Statistically-Indistinguishable Ensembles and the Evaluation of Climate Models

Corey Dethier

University of Notre Dame
Philosophy Department
corey.dethier@gmail.com

Feb 28, 2020

Intro

A problem

There are many different global climate models, and sometimes they don't agree.

A problem

There are many different global climate models, and sometimes they don't agree.

Example: global climate models deliver a range for "CO_2 sensitivity" of 2.1° C to 4.7° C (IPCC Working Group 1 2013, 817).

Seems to provide evidence that the true value is in this range.

The standing view

Both climate scientists and philosophers have registered skepticism.

- E.g.: Baumberger, Knutti, and Hadorn (2017), Justus (2012), Knutti, Allen, et al. (2008), Knutti, Furrer, et al. (2010), Parker (2011, 2018), Pirtle, Meyer, and Hamilton (2010), and Winsberg (2018)

The standing view

Both climate scientists and philosophers have registered skepticism.

- E.g.: Baumberger, Knutti, and Hadorn (2017), Justus (2012), Knutti, Allen, et al. (2008), Knutti, Furrer, et al. (2010), Parker (2011, 2018), Pirtle, Meyer, and Hamilton (2010), and Winsberg (2018)

The standard diagnosis: the group of models is a "ensemble of opportunity." Read: not like a random sample.

My thesis

I think there's a deeper problem.

My diagnosis: uncertainty about (constraints on) the space of possible models.

Recognizing this deeper problem helps us better understand and evaluate contemporary work within climate science.

Plan for the talk

1. (What's wrong with) The ensemble of opportunity diagnosis.
2. Understanding the statistically-indistinguishable paradigm.
3. Evaluating the statistically-indistinguishable paradigm.
4. Conclusion: "Are the models so out of touch? No, it's the meta-model that is wrong."

Ensembles of opportunity

How to draw conclusions of groups of models

Treat a group of models like a sample from a population—that is, use statistics.

How to draw conclusions of groups of models

Treat a group of models like a sample from a population—that is, use statistics.

The standard diagnosis: the method of construction of actual ensembles isn't like random sampling.

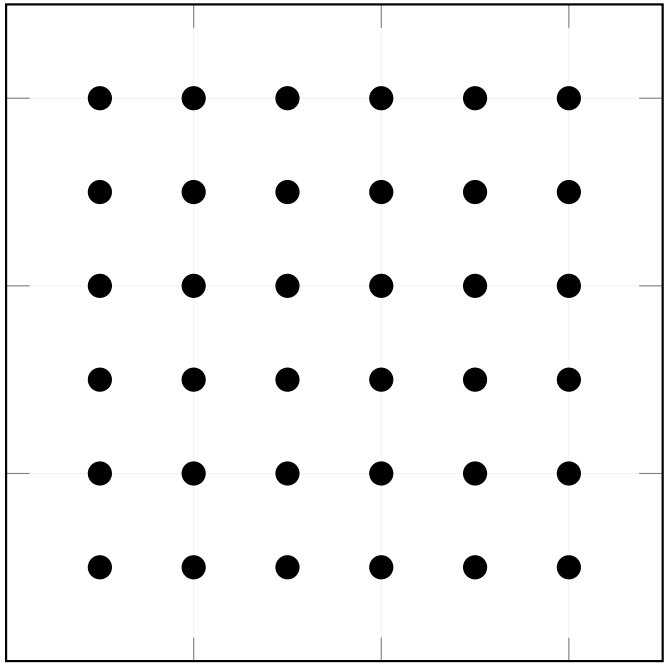
My diagnosis: there's uncertainty about the space of possible models.

A thorough method

Method 1: just build a model for every possibility.

Problems:

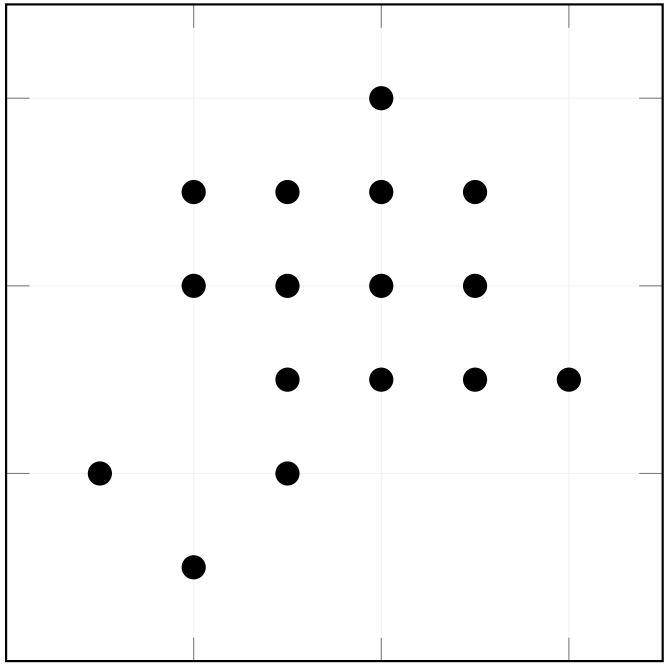
- Impractical.
- Only works if the possibilities are equally likely.



