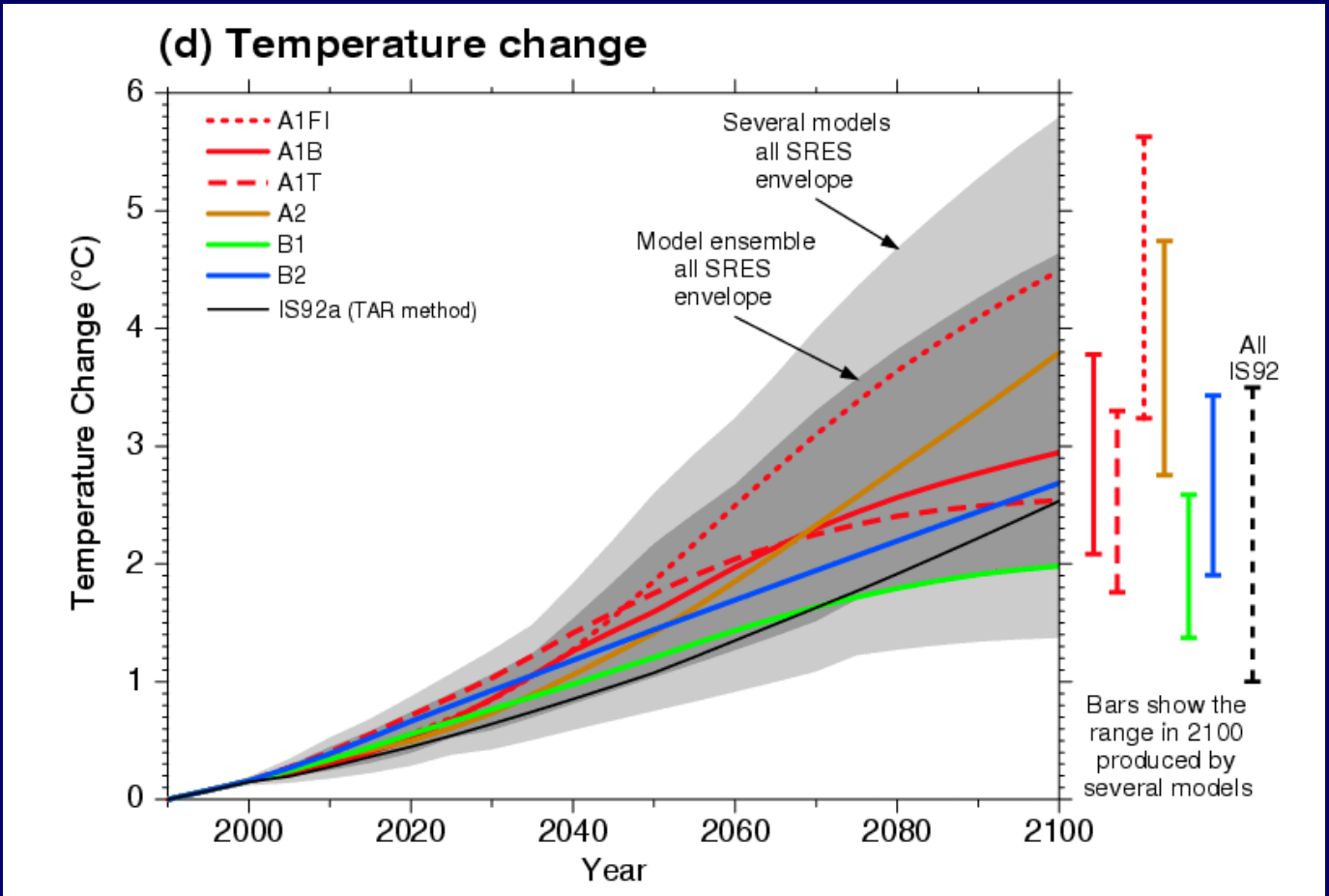
How can we (in)validate a probabilistic forecast?

Myles Allen
Department of Physics
University of Oxford
myles.allen@physics.ox.ac.uk

The problem

- The more interesting the forecasting situation (the larger the storm, the longer the lead time), the harder it is to verify or falsify a probabilistic forecast.
- “My forecast is more accurate than yours” is a meaningless claim.
- “My forecast is more stable than yours” (less likely to change with increasing model resolution, including more processes or changing expert opinion) is meaningful and testable.
- How can we speed up convergence and tell when a probabilistic forecast has converged?
- Implications for forecasting system design and the climate*prediction*.net experiment.

