



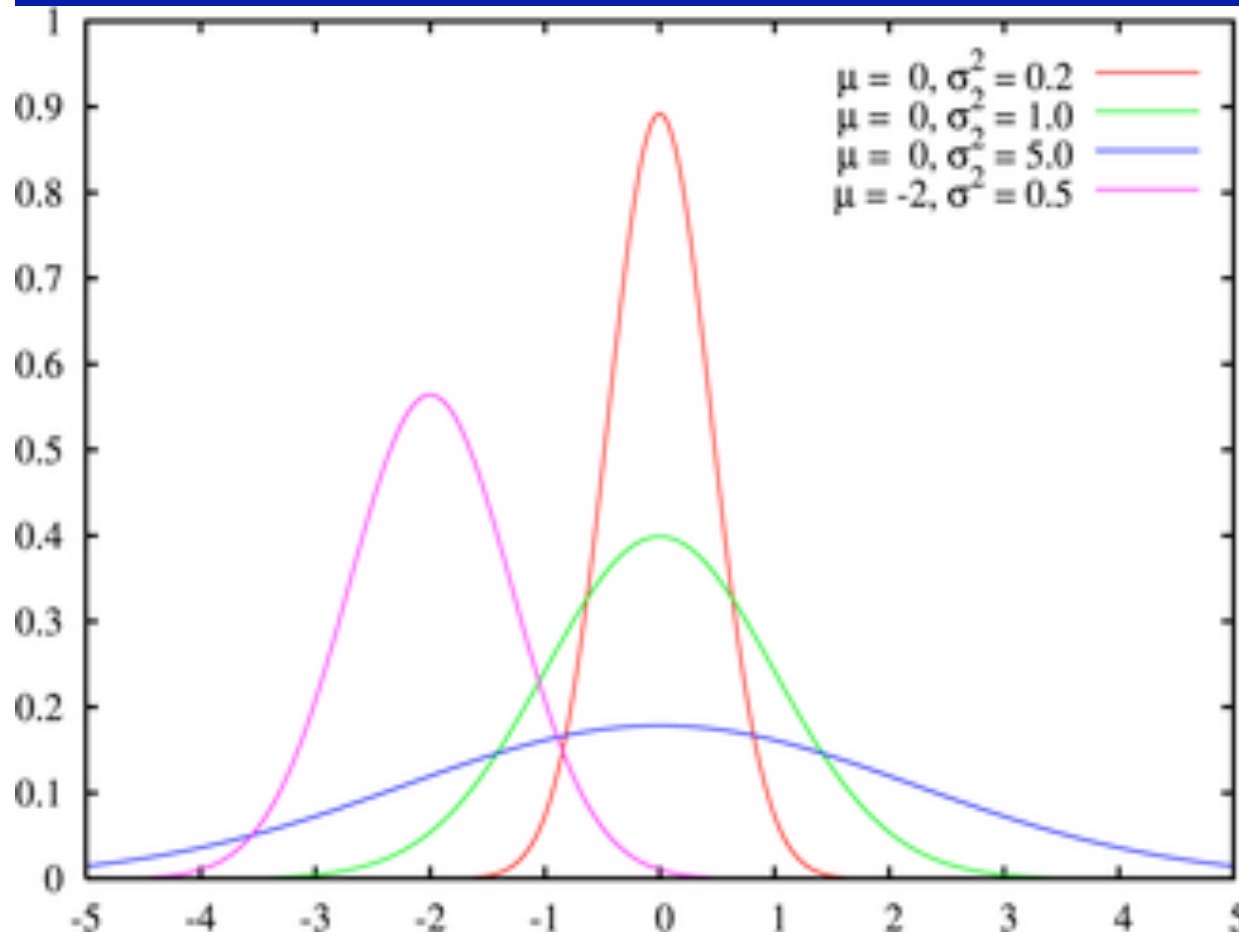
NCAR

# Probabilistic Forecasting

# pdfs and Histograms

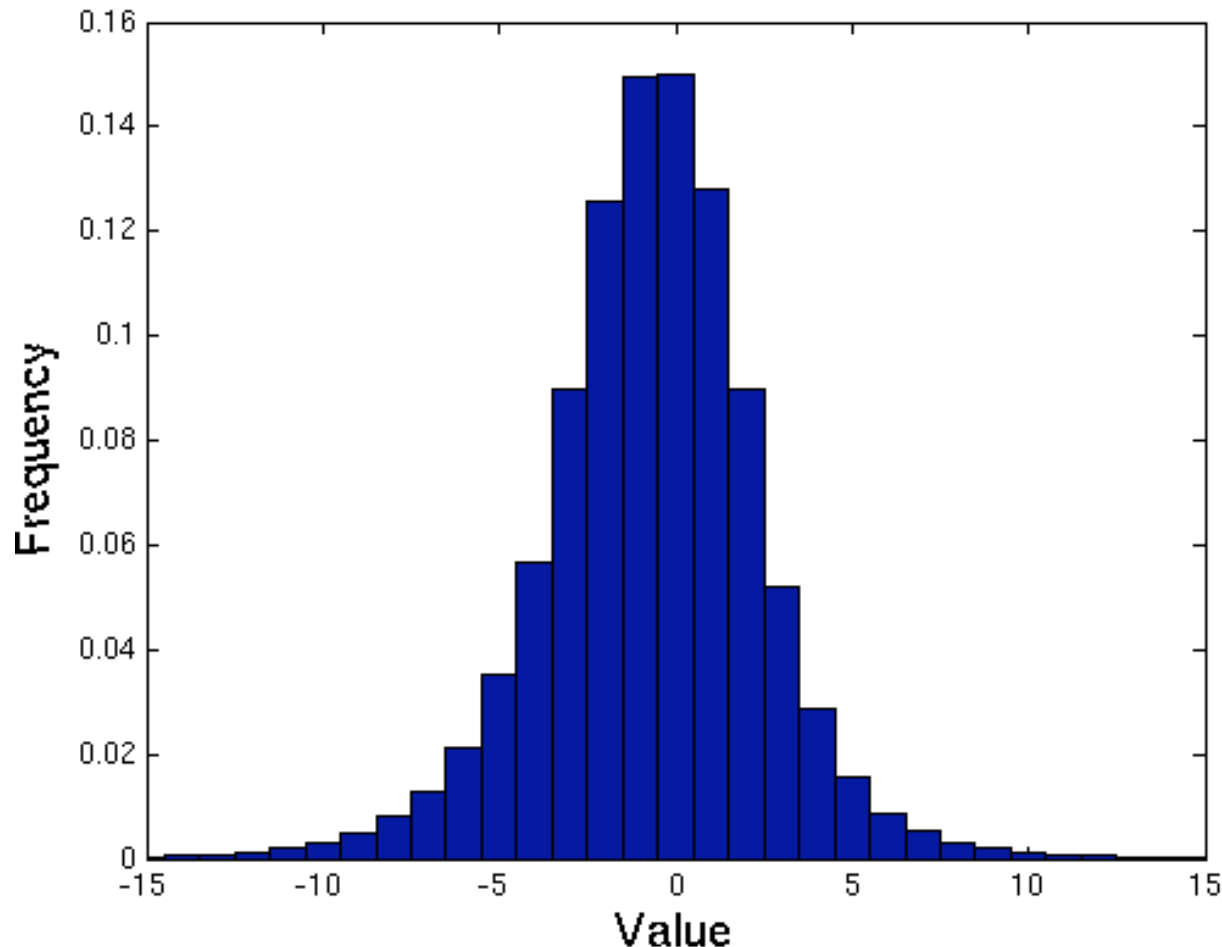


A few different normal (Gaussian) pdfs



