Multimodel Ensemble Forecasts

Introduction to statistical treatment
Performance independent methods

Michael K. Tippett

International Research Institute for Climate and Society
The Earth Institute, Columbia University

ERFS Climate Predictability Tool Training Workshop
May 4-9, 2009

Ideas

- Extract useful information from ensembles.
- Construct forecast probabilities.
- Simple combination of ensembles from different models

Methods that do not depend on model performance.

Performance independent methods

- Forecast probabilities are constructed without taking into account model performance.
- Historical observations not necessary.
- Variety of methods of extracting information from the ensemble.
- Assume *some* aspects of the ensemble information are correct and use them.

“Counting” methods

Pool ensemble members from different models.

$$P(\text{event}) = \text{Ensemble frequency of event}$$

Example:

$$P(T > 26.3^\circ C) = \frac{\#\text{ ensemble members with } T > 26.3^\circ C}{\#\text{ ensemble members}}$$

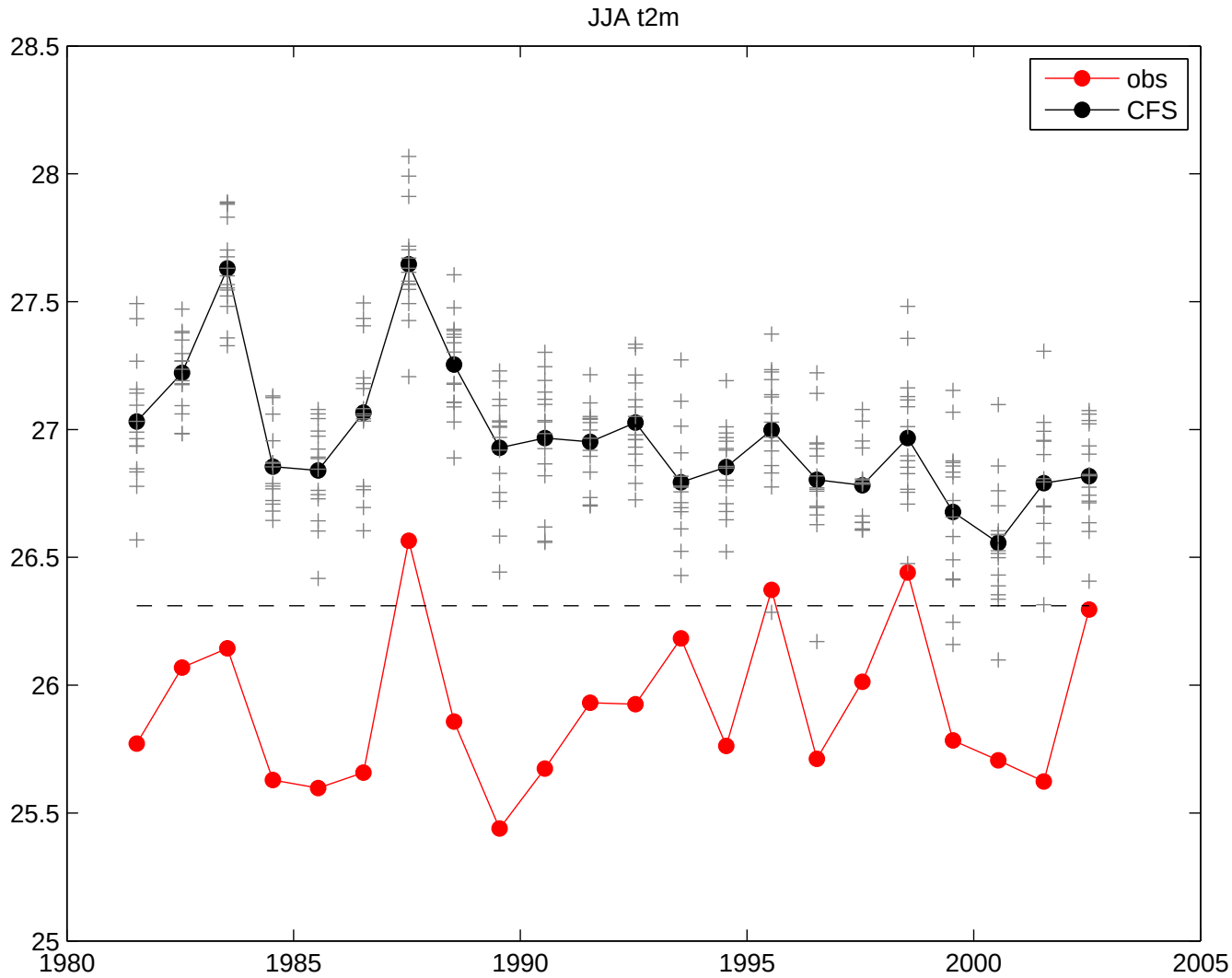
Assumes the ensemble members are equally like samples of the future state.
Does not account for ensemble bias.

- Mean bias.
- Spread bias.

Example

- JJA average temperature
- Averaged over large box 2N - 32N, 64E - 93E.

Model has a warm bias. Forecasts of $T > 26.3$ are 100%, 93%, 87%.



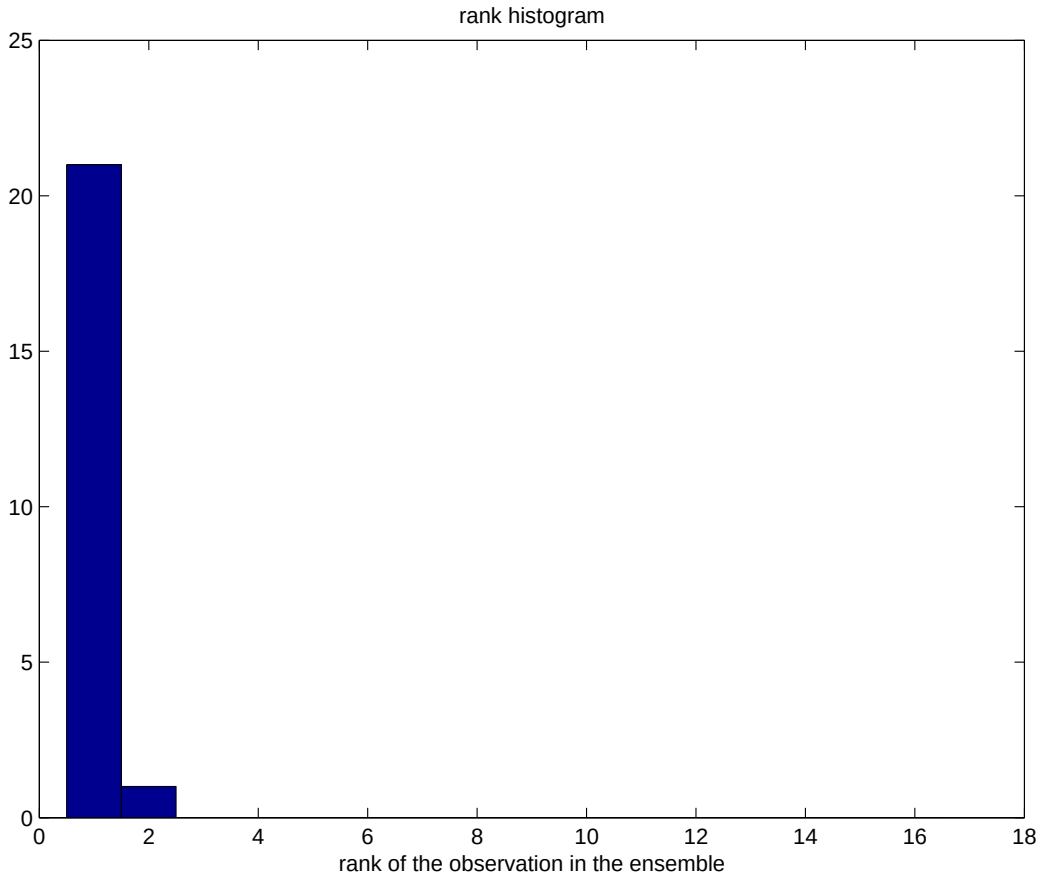
Measures of ensemble "goodness"

- Relation of ensemble mean with observations, e.g., correlation
- Are the observations like ensemble members? Rank histogram.

Compute rank histogram

- For each verification, pool ensemble members and observation.
- Sort by value from smallest to largest.
- Record the rank of the observation. Smallest, largest, 5th, etc.
- For a set of forecasts make a histogram of the observations ranks.
- If the observations are like ensemble members, histogram should be flat. All ranks equally likely.

Correlation between ensemble mean and observations is 0.46. Rank histogram indicates that the observations are almost always outside the ensemble.



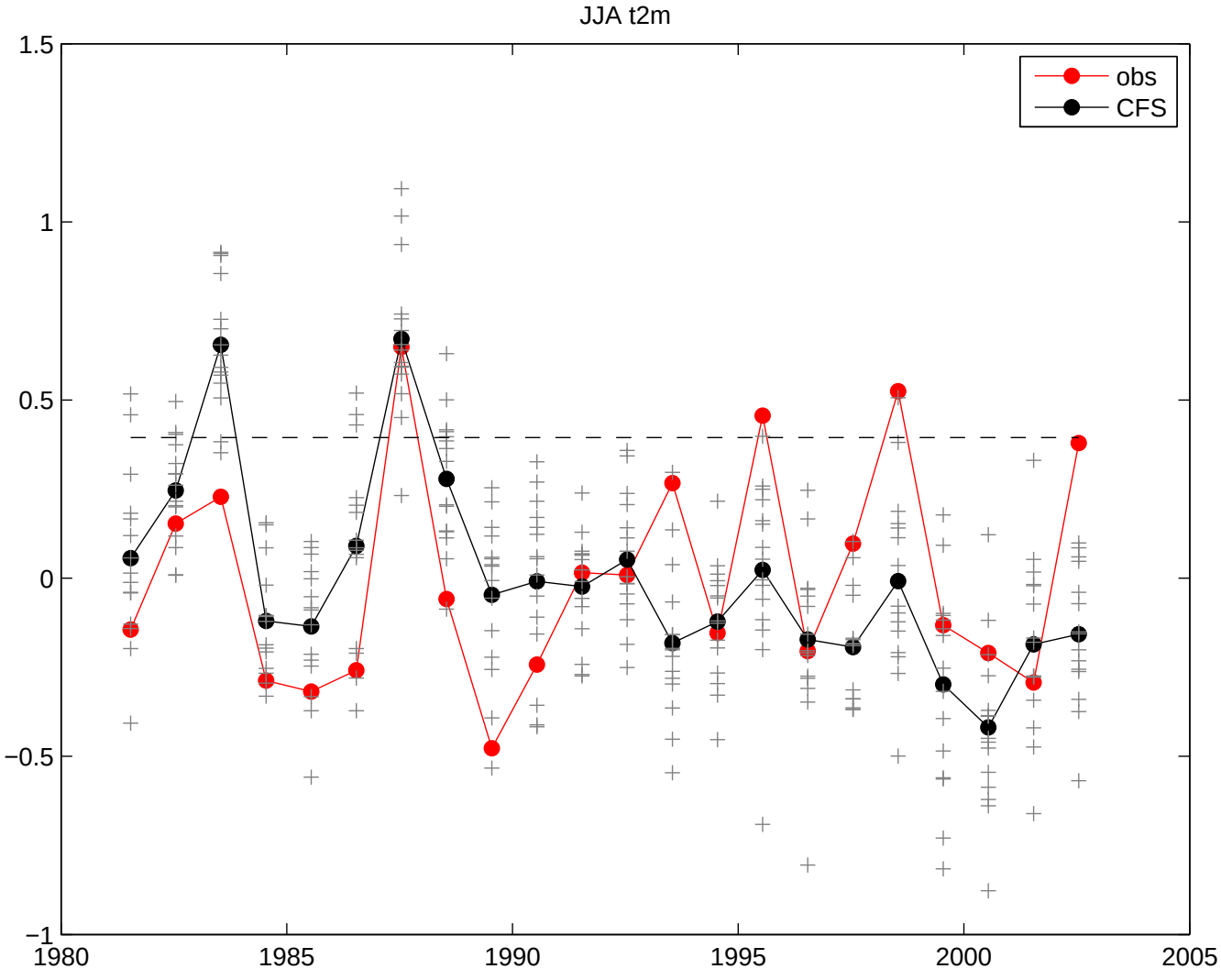
“Counting” methods

Define event with respect to model climatology.