An example of an “ensemble of opportunity”



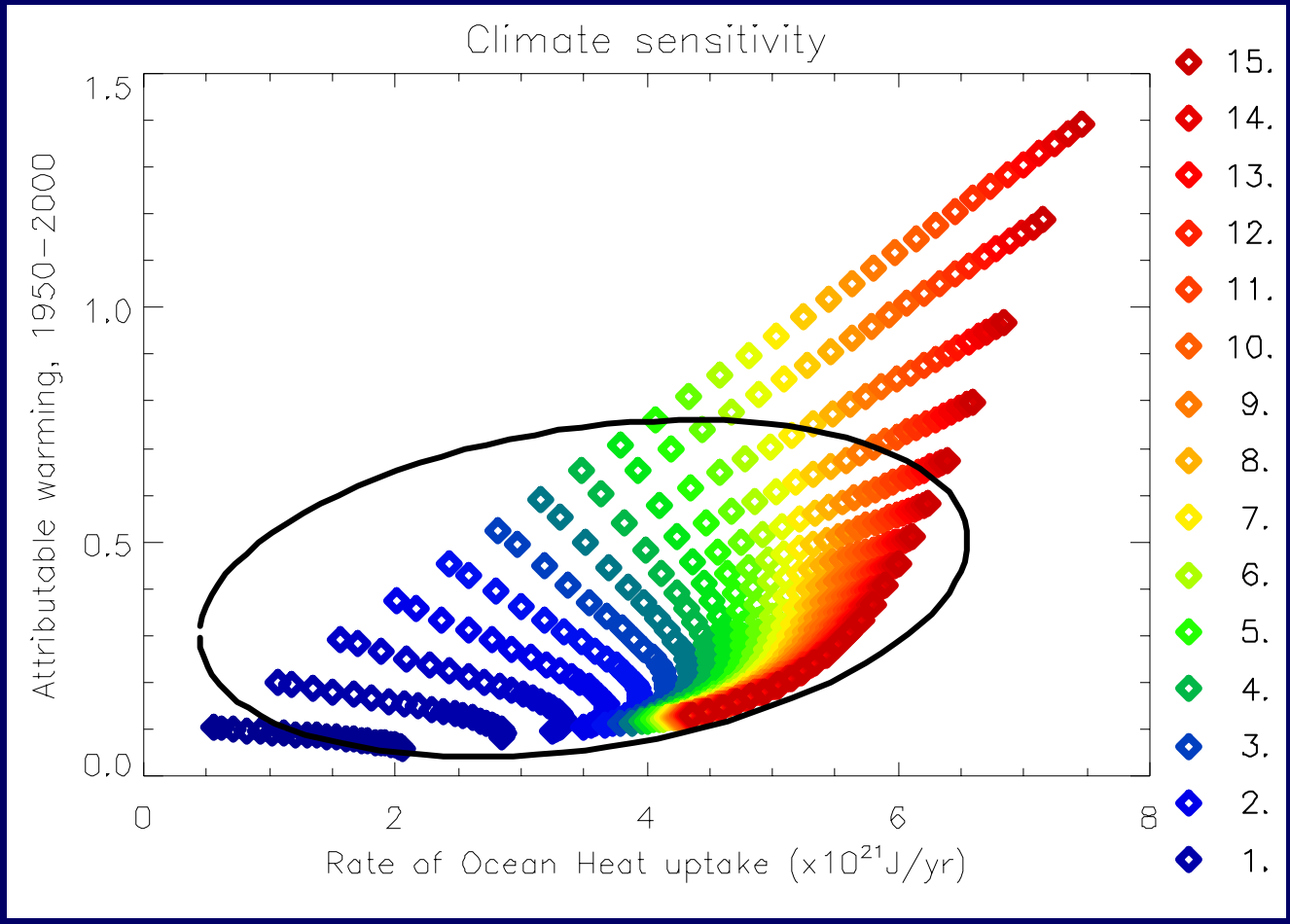
“Model error” in IPCC TAR addressed by providing a range of predictions from 7 models (not including most and least sensitive).

No interpretation given in terms of probability.

