Constructing Probability Forecasts

Michael K. Tippett

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

Statistical Methods in Seasonal Prediction, ICTP
Aug 2-13, 2010

Ideas

A forecast is completely desribed by a pdf What is a pdf and where do I get one?

- Regressions
  - OLS Gaussian
  - GLM Poisson
- Ensembles
  - Extract information from ensembles to construct forecast probabilities.

Ideas

A forecast is completely desribed by a pdf What is a pdf and where do I get one?

- Regressions
  - OLS Gaussian
  - GLM Poisson
- Ensembles
  - Extract information from ensembles to construct forecast probabilities.

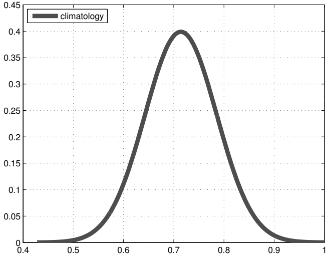
Climatological distribution

- Let Y be the quantity of interest, e.g., seasonal rainfall at some location.
- The climatological distribution of Y is the probability density function $p(y)$ estimated from past observations.
  - For seasonal forecasts, often based on a recent 30-year period.
  - For parametric distribution, estimate parameters, e.g., mean and variance of Gaussian.

$$p(y)\Delta y = \text{Prob}(y < Y < y + \Delta y)$$

$$\int_a^b p(y)\, dy = \text{Prob}(a < Y < b)$$

$$\int_{-\infty}^{\infty} p(y)\, dy = ??$$



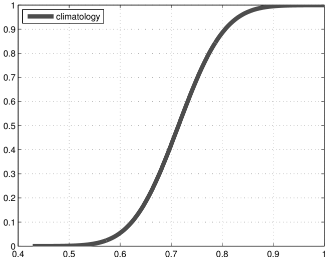
Climatological distribution

- Let Y be the quantity of interest, e.g., seasonal rainfall at some location.
- The climatological distribution of Y is the probability density function $p(y)$ estimated from past observations.
  - For seasonal forecasts, often based on a recent 30-year period.
  - For parametric distribution, estimate parameters, e.g., mean and variance of Gaussian.

Useful to define the cumulative density function (cdf).

$$F(y) \equiv \int_{-\infty}^{y} p(y')\, dy'$$

$$\text{Prob}(a < Y < b) = F(b) - F(a)$$



Climatological distribution: Example

- June-Aug temperature in Trieste.
- Period 1980-2009 30 years.
  - mean = 23.7
  - std = 1.1
- Answer questions like Prob($Y < 25$) $= 88\%$

Compute percentiles.
$q\%$-tile $= Y_q$ so that

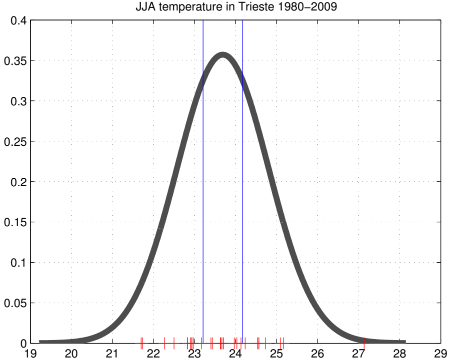
$$\text{Prob}(Y < Y_q) = q\%$$

More accurate than "counting"

Terciles

$$\text{Prob}(Y < 23.2) = 33\%$$

$$\text{Prob}(Y < 24.2) = 66\%$$



Climatological distribution: Example

- June-Aug temperature in Trieste.
- Period 1980-2009 30 years.
  - mean = 23.7
  - std = 1.1
- Answer questions like Prob($Y < 25) = 88\%$

Compute percentiles.
$q\%$-tile $= Y_q$ so that

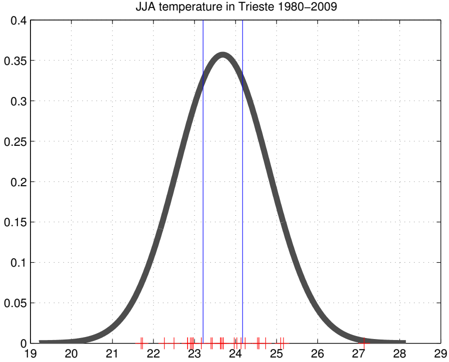
$$\text{Prob}(Y < Y_q) = q\%$$

More accurate than “counting”

Terciles

$$\text{Prob}(Y < 23.2) = 33\%$$

$$\text{Prob}(Y < 24.2) = 66\%$$



Climatological distribution: Example

- June-Aug temperature in Trieste.
- Period 1980-2009 30 years.
  - mean = 23.7
  - std = 1.1
- Answer questions like Prob($Y < 25$) $= 88\%$

Compute percentiles.
$q\%$-tile $= Y_q$ so that

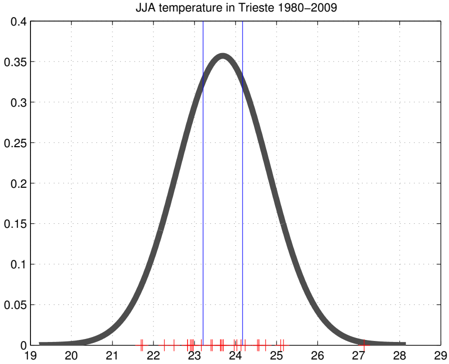
$$\text{Prob}(Y < Y_q) = q\%$$

More accurate than "counting"

Terciles

$$\text{Prob}(Y < 23.2) = 33\%$$

$$\text{Prob}(Y < 24.2) = 66\%$$



Climatological distribution: Example

- June-Aug temperature in Trieste.
- Period 1980-2009 30 years.
  - mean = 23.7
  - std = 1.1
- Answer questions like Prob($Y < 25$) $= 88\%$