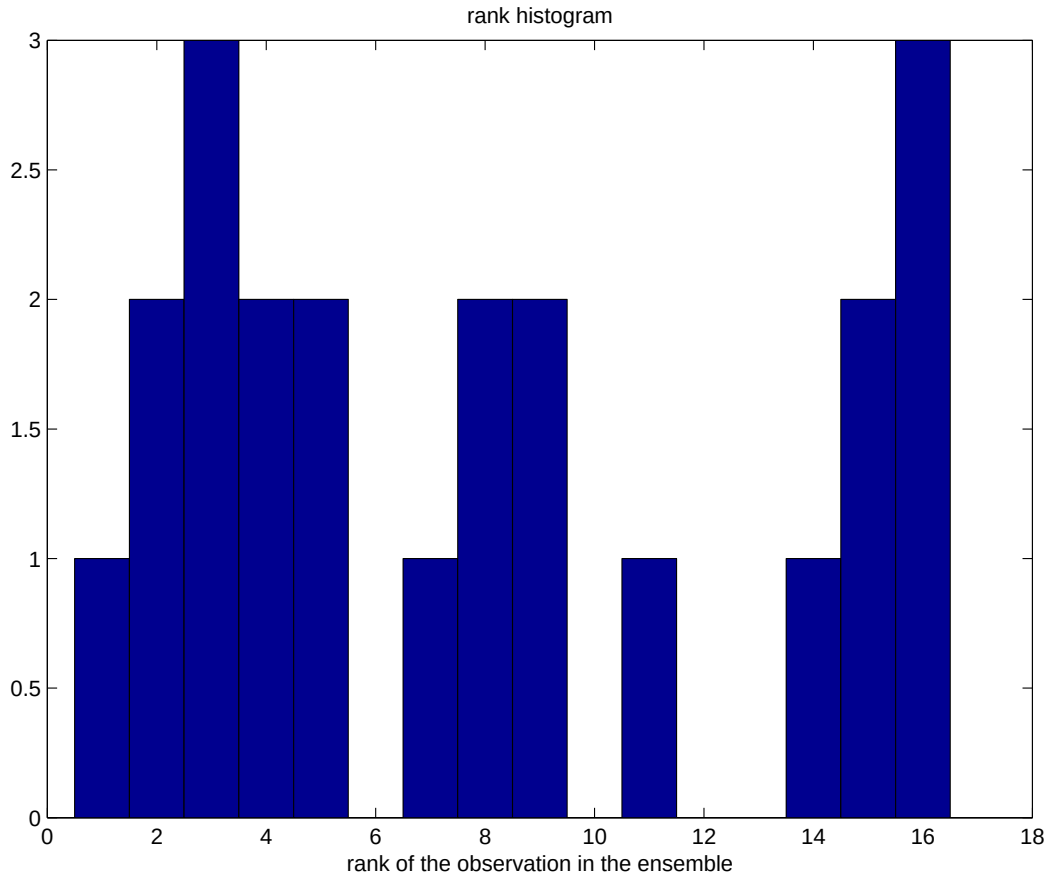
$$P(T > 0.4^{\circ}\text{+climatology}) = \frac{\text{\# ensemble members with } T_{\text{anom}} > 0.4^{\circ}}{\text{\# ensemble members}}$$

climatology = model climatology (history)
Accounts for bias of multimodel ensemble mean.
Does not account for bias of individual models, bias in spread.

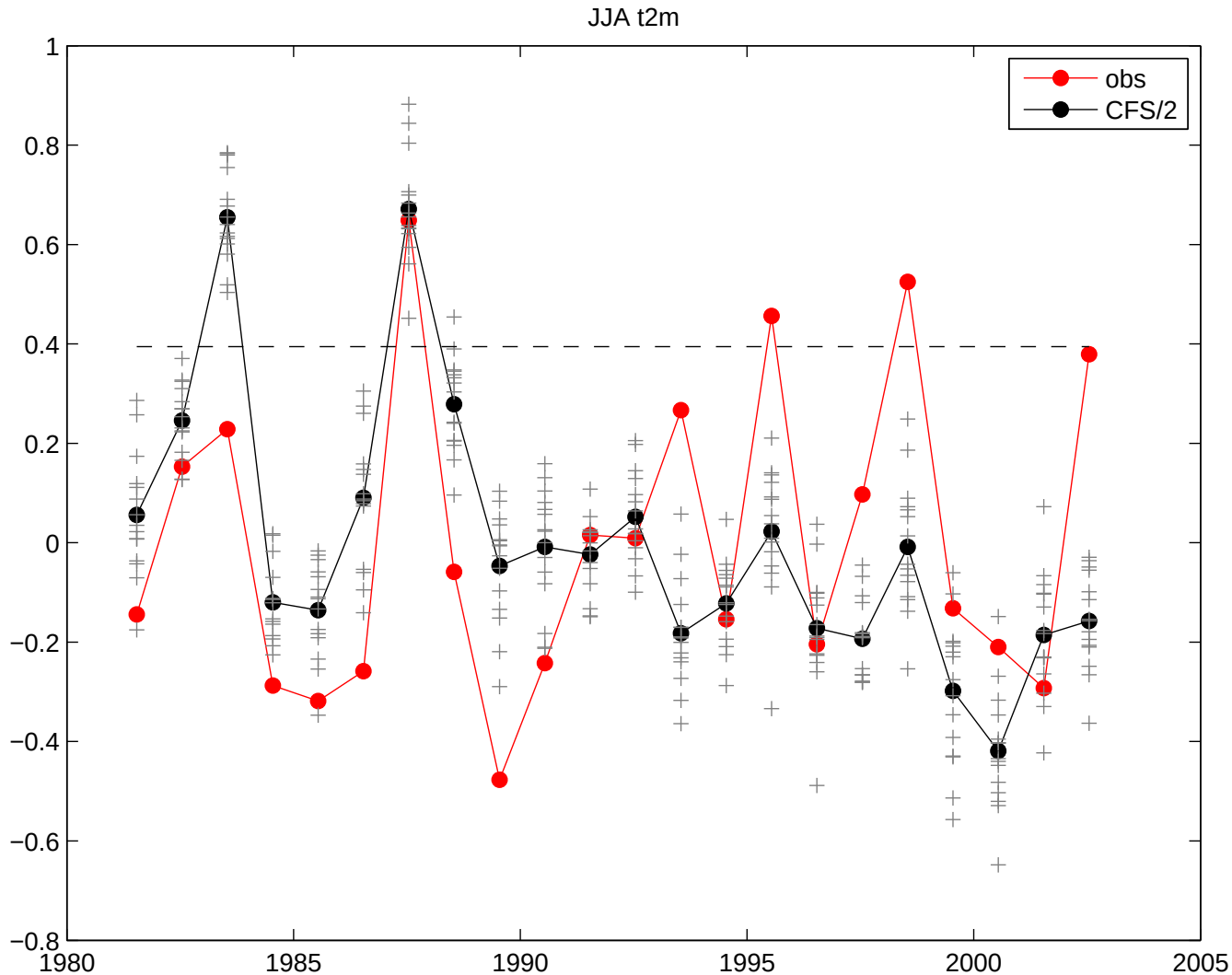
Model and observation anomalies.



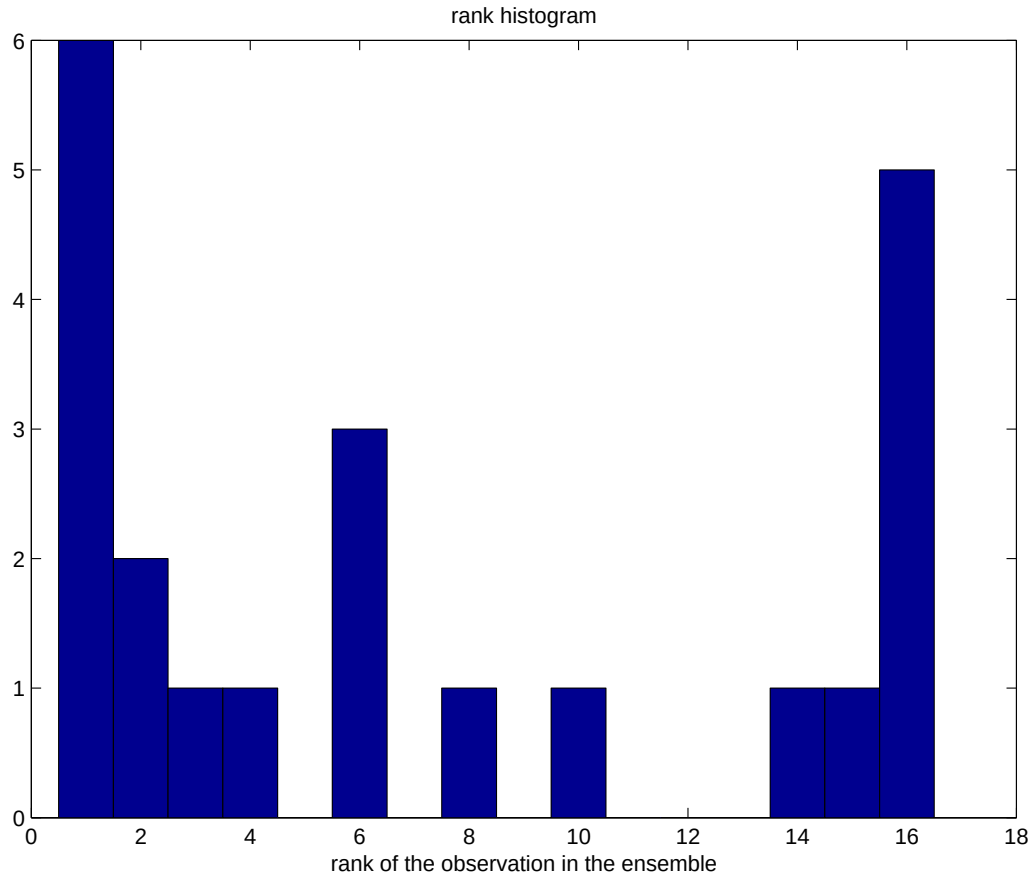
Correlation between ensemble mean and observation still 0.46.
Rank histogram indicates better spread.
Sample size is small.