But what about model "validation"

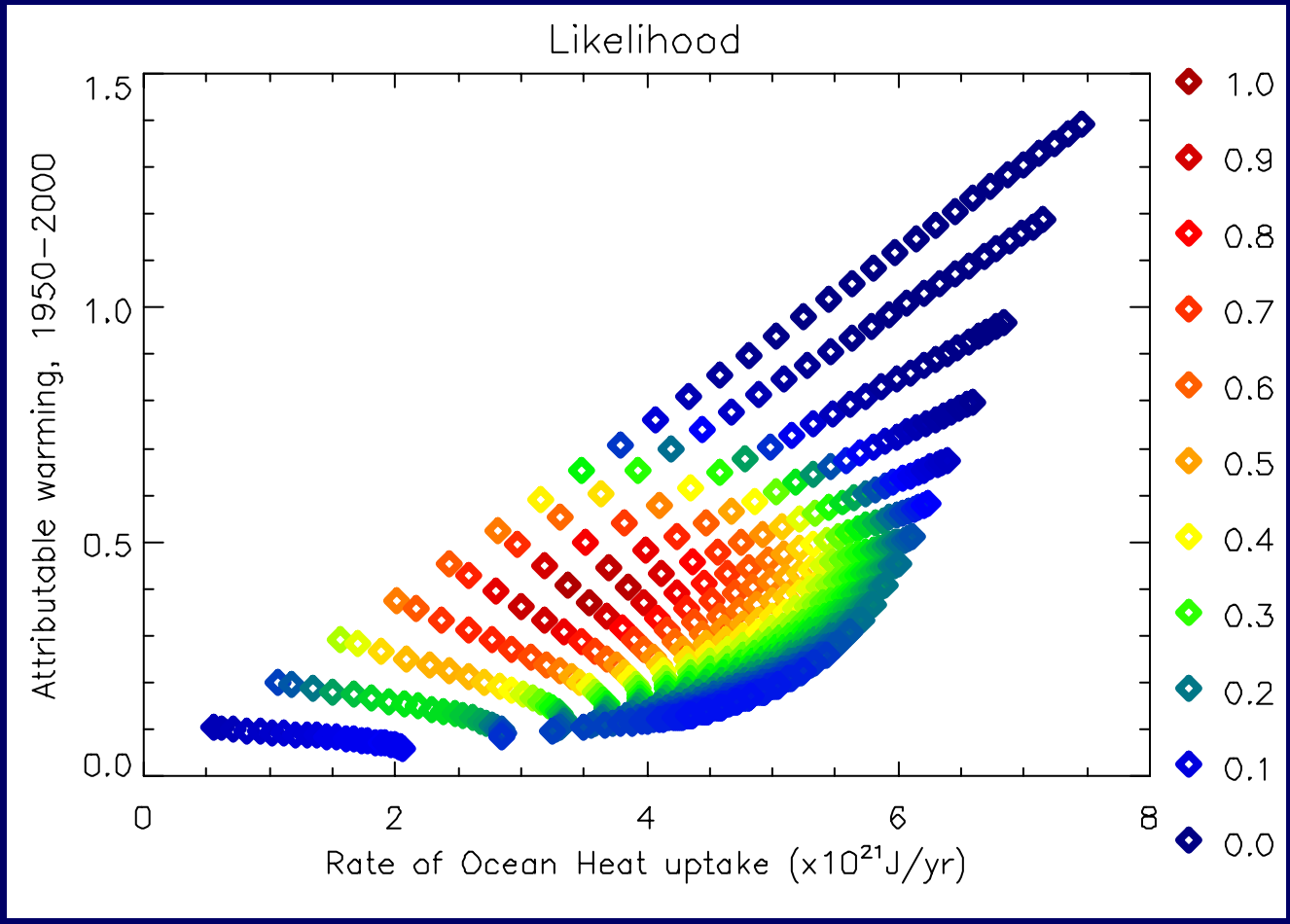
- If we reject models that compare "badly" with present-day climate, will the spread of results of the remainder provide an estimate of the forecast PDF?
- No.
- This strategy only works if we begin with an unbiased sample of "possible models".
- No way of beginning to do this, because there is no way of defining the "distance" between two models: no model-error analogue of the "analysis norm".
- Distance between models can only be defined in terms of outputs, necessitating mega-ensembles.

