Japan Uncertainty Modelling Project

Assessing the reliability of climate model ensembles using the Last Glacial Maximum

Julia Hargreaves
Research Institute for Global Change, JAMSTEC, Japan
James Annan, Andre Paul, Rumi Ohgaito, T. Yokohata, Ayako Abe-Ouchi.

KAKUSHIN
MoE GERP S-5 Getting a 'Feel' for Future Climate Change!
JAMSTEC Japan Agency for Marine-Earth Science and Technology
Japan Uncertainty Modelling Project

Outline of talk

Outline of talk

- For multi-model and single-model ensembles of climate models

Outline of talk

- For multi-model and single-model ensembles of climate models
  - ✳ Assessing Reliability

Outline of talk

- For multi-model and single-model ensembles of climate models
  - ✳ Assessing Reliability
    - Using modern climatology

Outline of talk

- For multi-model and single-model ensembles of climate models
  - ✳ Assessing Reliability
    - Using modern climatology
    - Using paleo-data

Outline of talk

- For multi-model and single-model ensembles of climate models
  - ✳ Assessing Reliability
    - Using modern climatology
    - Using paleo-data
  - ✳ Assessing Skill

Outline of talk

- For multi-model and single-model ensembles of climate models

  ✳ Assessing Reliability

    - Using modern climatology
    - Using paleo-data

  ✳ Assessing Skill

    - Using historical data

Outline of talk

- For multi-model and single-model ensembles of climate models

  ✳ Assessing Reliability

    - Using modern climatology
    - Using paleo-data

  ✳ Assessing Skill

    - Using historical data
    - Using paleo-data

Reliability and the rank histogram

Reliability and the rank histogram

- Informally: in a good ensemble, we would expect a new accurate observation to usually lie in the ensemble range.

Reliability and the rank histogram

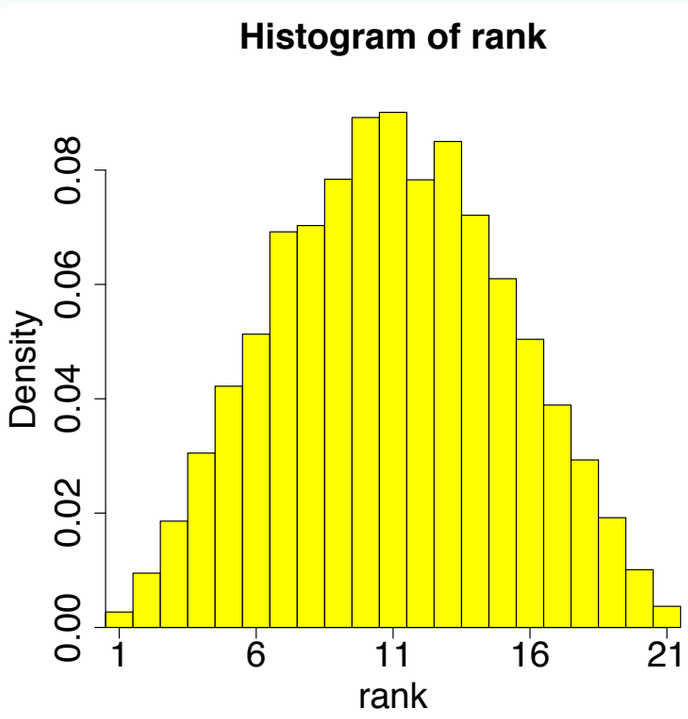
- Informally: in a good ensemble, we would expect a new accurate observation to usually lie in the ensemble range.
- Formally: for a reliable ensemble, the truth lies equiprobably in every position in the rank ordering.

Reliability and the rank histogram

- Informally: in a good ensemble, we would expect a new accurate observation to usually lie in the ensemble range.
- Formally: for a reliable ensemble, the truth lies equiprobably in every position in the rank ordering.
  - ✳ Leading to a flat histogram of the rank of each observation with the ensemble

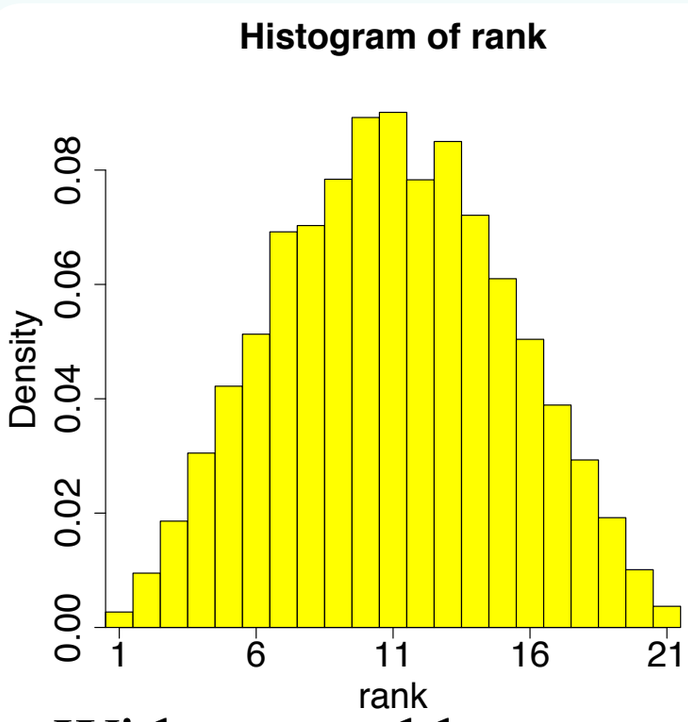
Reliability and the rank histogram

- Informally: in a good ensemble, we would expect a new accurate observation to usually lie in the ensemble range.
- Formally: for a reliable ensemble, the truth lies equiprobably in every position in the rank ordering.
  - ✳ Leading to a flat histogram of the rank of each observation with the ensemble



Reliability and the rank histogram

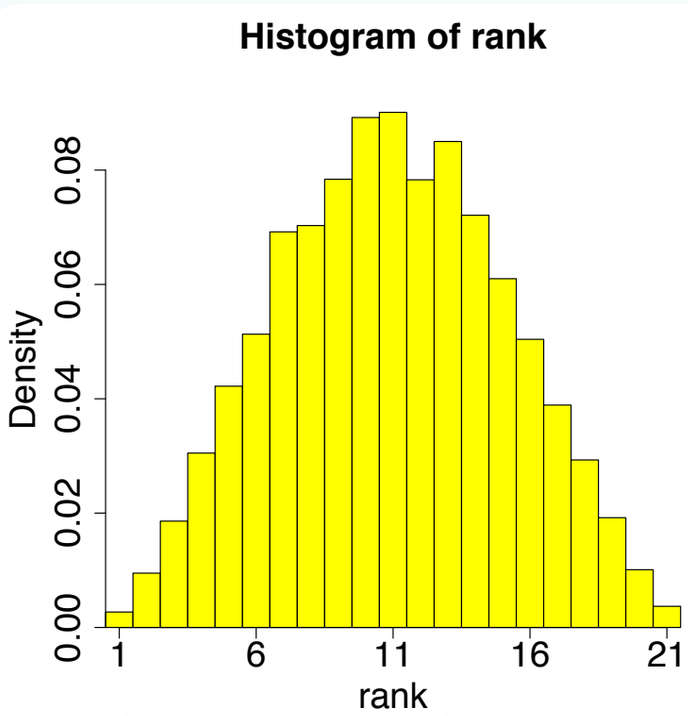
- Informally: in a good ensemble, we would expect a new accurate observation to usually lie in the ensemble range.
- Formally: for a reliable ensemble, the truth lies equiprobably in every position in the rank ordering.
  - ✳ Leading to a flat histogram of the rank of each observation with the ensemble



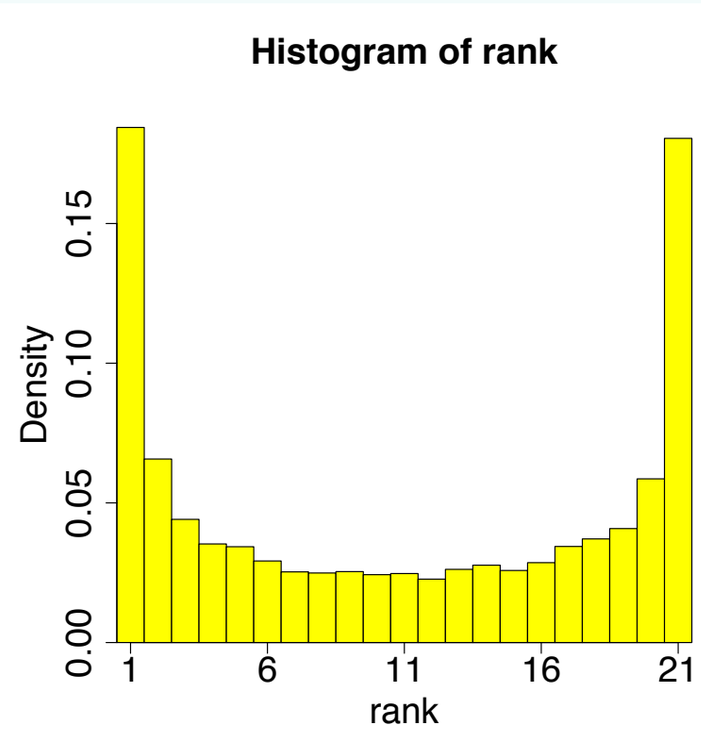
Wide ensemble: spread is too large, observation is near the centre.

Reliability and the rank histogram

- Informally: in a good ensemble, we would expect a new accurate observation to usually lie in the ensemble range.
- Formally: for a reliable ensemble, the truth lies equiprobably in every position in the rank ordering.
  - ✳ Leading to a flat histogram of the rank of each observation with the ensemble

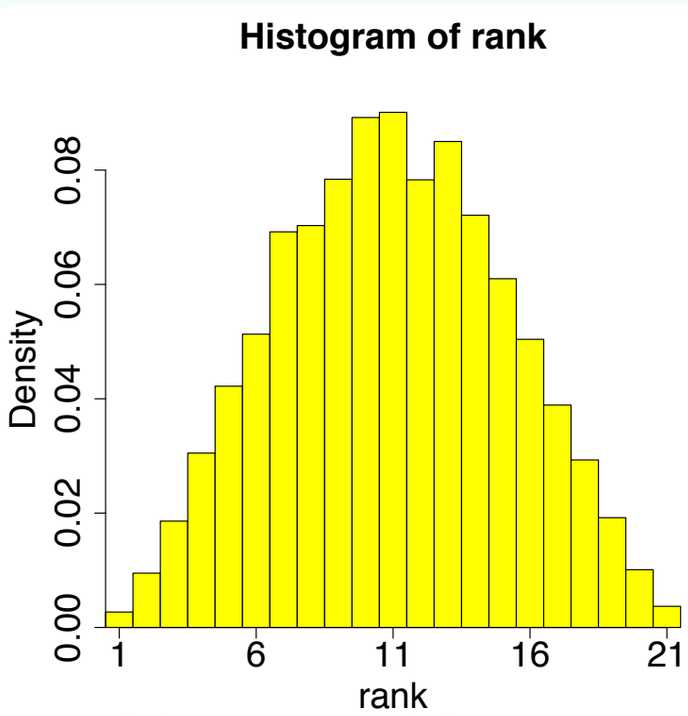


Wide ensemble: spread is too large, observation is near the centre.

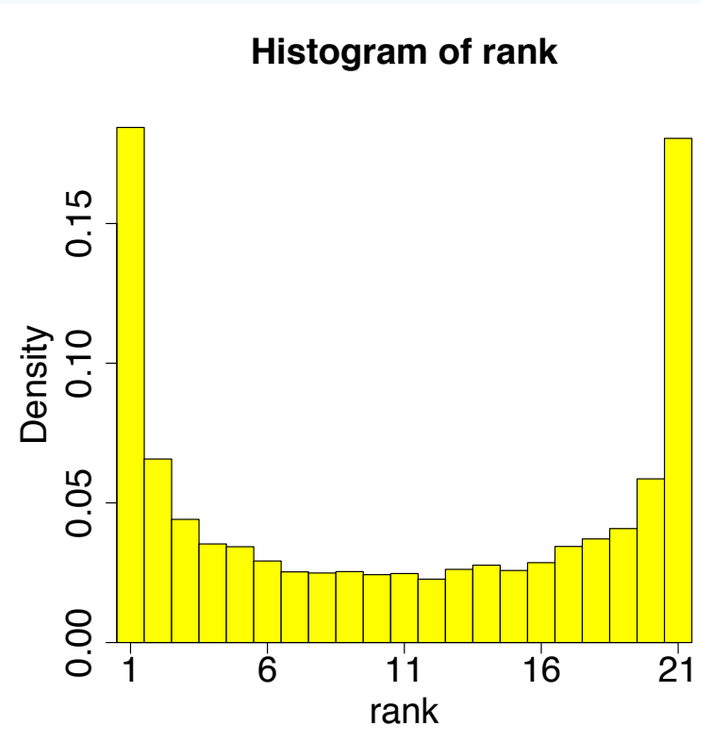


Reliability and the rank histogram

- Informally: in a good ensemble, we would expect a new accurate observation to usually lie in the ensemble range.
- Formally: for a reliable ensemble, the truth lies equiprobably in every position in the rank ordering.
  - ✳ Leading to a flat histogram of the rank of each observation with the ensemble



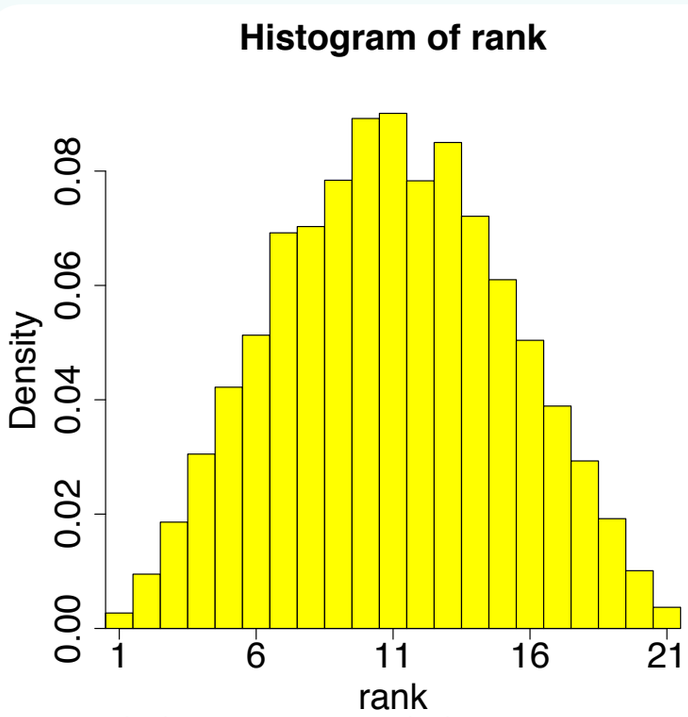
Wide ensemble: spread is too large, observation is near the centre.



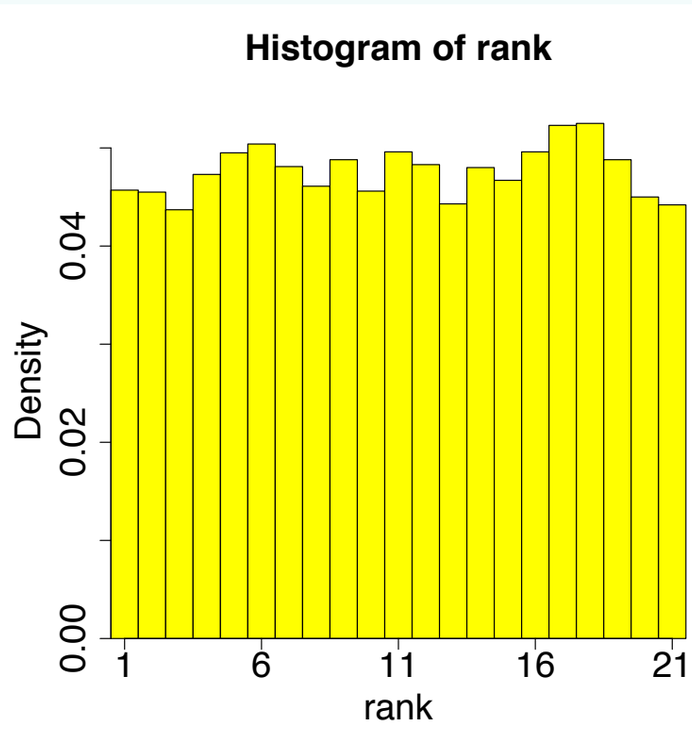
Narrow ensemble: observation is too often outside ensemble range.