Independent sampling

Method 2: build models that are representative of each component taken independently.

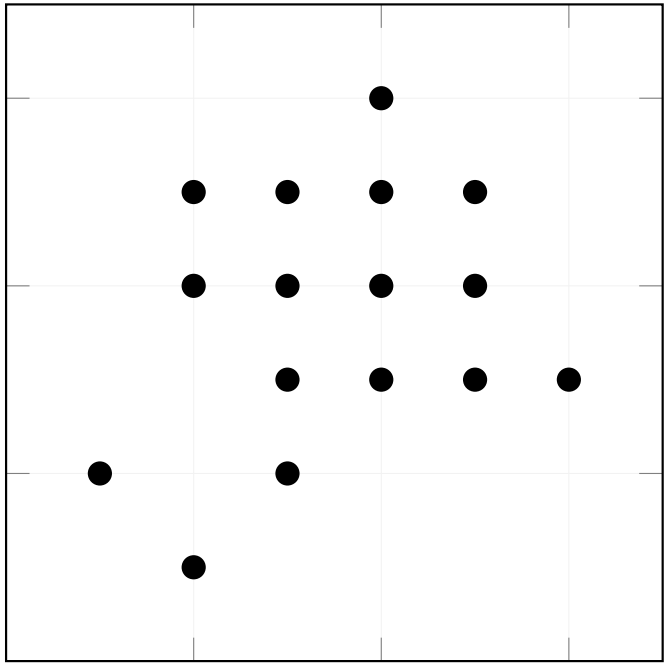
- Maybe what's intended by "principled."



Independent sampling

Method 2: build models that are representative of each component taken independently.

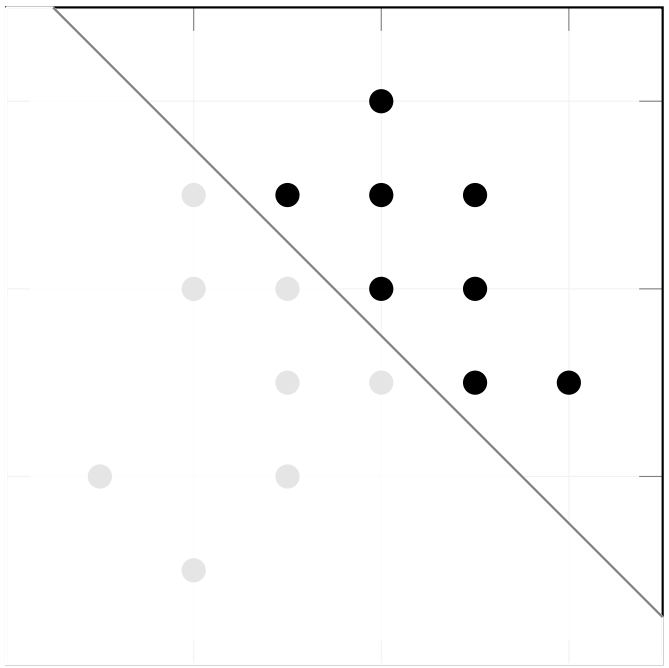
- Maybe what's intended by "principled."
- But only works if each component is independent.



Independent sampling

Method 2: build models that are representative of each component taken independently.

- Maybe what's intended by "principled."
- But only works if each component is independent.



The problem, then

Takeaway: in order to even say what a "principled" construction method is, we need background knowledge about the constraints on the set of models.

And that knowledge isn't being invoked in theoretical discussions of evaluation.

Understanding "statistically-indistinguishable" ensembles

Forgetting about construction

An alternative means of justifying inferences from a given ensemble: use proxies to check whether the ensemble is representative.