Model and observation anomalies. Ensemble spread too small.



Correlation between ensemble mean and observation still 0.46. Rank histogram indicates observations often lies outside the ensemble.



“Counting” methods

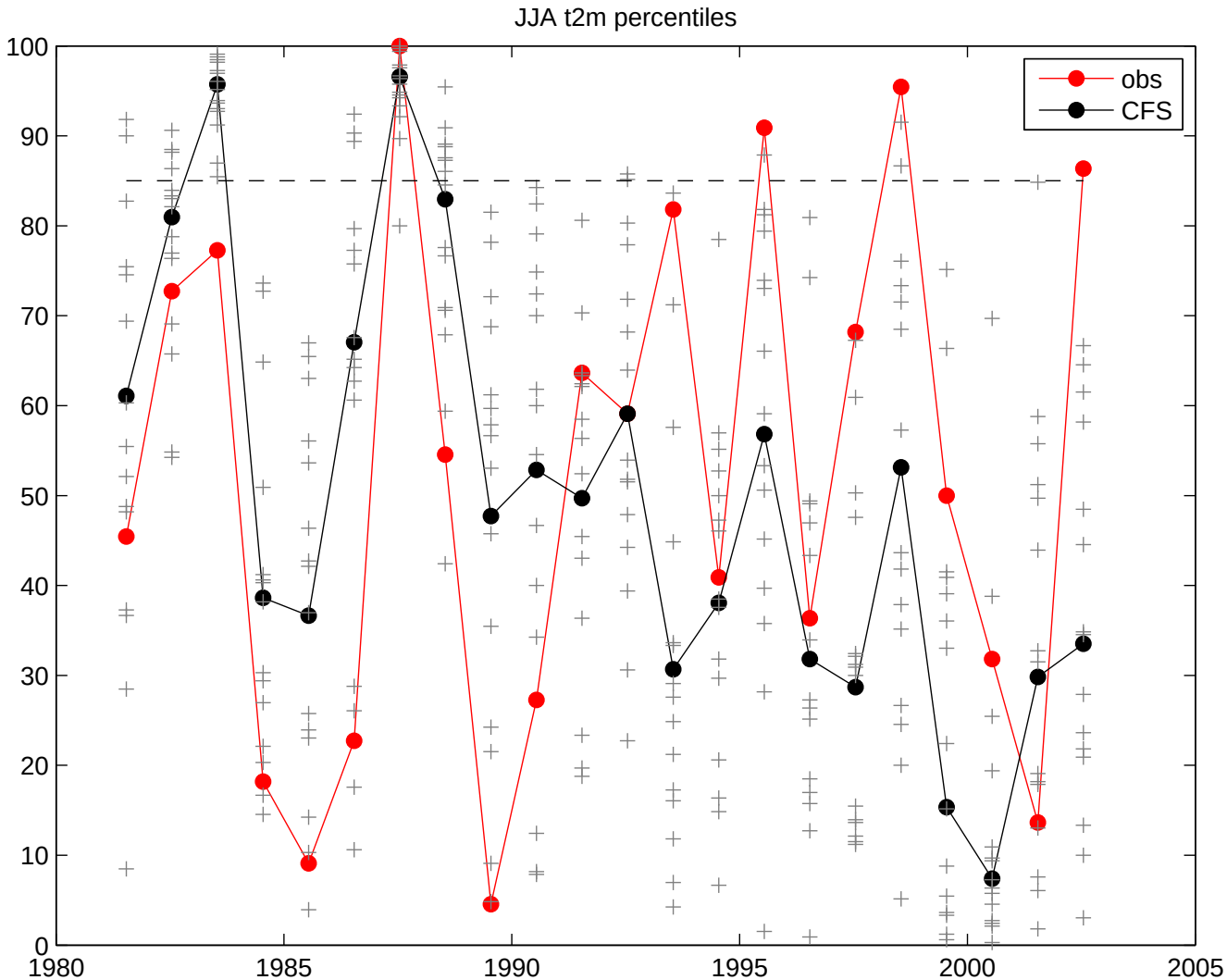
One way to account for errors in mean and spread:

Define event with respect to model history percentiles.

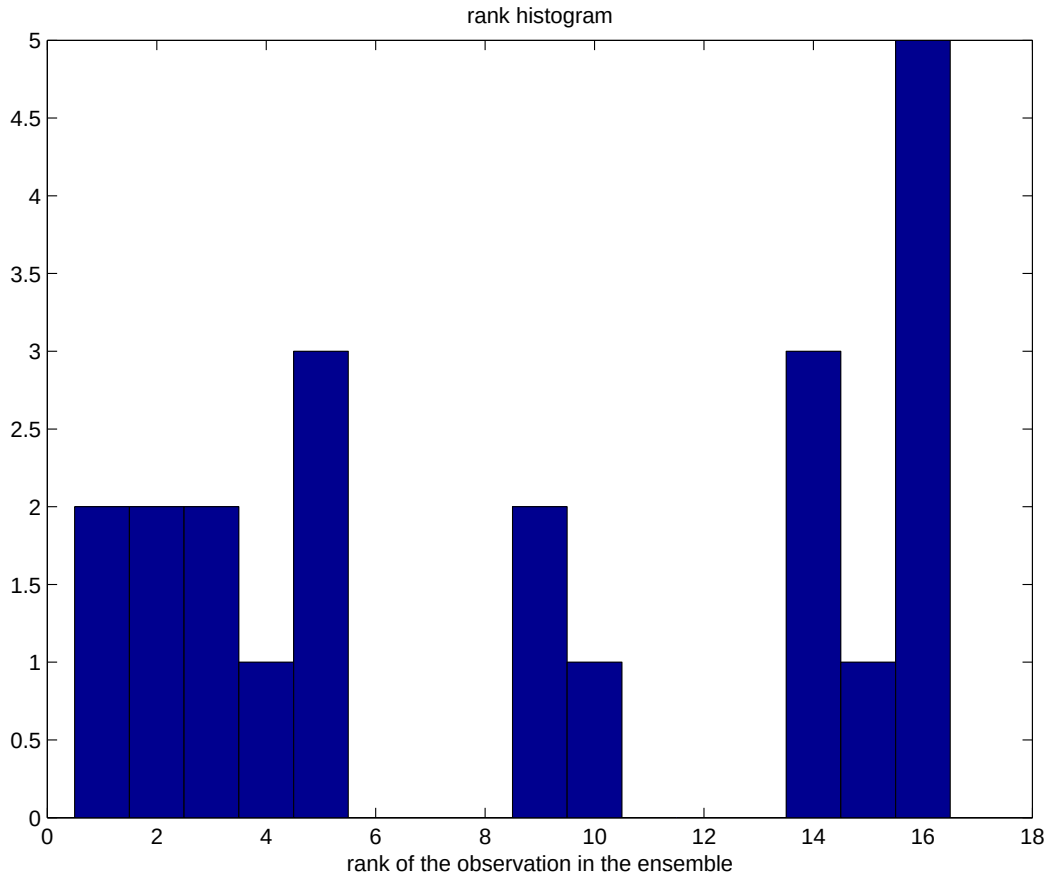
$$P(T > \text{85-th percentile}) = \frac{\#\text{ ensemble members with } T > 85\%}{\#\text{ ensemble members}}$$

percentile = model percentile (history)
Accounts for bias of multimodel ensemble.
Does not account for bias of individual models.