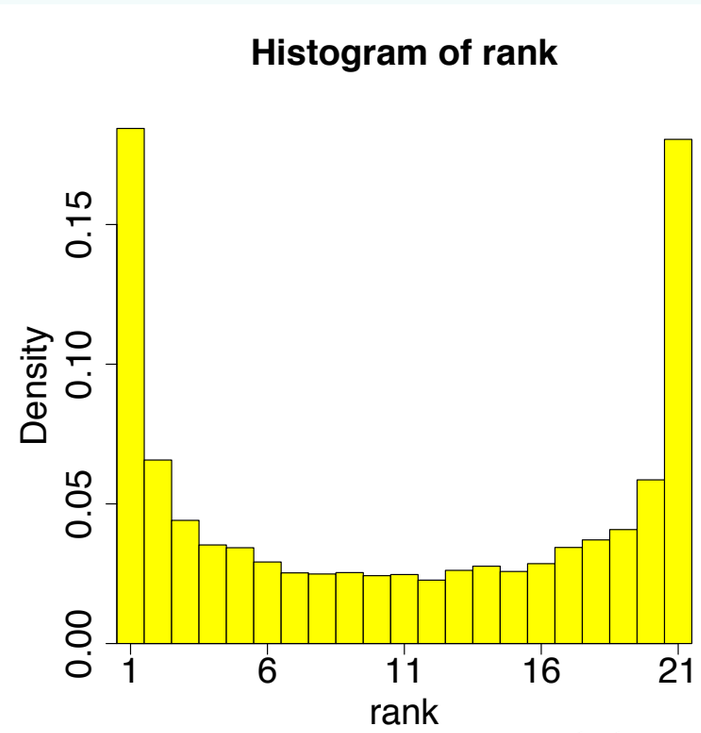
Reliability and the rank histogram

- Informally: in a good ensemble, we would expect a new accurate observation to usually lie in the ensemble range.
- Formally: for a reliable ensemble, the truth lies equiprobably in every position in the rank ordering.
  - ✳ Leading to a flat histogram of the rank of each observation with the ensemble



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



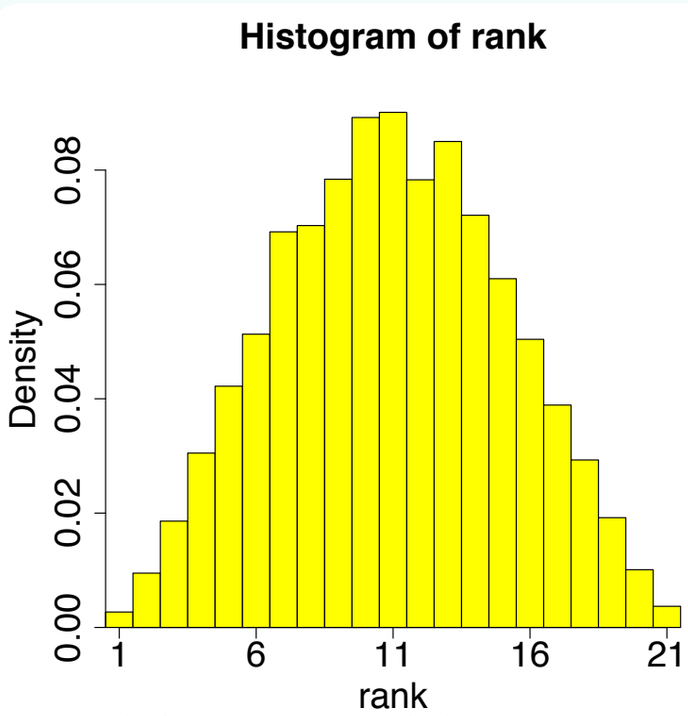


Narrow ensemble: observation is too often outside ensemble range.

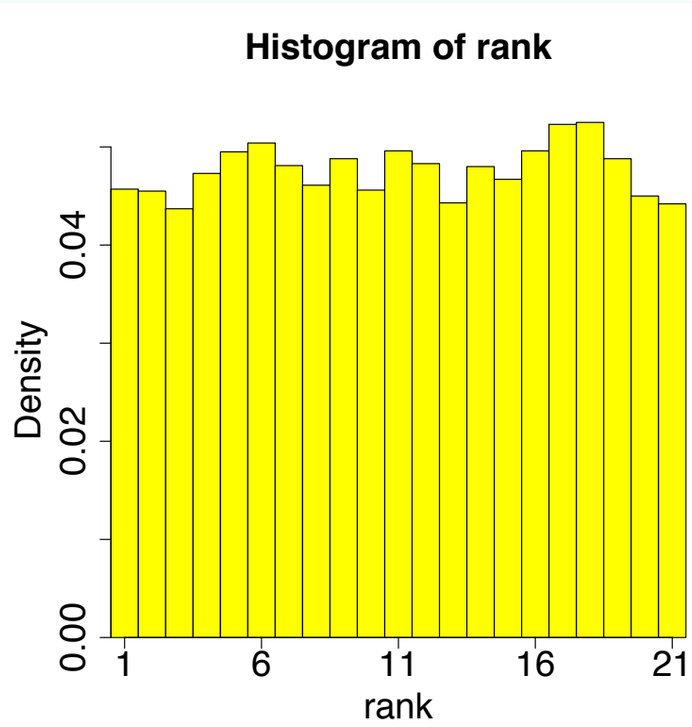
KAKUSHIN S-5 Getting a 'Feel' for Future Climate Change! JUMP Japan Uncertainty Modelling Project

Reliability and the rank histogram

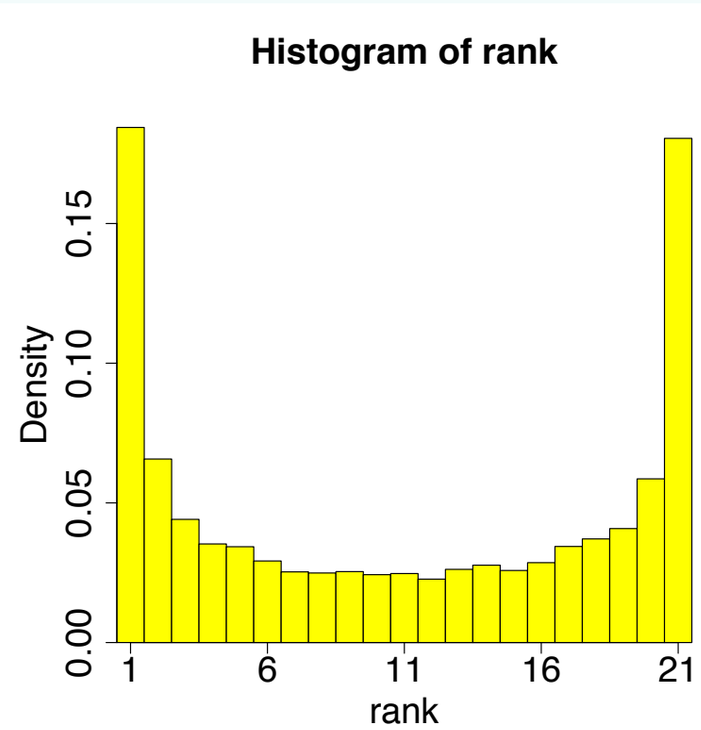
- Informally: in a good ensemble, we would expect a new accurate observation to usually lie in the ensemble range.
- Formally: for a reliable ensemble, the truth lies equiprobably in every position in the rank ordering.
  - ✳ Leading to a flat histogram of the rank of each observation with the ensemble



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.

Results for present day climatology

Results for present day climatology

- CMIP3 - AOGCMs and ASGCMs - Multi-model ensembles (MME).

Results for present day climatology

- CMIP3 - AOGCMs and ASGCMs - Multi-model ensembles (MME).
- Several single model ensembles with varied parameters (SME) - MIROC3.2.2, HADSM3, HADCM3, NCAR-CAM3.

Results for present day climatology

- CMIP3 - AOGCMs and ASGCMs - Multi-model ensembles (MME).
- Several single model ensembles with varied parameters (SME) - MIROC3.2.2, HADSM3, HADCM3, NCAR-CAM3.
- 12 fields of annual mean climate observations - temperature, precipitation, radiation, clouds.

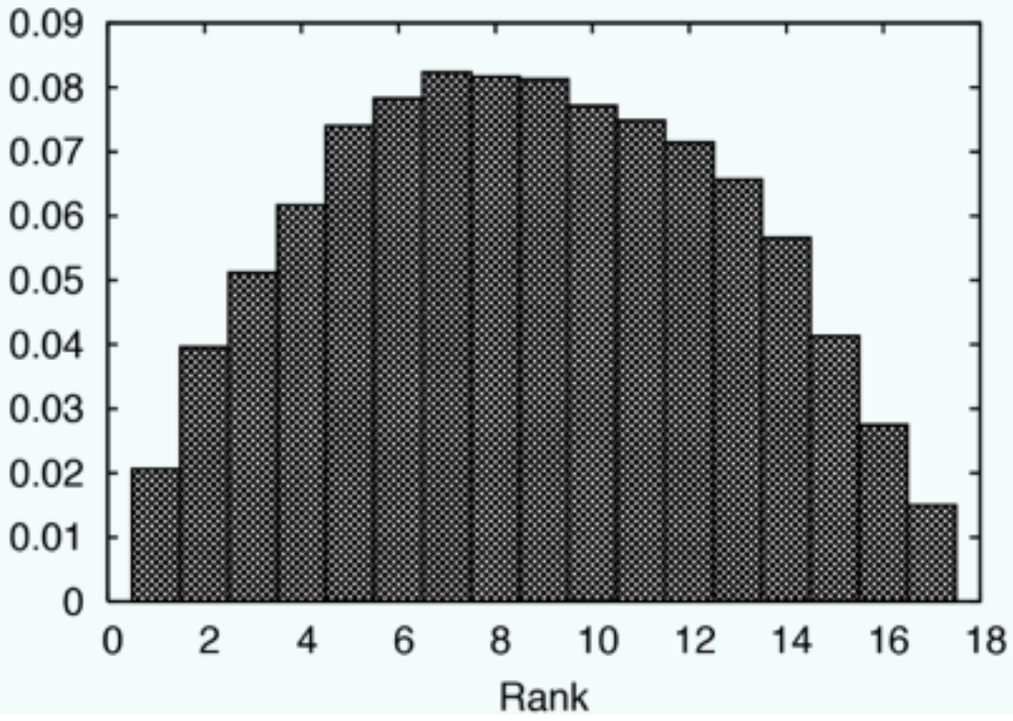
Results for present day climatology

- CMIP3 - AOGCMs and ASGCMs - Multi-model ensembles (MME).
- Several single model ensembles with varied parameters (SME) - MIROC3.2.2, HADSM3, HADCM3, NCAR-CAM3.
- 12 fields of annual mean climate observations - temperature, precipitation, radiation, clouds.
- Assessment over broad scales (globally integrated).

Results for present day climatology

- CMIP3 - AOGCMs and ASGCMs - Multi-model ensembles (MME).
- Several single model ensembles with varied parameters (SME) - MIROC3.2.2, HADSM3, HADCM3, NCAR-CAM3.
- 12 fields of annual mean climate observations - temperature, precipitation, radiation, clouds.
- Assessment over broad scales (globally integrated).

MME - CMIP3 AOGCMs

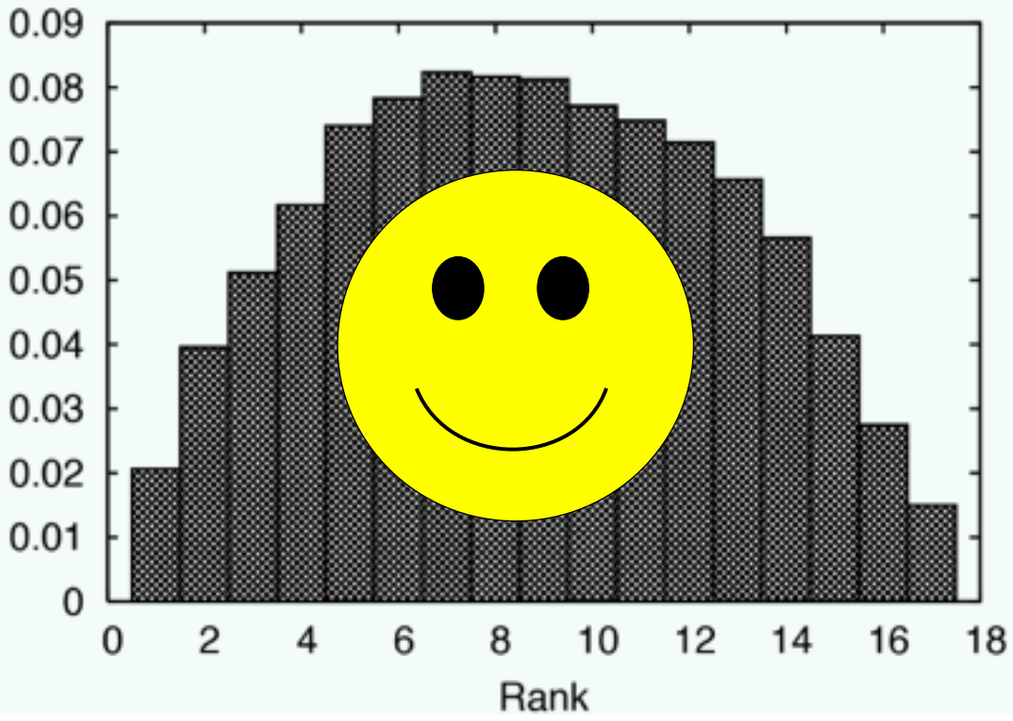


Yokohata et al, Climate Dynamics, 2011
Poster W247B - today!

Results for present day climatology

- CMIP3 - AOGCMs and ASGCMs - Multi-model ensembles (MME).
- Several single model ensembles with varied parameters (SME) - MIROC3.2.2, HADSM3, HADCM3, NCAR-CAM3.
- 12 fields of annual mean climate observations - temperature, precipitation, radiation, clouds.
- Assessment over broad scales (globally integrated).

MME - CMIP3 AOGCMs

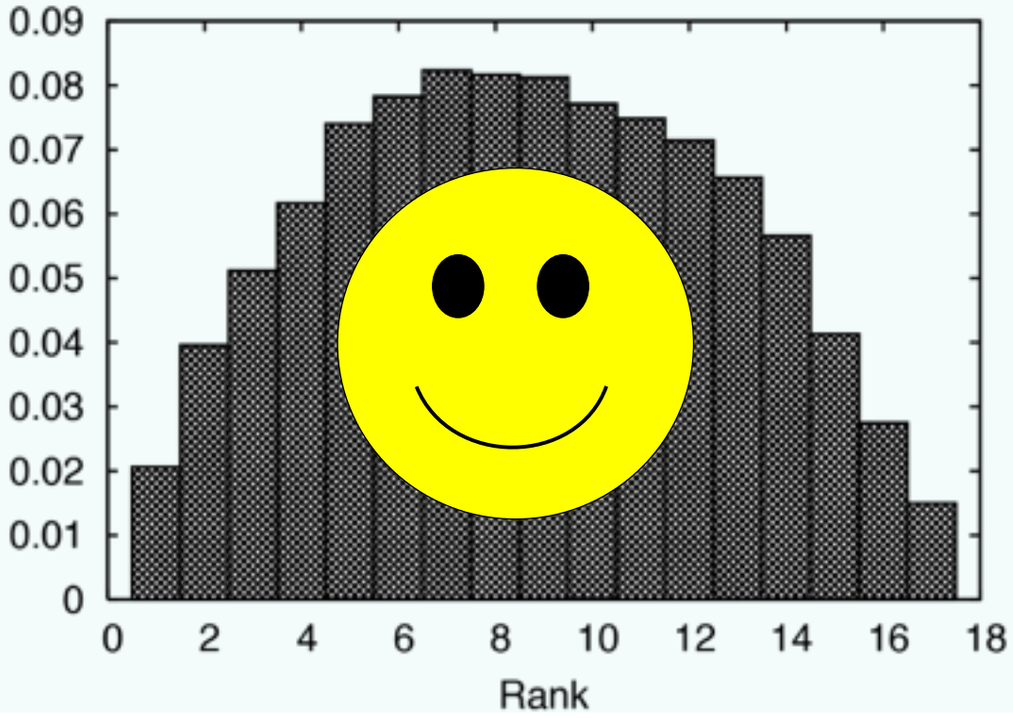


Yokohata et al, Climate Dynamics, 2011
Poster W247B - today!

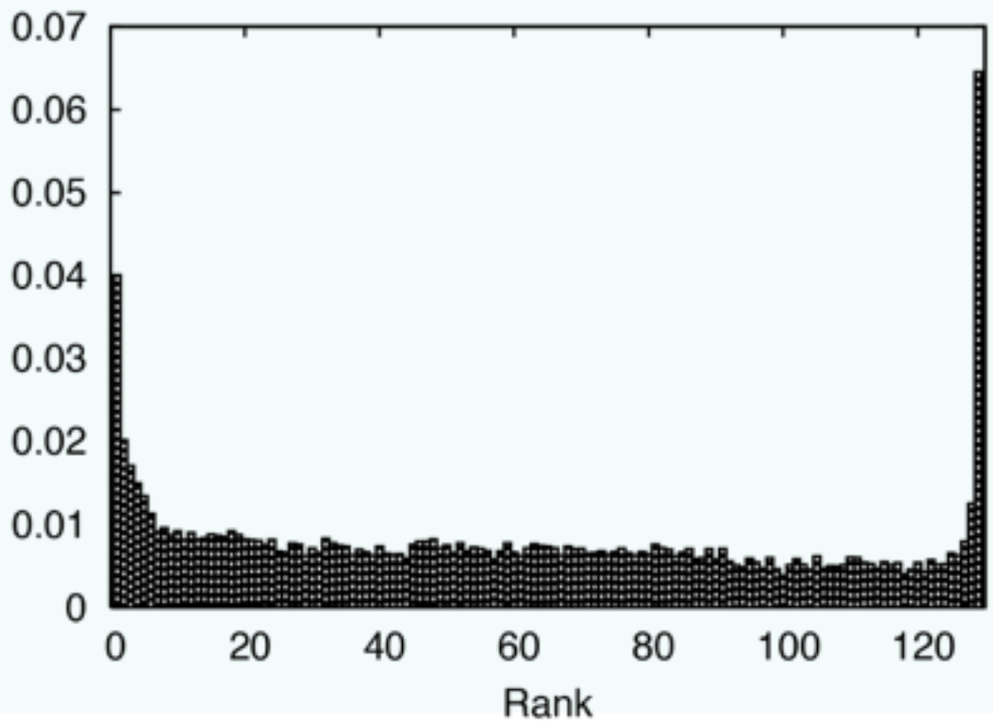
Results for present day climatology

- CMIP3 - AOGCMs and ASGCMs - Multi-model ensembles (MME).
- Several single model ensembles with varied parameters (SME) - MIROC3.2.2, HADSM3, HADCM3, NCAR-CAM3.
- 12 fields of annual mean climate observations - temperature, precipitation, radiation, clouds.
- Assessment over broad scales (globally integrated).

