Japan Uncertainty Modelling Project

Understanding and Interpreting Climate Model Ensembles

James Annan and Julia Hargreaves
Research Institute for Global Change, JAMSTEC
Japan

With contributions from T Yokohata, A Paul, R Ohgaito,
M Collins, C Jackson, M Tobis

KAKUSHIN

MoE GERF S-5 Getting a 'Feel' for Future Climate Change!

JAMSTEC Japan Agency for Marine-Earth Science and Technology

Japan Uncertainty Modelling Project

How can we use the (CMIP3) MME in probabilistic forecasting?

- MME samples some uncertainties in (physical) parameterisations
- Does it in some way "include" reality?
  - Is it too narrow?
  - Is it too broad?
- What do these statements mean, and how can we check them?

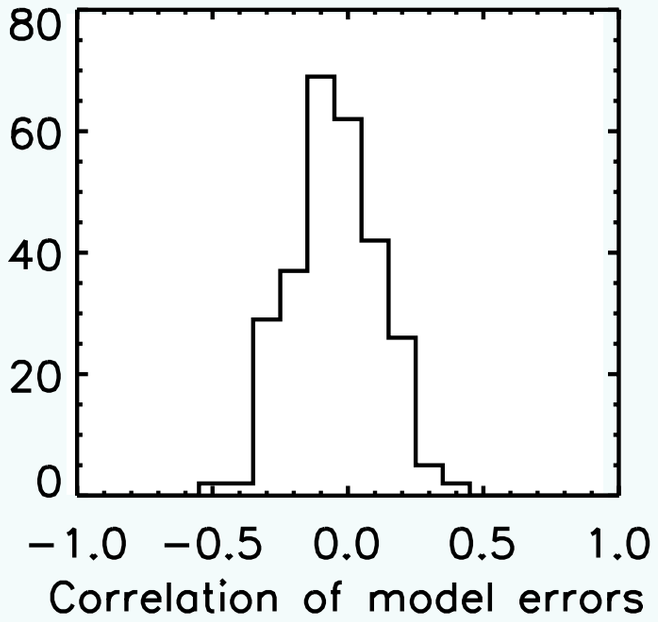
Paradigms for interpreting the CMIP3 "ensemble of opportunity"

(IPCC Expert Meeting, Boulder 2010)

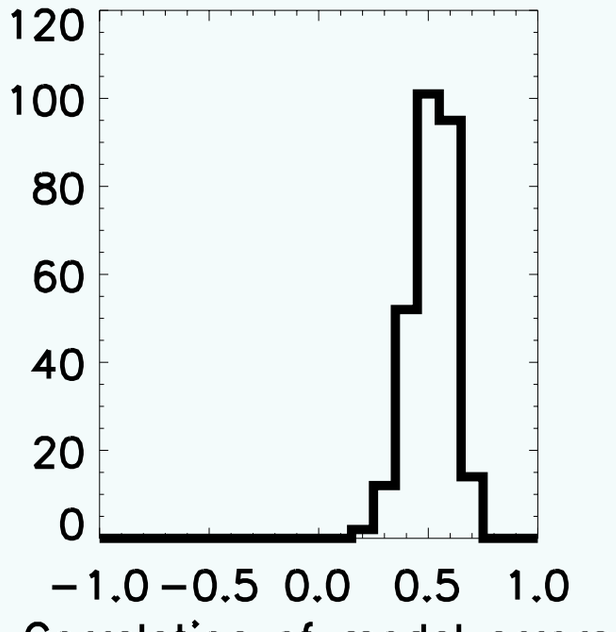
Truth-centred	**Statistically indistinguishable**
"each ensemble member is sampled from a distribution centered around the truth"	"each of the members is considered to be 'exchangeable' with the other members and with the real system"
$m_i = T + e_i \sim N(T,\sigma)$	$m_i = M + e_i \sim N(M,\sigma)$ $T = M + e_t \sim N(M,\sigma)$