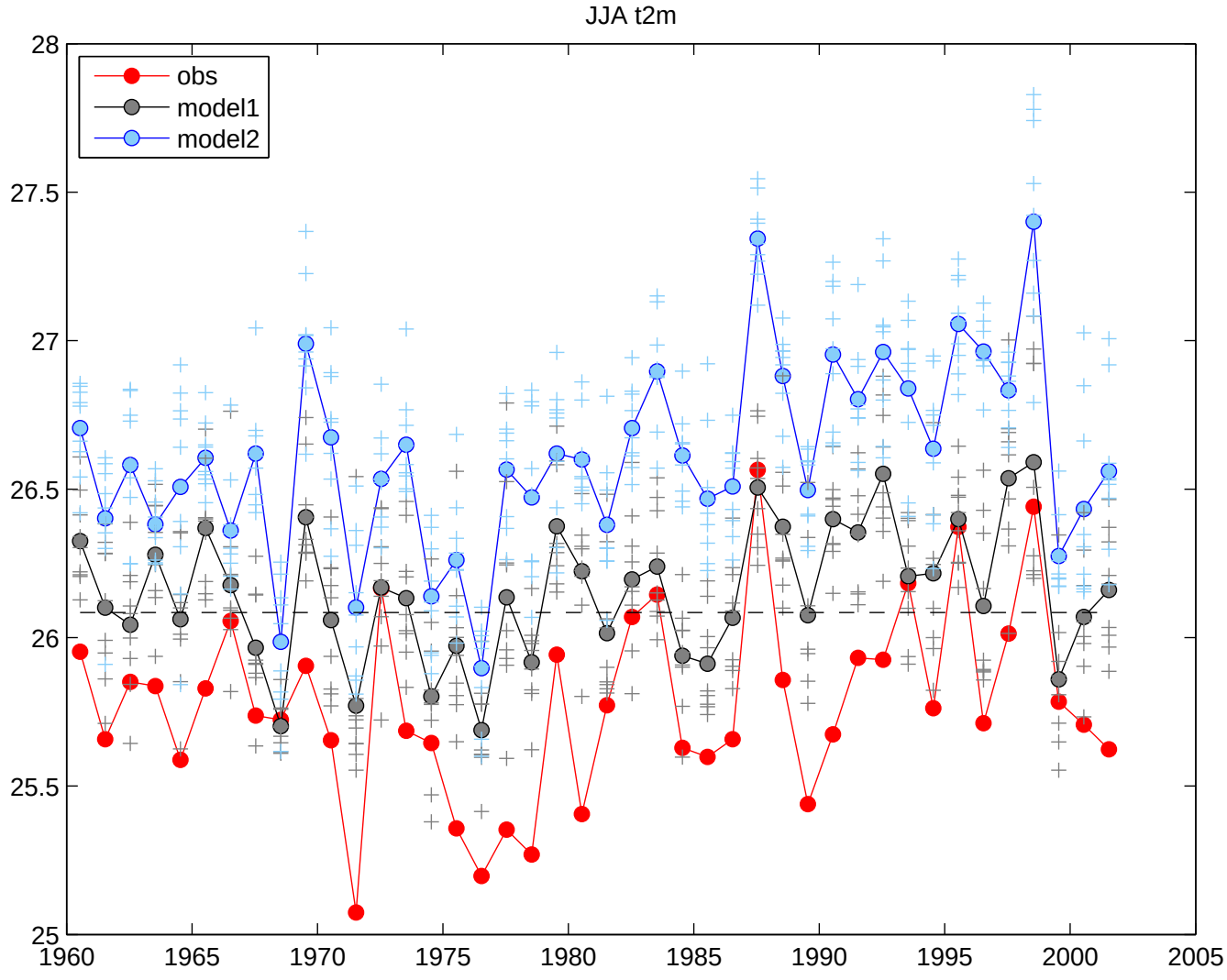
Model and observation percentiles



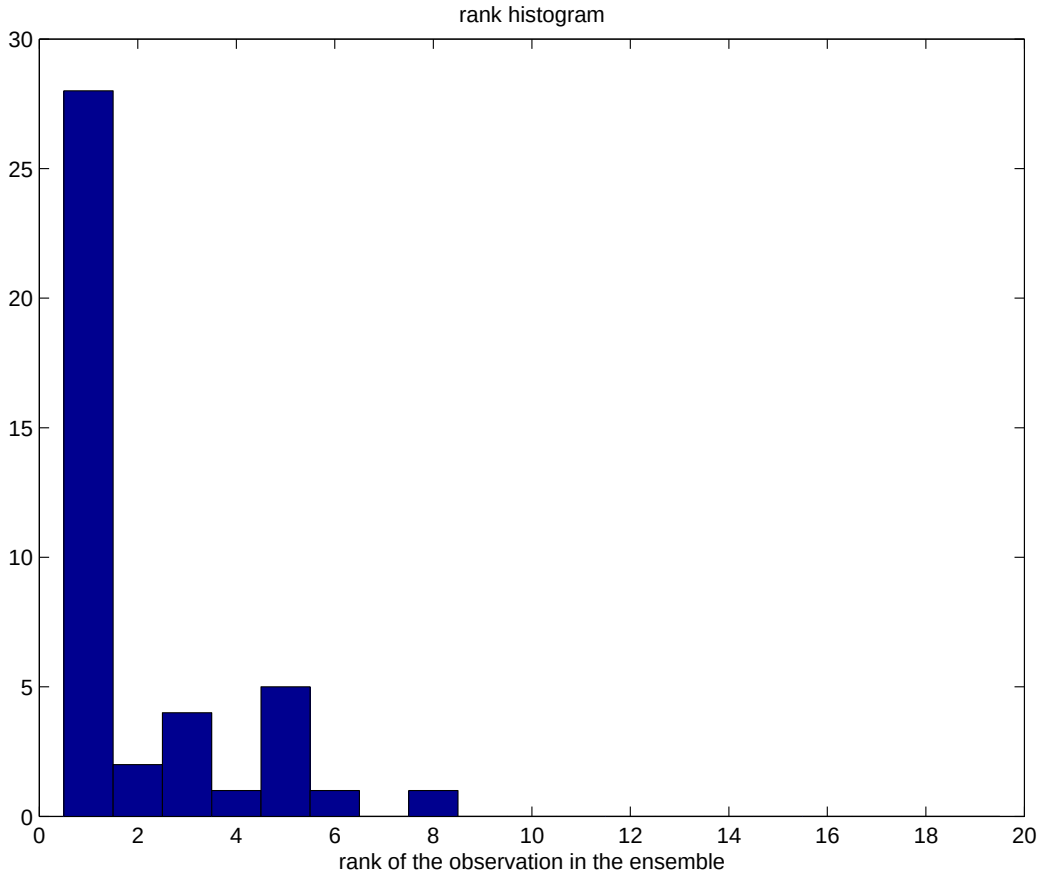
Correlation between ensemble mean and observation still 0.46.
Rank histogram indicates better spread.
Sample size is small.



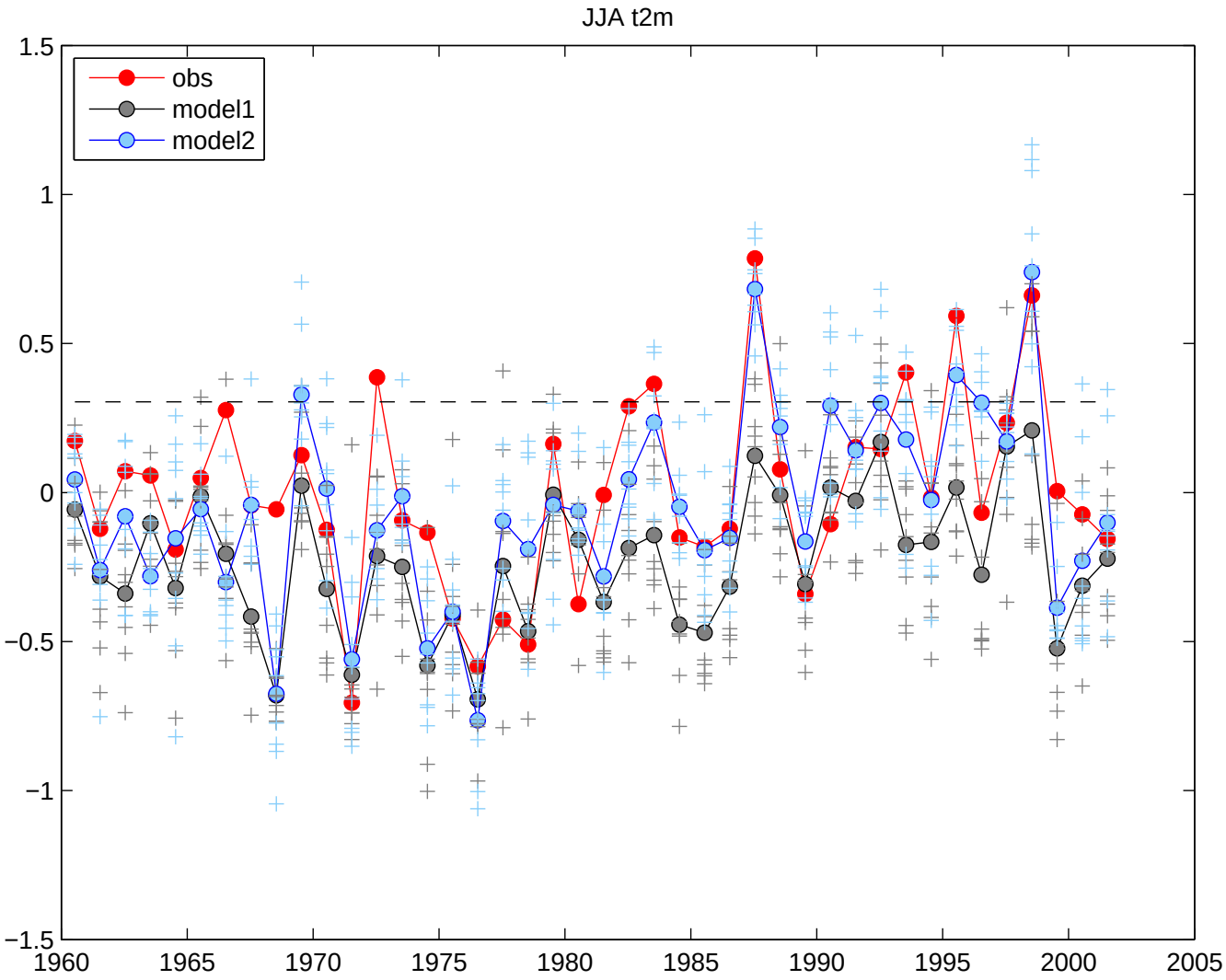
Two models with different mean biases.



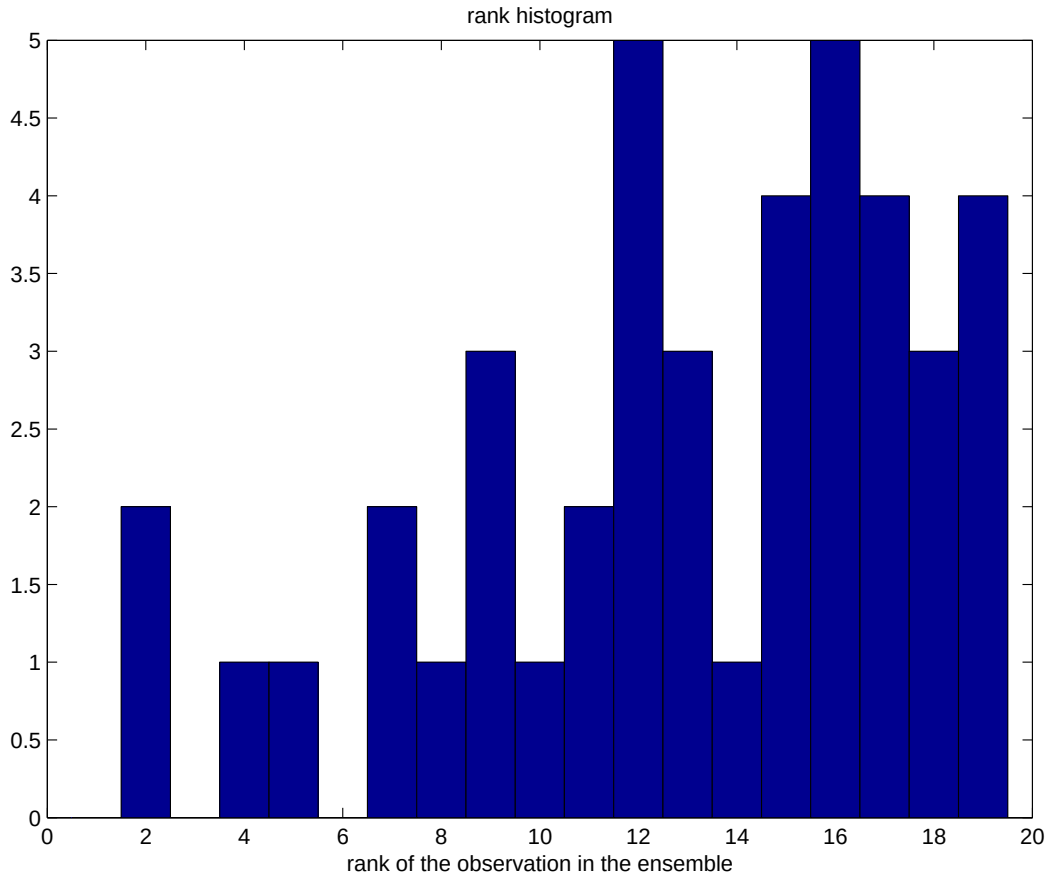
Correlation between ensemble mean and observation is 0.72.
Model 1 correlation = 0.68. Model 2 correlation 0.70.
Rank histogram indicates bias.



Pooled models with *multimodel* climatology removed.
Additional spread due to different means.