Compute percentiles.
$q\%$-tile $= Y_q$ so that

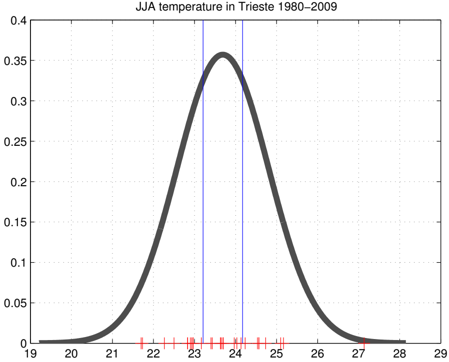
$$\text{Prob}(Y < Y_q) = q\%$$

More accurate than "counting"

Terciles

$$\text{Prob}(Y < 23.2) = 33\%$$

$$\text{Prob}(Y < 24.2) = 66\%$$



Climatological distribution: Example

- June-Aug temperature in Trieste.
- Period 1980-2009 30 years.
  - mean = 23.7
  - std = 1.1
- Answer questions like Prob($Y < 25$) $= 88\%$

Compute percentiles.
$q\%$-tile $= Y_q$ so that

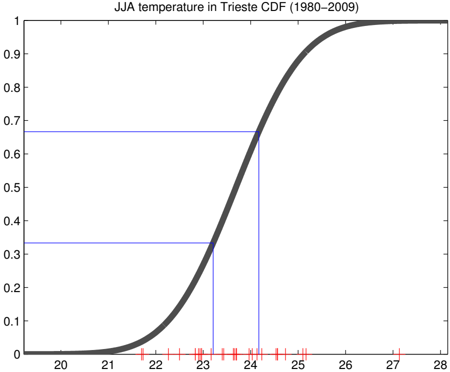
$$\text{Prob}(Y < Y_q) = q\%$$

More accurate than “counting”

Terciles

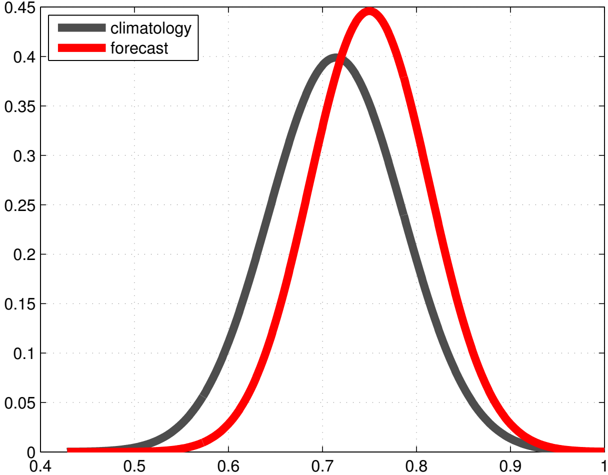
$$\text{Prob}(Y < 23.2) = 33\%$$

$$\text{Prob}(Y < 24.2) = 66\%$$



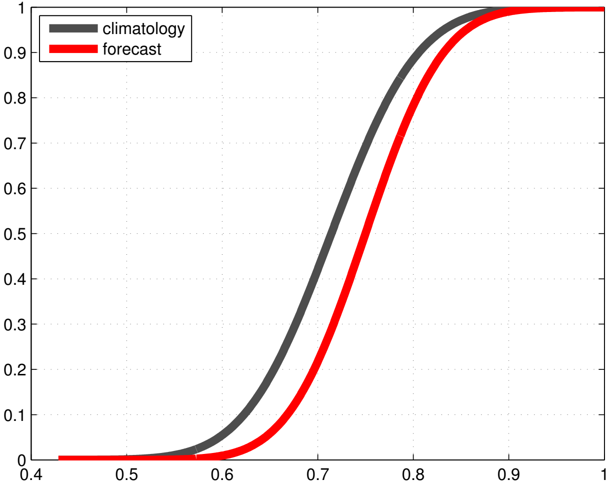
Forecast distribution

If the forecast is uncertain, as seasonal forecasts are, it is reasonable to describe it with a pdf.



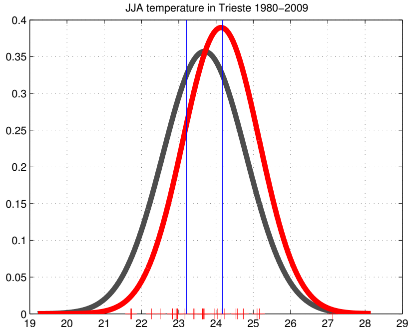
Forecast distribution

If the forecast is uncertain, as seasonal forecasts are, it is reasonable to describe it with a pdf. Or a cdf.



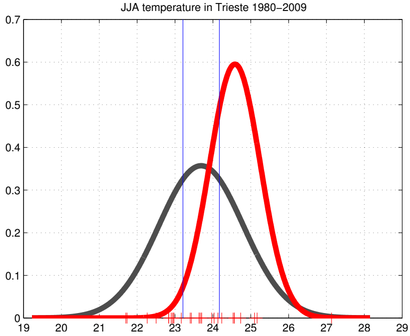
Forecast distribution: Example–modest skill

Forecast distribution tells the probability of exceeding any value.
Terciles:
Prob($Y < 23.2$) $= 18.2\%$
Prob($Y > 24.2$) $= 48.7\%$
Prob($23.2 < Y < 24.2$) $= 33.1\%$



Forecast distribution: Example–high skill

Forecast distribution tells the probability of exceeding any value.
Terciles:
Prob($Y < 23.2$) = 2.0%
Prob($Y > 24.2$) = 73.1%
Prob($23.2 < Y < 24.2$) = 25%



Should issue "narrow" forecast pdfs?