Climate sensitivity versus observable variables

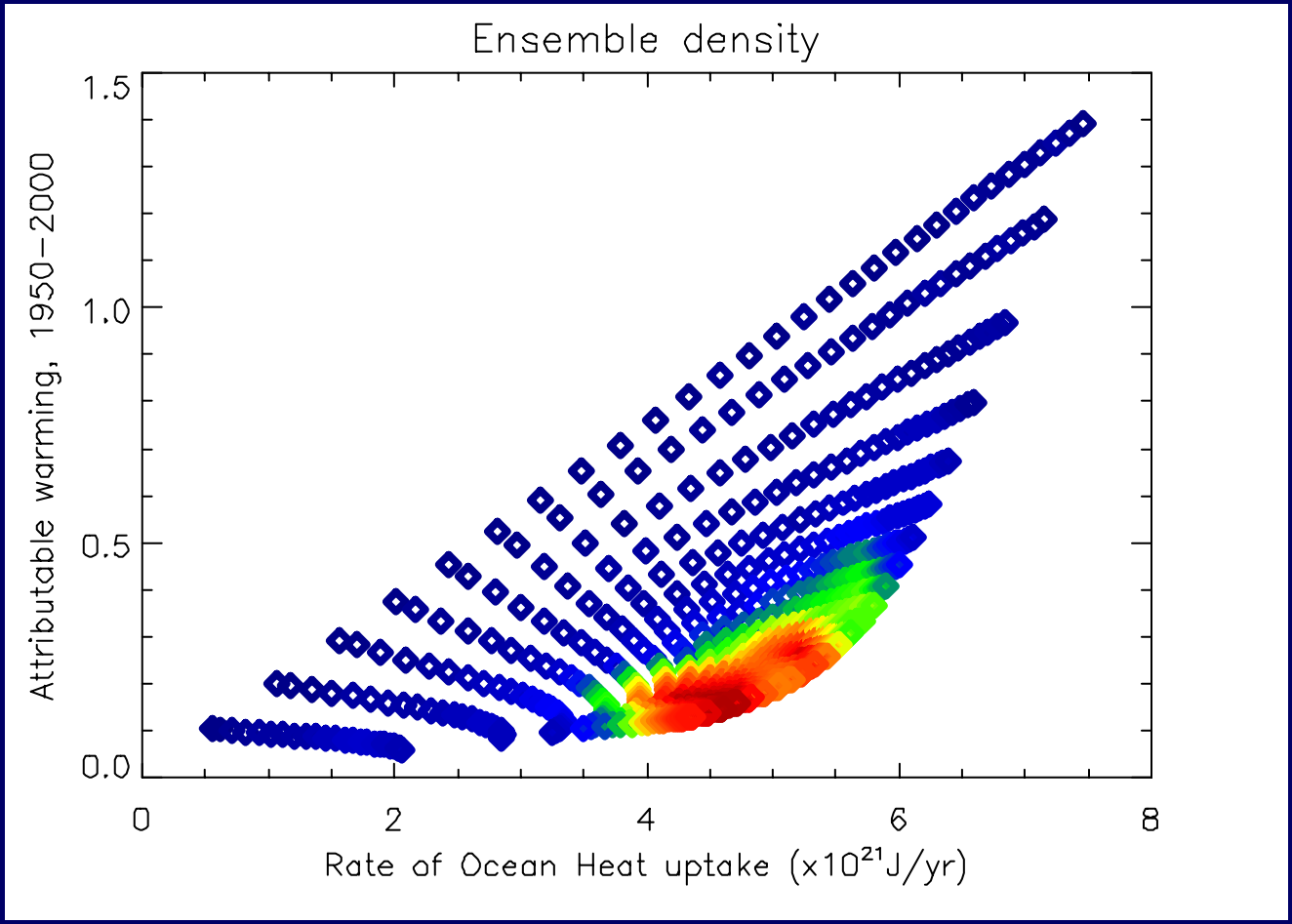


Data courtesy of Ben Booth: see