \psi'_{\text{future}})'$ for some vector ψ
- Can estimate $\skull_{\text{historic}}$ from difference between descriptor for observations and historical simulator consensus
- Then, since $\Lambda = \mathrm{E}\left(\skull\skull'\right)$, set $\hat{\Lambda}_{\text{historic}} = \hat{\skull}_{\text{historic}}\hat{\skull}'_{\text{historic}}$
- Judgements needed for remaining structure.

Example

To represent judgement “collective discrepancy will remain unchanged in future”, set

$$\hat{\Lambda} = \begin{pmatrix} \hat{\Lambda}_{\text{historic}} & \hat{\Lambda}_{\text{historic}} \\ \hat{\Lambda}_{\text{historic}} & \hat{\Lambda}_{\text{historic}} \end{pmatrix}$$

Toy example again

- Use earlier emulator

$$Y_t = \beta_0 + \beta_1 \cos\frac{2\pi t}{365} + \beta_2 \sin\frac{2\pi t}{365} + \beta_3 t + \varepsilon_t \quad , \varepsilon_t \sim N(0, \sigma^2)$$

fitted separately for each data source and time period.

Toy example again

- Use earlier emulator

$$Y_t = \beta_0 + \beta_1 \cos \frac{2\pi t}{365} + \beta_2 \sin \frac{2\pi t}{365} + \beta_3 t + \varepsilon_t \quad , \varepsilon_t \sim N(0, \sigma^2)$$

  fitted separately for each data source and time period.
- Assert complete prior ignorance: $\mu_0 = \mathbf{0}, \Sigma_0^{-1} = \mathbf{0}$ (could think about this more carefully for real applications)
- Treat simulators democratically: $\mathbf{C}_1 = \mathbf{C}_2 = \mathbf{C}$

Toy example: results

	Historical					Future				
	β_0	β_1	β_2	β_3	σ^2	β_0	β_1	β_2	β_3	σ^2
$\hat{\theta}_0$	11.98	5.10	-0.03	0.01	1.00	—	—	—	—	—
$\hat{\theta}_1$	7.99	4.73	-0.34	0.01	1.00	7.57	4.66	-0.39	0.02	1.01
$\hat{\theta}_2$	11.56	1.65	0.11	0.00	0.87	11.52	1.49	0.10	0.00	0.89
τ										

Toy example: results

	Historical					Future				
	β_0	β_1	β_2	β_3	σ^2	β_0	β_1	β_2	β_3	σ^2
$\hat{\theta}_0$	11.98	5.10	-0.03	0.01	1.00	—	—	—	—	—
$\hat{\theta}_1$	7.99	4.73	-0.34	0.01	1.00	7.57	4.66	-0.39	0.02	1.01
$\hat{\theta}_2$	11.56	1.65	0.11	0.00	0.87	11.52	1.49	0.10	0.00	0.89
τ	11.98	5.10	-0.03	0.01	1.00	11.75	4.98	-0.06	0.01	1.02
SD	0.08	0.07	0.04	0.00	0.04	0.16	0.10	0.05	0.00	0.04

Summary

- Framework handles uncertainty in transparent, coherent & logically consistent way
- Brings clarity to debate about how to weight simulators
- Shows value of different information sources — implications for design of future MMEs
- Provides probability statements about future reality
- Exploits strength, discounts weakness

Summary

- Framework handles uncertainty in transparent, coherent & logically consistent way
- Brings clarity to debate about how to weight simulators
- Shows value of different information sources — implications for design of future MMEs
- Provides probability statements about future reality
- Exploits strength, discounts weakness

☺ Thank you for your attention ☺

More details: technical report 311 at
`http://www.ucl.ac.uk/statistics/research/reports`