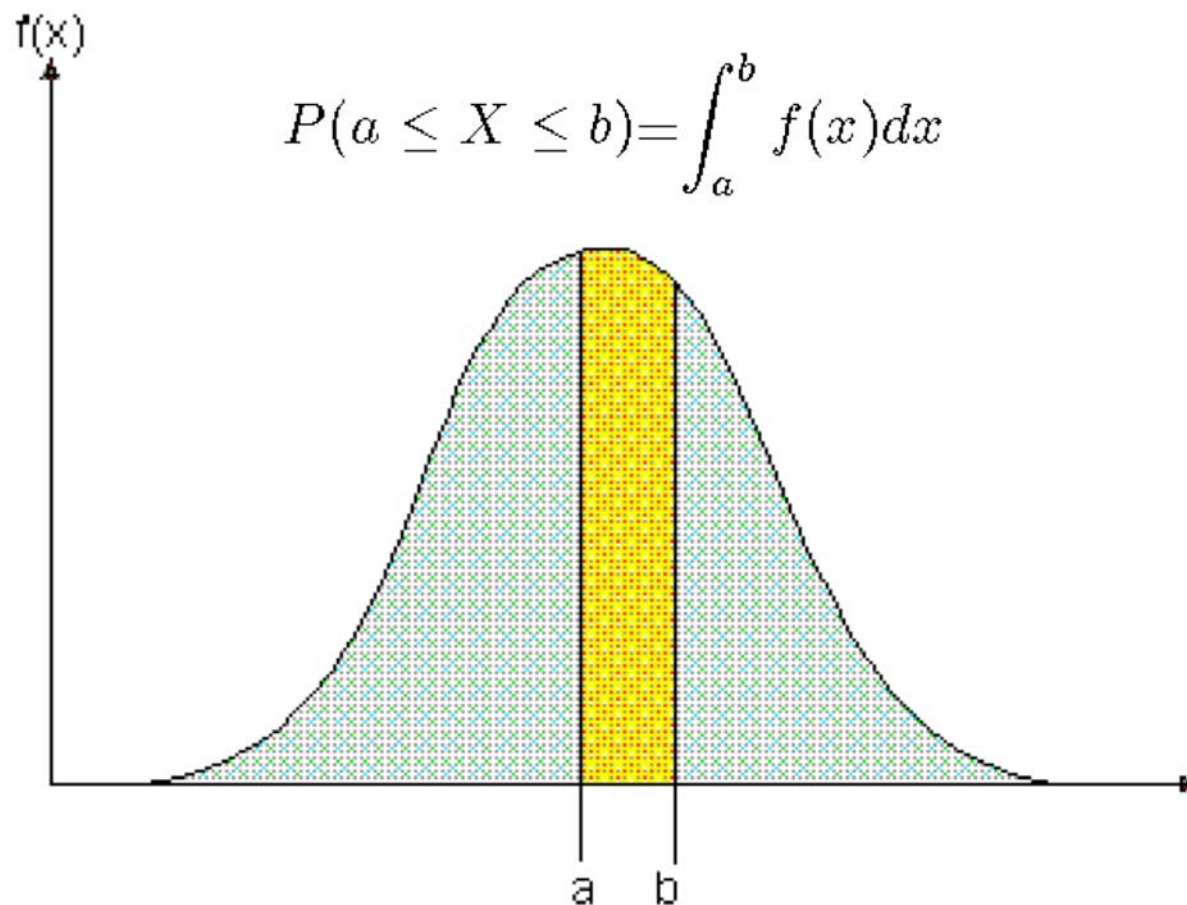
- Probability density functions (pdfs) are unobservable. They can only be estimated.
- They tell us the density, and must be integrated to get the probability.

# pdfs and Histograms



- Histograms are already integrated over the chosen bin width, and provide an estimated probability.
- One might fit a function to a histogram to arrive at a pdf.