Comparison of behaviours

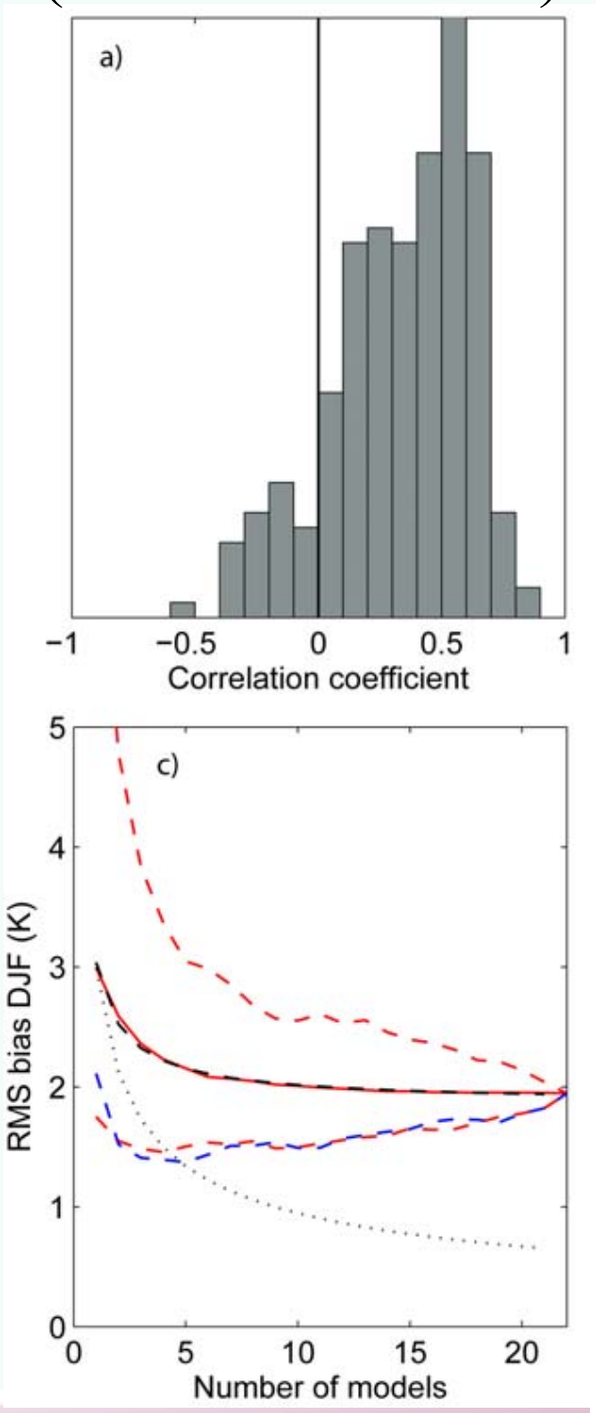
Synthetic
Truth-centred

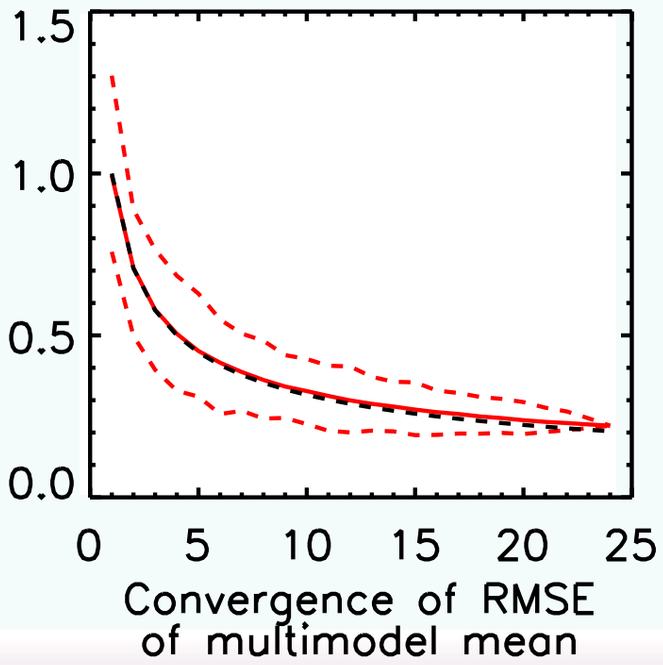


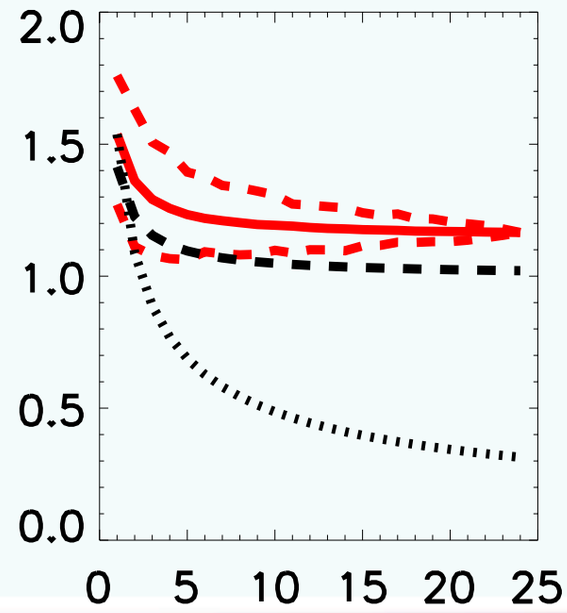
Statistically
Indistinguishable



CMIP3
(Knutti et al 2010)