MME - CMIP3 AOGCMs



Typical SME

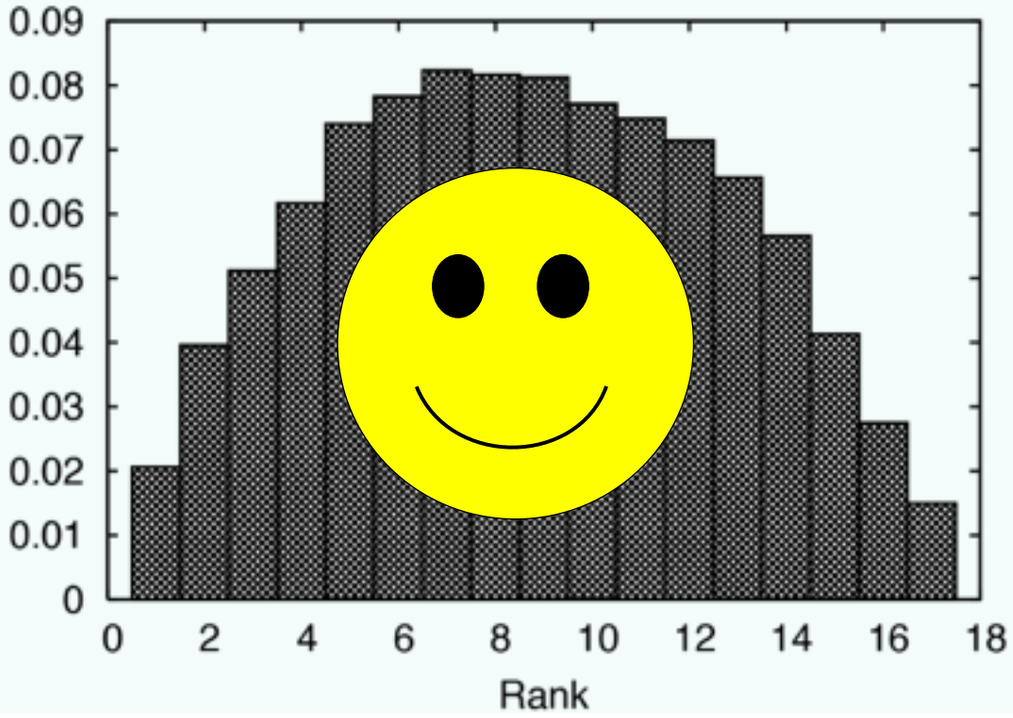


Yokohata et al, Climate Dynamics, 2011
Poster W247B - today!

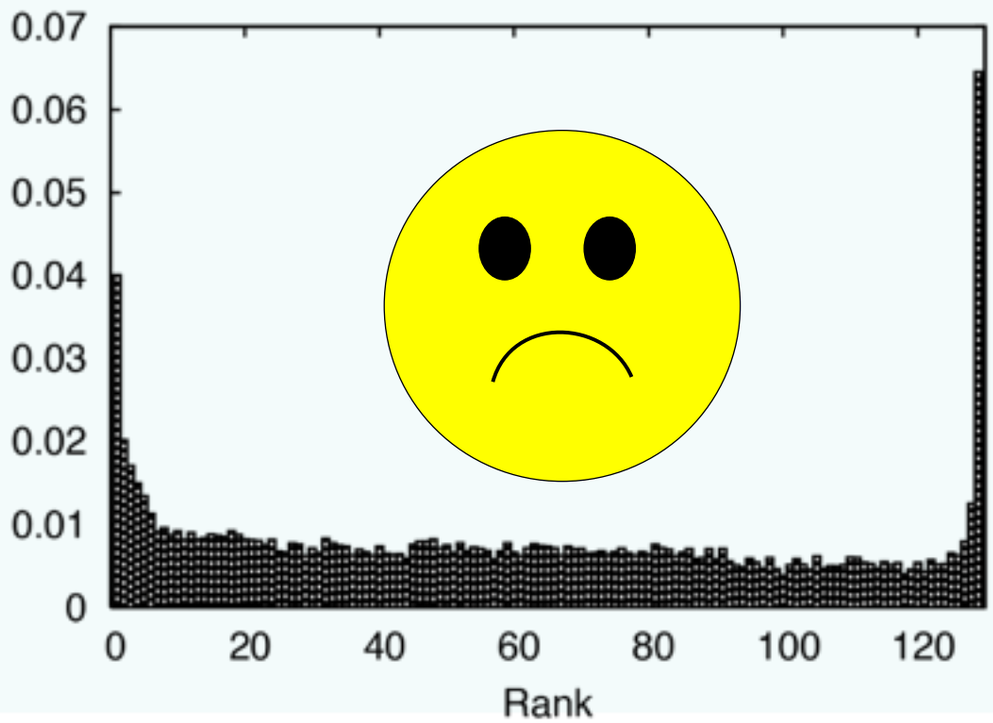
Results for present day climatology

- CMIP3 - AOGCMs and ASGCMs - Multi-model ensembles (MME).
- Several single model ensembles with varied parameters (SME) - MIROC3.2.2, HADSM3, HADCM3, NCAR-CAM3.
- 12 fields of annual mean climate observations - temperature, precipitation, radiation, clouds.
- Assessment over broad scales (globally integrated).

MME - CMIP3 AOGCMs



Typical SME



WCRP OSC

Yokohata et al, Climate Dynamics, 2011
Poster W247B - today!

KAKUSHIN

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

JUMP Japan Uncertainty Modelling Project

Out-of-sample evaluation

Out-of-sample evaluation

- Modern data may have been used in model building/ tuning.

Out-of-sample evaluation

- Modern data may have been used in model building/ tuning.
- Assessment of reliability for present day climate does not mean that projections of climate change are reliable.

Out-of-sample evaluation

- Modern data may have been used in model building/ tuning.
- Assessment of reliability for present day climate does not mean that projections of climate change are reliable.
- Analysis of paleoclimate helps us to assess our confidence.

The Last Glacial Maximum

The Last Glacial Maximum

- 21ka BP

The Last Glacial Maximum

- 21ka BP
- Large ice sheets over Northern Hemisphere

The Last Glacial Maximum

- 21ka BP
- Large ice sheets over Northern Hemisphere
- Atmospheric carbon dioxide significantly lower than pre-industrial, 185ppm. Other GHGs changed.

The Last Glacial Maximum

- 21ka BP
- Large ice sheets over Northern Hemisphere
- Atmospheric carbon dioxide significantly lower than pre-industrial, 185ppm. Other GHGs changed.
- Slightly different solar forcing

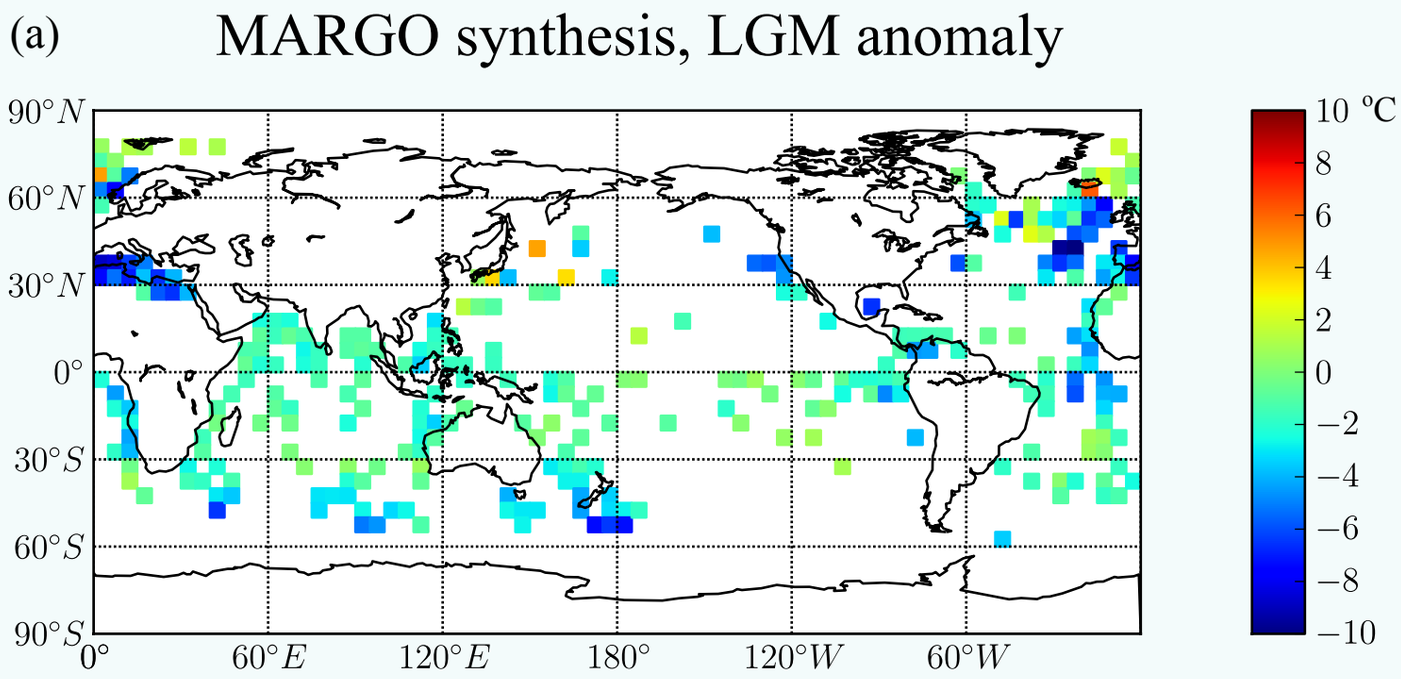
The Last Glacial Maximum

- 21ka BP
- Large ice sheets over Northern Hemisphere
- Atmospheric carbon dioxide significantly lower than pre-industrial, 185ppm. Other GHGs changed.
- Slightly different solar forcing
- More dust in atmosphere (not included in models)

Data

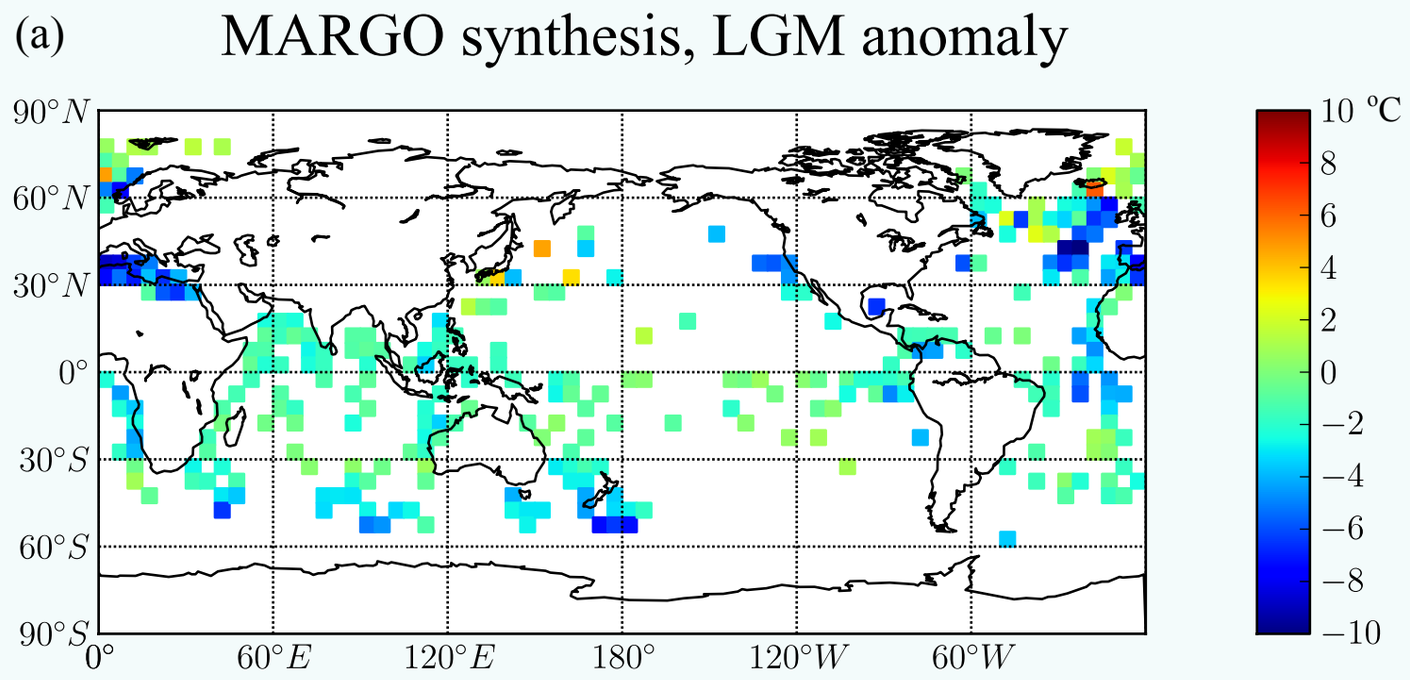
MARGO Project Members, Nature Geoscience, 2009

Data


(a) MARGO synthesis, LGM anomaly
90°N
60°N
30°N
0°
30°S
60°S
90°S
0°
60°E
120°E
180°
120°W
60°W
10 ºC
8
6
4
2
0
−2
−4
−6
−8
−10

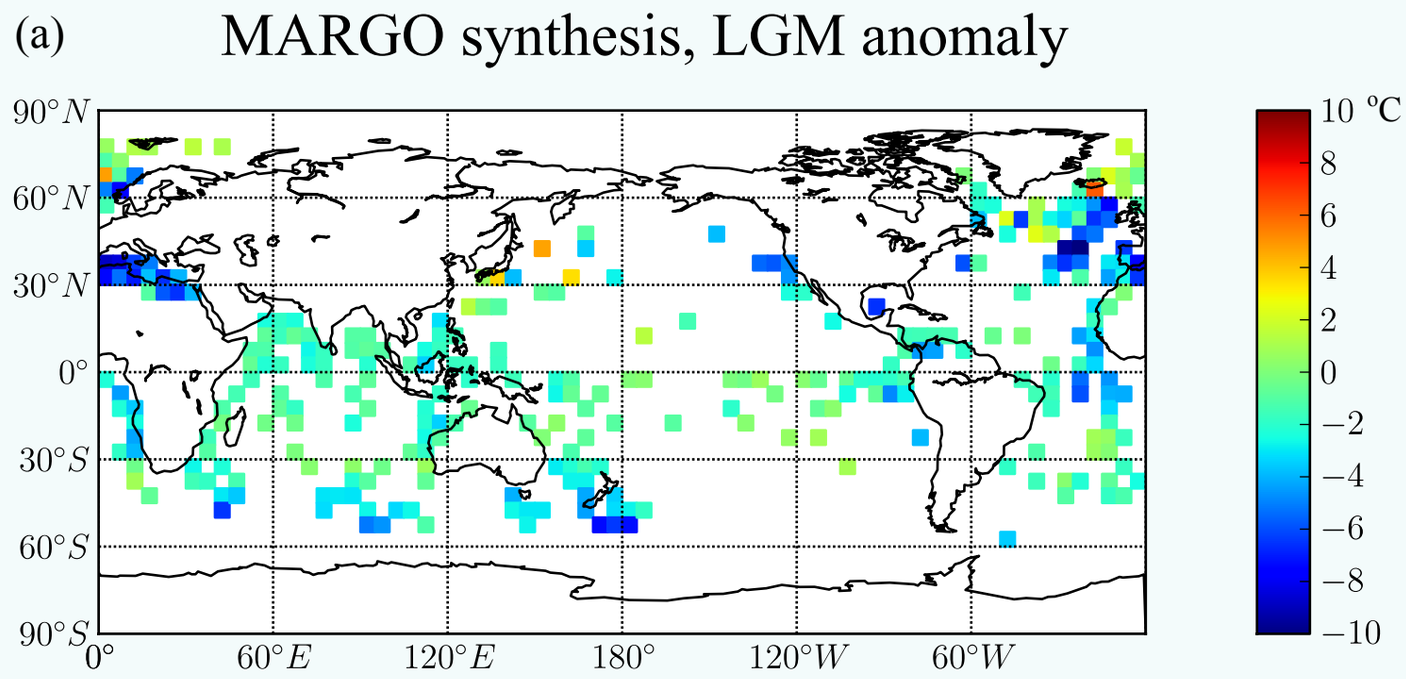
Data

- MARGO SST, combination of several different proxy data types.