Probability forecasts

The forecast pdf (cdf) contains all the information about the forecast.

- Mean.
- Uncertainty
- Probabilities of exceeding, not exceeding between two values.

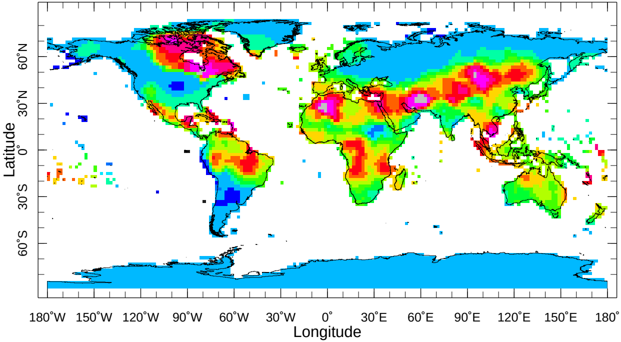
At a specific location, users may want to know the probability of exceeding some fixed value that is relevant to their application.

- Prob seasonal rainfall exceeds 100mm.
- Prob the longest dry spell is less than 12 days

Probability forecasts

When looking at many regions, with differing climatologies, it may be convenient to formulate probability forecasts in terms of percentiles.

Probability of temperature being above the 66%-tile. “Above” tercile category



Probabilistic forecasts

- Where do forecast pdf's (or at least categorical probabilities) come from? [Today]
- Are they any good? [Next week]
  - Are they what they say they are? (reliability)
    Do they reflect the forecast uncertainty?
    When the forecast probability of an event is 50%, does the event happen 50% of the time?
  - Are they better than nothing (climatology)? (resolution)
    Flipping a coin and forecasting the P(heads) = 50%, reliable, not so interesting.

Probability forecasts and regression

Conditional expectation

Suppose x and y are random variables.

- y forecast quantity
- x predictor

The best (mean square error sense) forecast of y given x is

$$E[y|x]$$

($|$ = "conditional on", $y|x$ = "y conditional on x")

- $E[y|x]$ is the *expected value* of y given the value of x.
- $p(y)$ is the climatological distribution.
- $p(y|x)$ is the forecast distribution (with mean $E[y|x]$)

What if x and y are independent?

By definition, $E[y|x]$ sounds like a good forecast.

Conditional expectation

Suppose x and y are random variables.

- y forecast quantity
- x predictor

The best (mean square error sense) forecast of y given x is

$$E[y|x]$$

($|$ = "conditional on", $y|x$ = "y conditional on x")

- $E[y|x]$ is the *expected value* of y given the value of x.
- $p(y)$ is the climatological distribution.
- $p(y|x)$ is the forecast distribution (with mean $E[y|x]$)

What if x and y are independent?

By definition, $E[y|x]$ sounds like a good forecast.

Conditional expectation

Suppose x and y are random variables.

- y forecast quantity
- x predictor

The best (mean square error sense) forecast of y given x is

$$E[y|x]$$

($|$ = "conditional on", $y|x$ = "y conditional on x")

- $E[y|x]$ is the *expected value* of y given the value of x.
- $p(y)$ is the climatological distribution.
- $p(y|x)$ is the forecast distribution (with mean $E[y|x]$)

What if x and y are independent?

By definition, $E[y|x]$ sounds like a good forecast.

Regression = conditional expectation

Suppose x and y have a joint normal distribution.

- x has a Gaussian distribution.
- y has a Gaussian distribution (climatological).
- (x, y) has a 2-D Gaussian distribution.

The forecast distribution $p(y|x)$ is Gaussian and

$$y|x \sim N\left(ax + b, (1 - \rho^2)\sigma_y^2\right)$$

where

$$a = \frac{\sigma_{xy}}{\sigma_x^2}\,, \quad b = \mu_y - a\mu_x\,, \quad \rho = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$$

Or

$$y|x = ax + b + \epsilon$$

with $\epsilon \sim N\left(0, (1 - \rho^2)\sigma_y^2\right)$. Variance is constant.

Moral

When you compute a regression $E[y|x]$,
you get a forecast pdf $p(y|x)$.

Not trivial to get a good estimate of the prediction error.

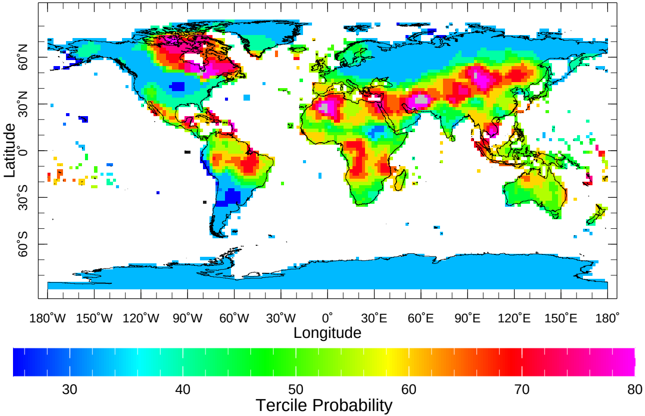
Moral

When you compute a regression $E[y|x]$,
you get a forecast pdf $p(y|x)$.

Not trivial to get a good estimate of the prediction error.