"Reliability of the CMIP3 ensemble" GRL 2010

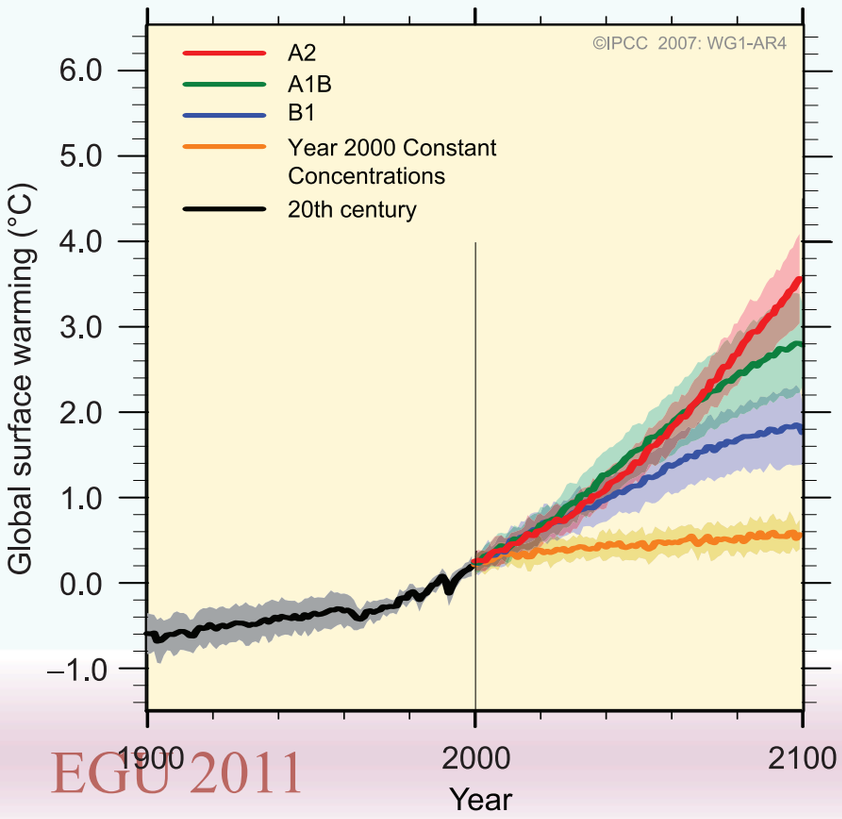
KAKUSHIN

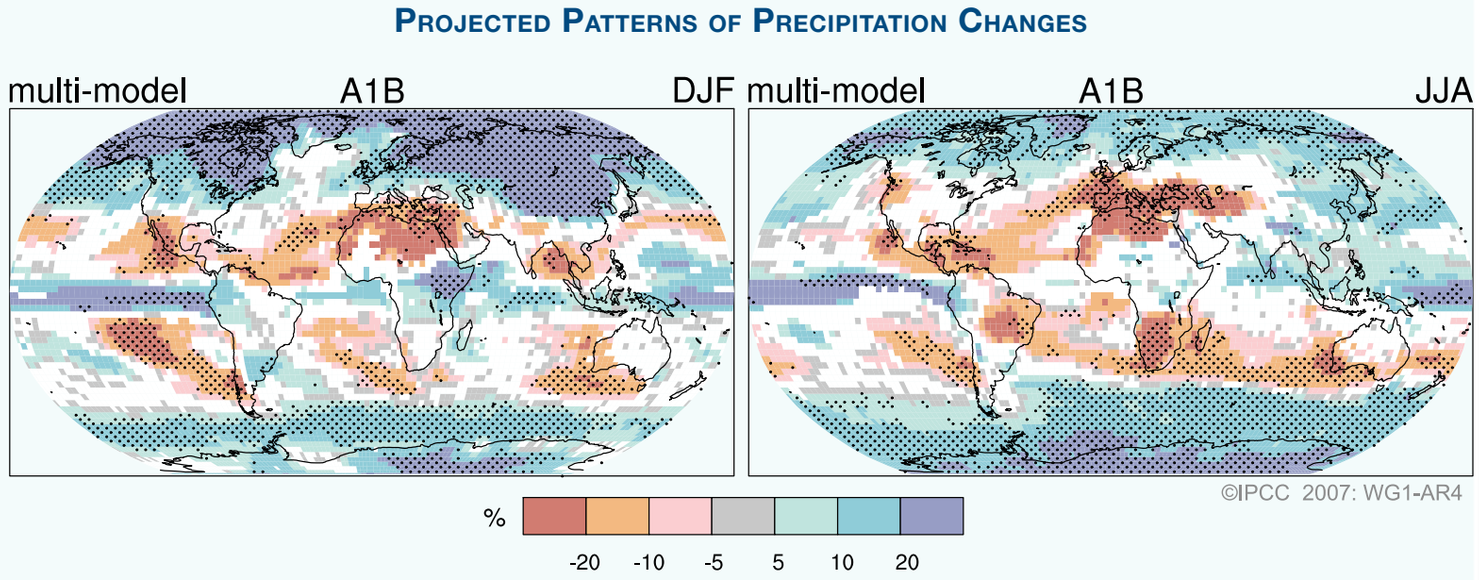
MoE GERF S-5 Getting a 'Feel' for Future Climate Change!