Data

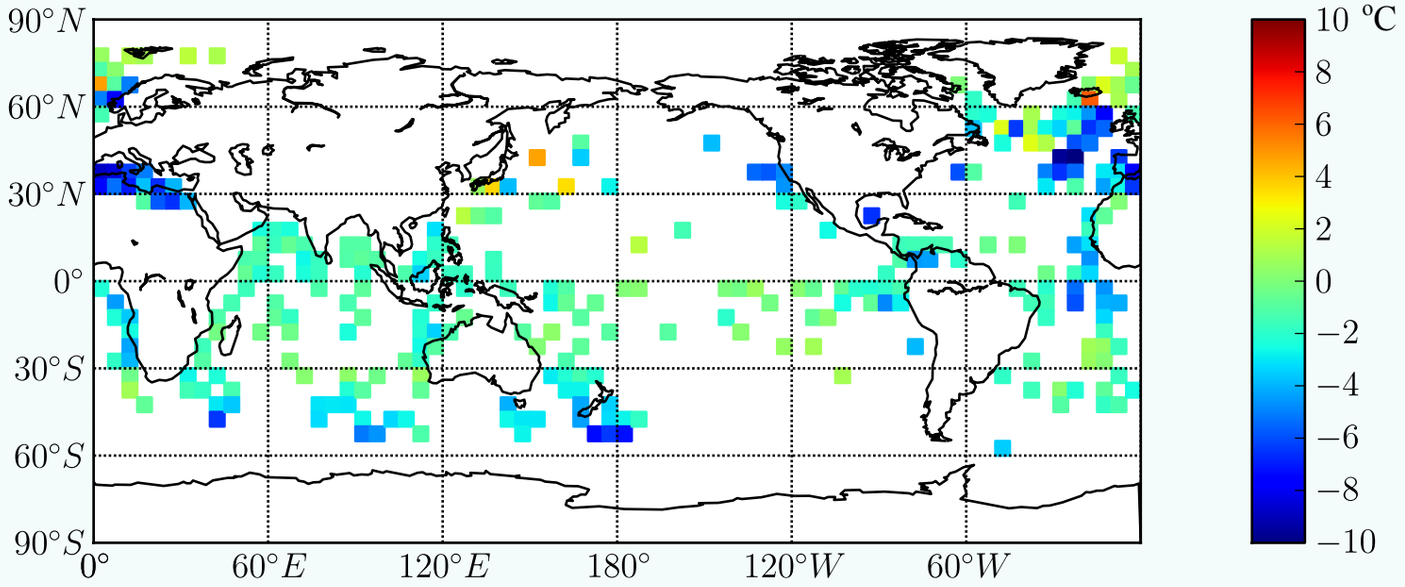
- MARGO SST, combination of several different proxy data types.
- About 315 points over the globe on 5 degree grid.



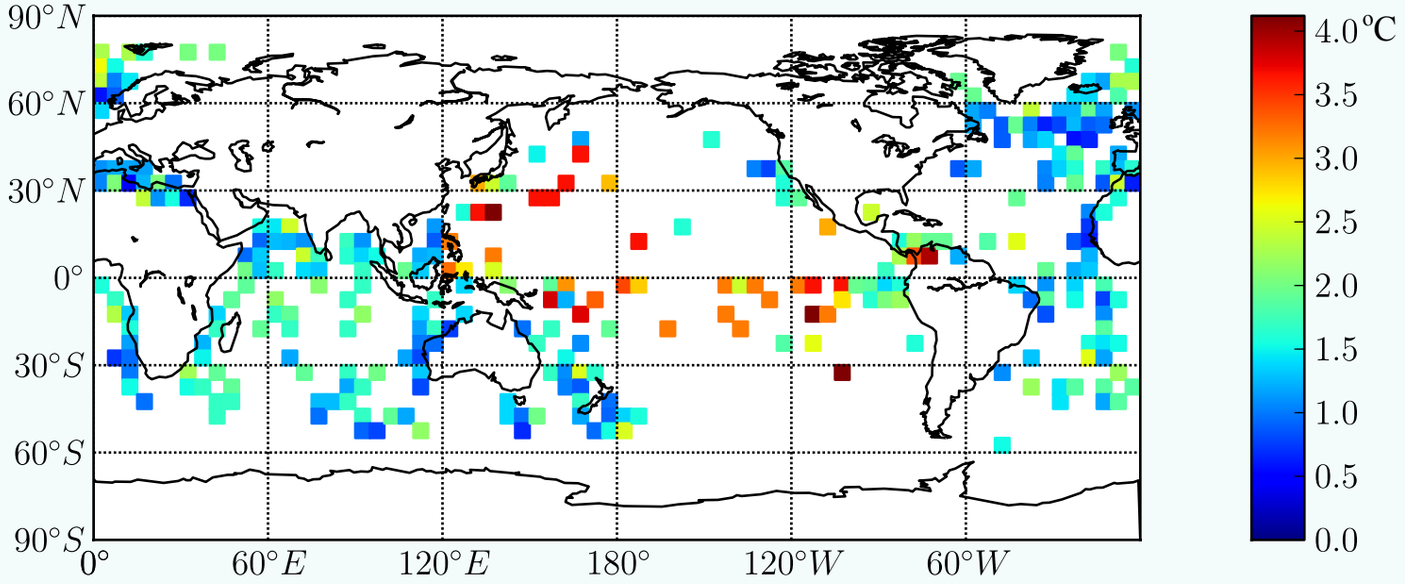
Data

- MARGO SST, combination of several different proxy data types.
- About 315 points over the globe on 5 degree grid.
- Includes an error estimate!

(a) MARGO synthesis, LGM anomaly



(b) MARGO synthesis, uncertainty in LGM anomaly



MARGO Project Members, Nature Geoscience, 2009

Model Ensembles

Model Ensembles

- "Paleoclimate Modelling Intercomparison Project" - PMIP

Model Ensembles

- “Paleoclimate Modelling Intercomparison Project” - PMIP
- MME - PMIP2 - AOGCMs (dynamical ocean)

Model Ensembles

- "Paleoclimate Modelling Intercomparison Project" - PMIP
- MME - PMIP2 - AOGCMs (dynamical ocean)
  - Sea surface temperature (SST) for 9 models

Model Ensembles

- "Paleoclimate Modelling Intercomparison Project" - PMIP
- MME - PMIP2 - AOGCMs (dynamical ocean)
  - ✳ Sea surface temperature (SST) for 9 models
- SME - MIROC3.2.2 T21 slab ocean model version

Model Ensembles

- "Paleoclimate Modelling Intercomparison Project" - PMIP
- MME - PMIP2 - AOGCMs (dynamical ocean)
  - ✳ Sea surface temperature (SST) for 9 models
- SME - MIROC3.2.2 T21 slab ocean model version
  - ✳ 40 member ensemble with 13 parameters varied tuned to modern climate using EnKF.

Including data error

Including data error

- Unlike most data derived from paleoclimates, MARGO data have an uncertainty included, ranging from 0.6C to 4.2C (average 1.6C).

Including data error

- Unlike most data derived from paleoclimates, MARGO data have an uncertainty included, ranging from 0.6C to 4.2C (average 1.6C).
- Assume this error estimate to be the 1standard deviation of a Normal distribution.

Including data error

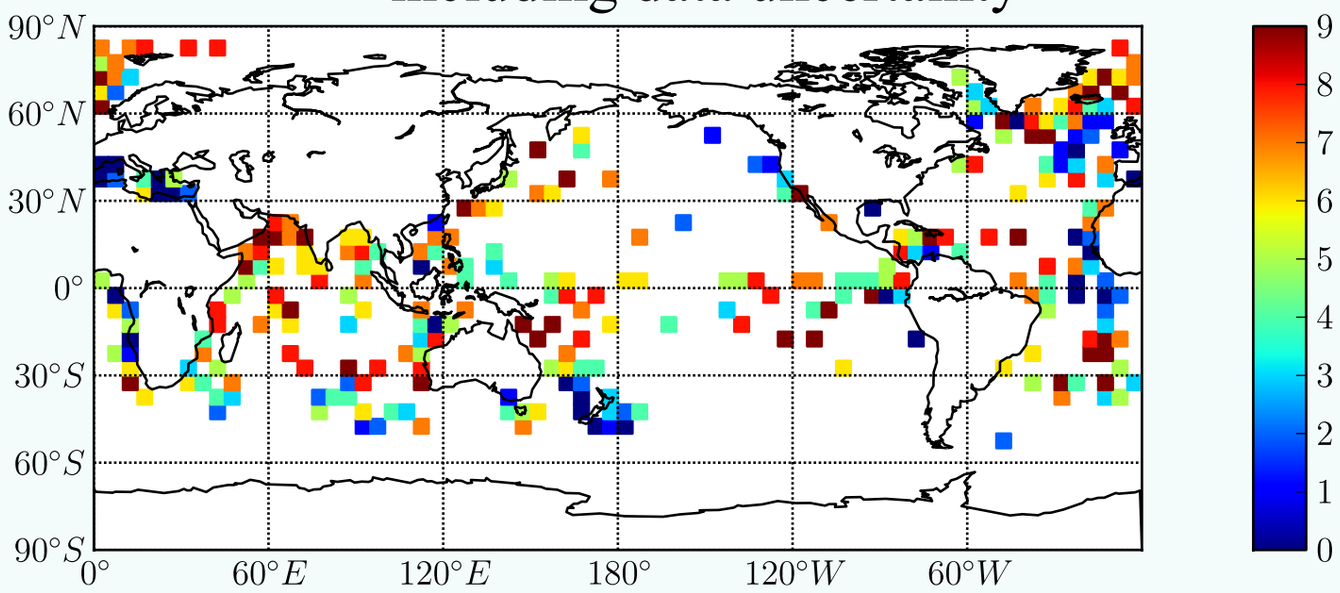
- Unlike most data derived from paleoclimates, MARGO data have an uncertainty included, ranging from 0.6C to 4.2C (average 1.6C).
- Assume this error estimate to be the 1standard deviation of a Normal distribution.
- Apply a random error of this size to each model point, and recalculate the rank histogram

Reliability with data error included

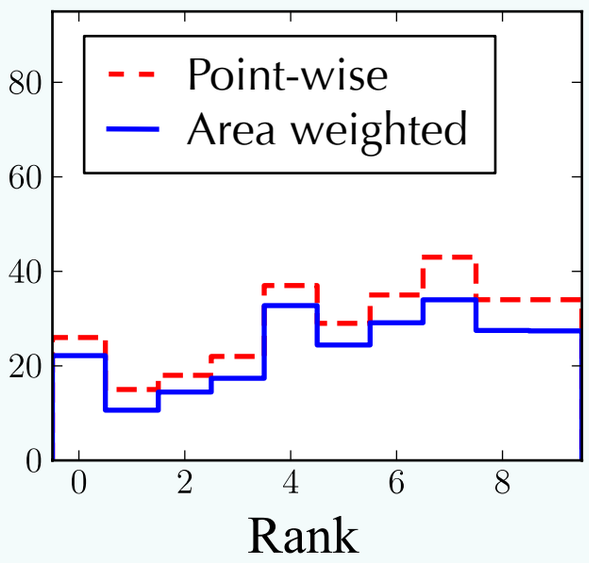
Hargreaves et al, Climate of the Past, 2011

Reliability with data error included

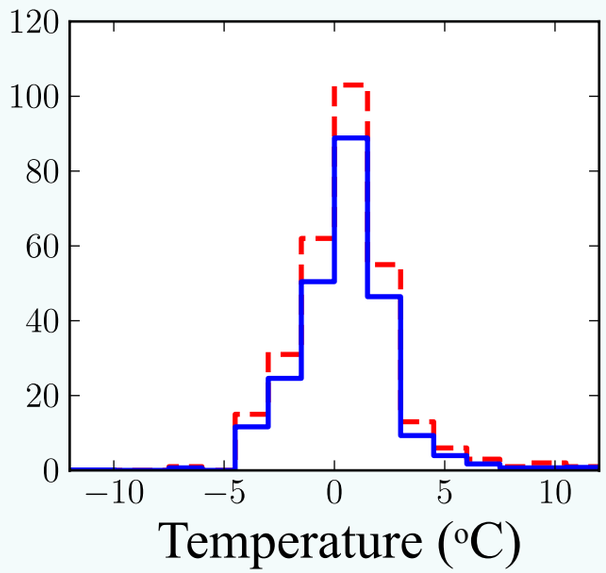
Rank of MARGO data in PMIP2 ensemble, including data uncertainty



(b) Rank histogram

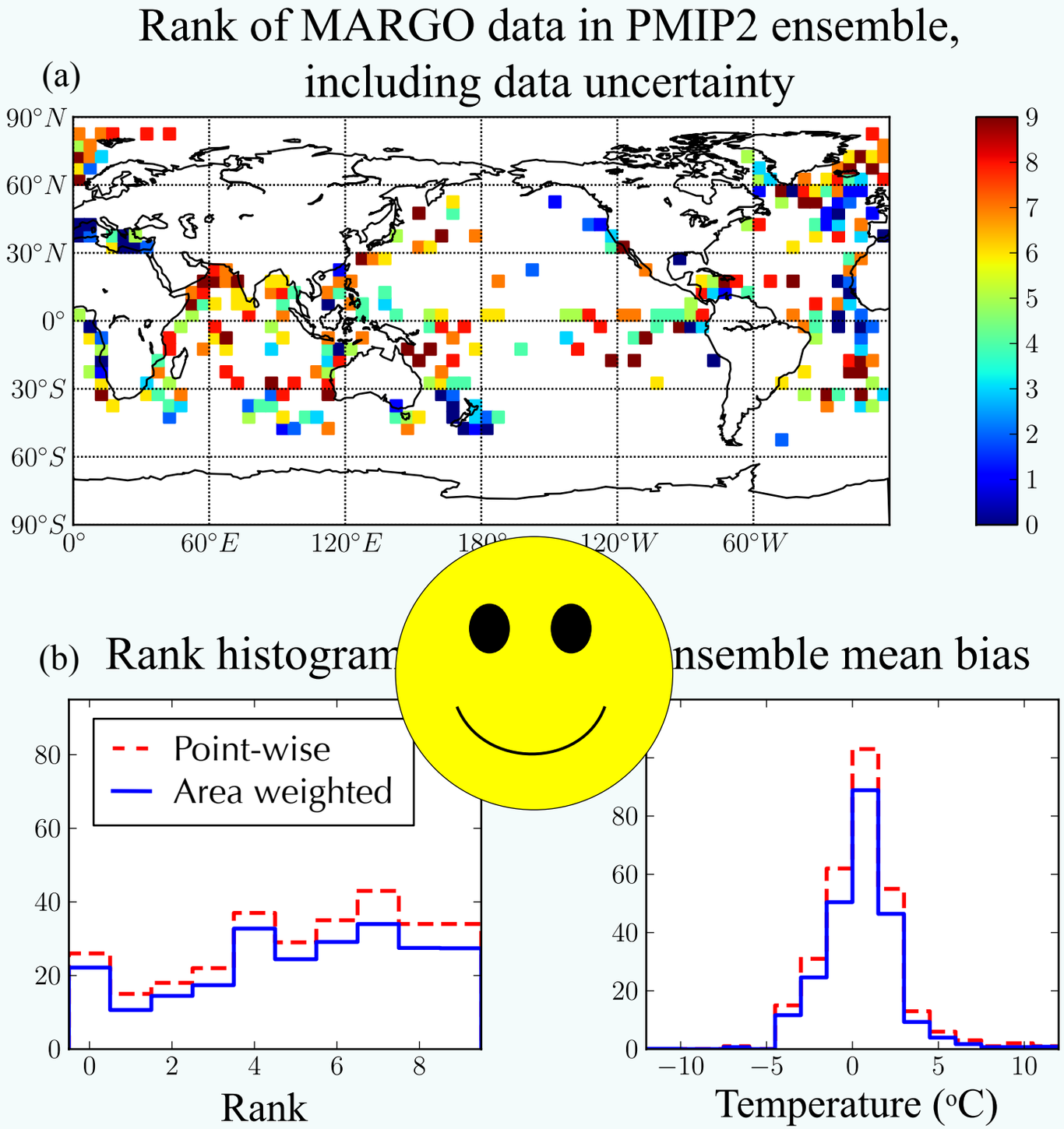


(c) Ensemble mean bias



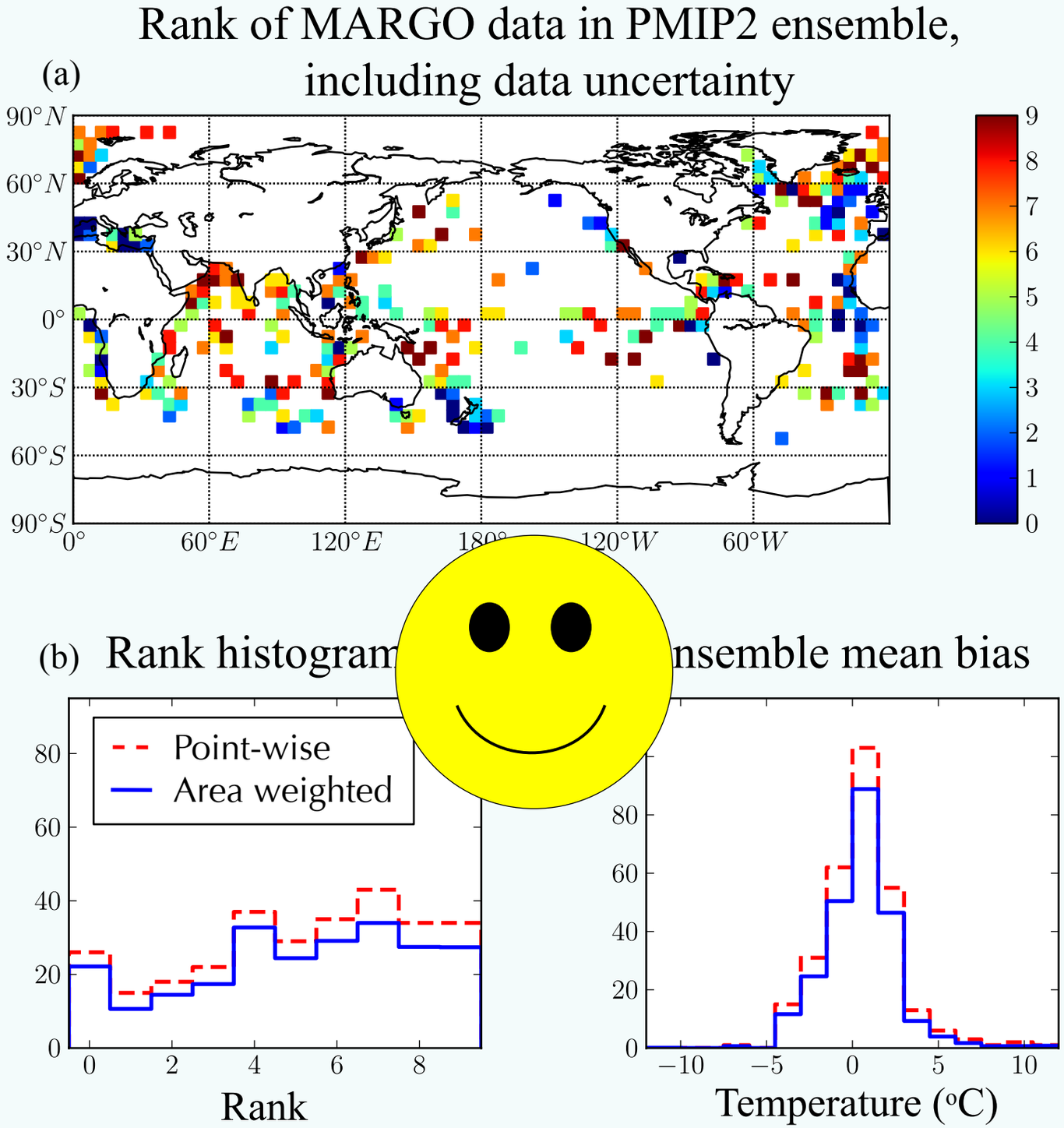
Hargreaves et al, Climate of the Past, 2011