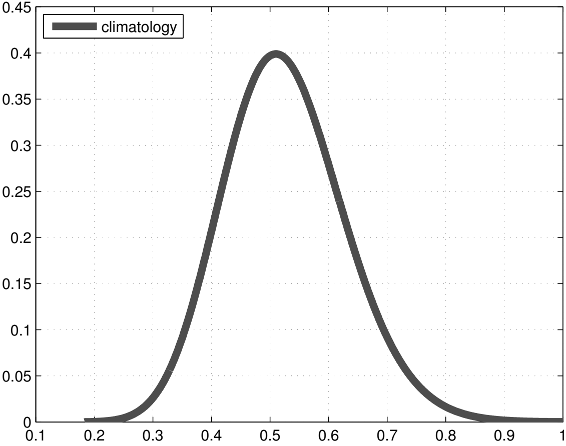
Regression example

IRI temperature forecast.
Probability of temperature being above the 66%-tile. “Above” tercile category



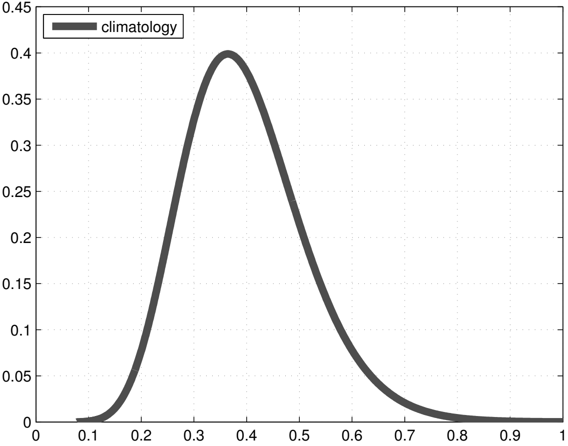
What if the forecast quantity is not Gaussian?

(Can seasonal precipitation really have a Gaussian distribution?)



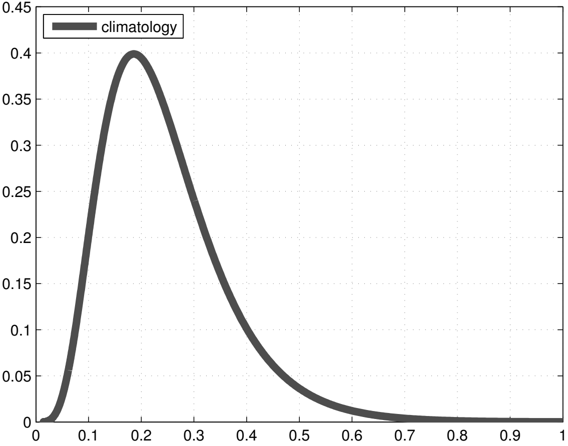
What if the forecast quantity is not Gaussian?

(Can seasonal precipitation really have a Gaussian distribution?)



What if the forecast quantity is not Gaussian?

(Can seasonal precipitation really have a Gaussian distribution?)



Box-Cox Transformation

A power transformation to make the data "more" Gaussian.

$$y - \rightarrow \begin{cases} \dfrac{y^\lambda - 1}{\lambda} & \lambda \neq 0 \\ \log y & \lambda = 0 \end{cases}$$

λ is chosen to maximize a likelihood function.

- This "fixes" the examples shown before.
- Transform data.
- Use Guassian methods–regression.
- Compute pdfs or probabilities.
- Invert transform.

What if the forecast quantity is *really* not Gaussian

- 1's and 0's. Binomial distribution.
- Count data. Poisson distribution.

Generalized linear models.

- Conditional expectation is modeled as a nonlinear function of the predictors.

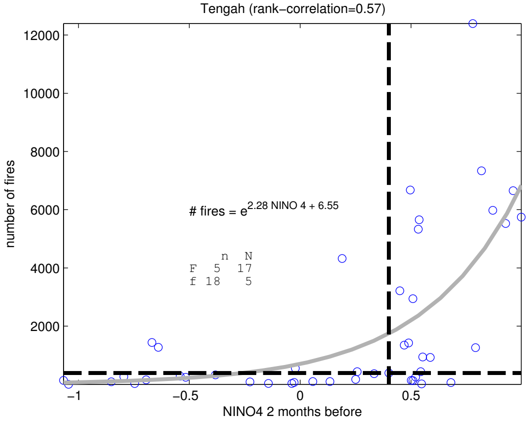
$$E(y|x) = g(x\beta)$$

Coefficients β are estimated by maximum likelihood.

Poisson regression example: Fire

y = # of fires in the Indonesian province Kalimantan Tengah.
x = NINO 4 two months before.

$$E[y|x] = \lambda = \log(ax+b)\,, \qquad p(y|x=n) = \frac{e^{-\lambda(x)}\lambda(x)^n}{n!}$$

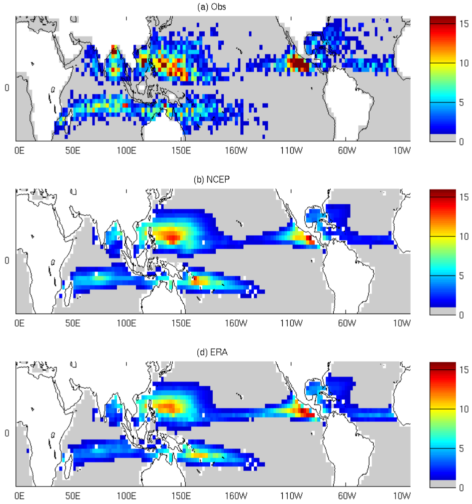


Poisson regression example: Tropical cyclone

y = #TC genesis events

x =

- absolute vorticity
- relative humidity
- relative SST
- vertical shear



fit to 40-yr climatology

Probability forecasts and ensembles

Constructing probability forecasts from ensembles

- Sample some methods of extracting information from the ensemble.
- Assume *some* aspects of the ensemble information are correct and use them.

Next week, how to correct ensemble defficiencies.

Can use model output at predictor in regressions.