JUMP Japan Uncertainty Modelling Project

How can CMIP3 be understood within the statistically indistinguishable paradigm?

- Our ensemble represents our uncertain beliefs about the system
  - Ensemble is centred on our "mean beliefs" (not reality)
- If our uncertainty (ie ensemble range) is well-calibrated relative to error in the mean, then reality and models will be a similar distance from the mean
- Corollary: natural "counting" interpretation will be *reliable*





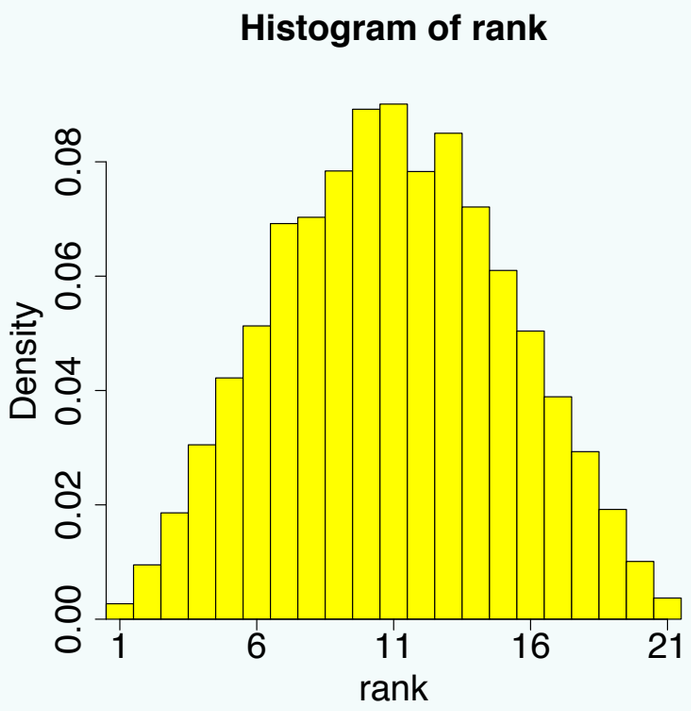
KAKUSHIN

MoE GERF S-5 Getting a 'Feel' for Future Climate Change!

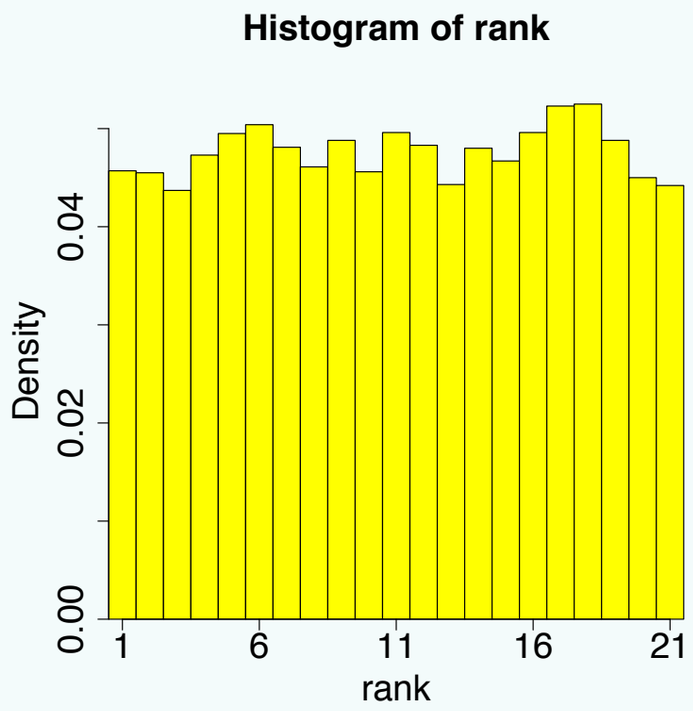
JUMP Japan Uncertainty Modelling Project