# pdfs and Histograms

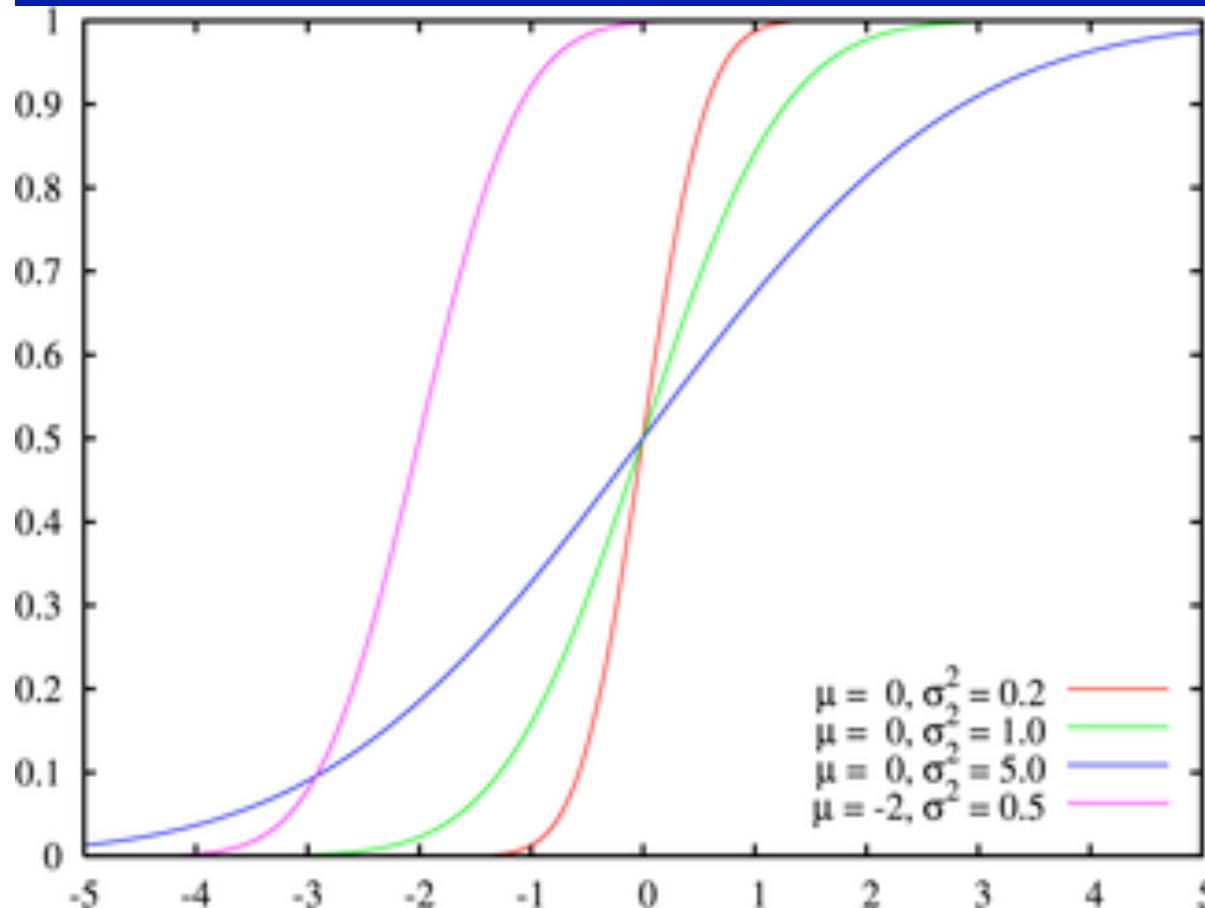


- Probability density functions (pdfs) are unobservable. They can only be estimated.
- They tell us the density, and must be integrated to get the probability.