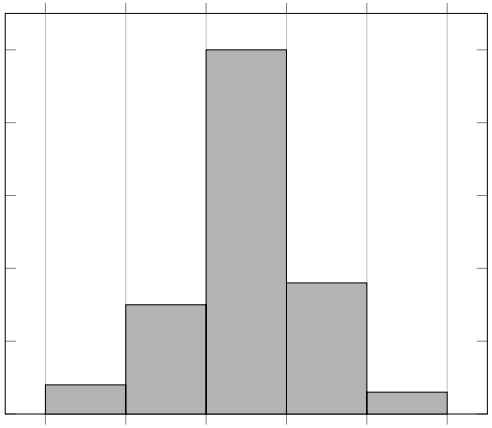
Forgetting about construction

An alternative means of justifying inferences from a given ensemble: use proxies to check whether the ensemble is representative.

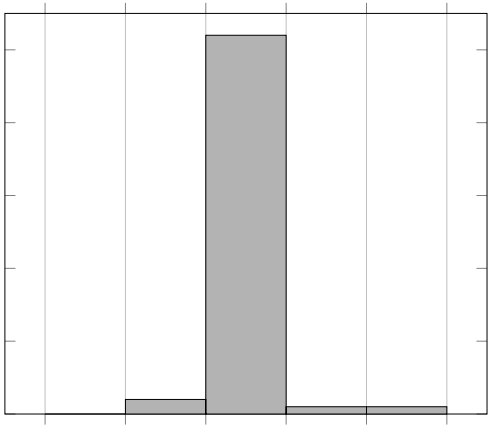
A different problem: proxies indicate that extant ensembles aren't representative.

First, the problem

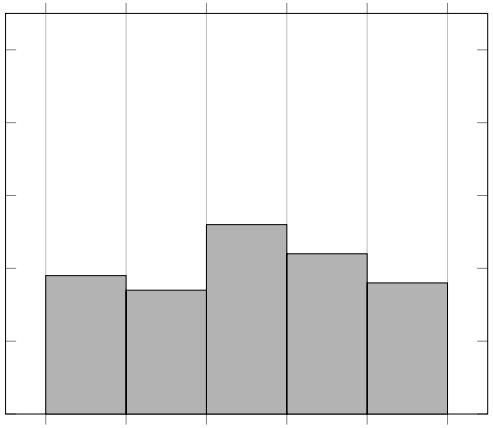
The problem, *very* roughly pictured:



(a) Ensemble is representative



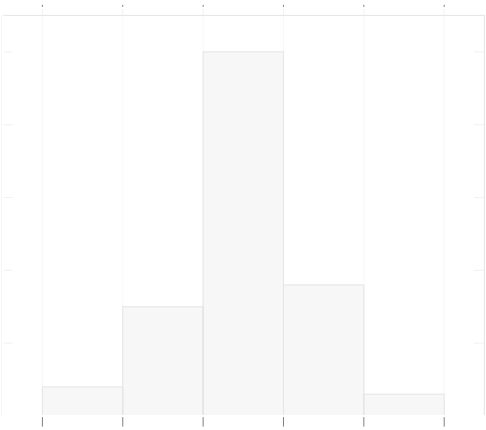
(b) Ensemble is too wide



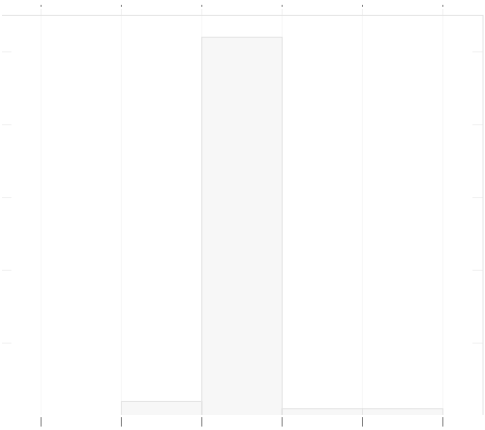
(c) Ensemble is too narrow

First, the problem

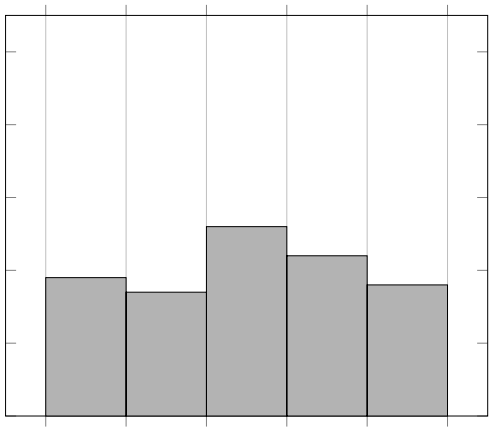
The problem, *very* roughly pictured:



(a) Ensemble is representative



(b) Ensemble is too wide



(c) Ensemble is too narrow

The solution

A number of climate scientists—most prominently Annan and Hargreaves (2010, 2011, 2017)—have argued that this result is misleading, because it relies on a particular statistical "paradigm."