Evaluating reliability with the Rank Histogram (Talagrand diagram)

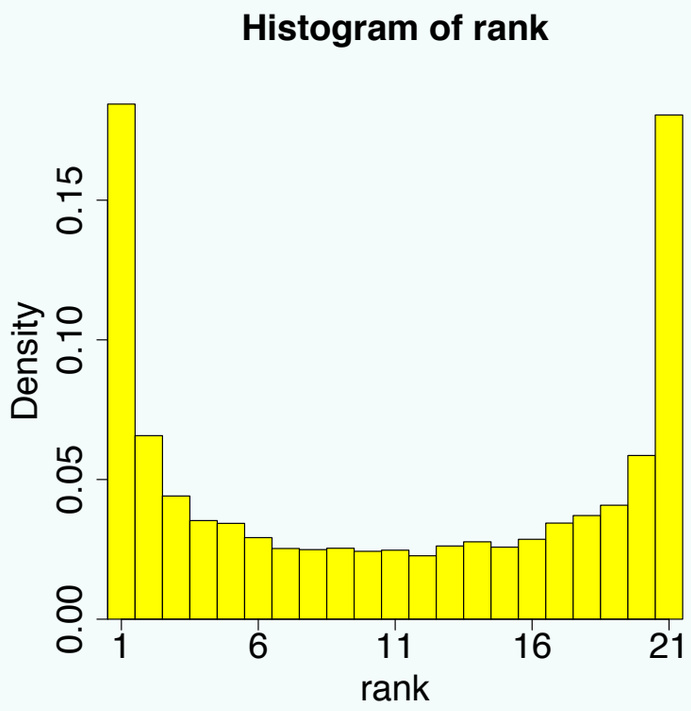
- For a reliable ensemble, the truth lies equiprobably at each position in the rank ordering of ensemble plus truth
- Histogram of ranks of observations should be flat



Wide ensemble: spread is too large, observation is near the centre.



Reliable: ensemble represents our uncertainty.



Narrow ensemble: observation is too often outside ensemble range.

Japan Uncertainty Modelling Project

Real data and models - MMEs and SMEs

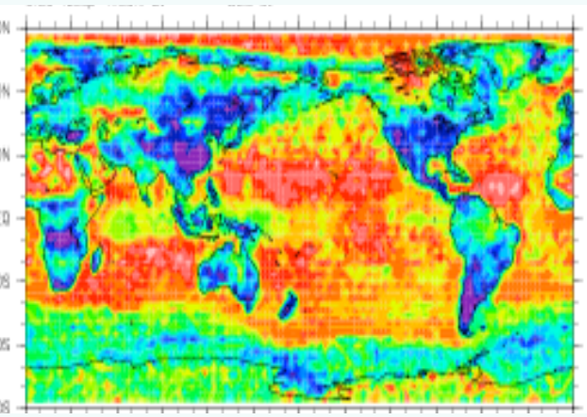
(Yokohata et al submitted - see poster XL115 today)

- Reliability of CMIP3 MME and several single model ensembles
  MIROC, HadCM3/SM3, NCAR CAM3

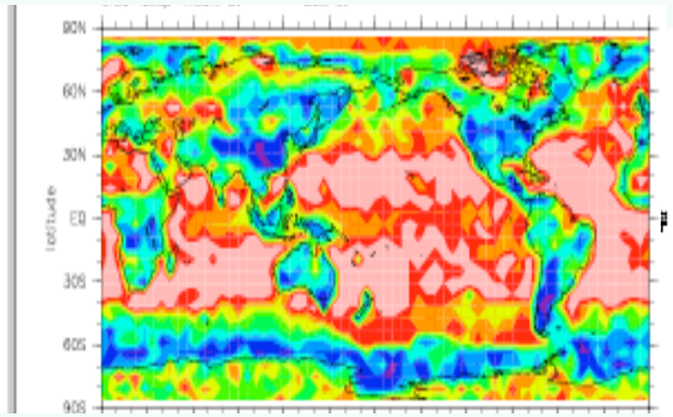
- Many observations relating to radiation balance and clouds