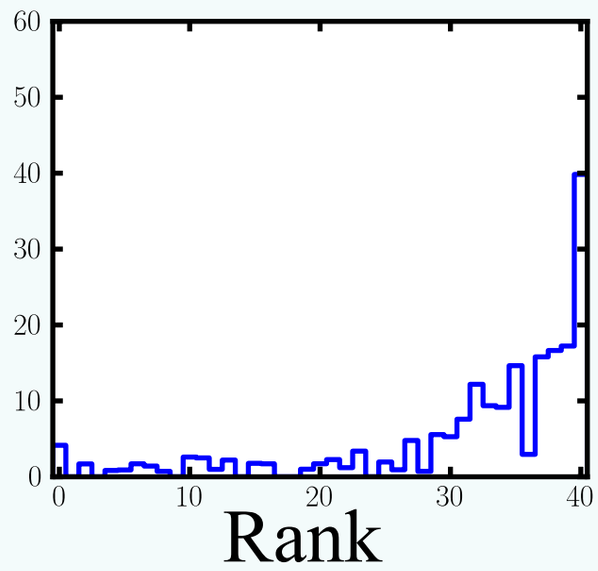
Reliability with data error included



Hargreaves et al, Climate of the Past, 2011

Reliability with data error included

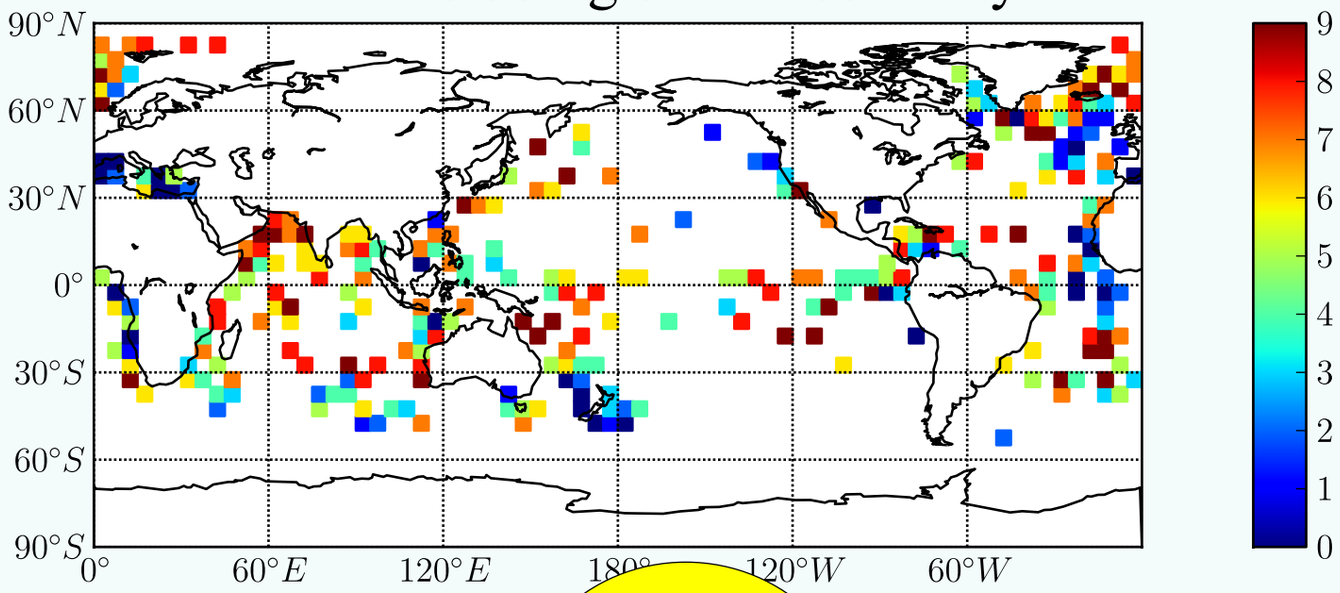


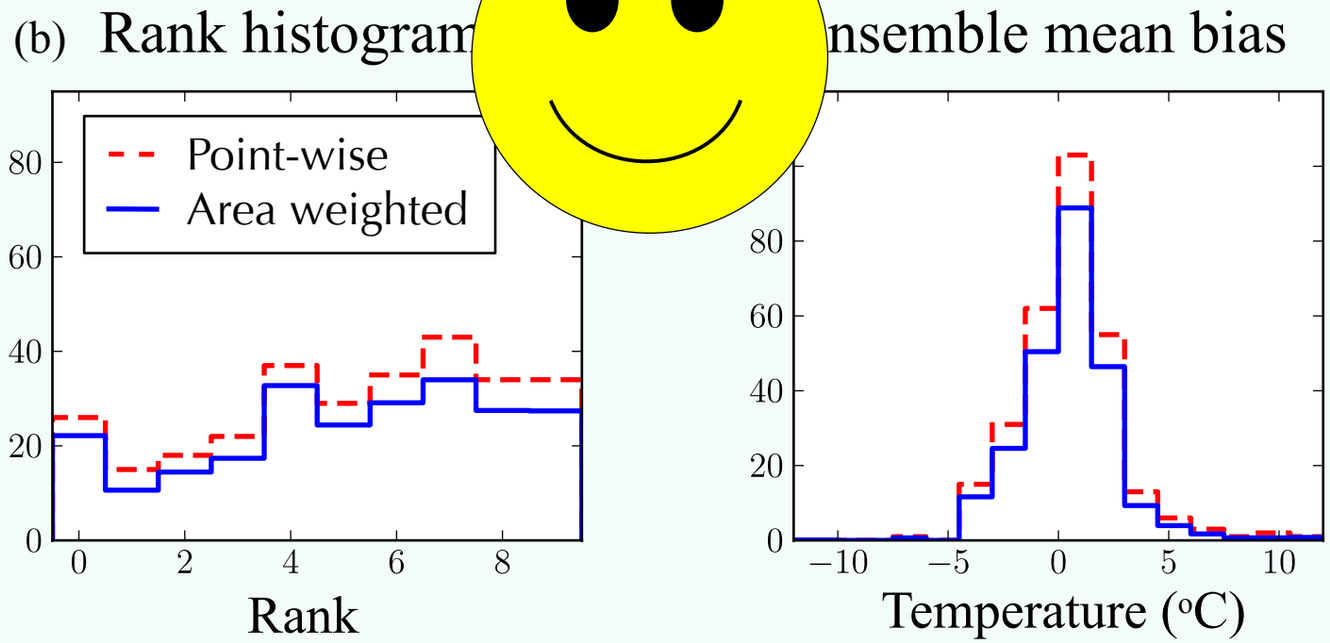


Hargreaves et al, Climate of the Past, 2011

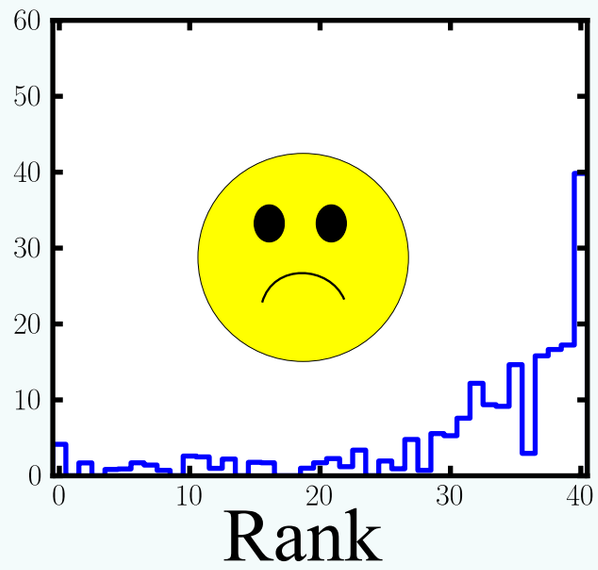
Reliability with data error included

Rank of MARGO data in PMIP2 ensemble, including data uncertainty





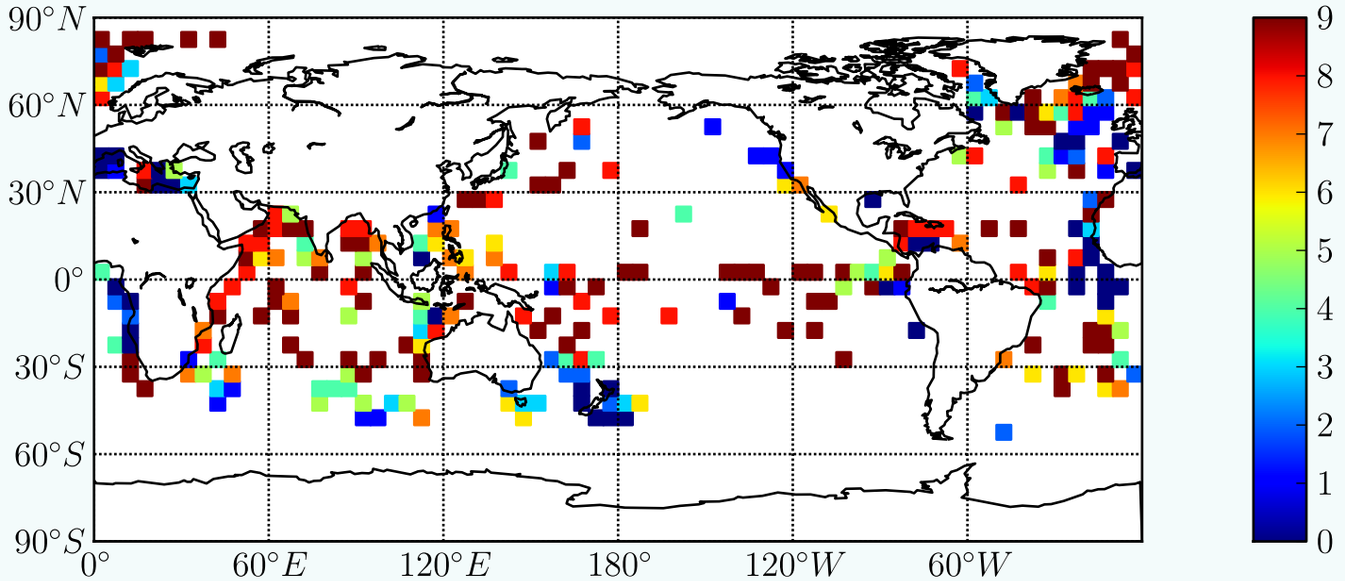
SME - MIROC



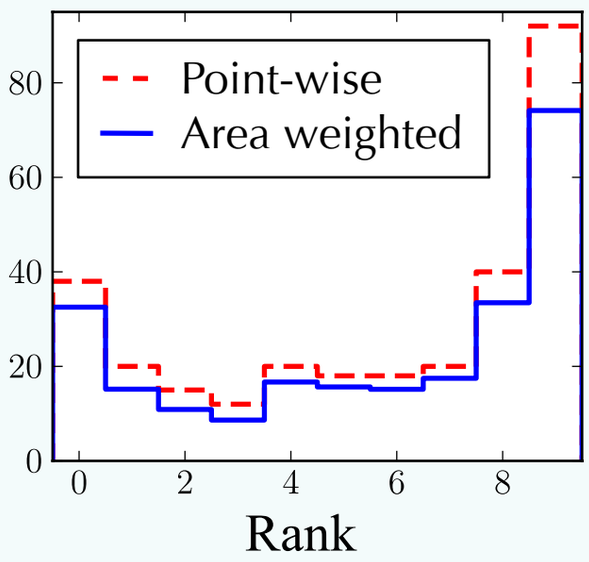
Hargreaves et al, Climate of the Past, 2011

Data error not included, PMIP2

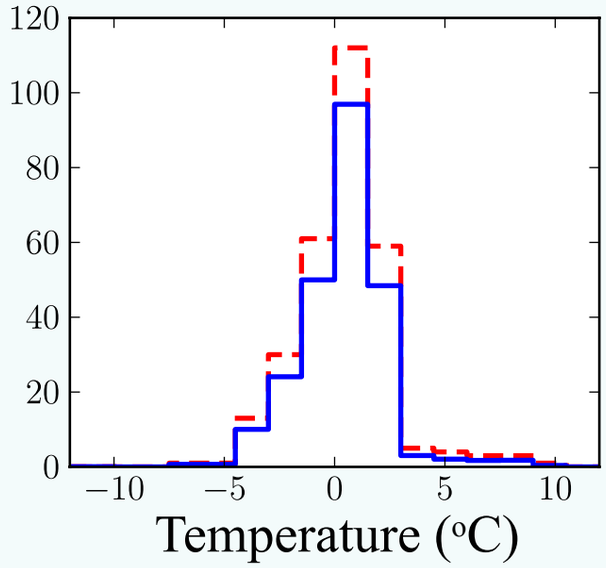
(a) Rank of MARGO data in PMIP2 ensemble



(b) Rank histogram



(c) Ensemble mean bias

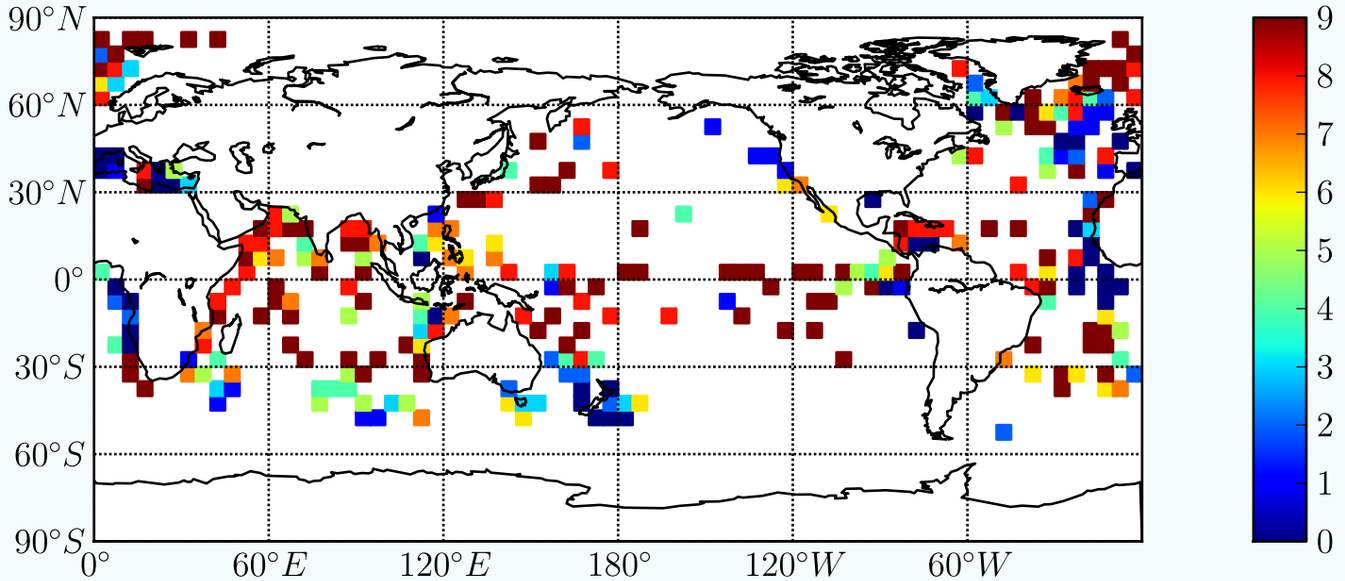


Hargreaves et al, Climate of the Past, 2011

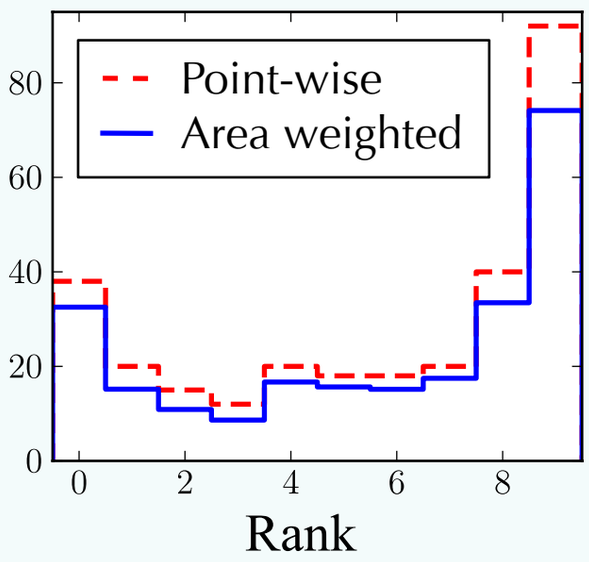
Data error not included, PMIP2

Statistically unreliable at 5% level - PMIP2 too sensitive.