# cdfs and thresholds

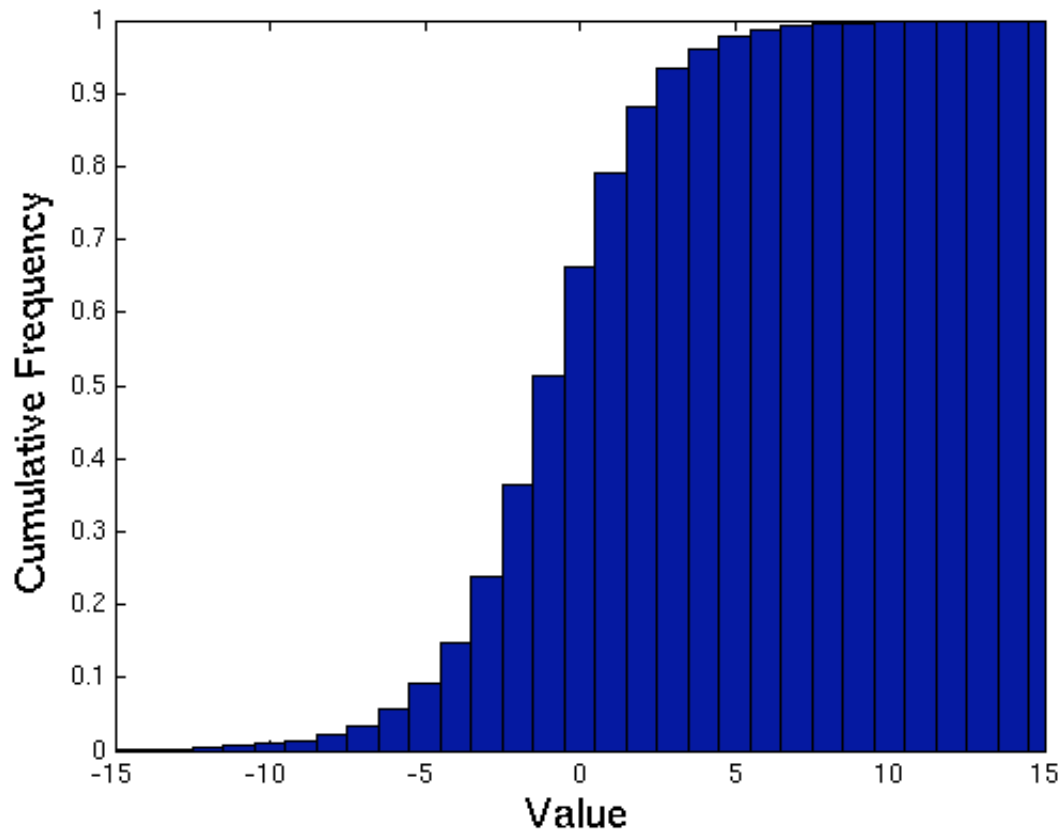


A few different normal (Gaussian) cdfs



- Can integrate from from one point to infinity to get the cumulative distribution function (cdf)

# cdfs and thresholds



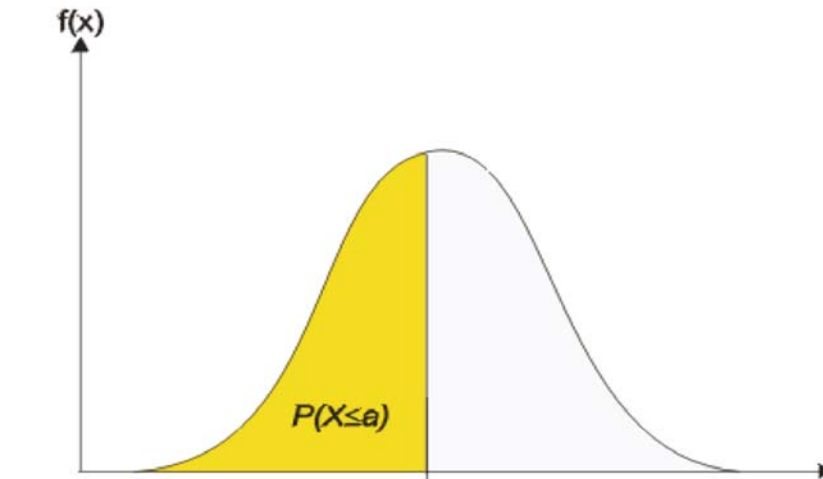
- Histograms can also be accumulated.
- One might fit a function to a cumulative histogram to arrive at a cdf.