"Counting" method

Pool ensemble members from different models.

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

Example:
JJA average temperature averaged over 2N - 32N, 64E - 93E.

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

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

“Counting” method

Pool ensemble members from different models.

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

Example:
JJA average temperature averaged over 2N - 32N, 64E - 93E.

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

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

“Counting” method

Pool ensemble members from different models.

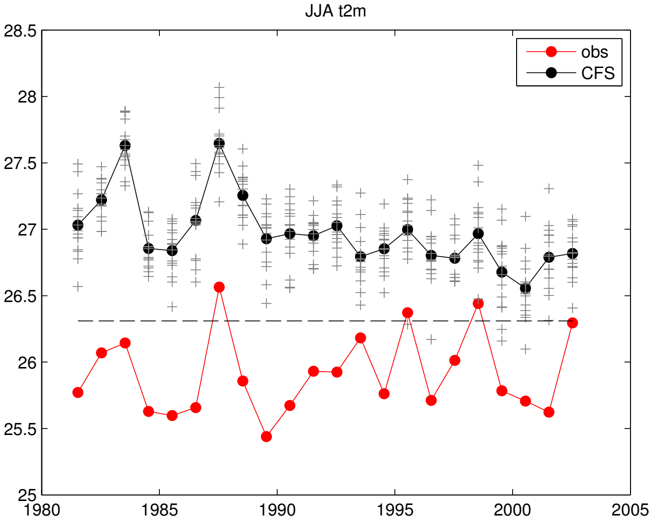
$$P(\text{event}) = \text{Ensemble frequency of event}$$

Example:
JJA average temperature averaged over 2N - 32N, 64E - 93E.

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

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

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



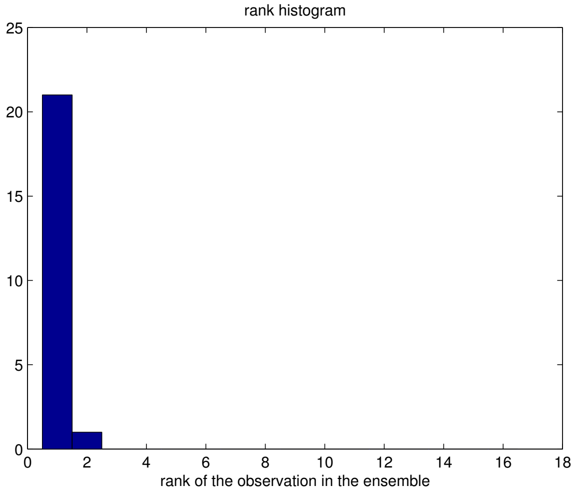
Measures of ensemble "goodness"

- 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.
- 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 below the ensemble.



"Counting" method

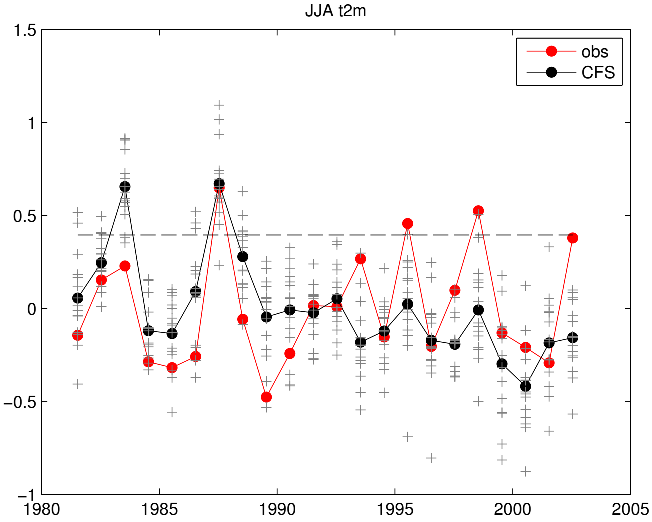
Refinement: Define event with respect to model climatology.

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

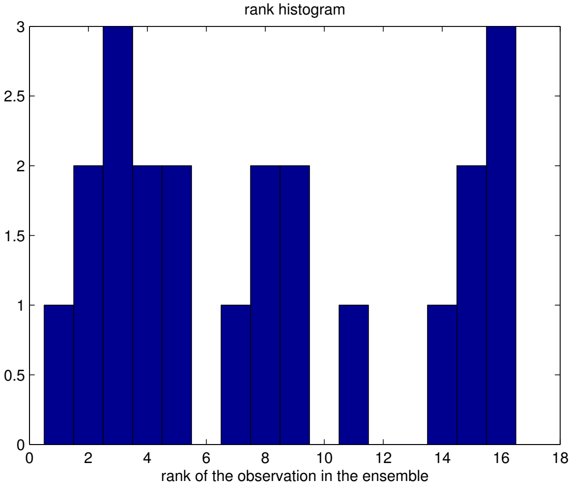
climatology = model climatology (historical)

- 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.



"Counting" method

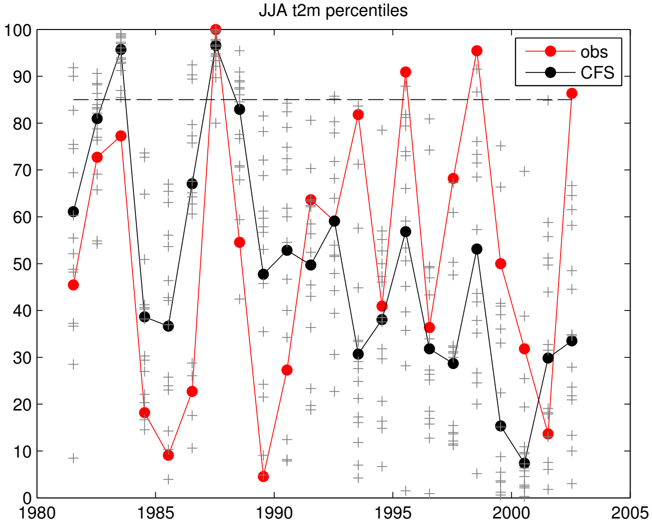
A way to account for errors in mean and spread:

Define event with respect to model percentiles.