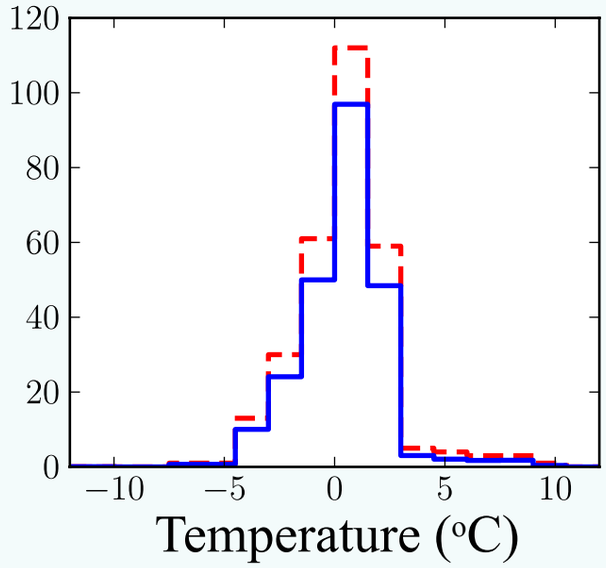
(a) Rank of MARGO data in PMIP2 ensemble



(b) Rank histogram



(c) Ensemble mean bias



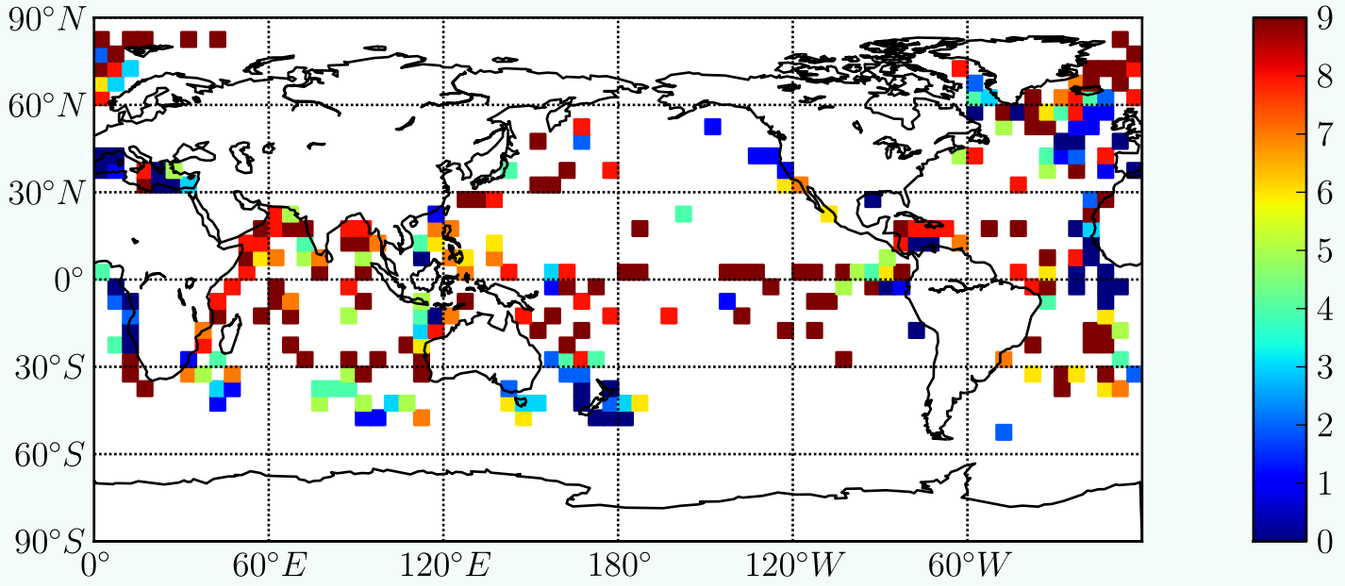
Hargreaves et al, Climate of the Past, 2011

Data error not included, PMIP2

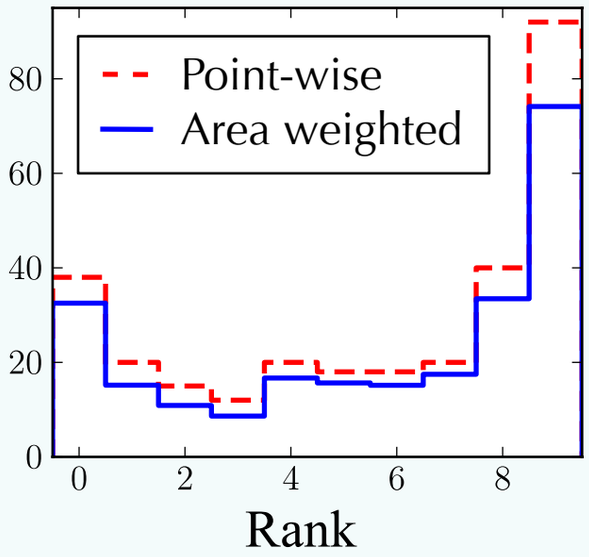
Statistically unreliable at 5% level - PMIP2 too sensitive.

Inclusion of data error very important in analyses of paleoclimate models

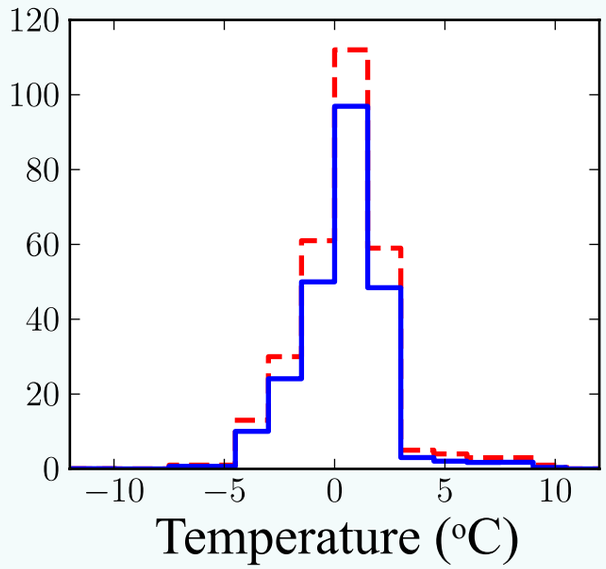
(a) Rank of MARGO data in PMIP2 ensemble



(b) Rank histogram



(c) Ensemble mean bias



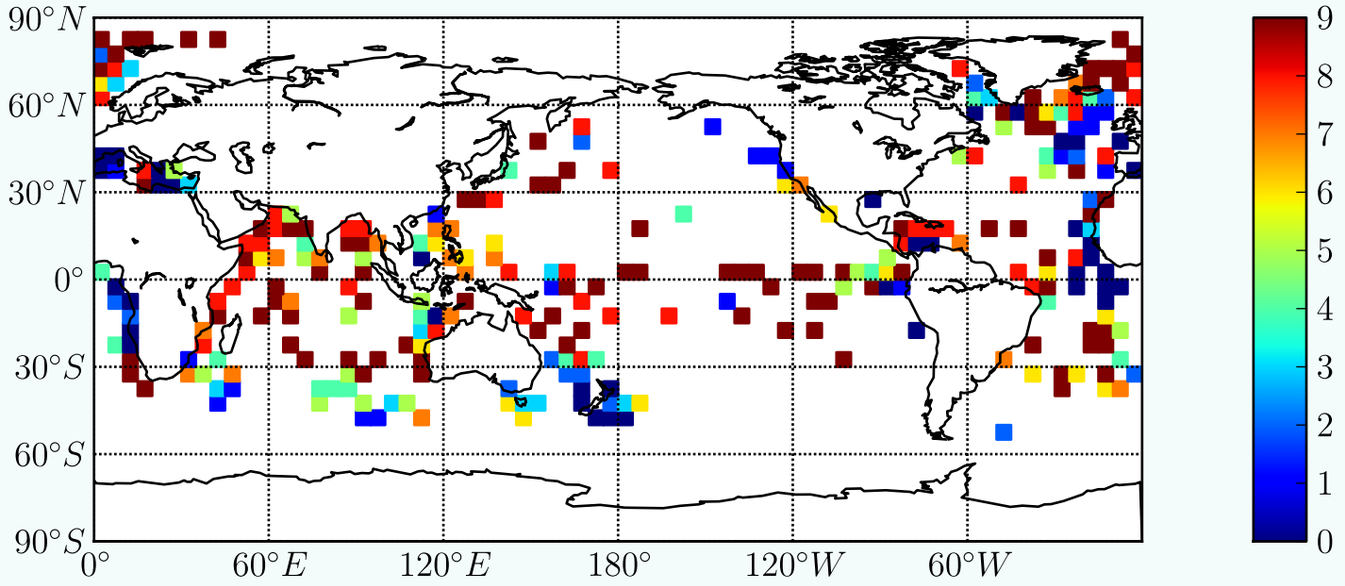
Hargreaves et al, Climate of the Past, 2011

Data error not included, PMIP2

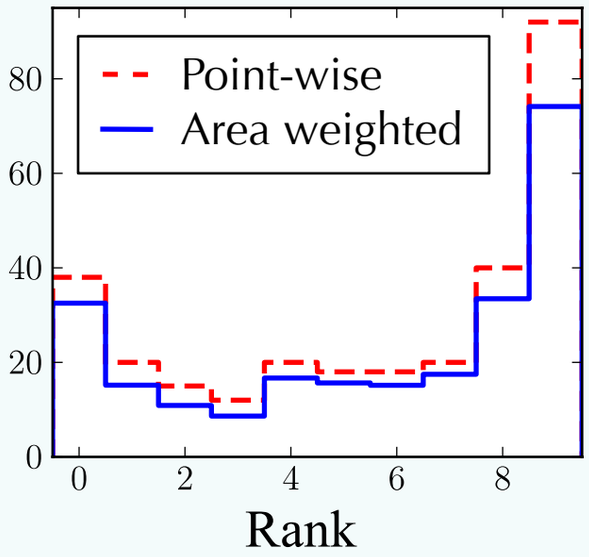
Statistically unreliable at 5% level - PMIP2 too sensitive.

Inclusion of data error very important in analyses of paleoclimate models

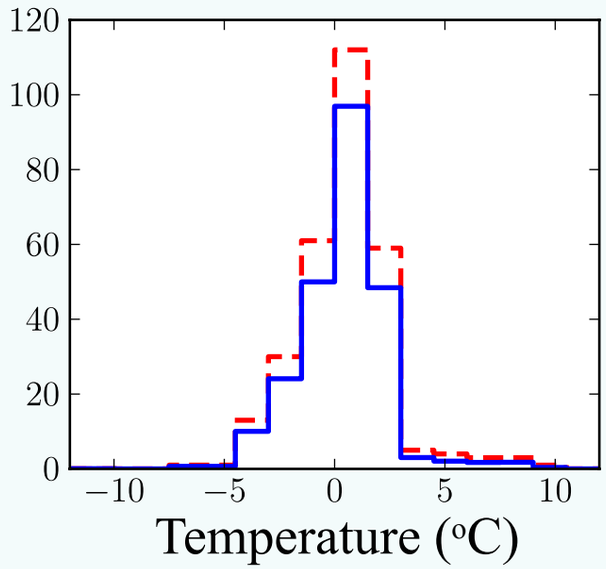
(a) Rank of MARGO data in PMIP2 ensemble



(b) Rank histogram



(c) Ensemble mean bias



Hargreaves et al, Climate of the Past, 2011

Do Climate Models have skill?

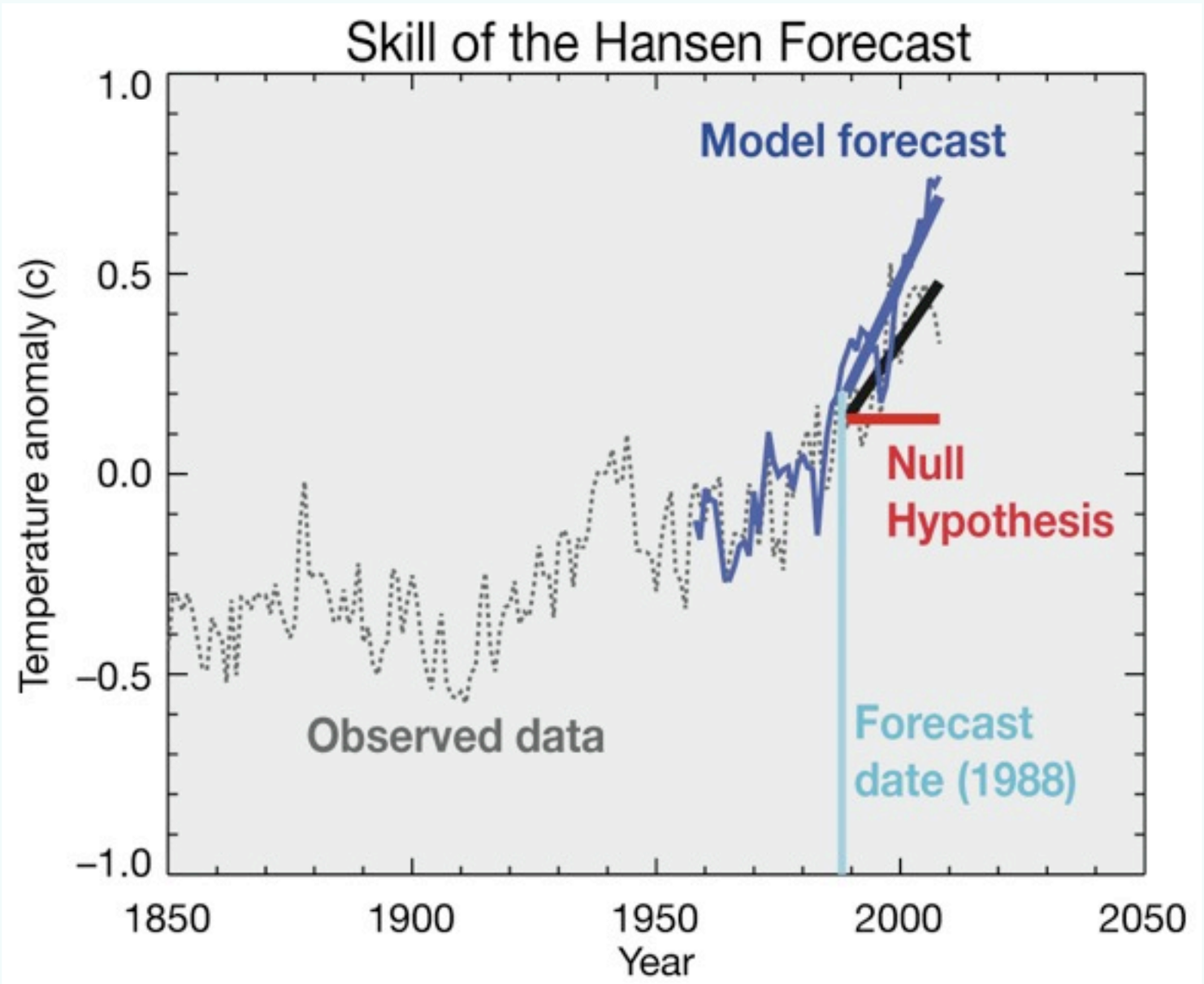
Do Climate Models have skill?

- Numerical Weather Prediction: skill is forecast performance in comparison to some other reference technique. eg. null hypothesis such as persistence.

Do Climate Models have skill?

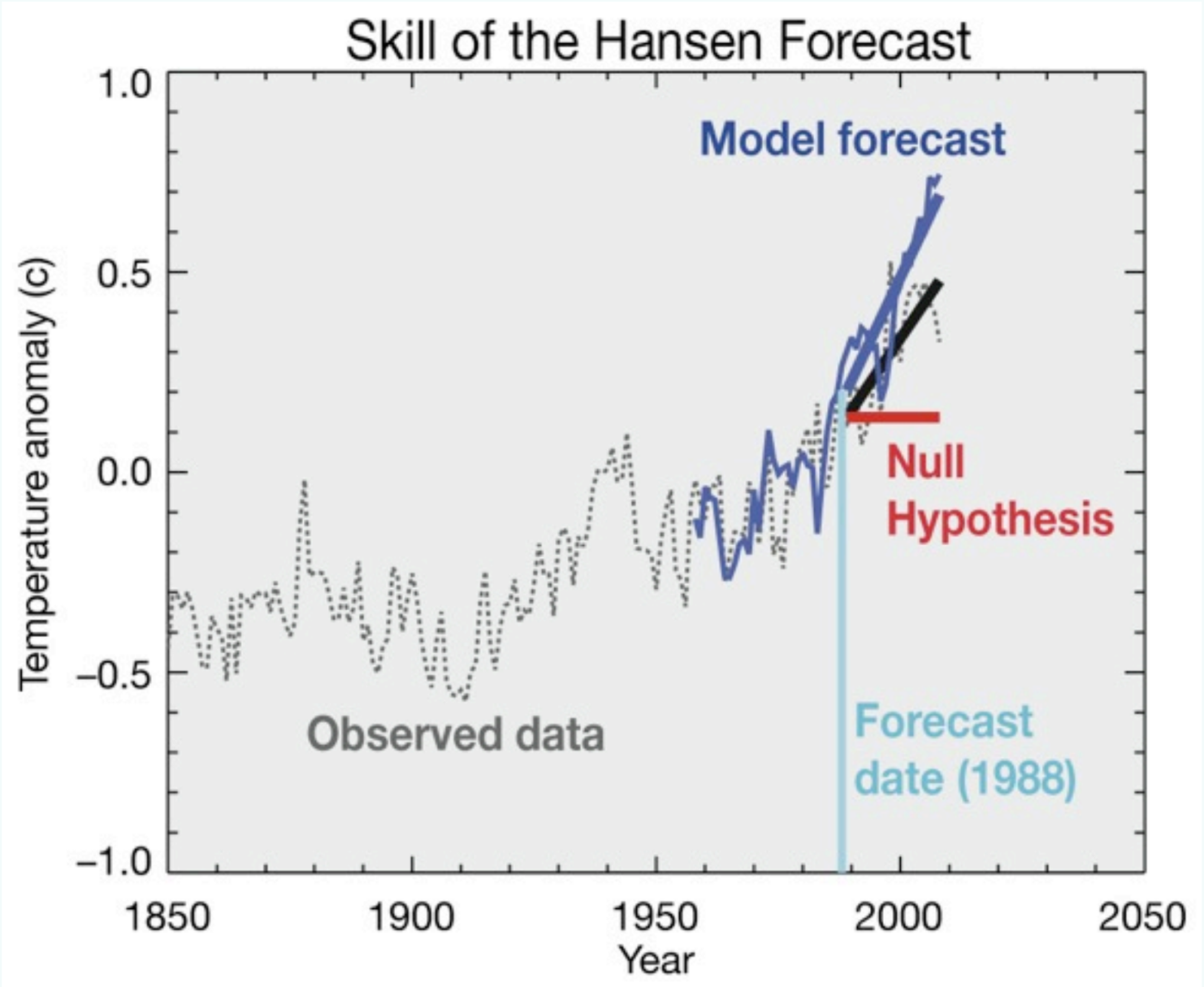
- Numerical Weather Prediction: skill is forecast performance in comparison to some other reference technique. eg. null hypothesis such as persistence.
- Climate prediction: Forecasts are not yet realised, so skill calculations generally not possible.