# pdfs and cdfs

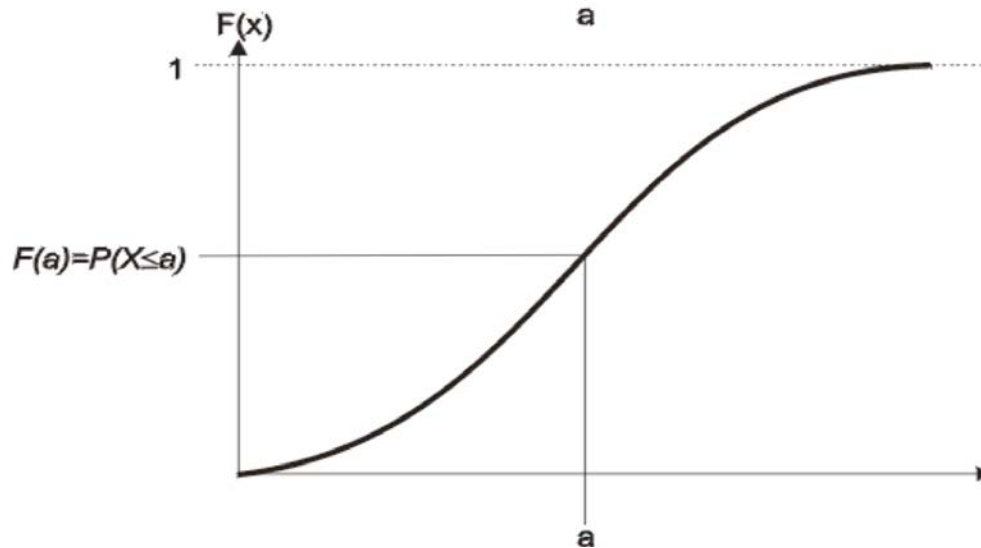


NCAR

Random Variable X



$$F(x) = P(X \leq x) = \int_0^x f(s) ds$$



# Verifying probabilistic forecasts for usefulness



NCAR

- Reliability: agreement between forecast frequency/probability and observed frequency
- Resolution: ability of a forecast to discriminate between events
- Sharpness: tendency to forecast event probabilities of 0 or 1 instead of clustering around the mean