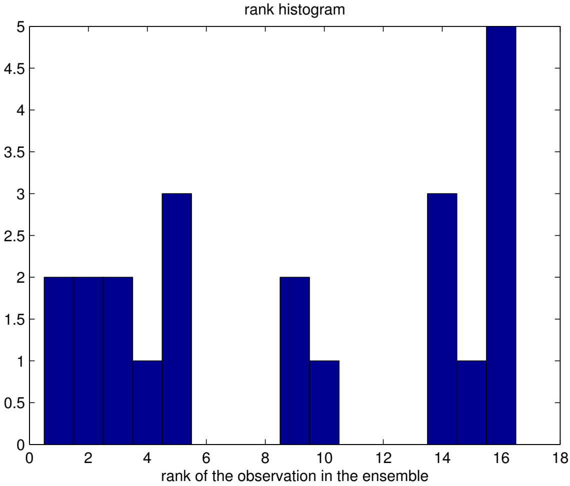
$$P(T > \text{85-th percentile}) = \frac{\text{\# ensemble members with } T > 85\%}{\text{\# ensemble members}}$$

percentile = model percentile (historical)

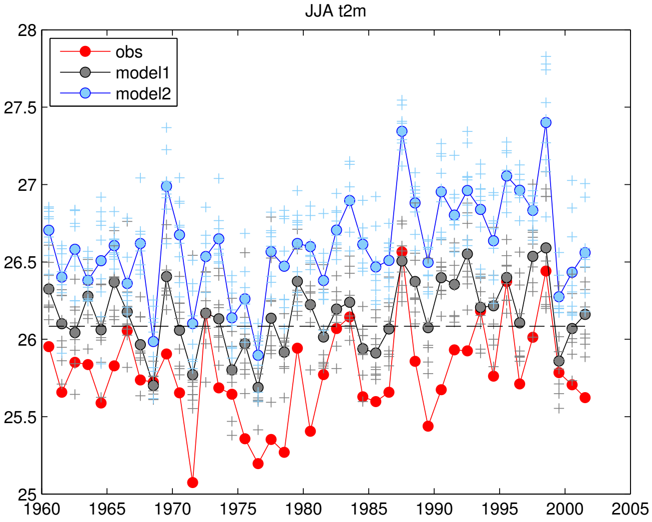
Model and observation percentiles



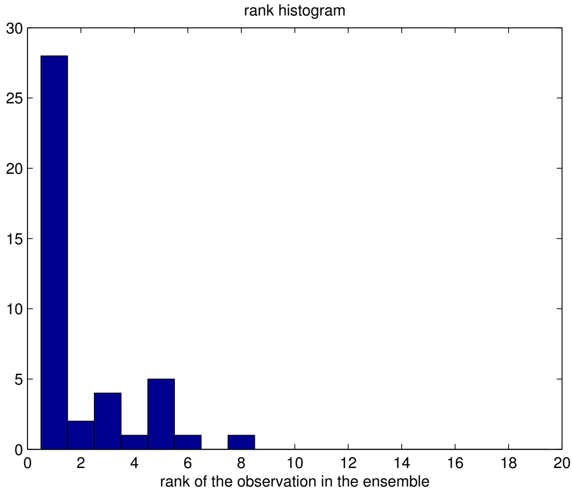
Correlation between ensemble mean and observation still 0.46.
Rank histogram indicates better spread.
Sample size is small.



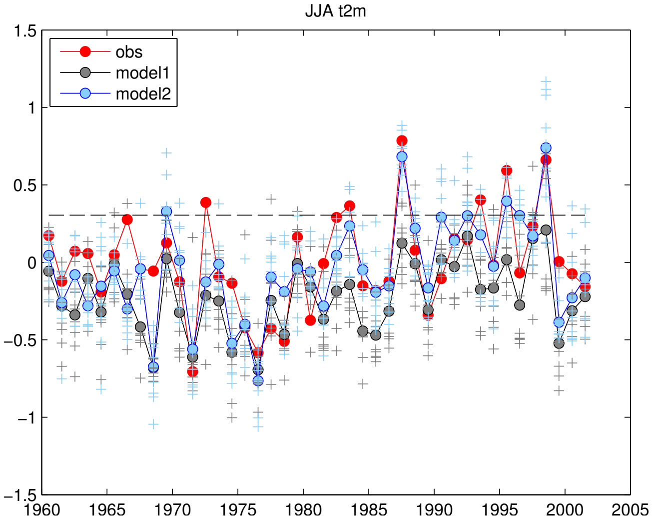
Two DEMETER models with different mean biases.



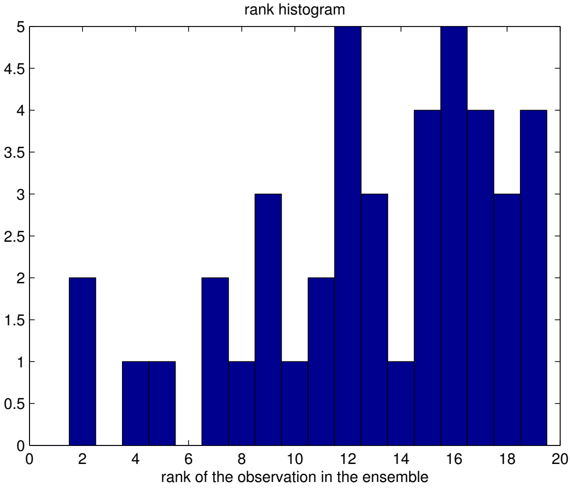
Correlation between ensemble mean and observation is 0.72.
Model 1 correlation = 0.68. Model 2 correlation 0.70.
Rank histogram indicates bias.