Skill analysis of a climate forecast



This Figure re-created by Hans von Storch's draftsman, Gruss Beate Gardeike

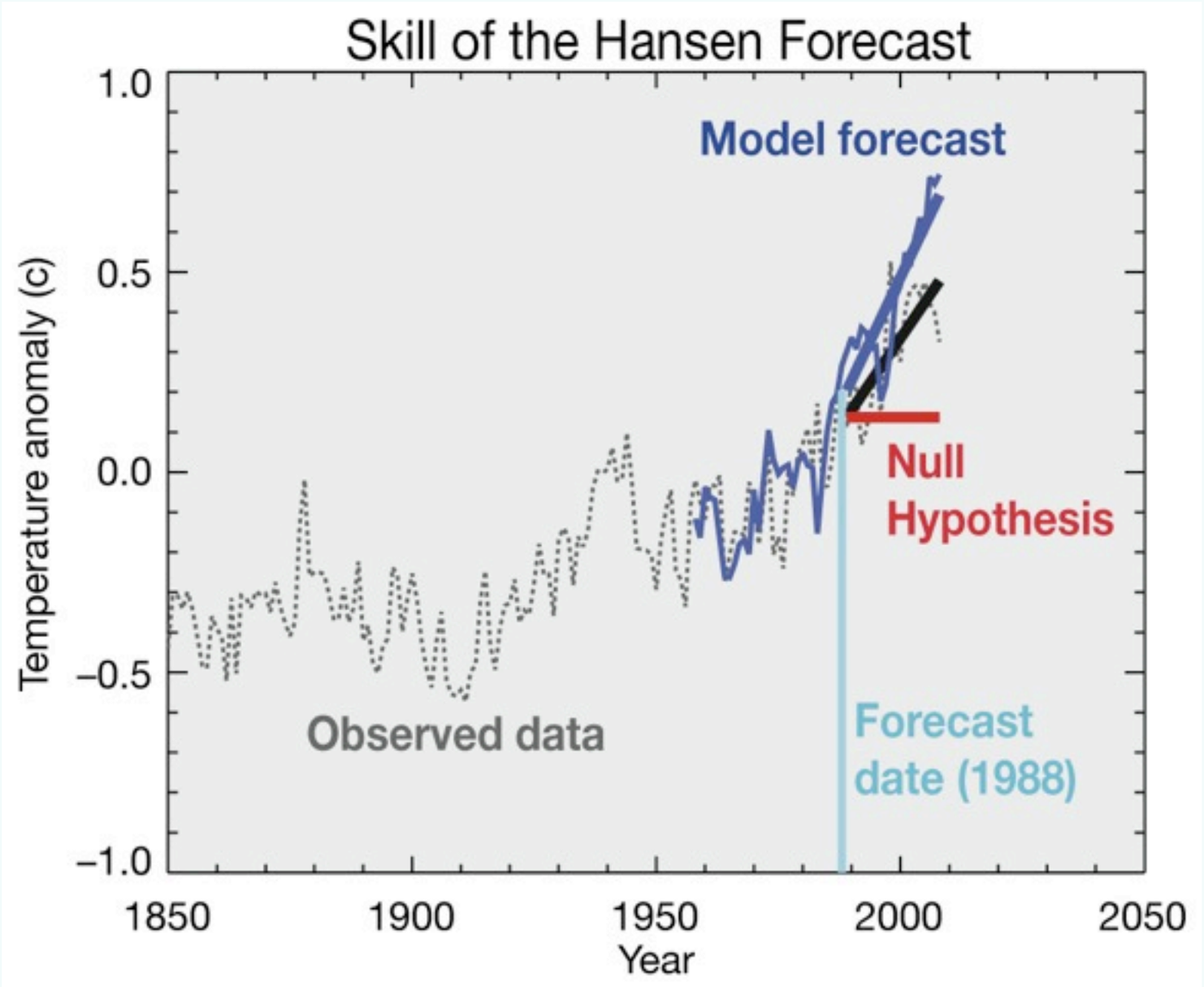
Skill analysis of a climate forecast



- 1989-2008:
  Hansen trend = 0.26C,
  HadCRUT trend=0.18C per decade.

This Figure re-created by Hans von Storch's draftsman, Gruss Beate Gardeike

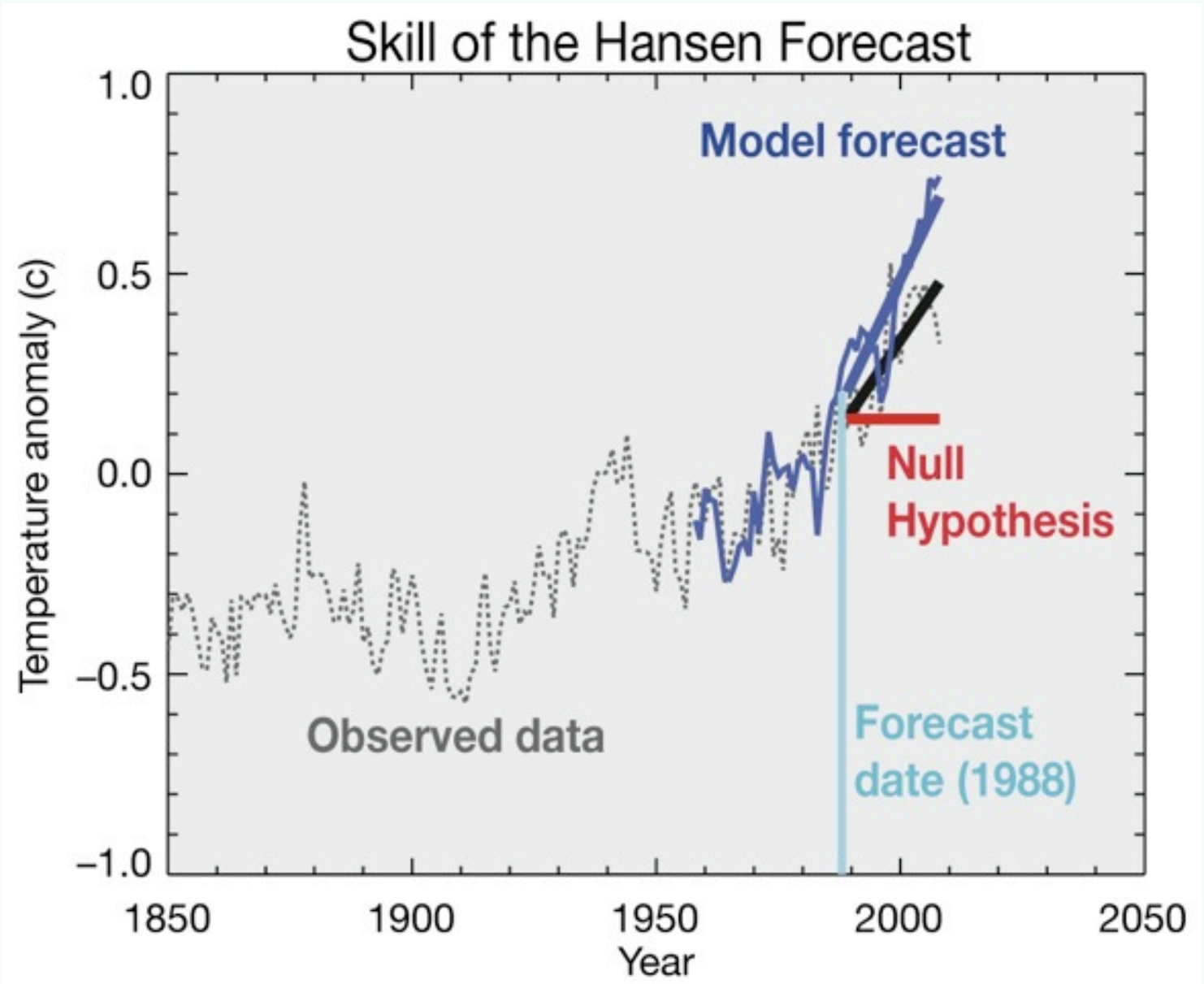
Skill analysis of a climate forecast



- 1989-2008:
  Hansen trend = 0.26C,
  HadCRUT trend=0.18C per decade.
- Null hypothesis:
  persistence over 20 years

This Figure re-created by Hans von Storch's draftsman, Gruss Beate Gardeike

Skill analysis of a climate forecast

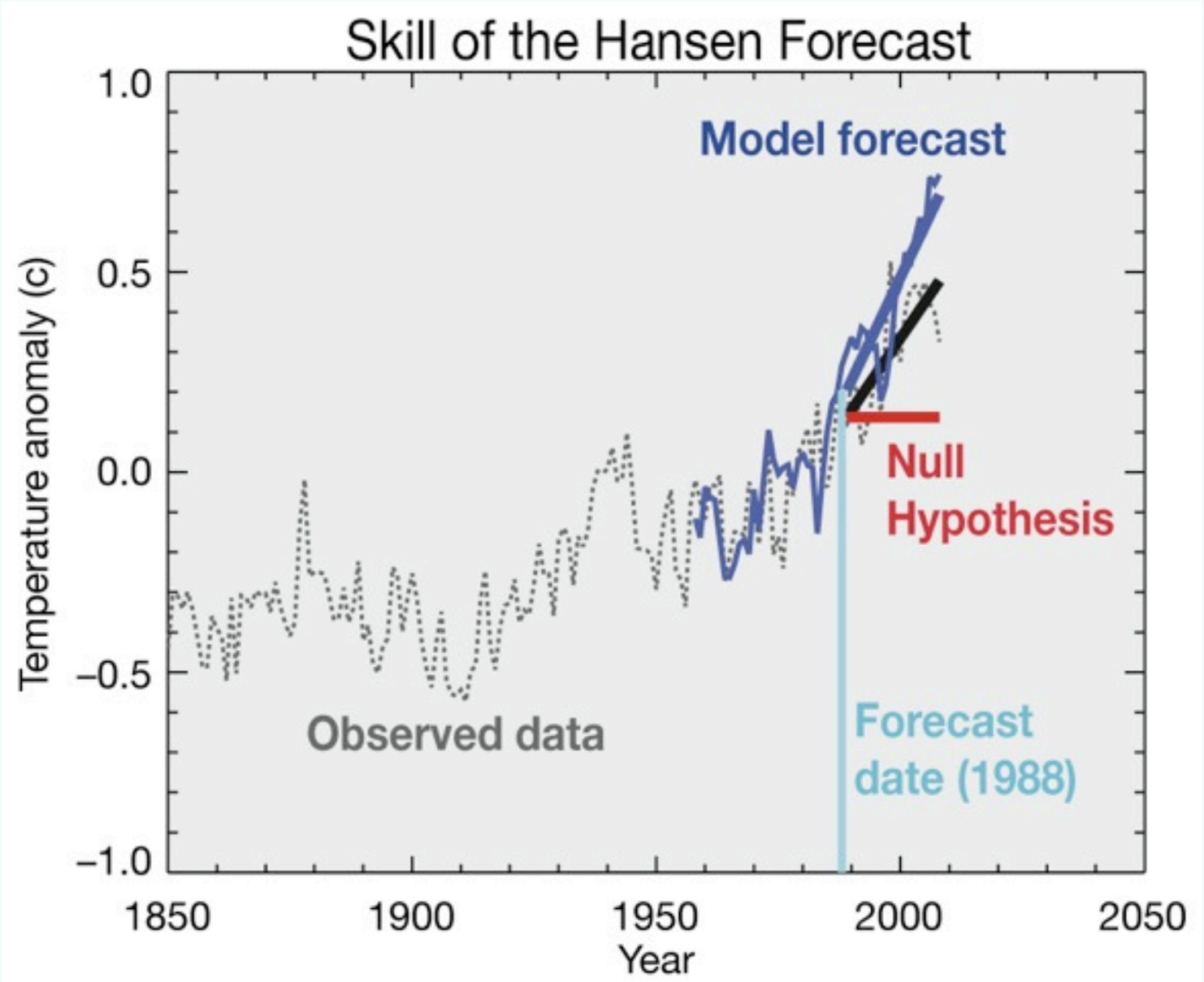


- 1989-2008:
  Hansen trend = 0.26C,
  HadCRUT trend=0.18C per decade.
- Null hypothesis:
  persistence over 20 years
- Skill=0.56 -> skillful!

This Figure re-created by Hans von Storch's draftsman, Gruss Beate Gardeike

Hargreaves, WIREs 2010

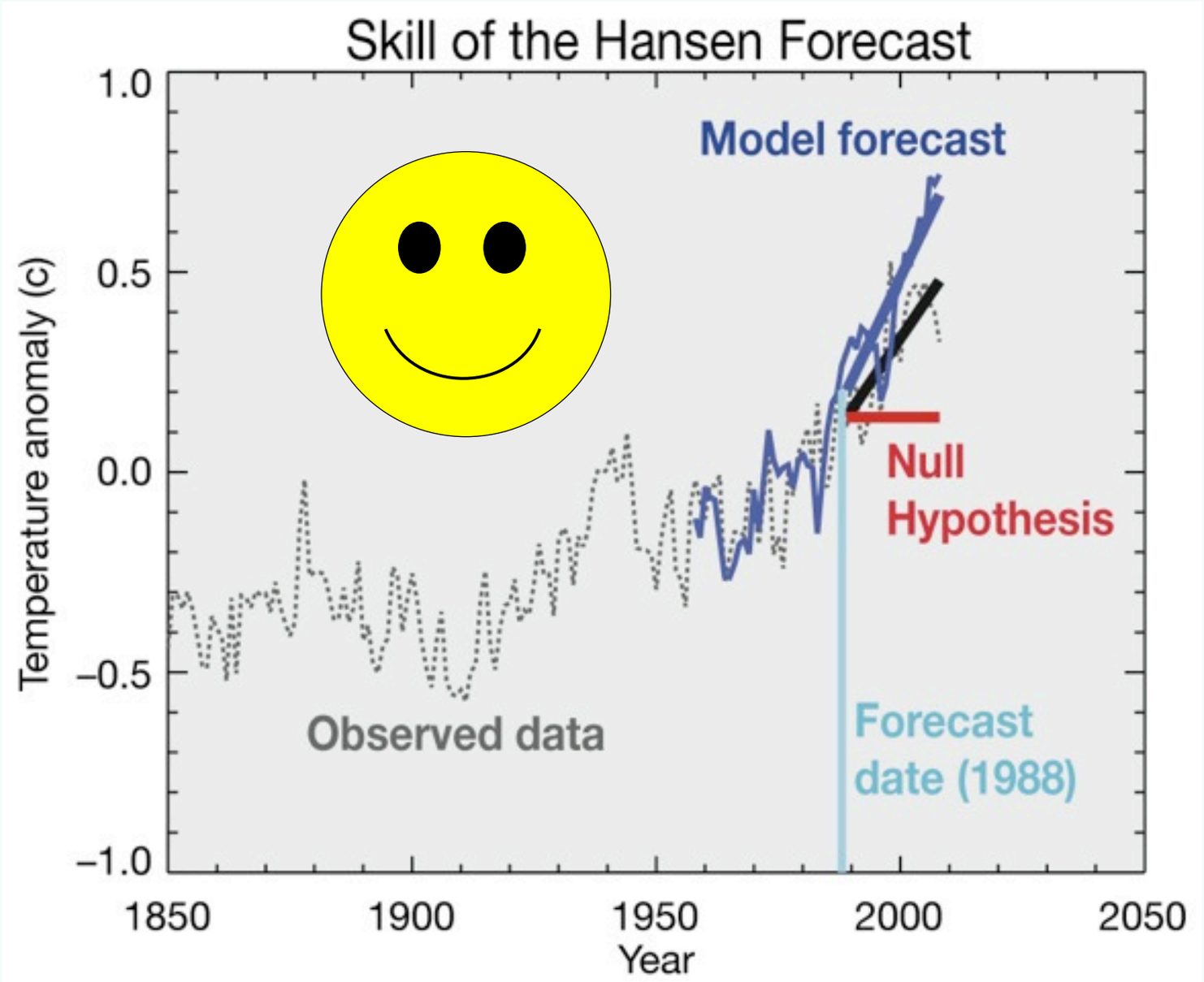
Skill analysis of a climate forecast



- 1989-2008:
  Hansen trend = 0.26C,
  HadCRUT trend=0.18C per decade.
- Null hypothesis:
  persistence over 20 years
- Skill=0.56 -> skillful!
- Bootstrapped statistical significance:
  P<5% that natural variability would have measured up so well against the forecast.