Correlation between ensemble mean and observation still 0.72.
Rank histogram indicates bias.
Observation tends to be too warm compared to the ensemble.



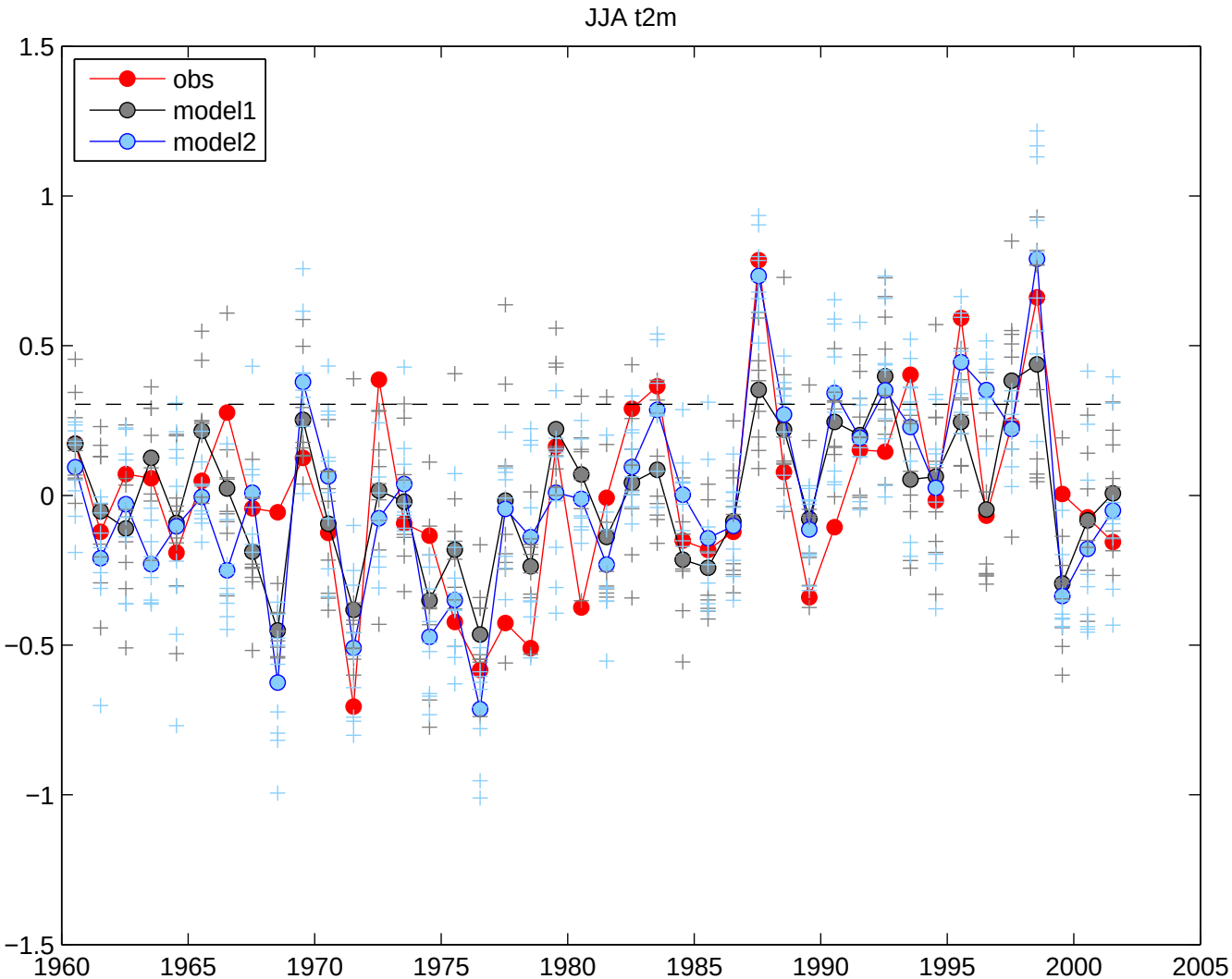
“Counting” methods

How to account for bias of individual models?
Pool *anomalies* of each model ensemble.
Define events in terms of anomalies

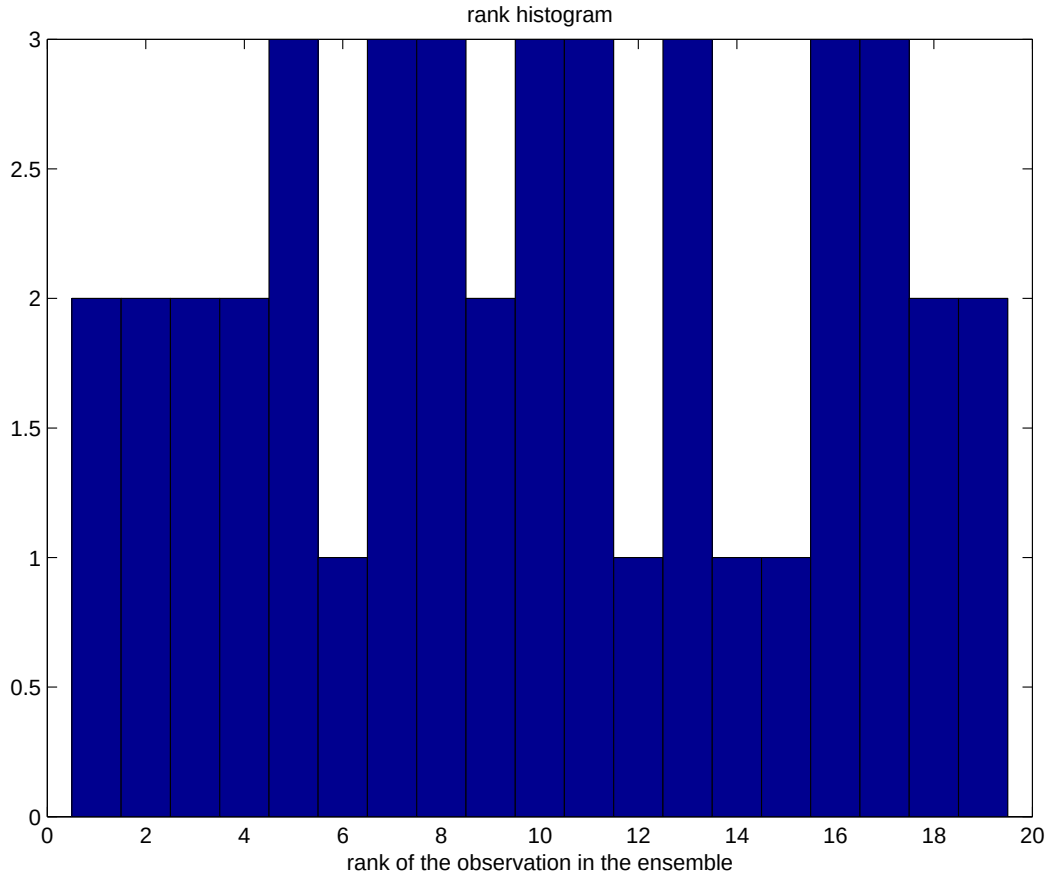
$$P(\text{event}) = \text{Ensemble frequency of event}$$

Accounts for mean biases of individual models.

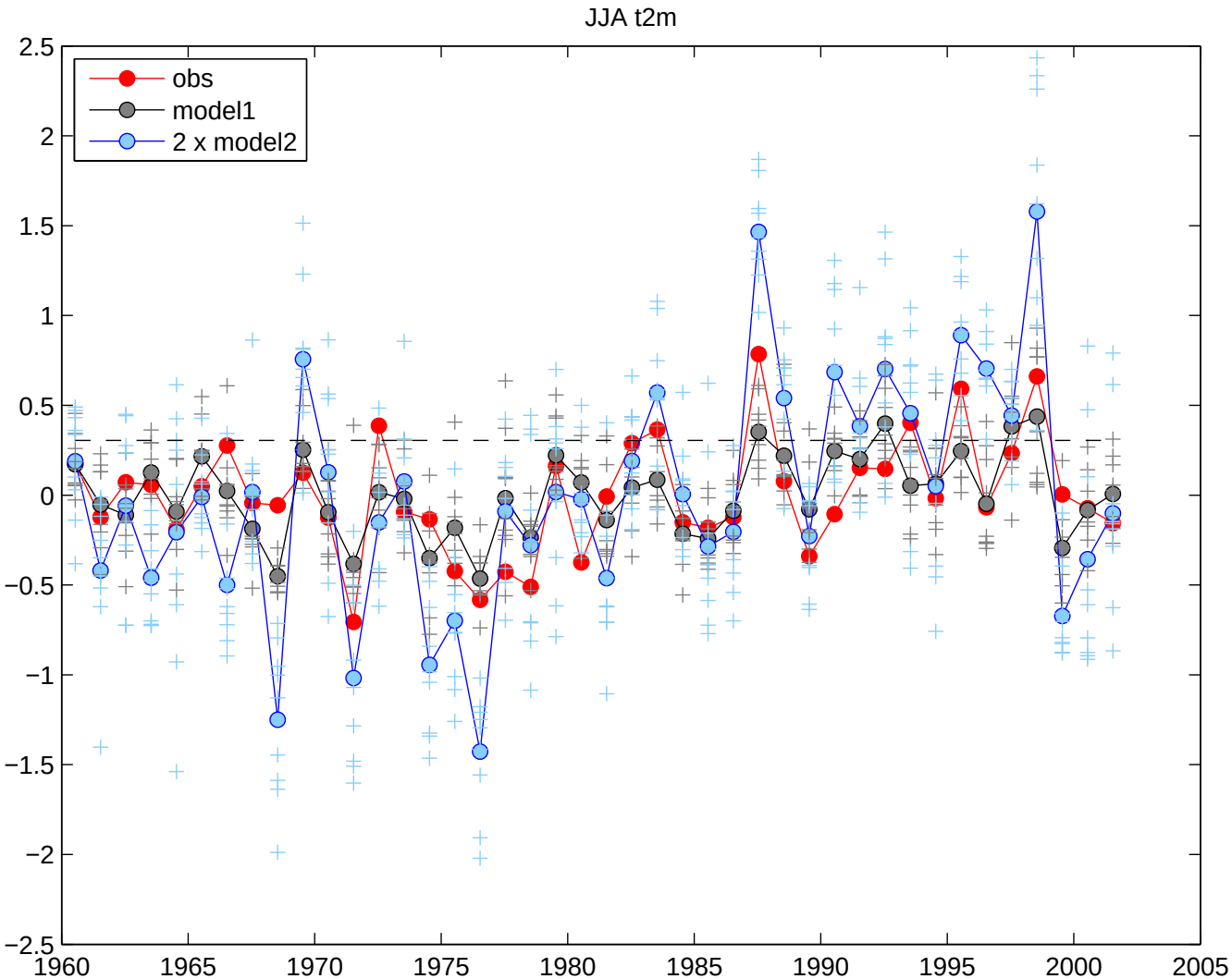
Pooled anomalies.