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.



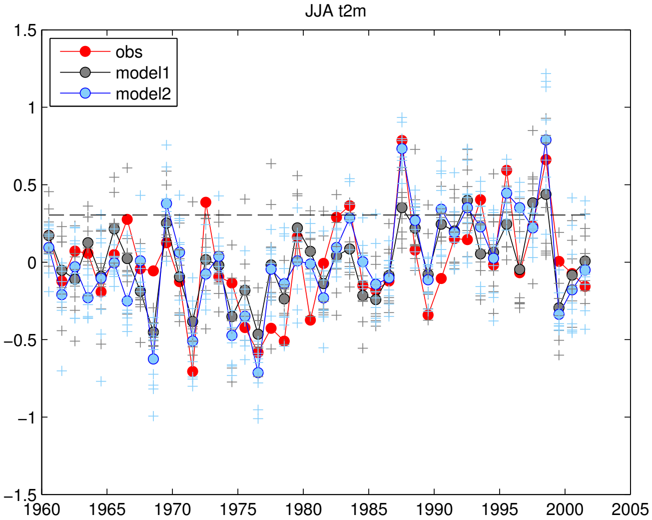
“Counting” methods

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

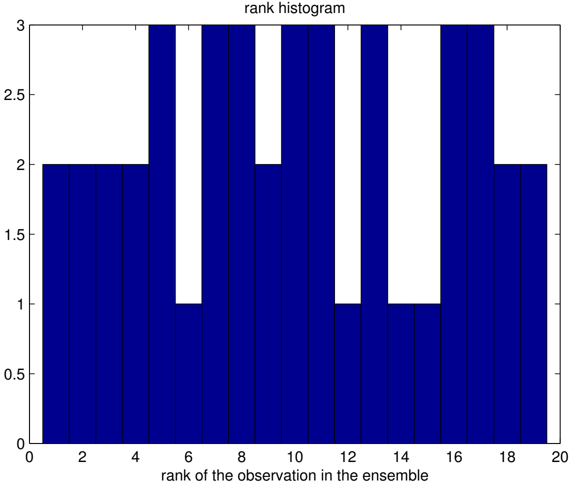
$$P(\text{event}) = \text{Ensemble frequency of event}$$

Accounts for mean biases of individual models.

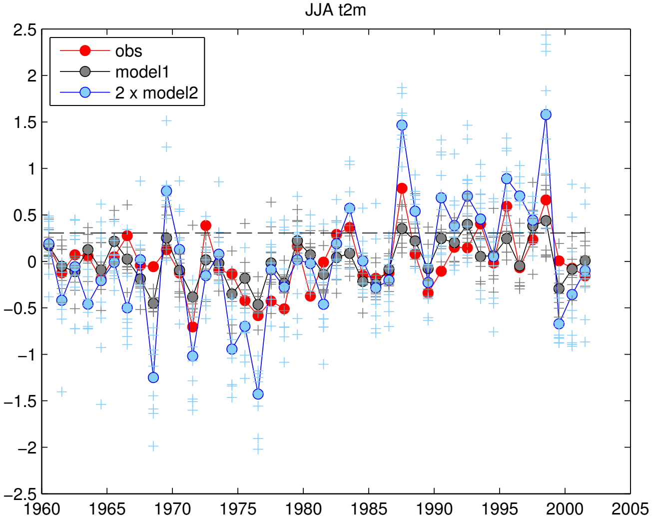
Pooled *anomalies*.



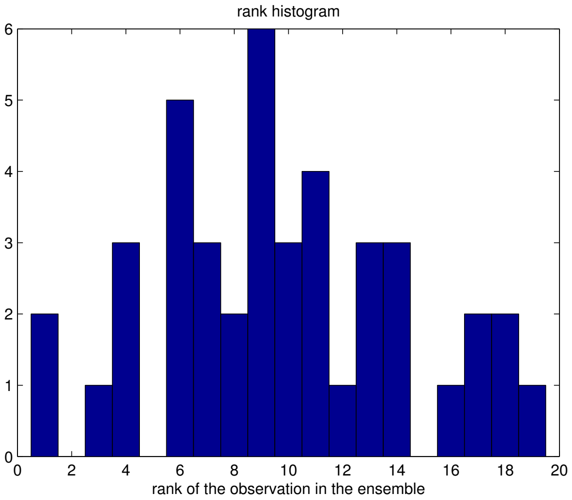
Correlation between ensemble mean and observation still 0.72.
Rank histogram indicates reduced bias.



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.