This Figure re-created by Hans von Storch's draftsman, Gruss Beate Gardeike

Skill analysis of a climate forecast



- 1989-2008:
  Hansen trend = 0.26C,
  HadCRUT trend=0.18C per decade.
- Null hypothesis:
  persistence over 20 years
- Skill=0.56 -> skillful!
- Bootstrapped statistical significance:
  P<5% that natural variability would have measured up so well against the forecast.

This Figure re-created by Hans von Storch's draftsman, Gruss Beate Gardeike

Skill at the LGM

Two null hypotheses

Skill at the LGM

Two null hypotheses

1. Persistence: LGM = modern

Skill at the LGM

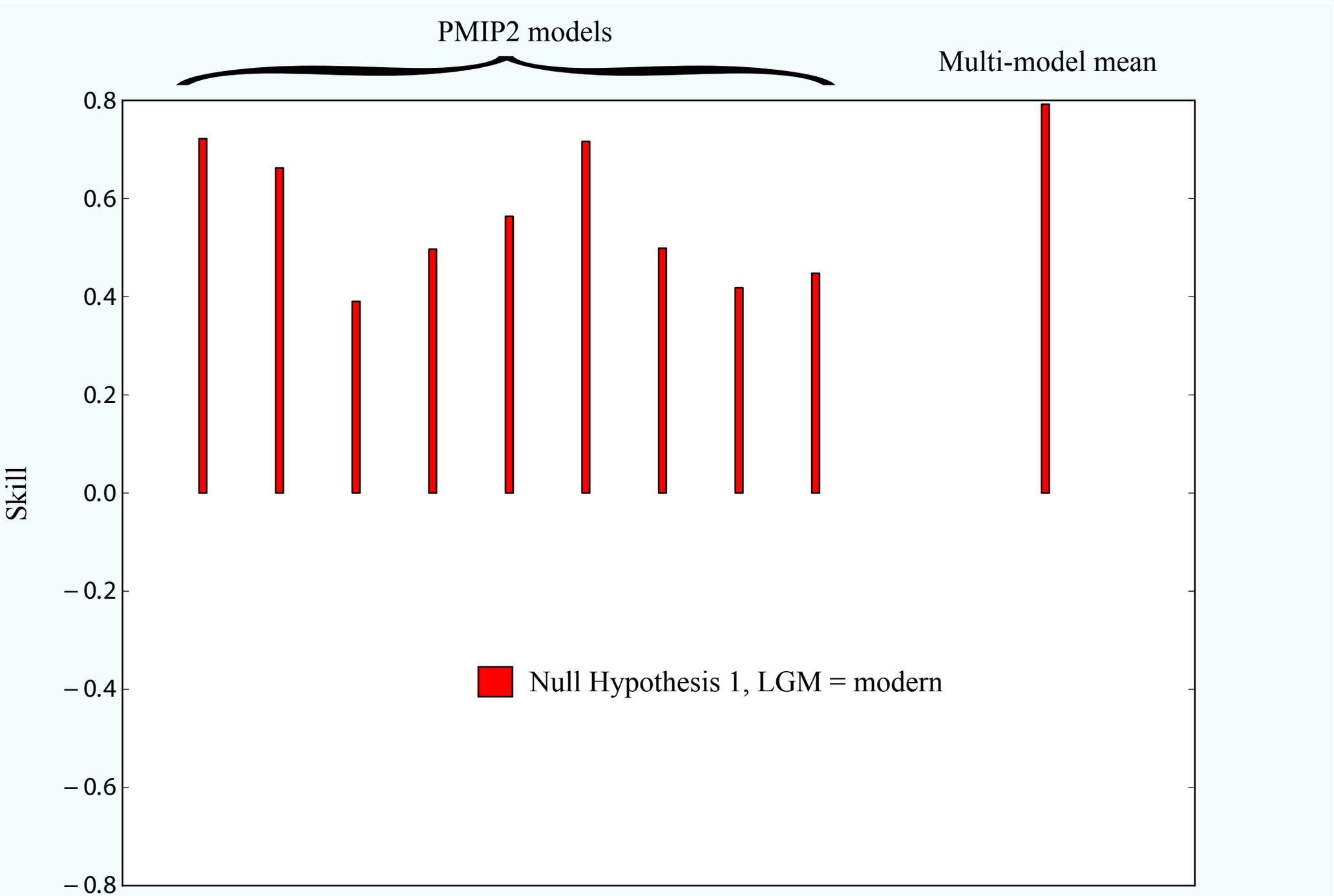
Two null hypotheses

1. Persistence: LGM = modern

2. Testing the pattern: LGM = mean model field at LGM

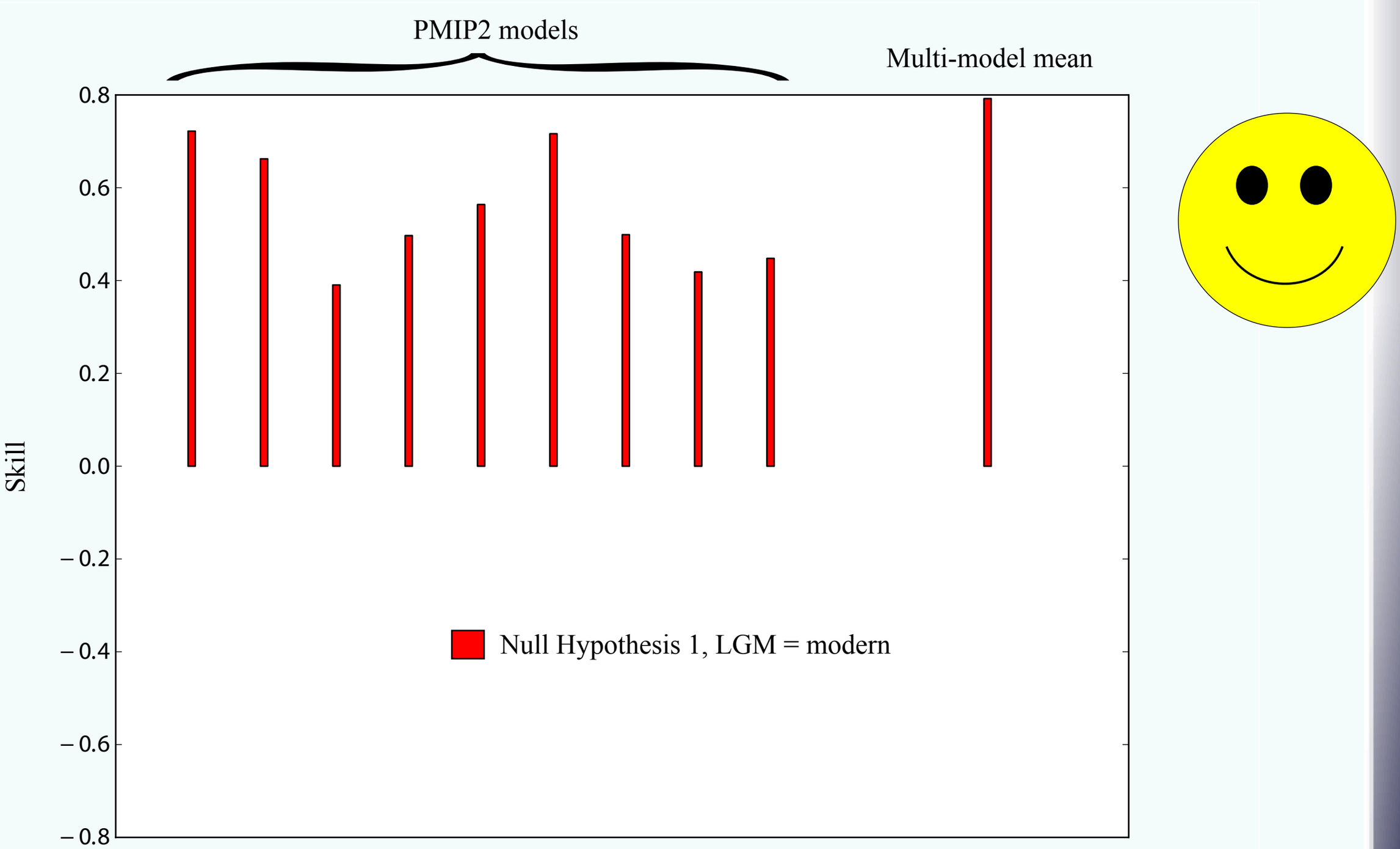
Skill at the LGM

Skill at the LGM



Preliminary results

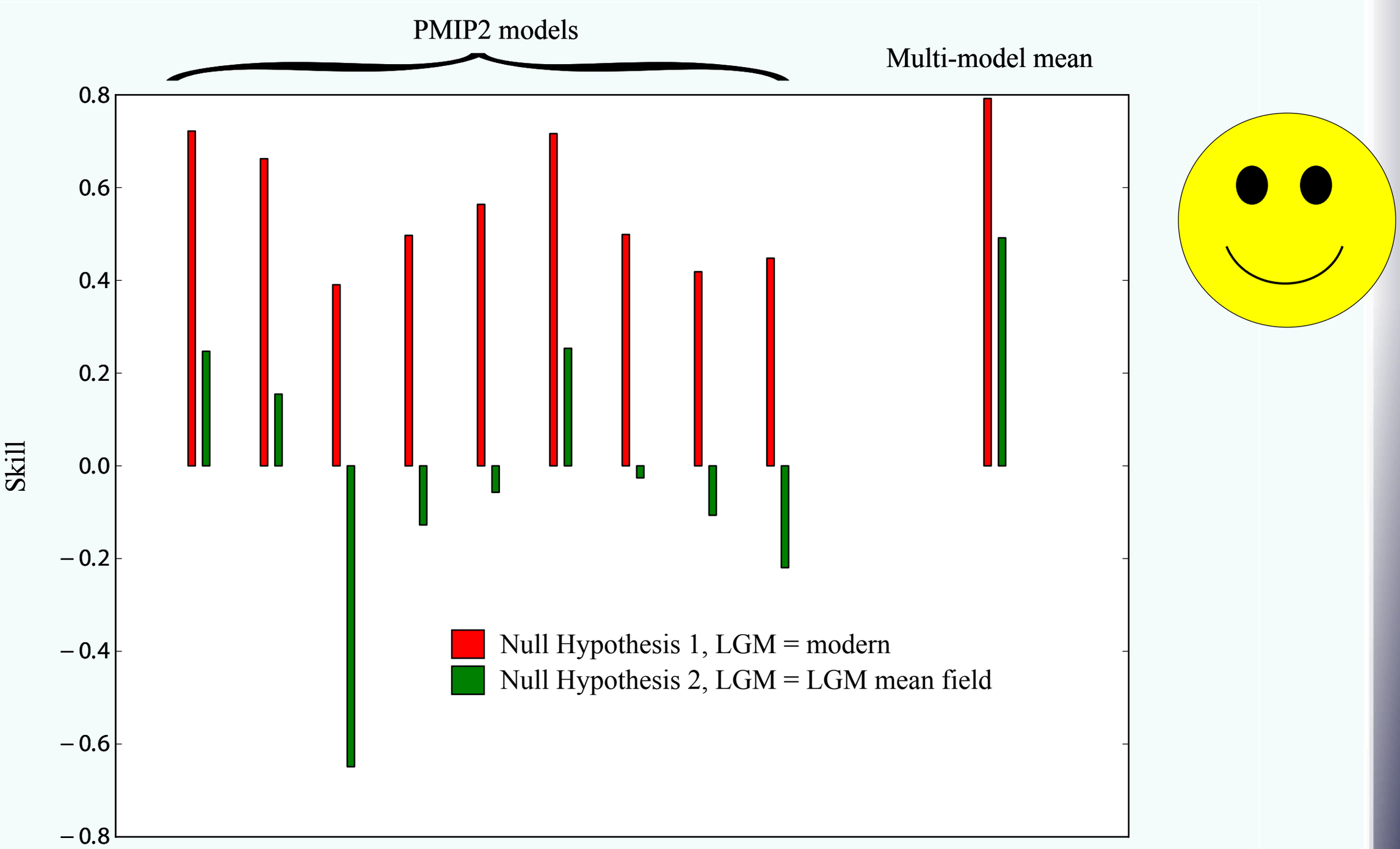
Skill at the LGM



Preliminary results

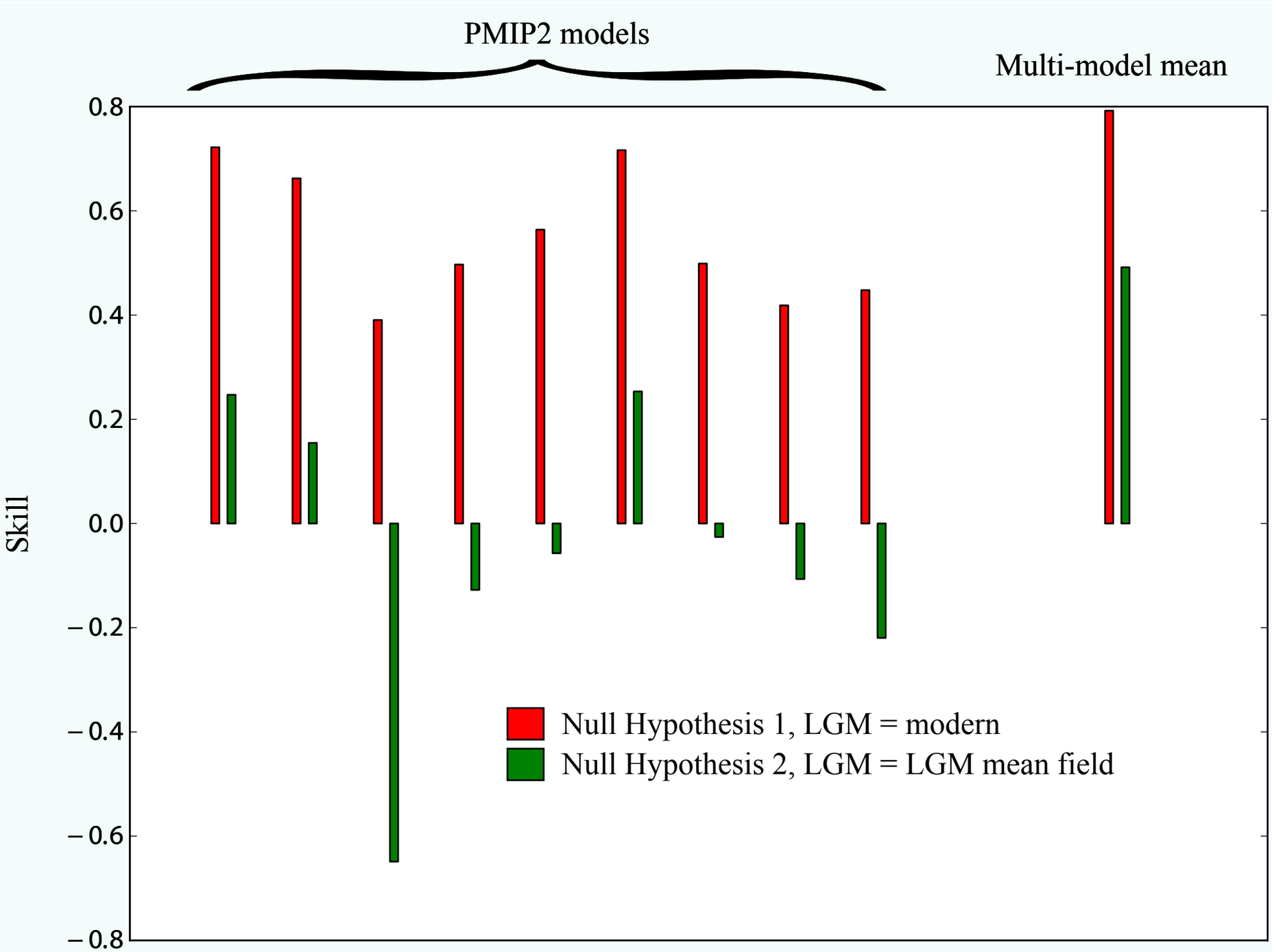
Japan Uncertainty Modelling Project

Skill at the LGM



Preliminary results

Skill at the LGM



Preliminary results

Japan Uncertainty Modelling Project

Conclusions

Conclusions

- CMIP3 is consistent with modern climatology

Conclusions

- CMIP3 is consistent with modern climatology
- PMIP2 is consistent with MARGO data.

Conclusions

- CMIP3 is consistent with modern climatology
- PMIP2 is consistent with MARGO data.
- Including data uncertainty is very important for analysis of paleoclimates.

Conclusions

- CMIP3 is consistent with modern climatology
- PMIP2 is consistent with MARGO data.
- Including data uncertainty is very important for analysis of paleoclimates.
- Ensembles derived from a single model are usually unreliable.

Conclusions

- CMIP3 is consistent with modern climatology
- PMIP2 is consistent with MARGO data.
- Including data uncertainty is very important for analysis of paleoclimates.
- Ensembles derived from a single model are usually unreliable.
- Climate models have some skill at the LGM, but not as much as we might like!

Workshop:

Using paleo-climate model-data comparisons to constrain future climate projections

Mar 1-3, 2012

Bishop Museum, Honolulu, Hawaii

(immediately before CMEP workshop)

Contacts: Gavin Schmidt, Valerie Masson-Delmotte

PAGES

Climate Variability & Predictability
CLIVAR
World Climate Research Programme

NOAA
National Oceanic and Atmospheric Administration
U.S. Department of Commerce