# Complementary metrics



NCAR

- Forecast conditioned on the observations

$$p(x^f | y^o)$$

- Observations conditioned on the forecasts

$$p(y^o | x^f)$$

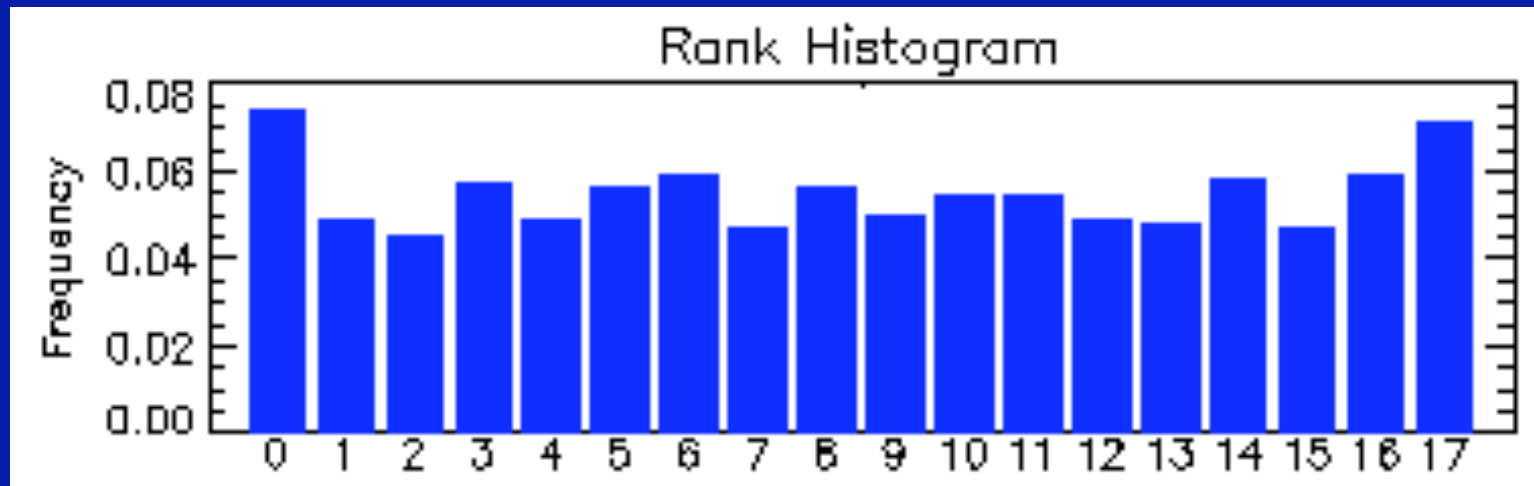
# Reliability



NCAR

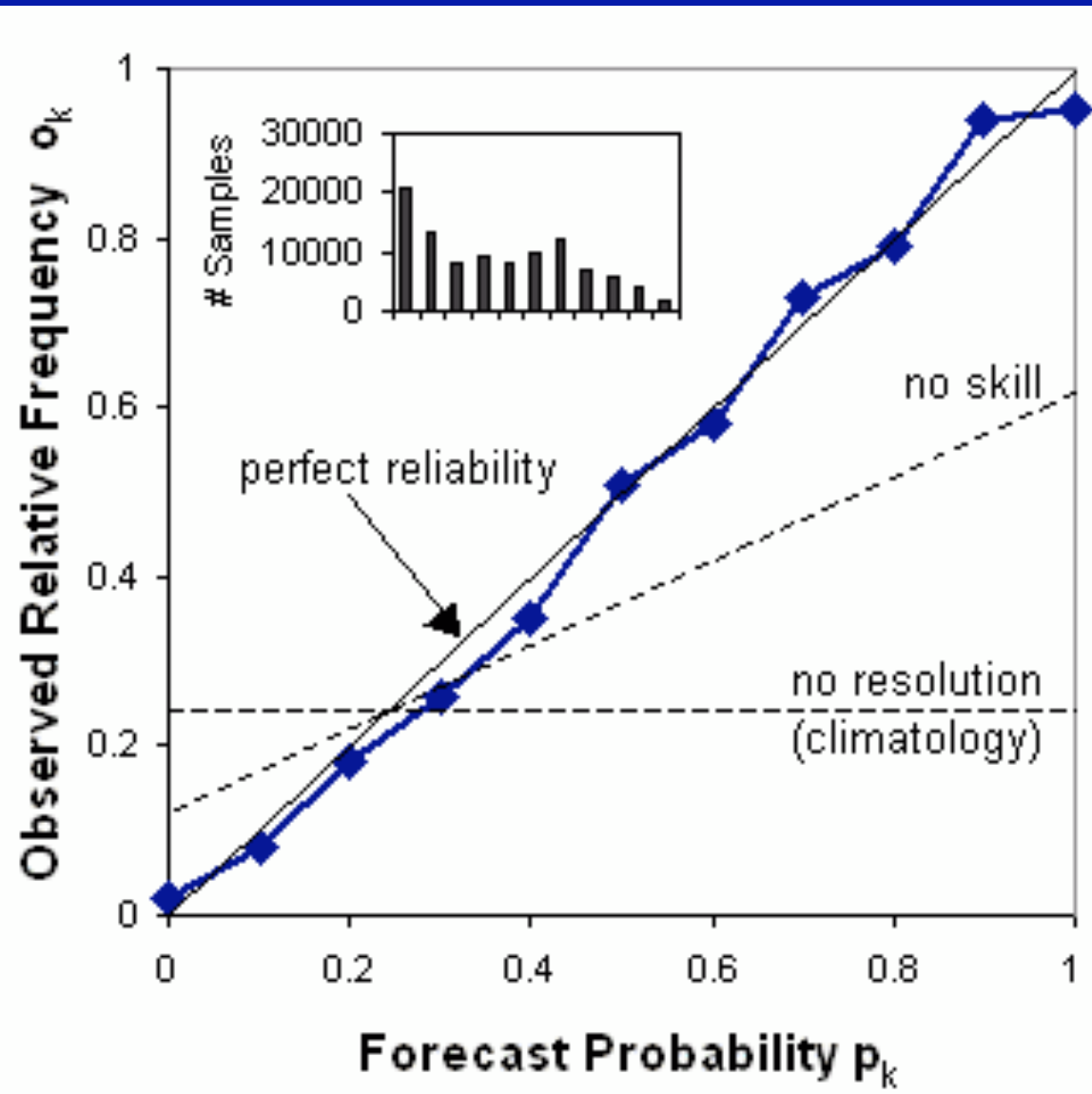
- Rank Histogram: How well does the ensemble spread in the forecast represent uncertainty, on average?
- Reliability Diagram: How well do the predicted probabilities of an event correspond to their observed frequencies?