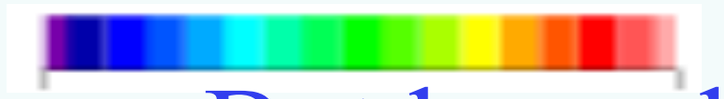
Clear sky SW Radiation

CMIP3-AOGCM

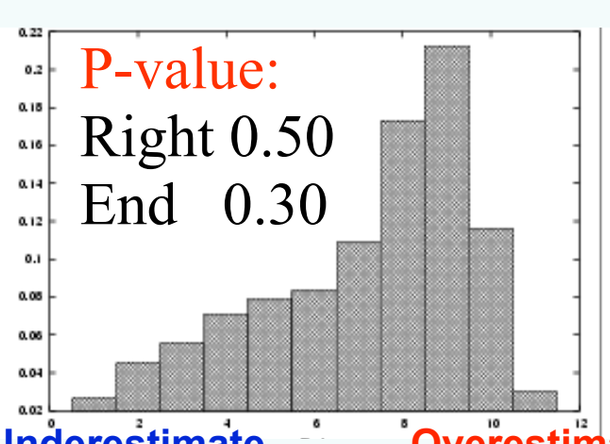


CMIP3-ASGCM

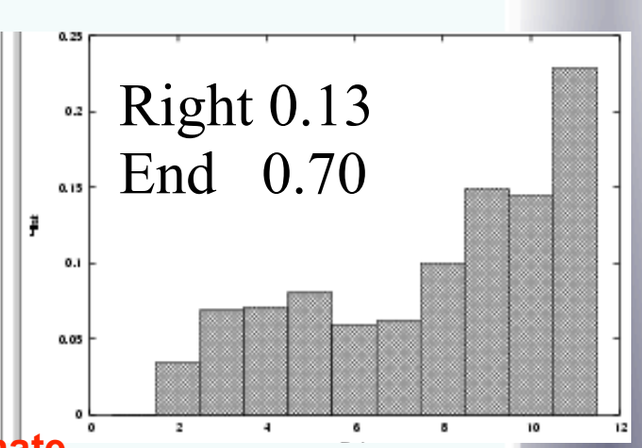




CMIP3-AOGCM



CMIP3-ASGCM



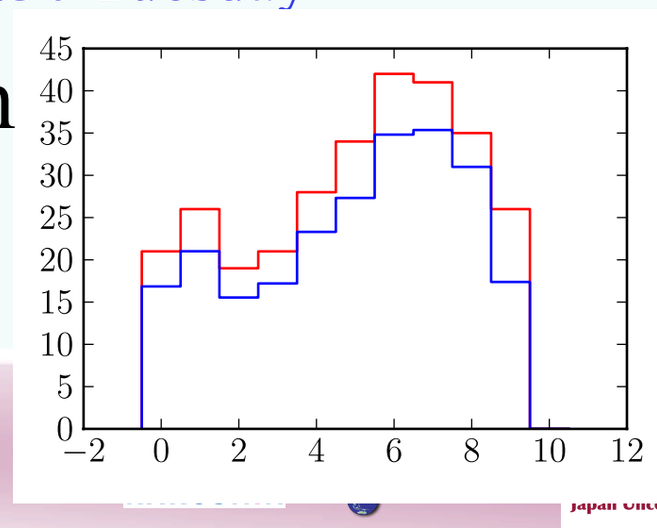
Underestimate Overestimate

But how about climate *change*?

Hargreaves et al submitted - see talk, CL1.2 15:30 Tuesday

Climate change at the Last Glacial Maximum

Rank Histogram of temperature anomalies
MARGO vs PMIP2



Japan Uncertainty Modelling Project

"But if the ensemble isn't centred on the truth, why is the multi-model mean so good?"

$$(1/n)\sum||m_i-O||^2 = (1/n)\sum||m_i-M||^2 + ||O-M||^2$$

Average of model errors = Ensemble spread + Error of multi-model mean
(eg Stephenson and Doblas-Reyes 2000, Epstein 1969, Leith 1974)