Poster by Kettleborough et al

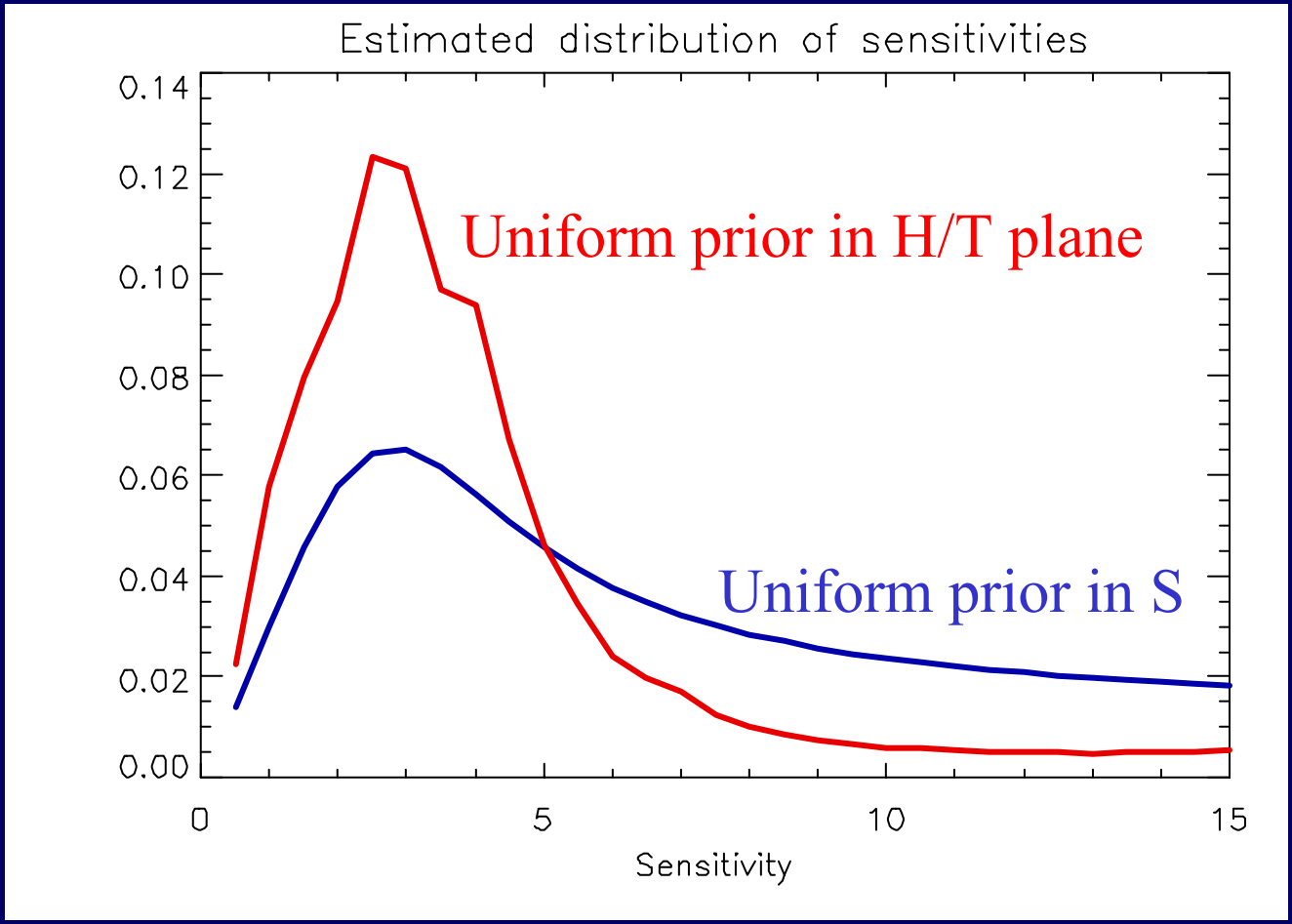
Estimated likelihood by comparison with "observations"