The solution

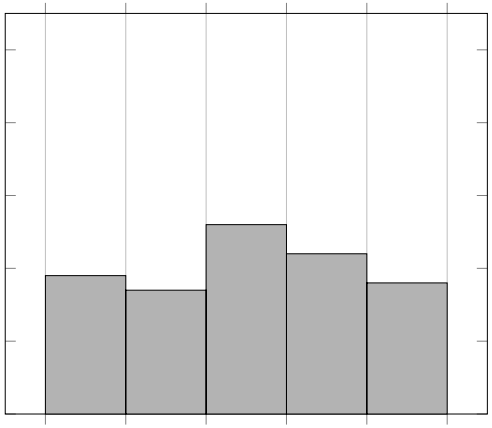
A number of climate scientists—most prominently Annan and Hargreaves (2010, 2011, 2017)—have argued that this result is misleading, because it relies on a particular statistical "paradigm."

"Truth-centered" paradigm: ensemble-proxy relationship is *like* that between a sample and a population *mean*.

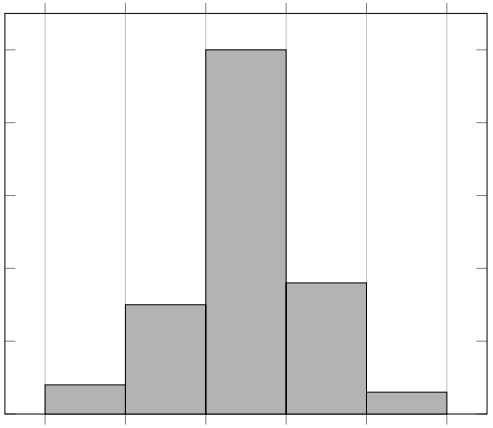
"Statistically indistinguishable" paradigm: ensemble-proxy relationship is *like* that between a sample and a population *member*.

The statistically-indistinguishable advantage

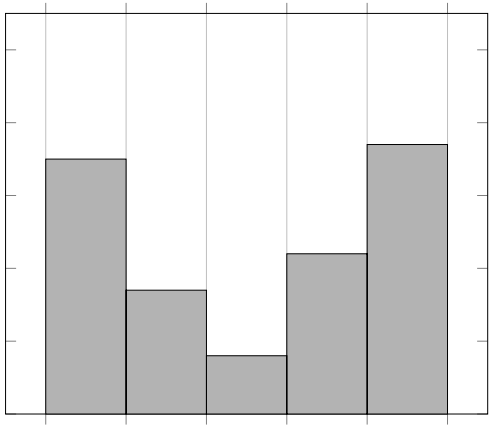
Given the SI paradigm:



(a) Ensemble is representative



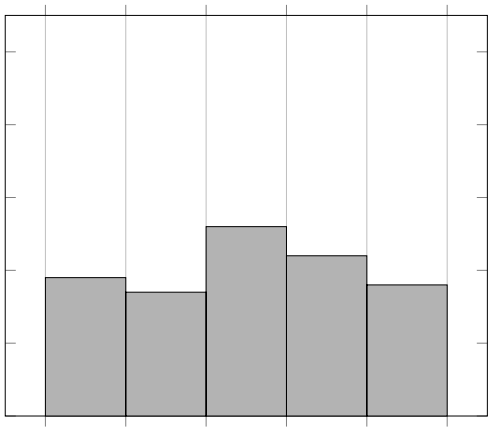
(b) Ensemble is too wide



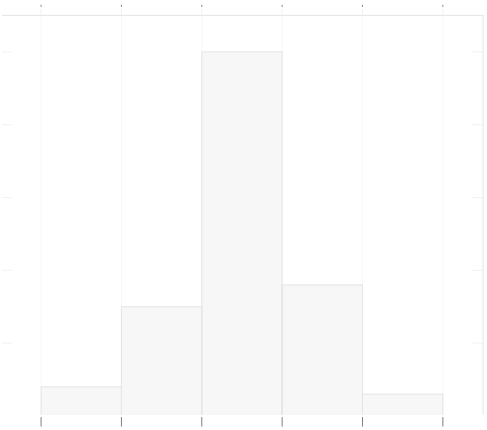
(c) Ensemble is too narrow

The statistically-indistinguishable advantage

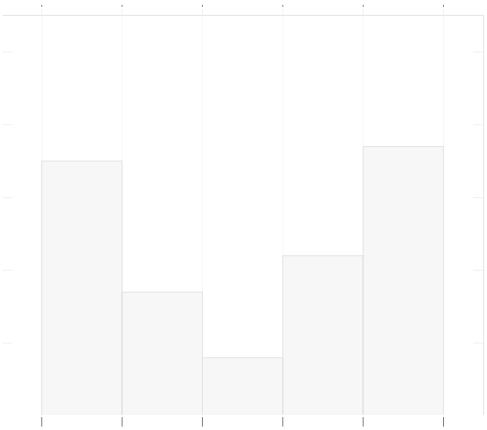
Given the SI paradigm:



(a) Ensemble is representative



(b) Ensemble is too wide



(c) Ensemble is too narrow

Understanding the framework

The upshot: if SI is the right paradigm, we can draw *some* conclusions from groups of models.

Not because we have a new construction method.

But because model evaluation provides us with sufficient background knowledge about the relationship between ensemble and world to justify said conclusions.

Evaluating "statistically-indistinguishable" ensembles

Are they right?

Are they right?

Yes and no.

More specifically: I don't think this buys all the inferences we want—particularly when it comes to the future.

Paradigms and predictions

Evaluation provides justification iff the proxy and the target can be assumed to be similar.

Paradigms and predictions

Evaluation provides justification iff the proxy and the target can be assumed to be similar.

In the context of future predictions about the climate, however, the assumption that the proxy (contemporary climate) is like the future in any sense is substantive.

Whence the extra power?

Recall: the truth-centered worry was the existence of models more extreme than extant ensembles.

Whence the extra power?

Recall: the truth-centered worry was the existence of models more extreme than extant ensembles.

If we take the shift in paradigm to provide us with (extra) justification for future predictions, we essentially rule this worry out by fiat.

That is: by way of an assumption about the nature of the space of possible models.

The main point

Note that this assumption may well be justified.

The main point

Note that this assumption may well be justified.

My point is that the evaluation of the SI paradigm turns on our knowledge about the space of possible models.

And doesn't have anything much to do with construction methods.

Outro

The takeaway

I've argued that the problem that we face is uncertainty about the space of possible models.

I could be wrong—particularly about the evaluative point.

The takeaway

I've argued that the problem that we face is uncertainty about the space of possible models.

I could be wrong—particularly about the evaluative point.

Maybe we still haven't identified the right the meta-"paradigm"; after all, both SI and the traditional alternative assume that the ensemble is like a random sample of something.

Thank you

Thank you!

- Annan, James D. and Julia C. Hargreaves (2010). Reliability of the CMIP3 Ensemble. *Geophysical Research Letters* 37: 1–5.
- – (2011). Understanding the CMIP3 Model Ensemble. *Journal of Climate* 24: 4529–38.
- – (2017). On the Meaning of Independence in Climate Science. *Earth Systems Dynamics* 8: 211–24.
- Baumberger, Christoph, Reto Knutti, and Gertrude Hirsch Hadorn (2017). Building Confidence in Climate Model Projections: An Analysis of Inferences From Fit. *Wiley Interdisciplinary Reviews: Climate Change* 8.3: e454.
- IPCC Working Group 1 (2013). *Climate Change 2013: The Physical Science Basis*. Ed. by Thomas F. Stocker et al. Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge: Cambridge University Press.
- Justus, James (2012). The Elusive Basis of Inferential Robustness. *Philosophy of Science* 79.5: 795–807.

- Knutti, Reto, Myles R. Allen, et al. (2008). A Review of Uncertainties in Global Temperature Projections over the Twenty-First Century. *Journal of Climate* 21.11: 2651–63.
- Knutti, Reto, Reinhard Furrer, et al. (2010). Challenges in Combining Projections from Multiple Climate Models. *Journal of Climate* 25.10: 2739–58.
- Parker, Wendy S. (2011). When Climate Models Agree: The Significance of Robust Model Predictions. *Philosophy of Science* 78.4: 579–600.
- – (2018). The Significance of Robust Climate Projections. In: *Climate Modeling: Philosophical and Conceptual Issues*. Ed. by Elisabeth A. Lloyd and Eric Winsberg. Cham: Palgrave Macmillan: 273–96.
- Pirtle, Zach, Ryan Meyer, and Andrew Hamilton (2010). What Does it Mean when Climate Models Agree? A Case for Assessing Independence Among General Circulation Models. *Environmental Science & Policy* 13.5: 351–61.

Winsberg, Eric (2018). What does Robustness Teach us in Climate Science: A Re-Appraisal. *Synthese* (online first).