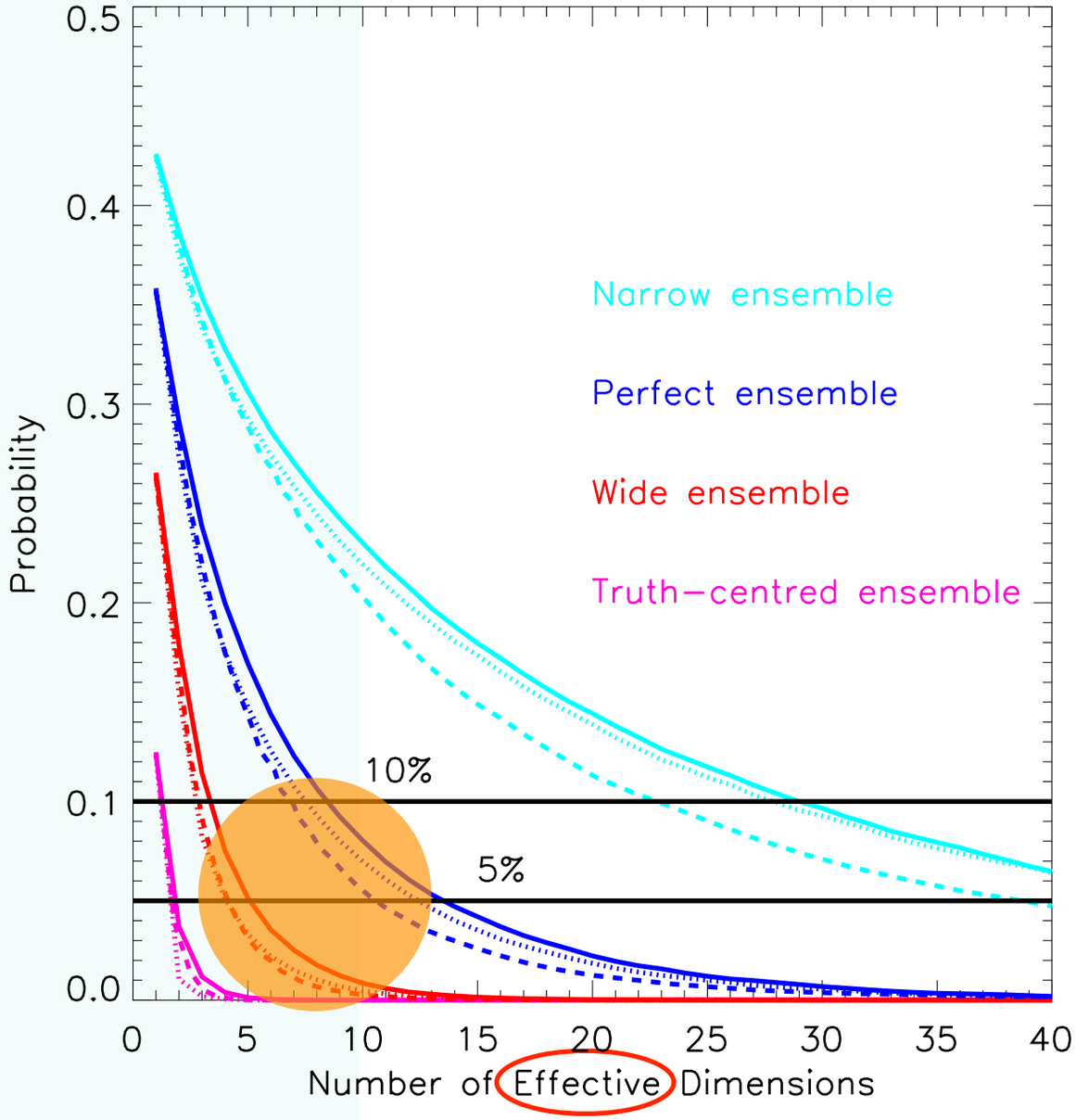
Where the m_i are the models, $M=(1/n)\sum m_i$ is their mean and O are the obs

Follow-up: When is the mean better than the *best* model, and why?

(Annan and Hargreaves, J Clim in press)

- Lambert and Boer 2001 Cli Dyn: "the mean model is generally the best model"
- Glecker et al 2008 JGR : "in most cases the mean and median models score best"
- Pierce et al 2009 PNAS: "Although MM's superiority has been found in previous studies focusing on the mean climate, the reasons for this have not generally been elucidated"
- Statistically Indistinguishable paradigm can provide some insights

Probability of a given model being better than the mean



Solid: Isotropic (spherical) Gaussian
Dotted: exponential eigenvalues
Dashed: 1/√i distribution

(T.Palmer, ECMWF Newsletter #106)