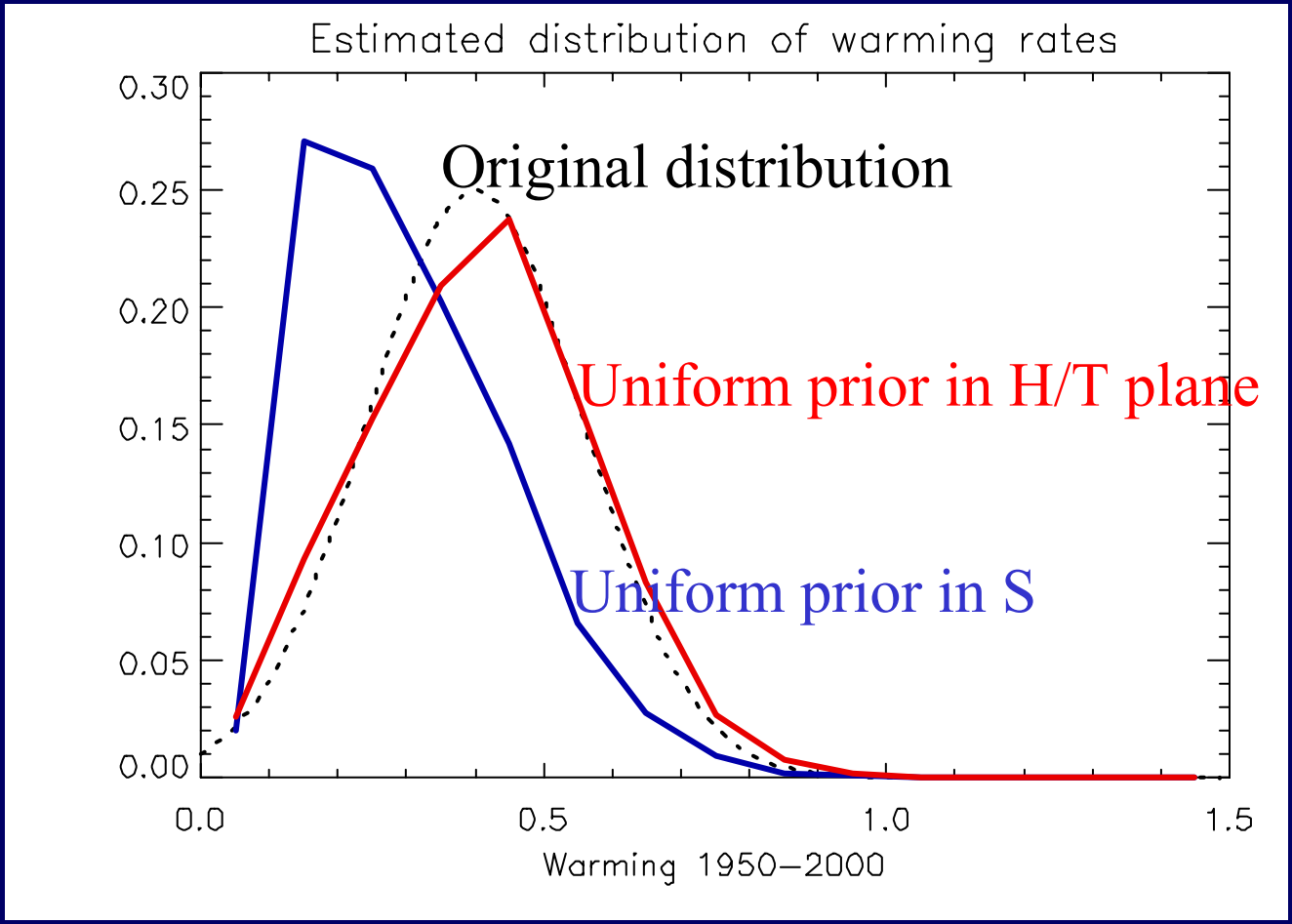
Density of ensemble in warming/heat-uptake plane