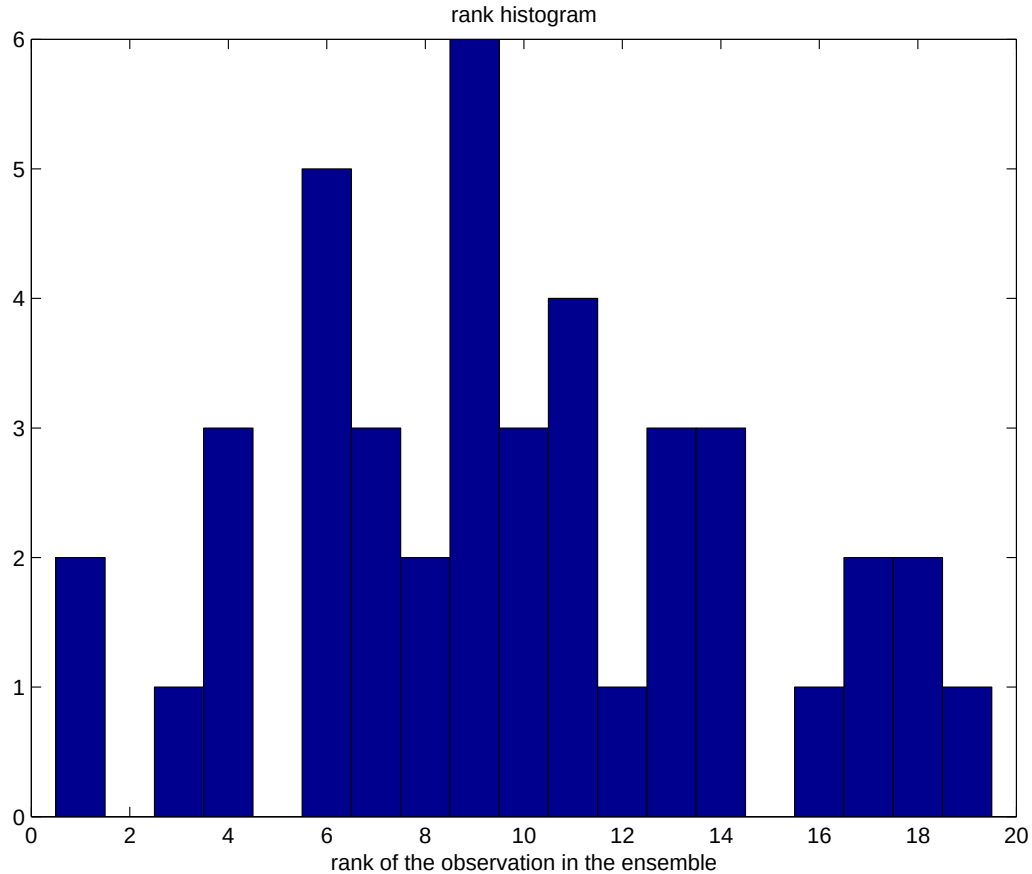
Correlation between ensemble mean and observation still 0.72.
Rank histogram indicates reduced bias.



What if one model has significantly more variability than another?



Correlation between ensemble mean and observation still 0.72.
Rank histogram indicates ensemble spread is too large.



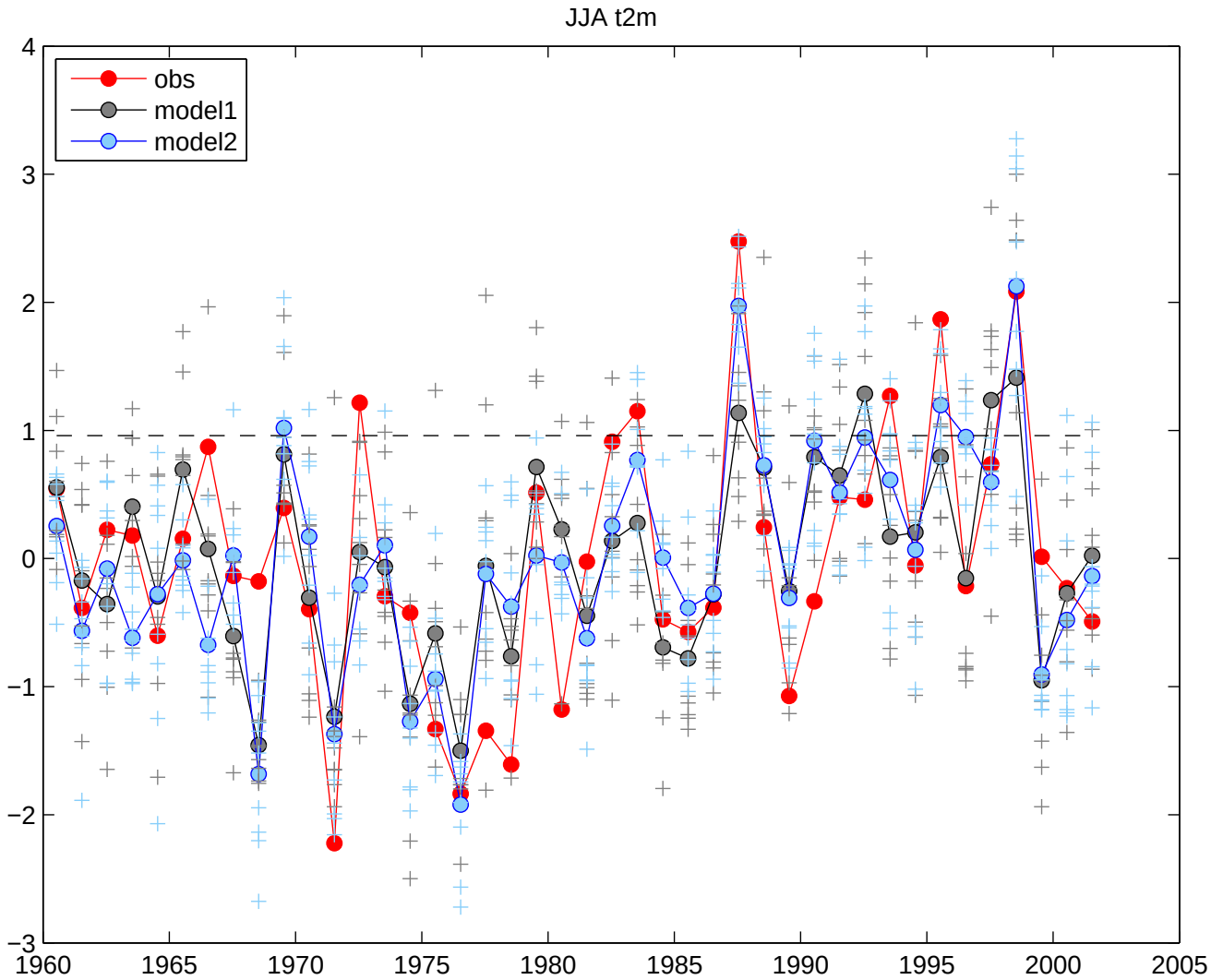
“Counting” methods

Pool *standardized anomalies* of each model ensemble.
Define events in terms of standardized anomalies

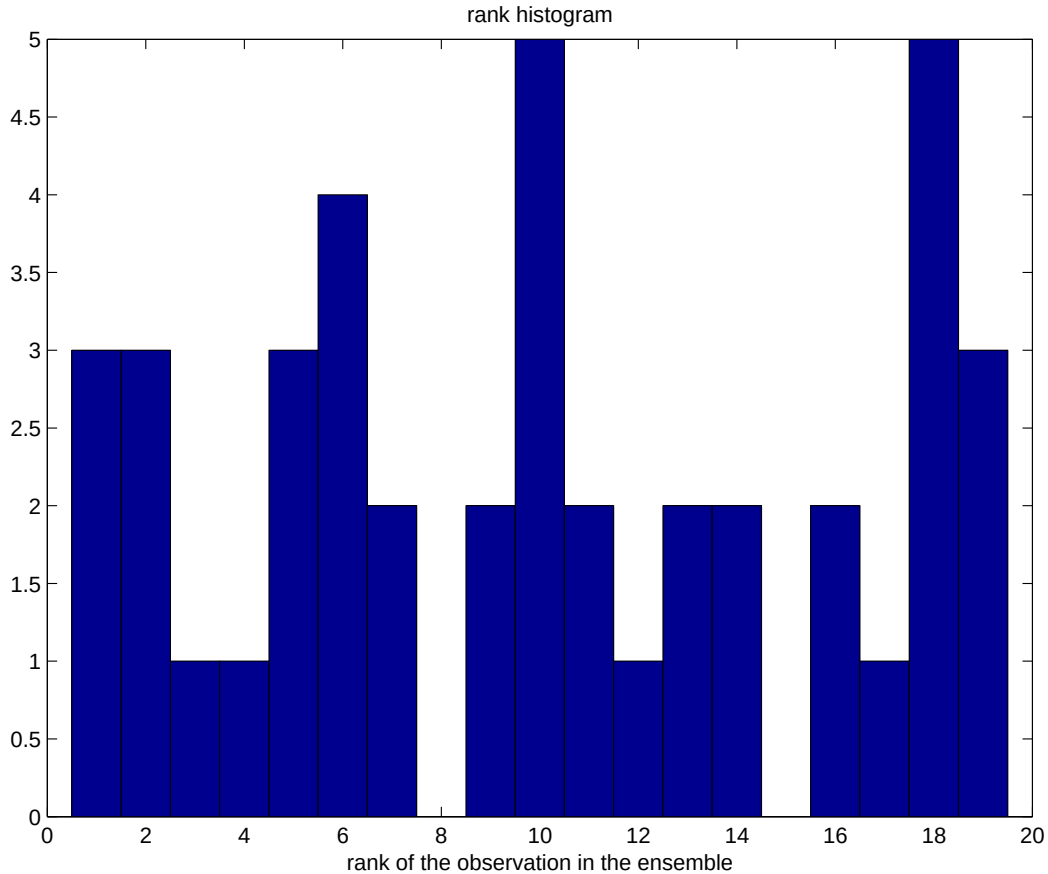
$$P(\text{event}) = \text{Ensemble frequency of event}$$

Accounts for mean biases and spread biases of individual models.

Pooled standardized anomalies.



Correlation between ensemble mean and observation still 0.72.
Rank histogram indicates less bias in ensemble spread.