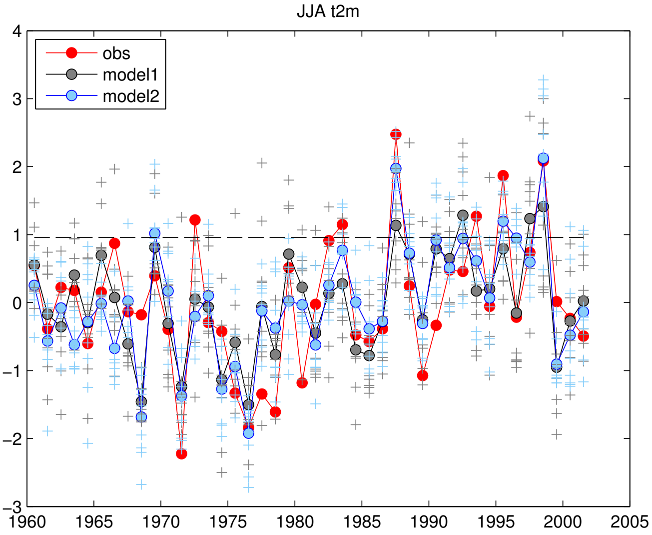
“Counting” methods

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

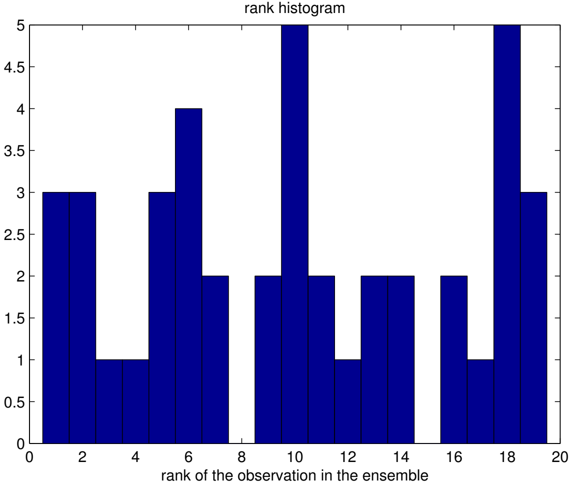
$$P(\text{event}) = \text{Ensemble frequency of event}$$

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.

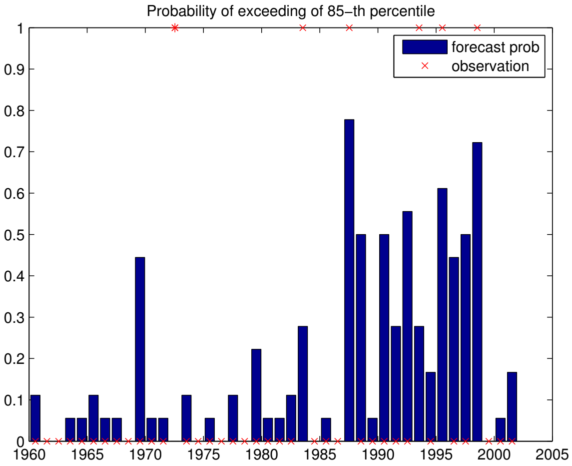


Sampling error

- 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 $P = 1$.

Rare event

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

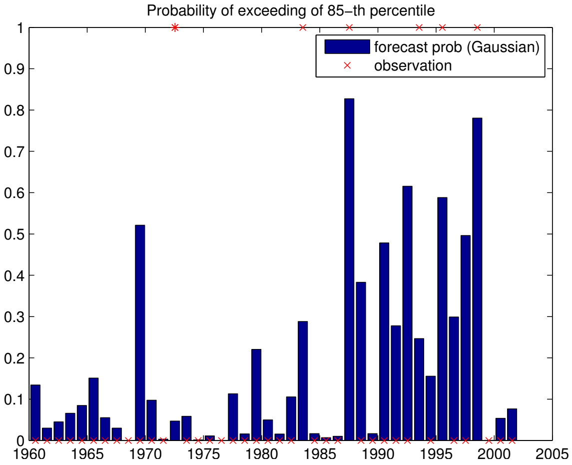
Example. $p(T)$ Gaussian.

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

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

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

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



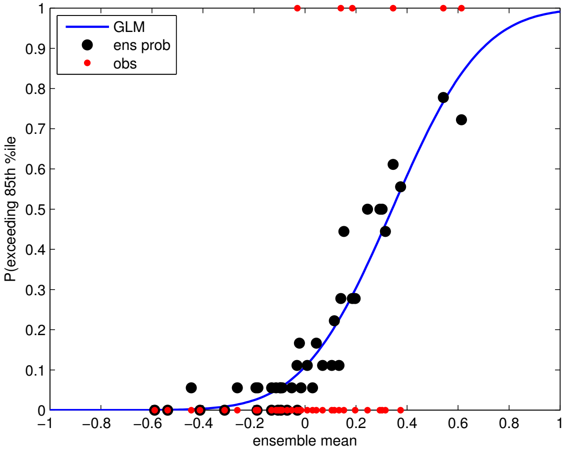
Using a Generalized Linear Model to remove noise

Form a GLM regression between

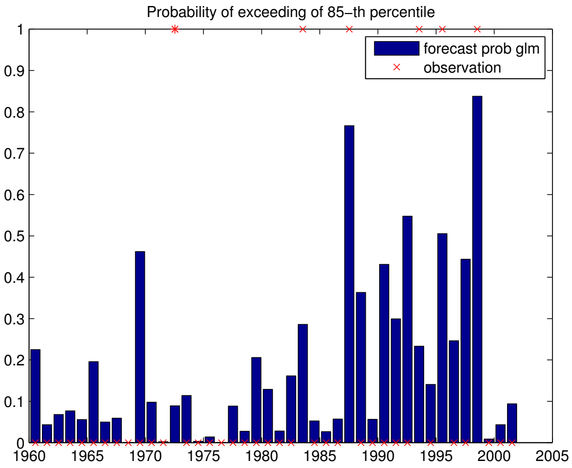
- ensemble based probabilities
- ensemble mean.

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

- Climatological distribution describes the observational record.
  (Trends?)
- The forecast pdf is a complete description of the forecast information.
- In regression, the conditional distribution is the forecast distribution.
  - OLS for Gaussian data.
  - Poisson for count data.
- Ensembles provide samples from the forecast distribution.
  - Model biases.
  - Small ensemble size.