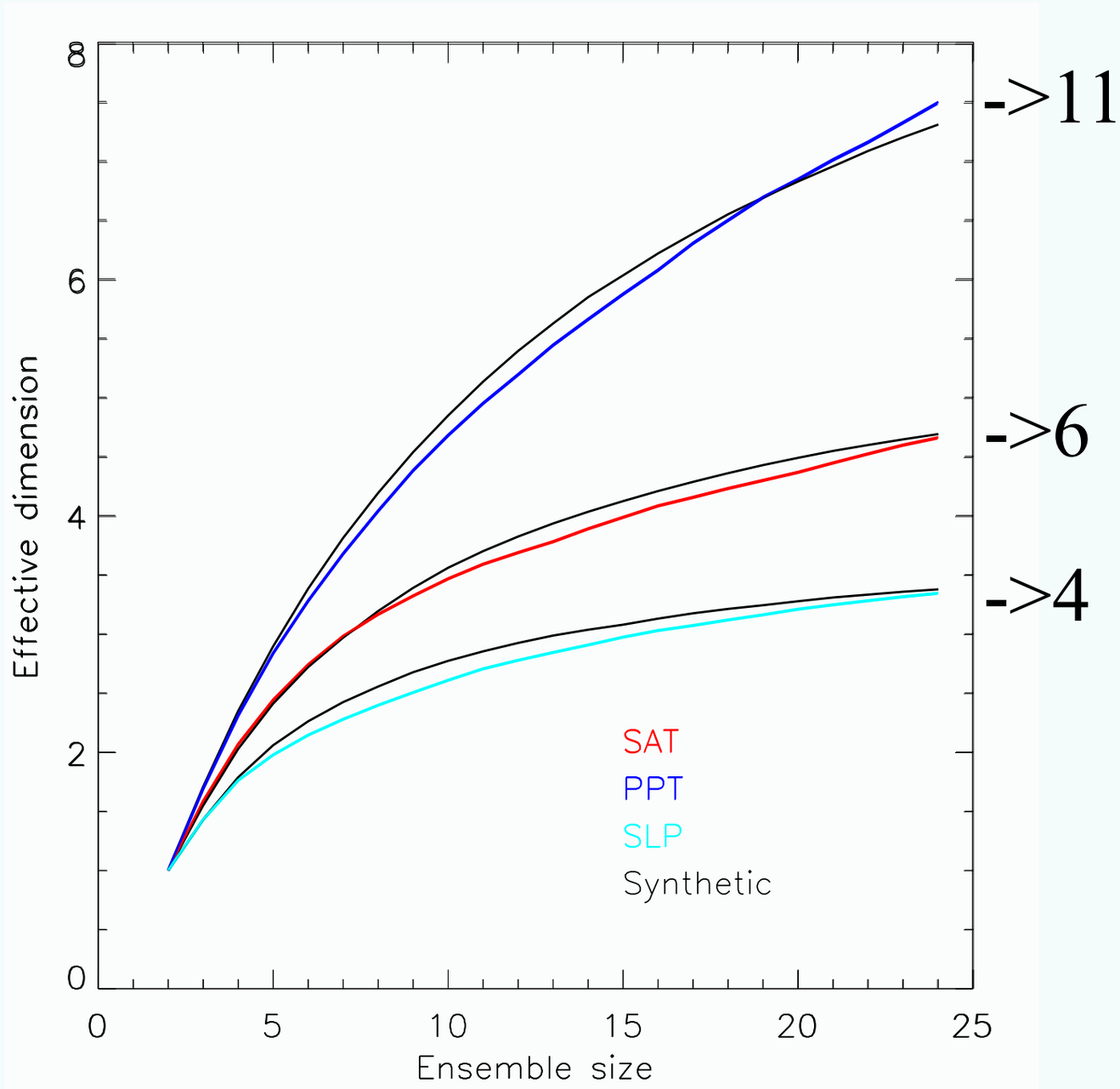
CMIP3 results (2D fields)

- Strongly depends on relative width and effective dimension
- Doesn't depend on shape of distribution

How many effective dimensions are there?

- We've seen this is critical for analysing CMIP3 behaviour
  - Annan and Hargreaves GRL 2010:
    N_{ef} = 40 (NWP) or 5 (cross-validation)
  - These imply very different interpretations of the results
- CMIP3 ensemble: N_{ef} = 4.6, 7.5 and 3.4 (SAT, PPT, SLP)
  (based on Bretherton et al EOF approach, supported by cross-validation)
- Finite sample gives lower N_{ef} than infinite ensemble from same distribution
  - More models will sample more dimensions

Test subsets of models



Comparison with synthetic data of known effective dimension

Summary

- Multi-model ensemble fits the statistically indistinguishable paradigm fairly well.
- ...Even for out of sample data (PMIP and the LGM).
- Many properties of the ensemble can be easily explained within this framework.
- “Reliability” is a key concept in evaluating the MME.
- Provides a basis for use of the MME in probabilistic prediction.
- Ensemble size is far from saturated

Bibliography

- J.C. Hargreaves “Skill and uncertainty in climate models” WIRES 2010
- J.D. Annan and J.C. Hargreaves, “Reliability of the CMIP3 ensemble”, GRL 2010
- J.D. Annan and J.C. Hargreaves “Understanding the CMIP3 multi-model ensemble”, In Press, J Clim
- T. Yokohata et al, “Reliability of structurally different perturbed physics and multi-model ensembles”, submitted to Climate Dynamics
- J.C. Hargreaves et al, Are the PMIP climate models consistent with the MARGO data synthesis for the Last Glacial Maximum? Clim Past Discussions.

All papers are online: Google “James Annan”, and go to his work page.

The End