Reminder

Using:

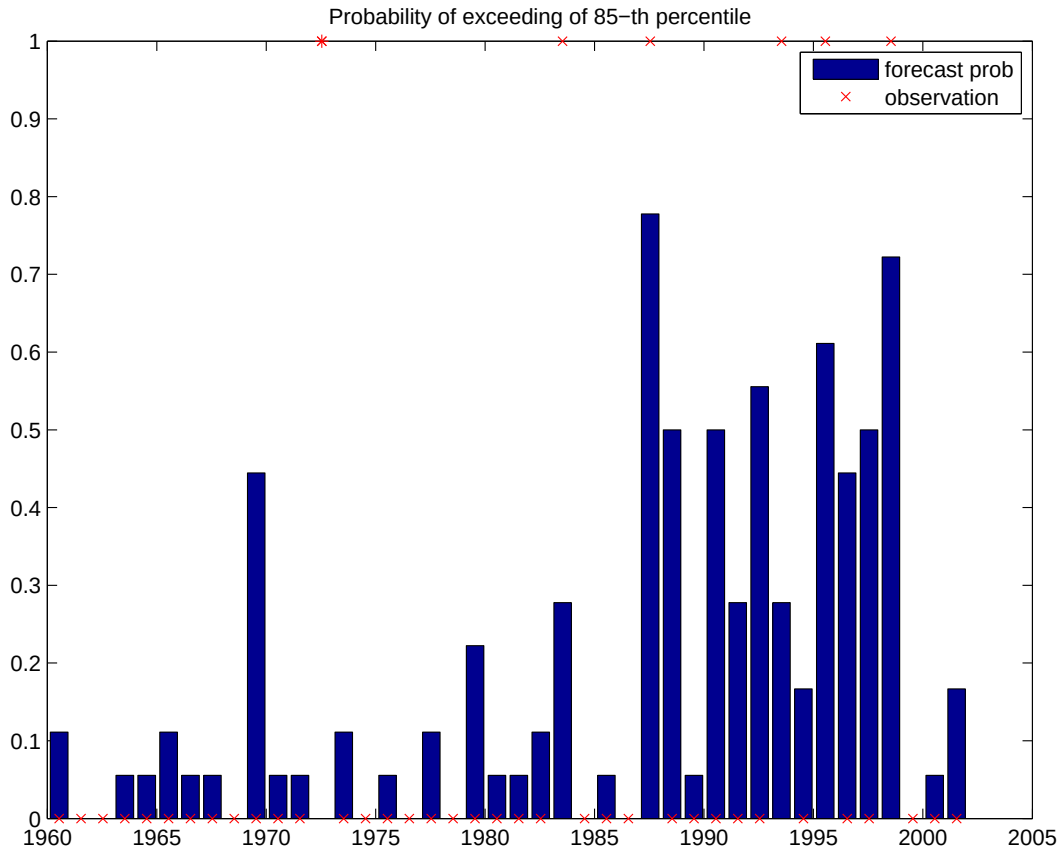
- anomalies
- standardized anomalies
- percentiles

corrects the only climatology.
May not help individual forecasts.
Does not make the spread of individual forecasts consistent with forecast errors.

Sampling error

- Even if the models are perfect, ensemble size limits accuracy of probabilities.
- $P(\text{event}) = 0$ may occur simply because the event is rare compared to the ensemble size, not because it is impossible.
- The only time a probabilistic forecast can be "wrong" is when it is deterministic. Must avoid $P = 0$ or $P = 1$.

1972 forecast zero probability but event occurred.
Brier skill score 0.46.



Continuous pdf

Replace ensemble with a continuous pdf f

- Parametric – e.g., Gaussian with mean μ and variance σ^2 from ensemble
- Nonparametric – kernel density estimator

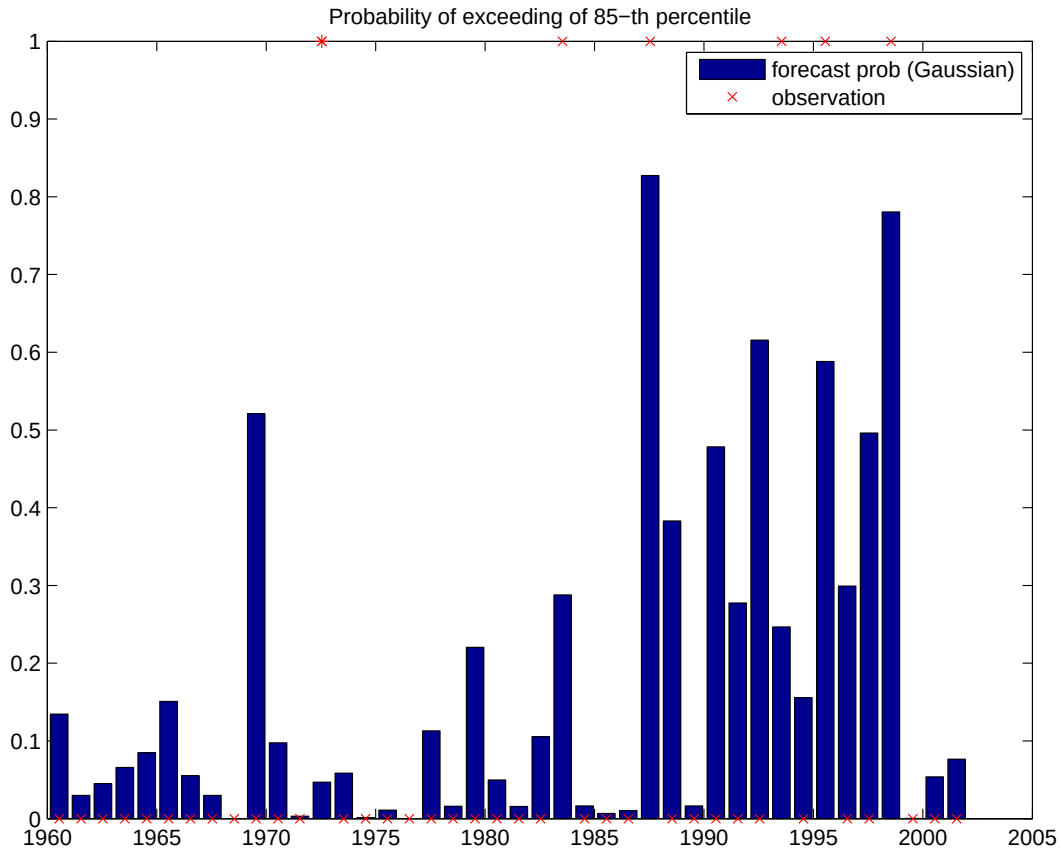
Example. $f(T)$ Gaussian.

$$f = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{(T-\mu)^2}{2\sigma^2}}$$

$$P(T > T_0) = 1 - \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{T_0} e^{\frac{(T-\mu)^2}{2\sigma^2}} \, dT$$

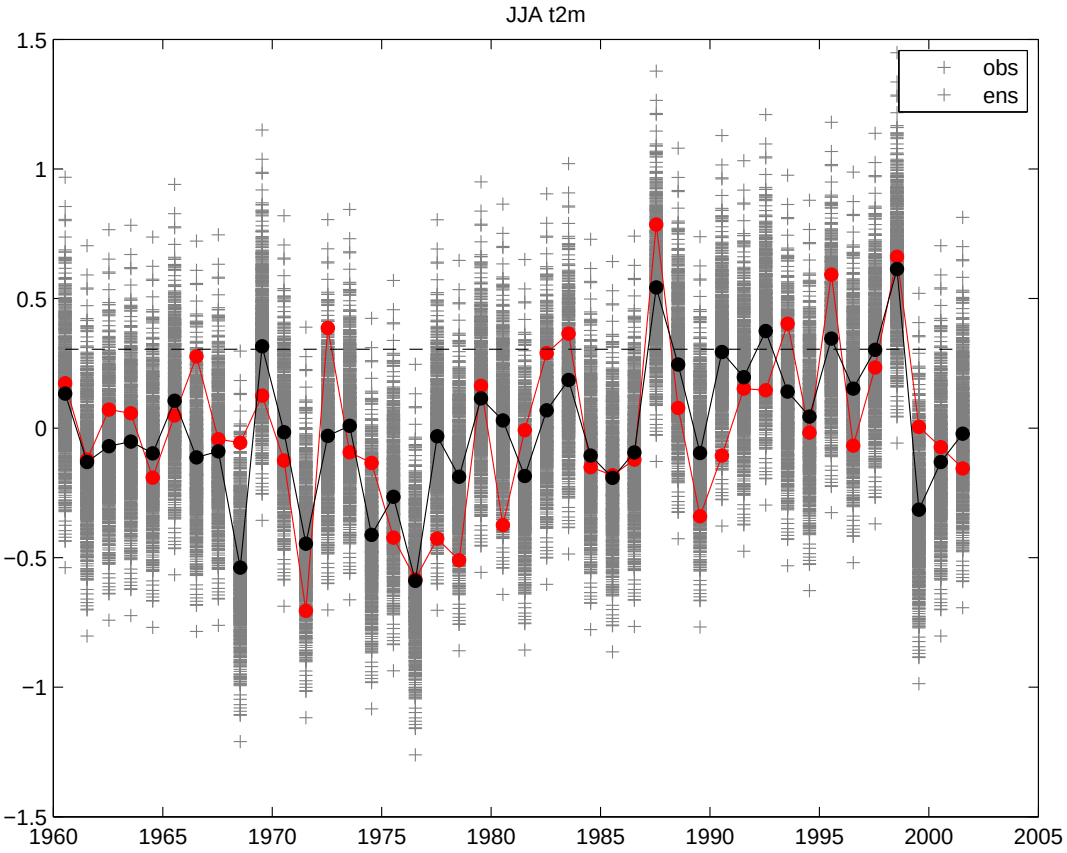
Probabilities vary continuously with thresholds.
Probability of rare events is not zero.
Access to forecast pdf.

1972 forecast is small but not zero probability when event occurred.
Brier skill score 0.52 (increased from 0.46).

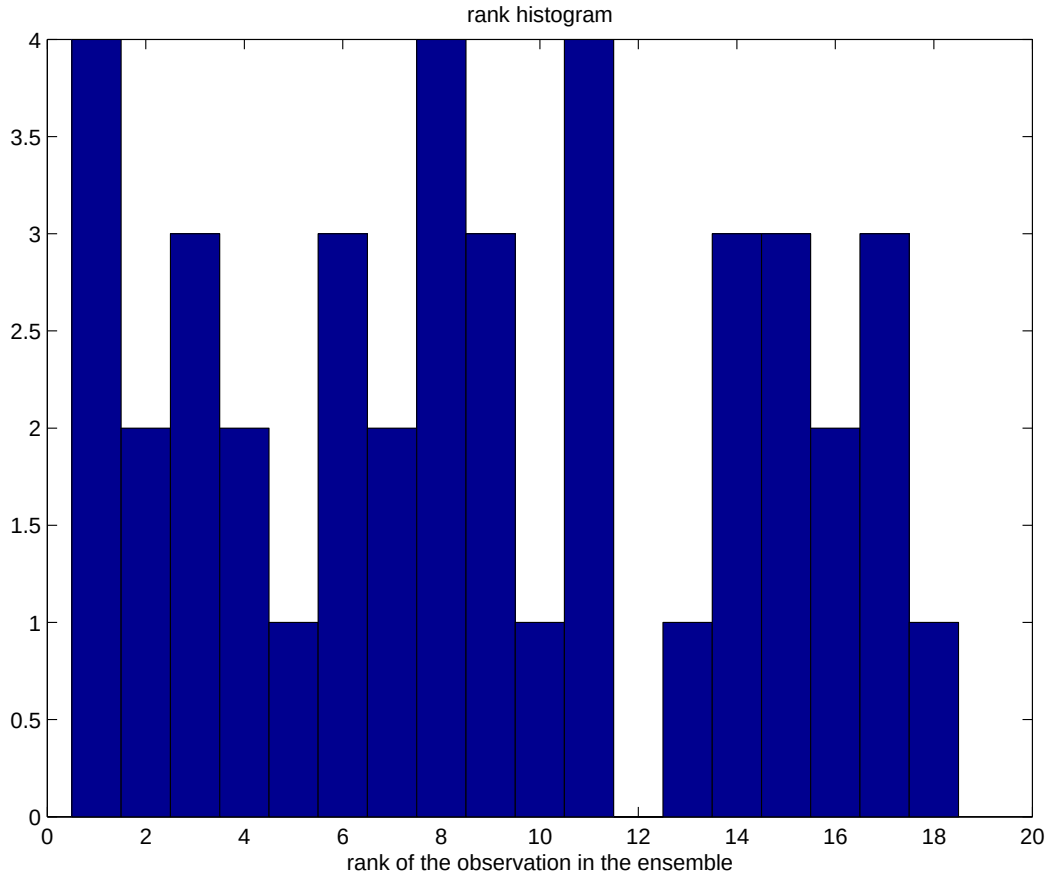


- In some problems, there is evidence that the forecast mean varies from year to year but the forecast spread is constant.
- If so, a much larger ensemble can be formed by centering the ensembles of past forecasts on the ensemble mean of the current forecast.