Estimated distribution of sensitivities from likelihood-weighted ensemble

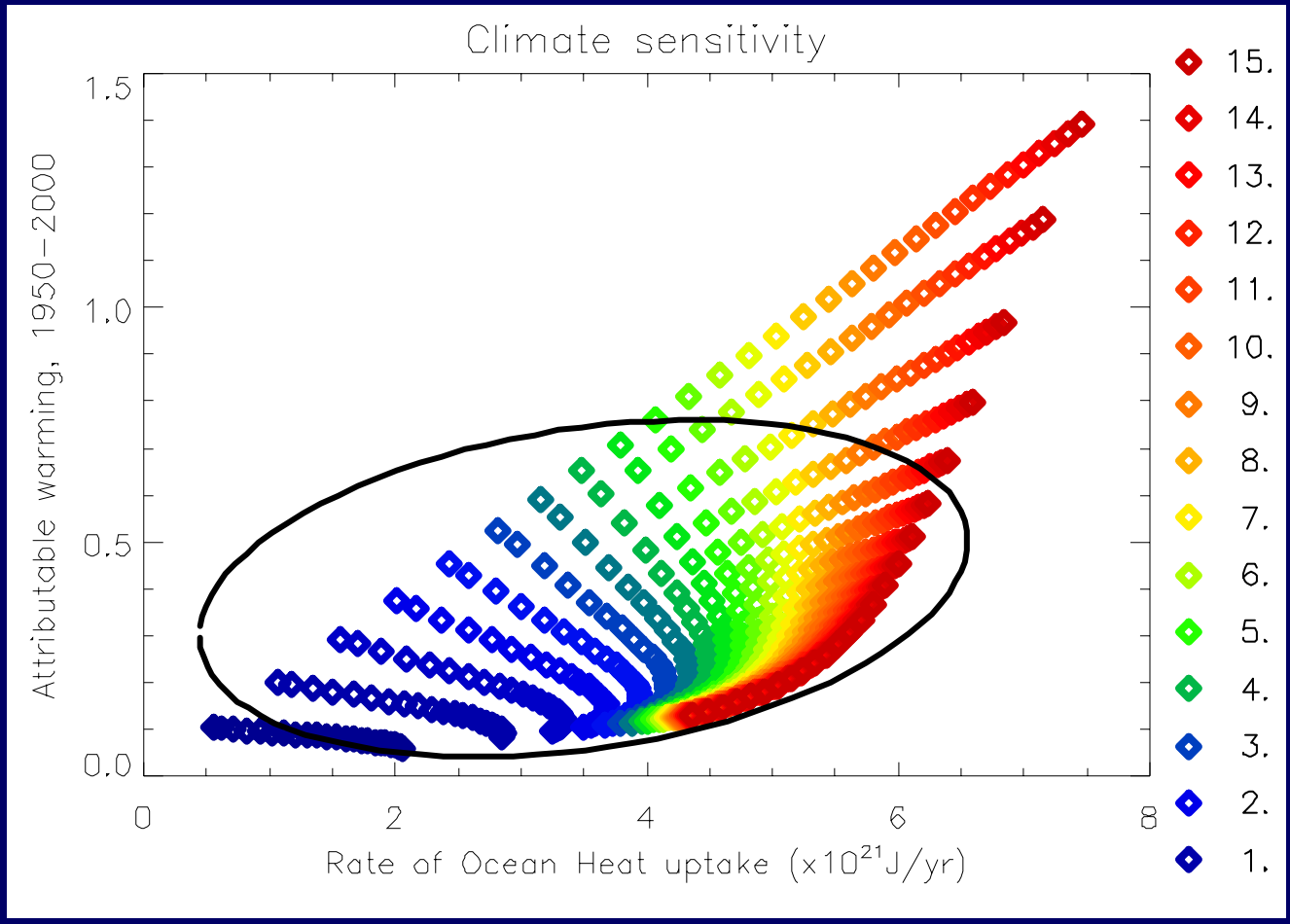


Estimated distribution of warming rates from likelihood-weighted ensemble



All priors are equal, but some are more equal than others...

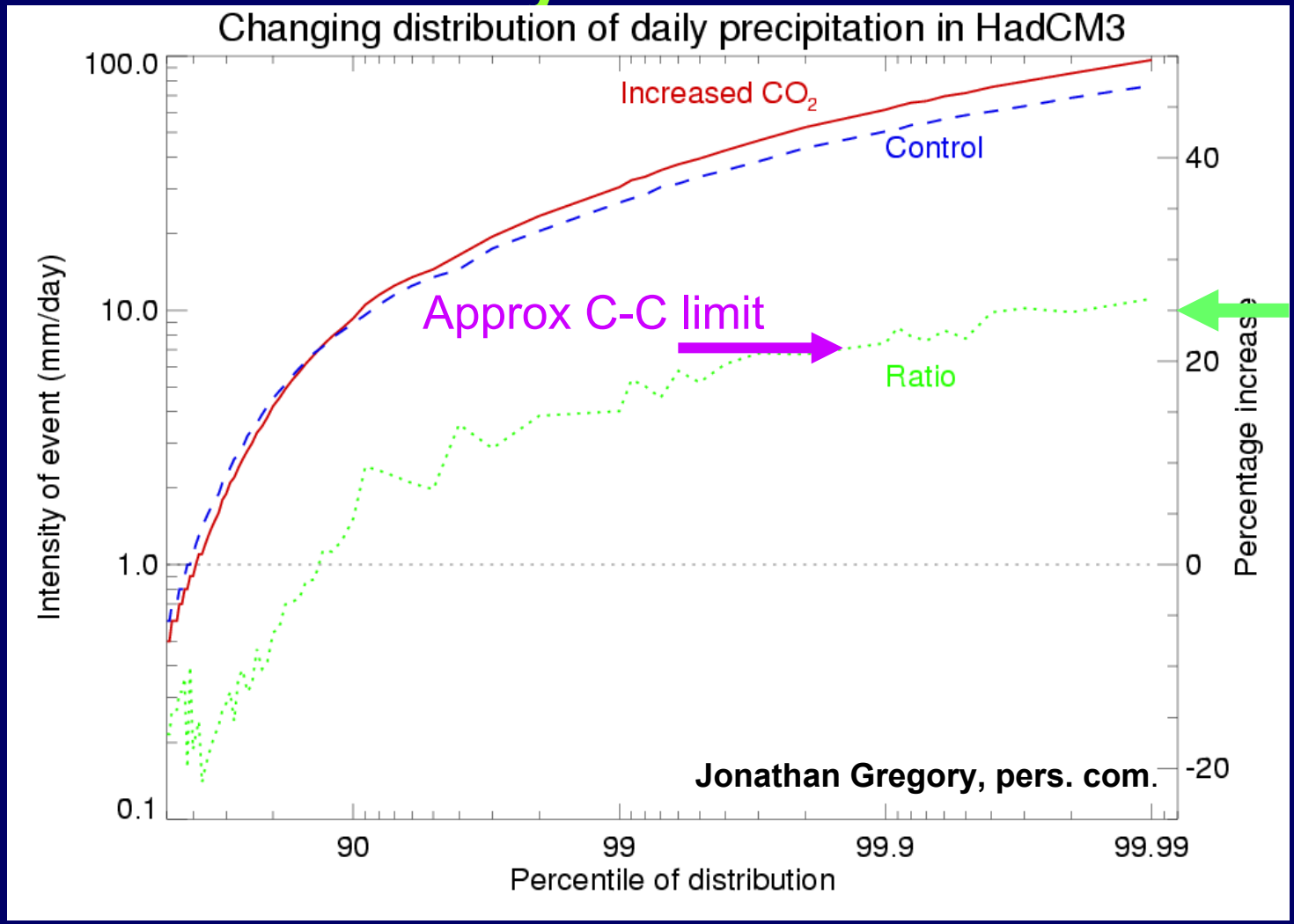
Can/should perturbed-physics ensembles change our views on current warming rates?



Summary

- We cannot rely on ensembles of opportunity.
- "Better" sampling strategies for model perturbations are not the answer: no way to sample "all AOGCMs".
- Constrained, perturbed-physics ensembles:
  - Perturb everything you can.
  - Find consistent relationships between observable and forecast variables.
  - Check relationships have converged as you make more perturbations & include different (resolution) models.
  - Provide physical interpretation if possible.
  - Resample (or weight) ensemble to make consistent with observations in space spanned by observable variables.
  - Infer forecast distribution from the re-sampled ensemble.

Constraints on impact-relevant forecast variables likely to be much harder to find



GFDL hurricane model (Knutson et al, 2001)

A methodology for the treatment of model error in climate forecasting

- Constraints on forecast regional changes likely to be more subtle, noisy and many-one.
- Initial-condition AOGCM ensembles required for comparison with observations.
- Perturbations interact non-linearly, so combinations must be explored.
- “Analysis” and forecast both depend on uncertain forcing, increasing ensemble size.
- Ideal for distributed computing: Windows HadCM3 under test (Stainforth et al, 2002).

www.climate*prediction*.net

- Use “slab” integrations to identify parameter perturbations that change response to CO_2 without changing control climate (c.f. SVs).
- Launch coupled (flux adjusted) ensemble simulations of 1950-2000 and weight by comparison with observations.
- Run on to 2050 under a range of natural and anthropogenic forcing scenarios.
- Establish which forecast variables depend on perturbations, which on observations.

First results from climate*prediction*.net