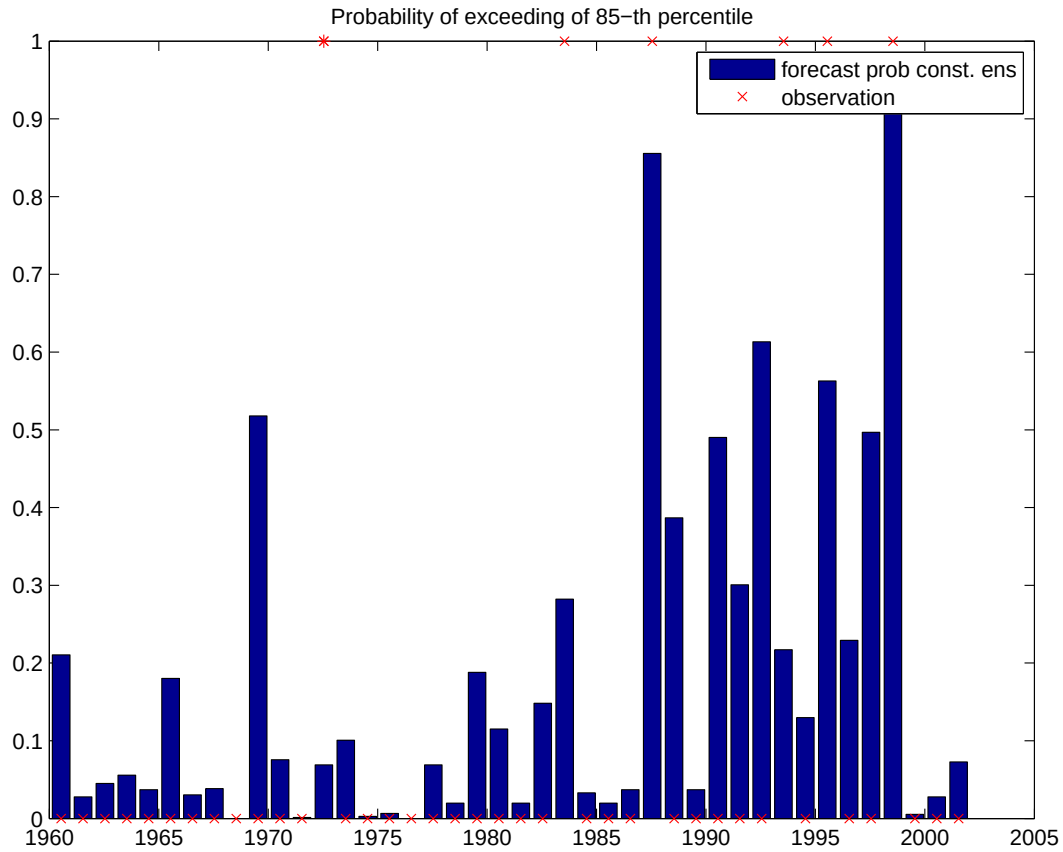
Constant ensemble.



Constant ensemble.

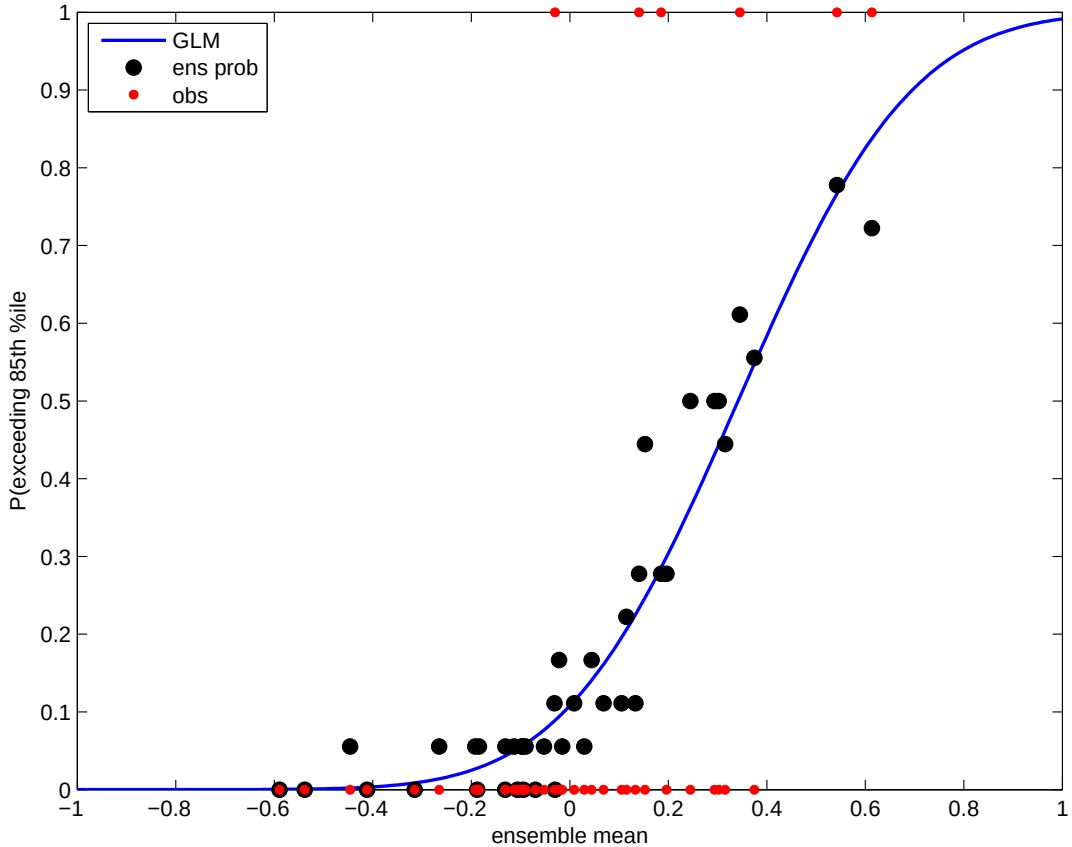


Constant ensemble.
Brier skill score 0.45.

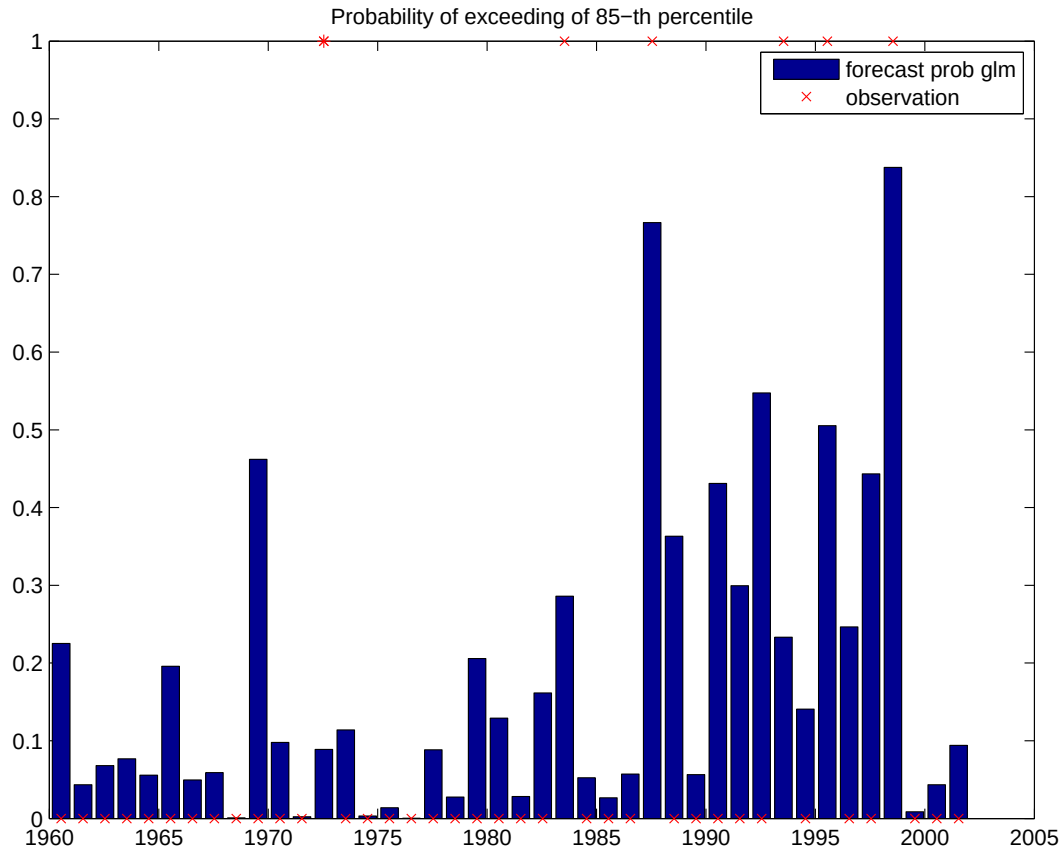


Form a GLM regression between ensemble based probabilities and the ensemble mean. Predict the probabilities from the ensemble mean.

Regression curve relating ensemble mean and probability of exceeding 85%ile.



GLM.
Brier skill score 0.48.



Summary

Constructing probability forecasts using methods that are performance independent.
Robust. Don't require long histories.
Counting

- Values
- Anomalies – multimodel, individual model.
- Standardized anomalies.
- Percentiles

Sampling error of counting

- Gaussian
- Constant (large) ensemble.
- GLM