# Rank Histogram



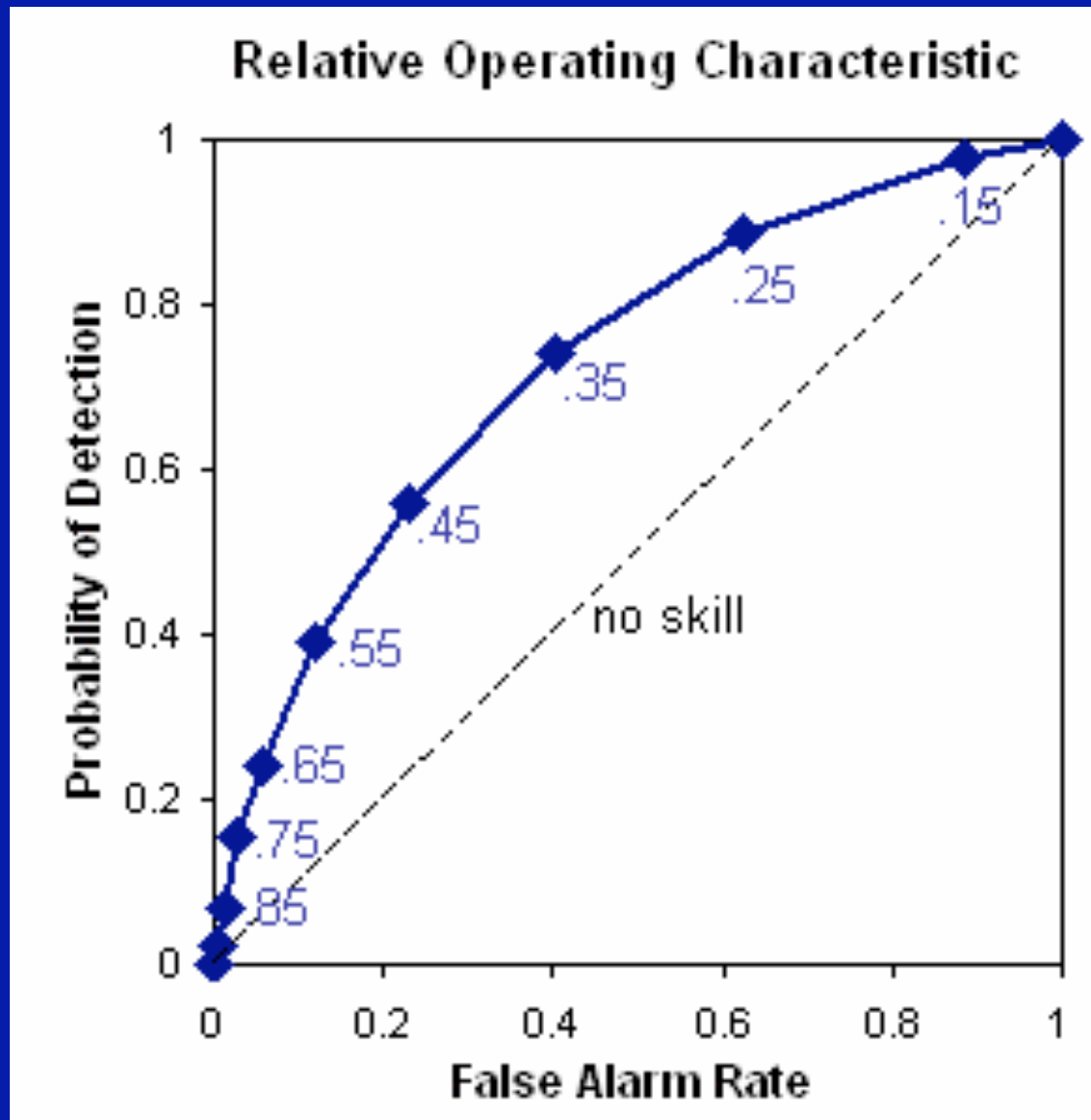
- **U-shaped:** observations usually outside of ensemble envelope; underdispersive ensemble
- **Flat:** observations usually indistinguishable from the members of the ensemble
- **Humped:** observations usually in the middle of the ensemble; overdispersive ensemble

# Reliability Diagram



- Given that  $X$  was *predicted* with probability  $Y$ , what was the outcome?
- How well do the observations of an event correspond to the predicted probabilities?
- A forecast of climatology has no resolution.

# Resolution



- Given that  $X$  was observed with probability  $Y$ , what was the forecast?
- How well did the probability forecast predict the category bin containing the observation?

# Calibration



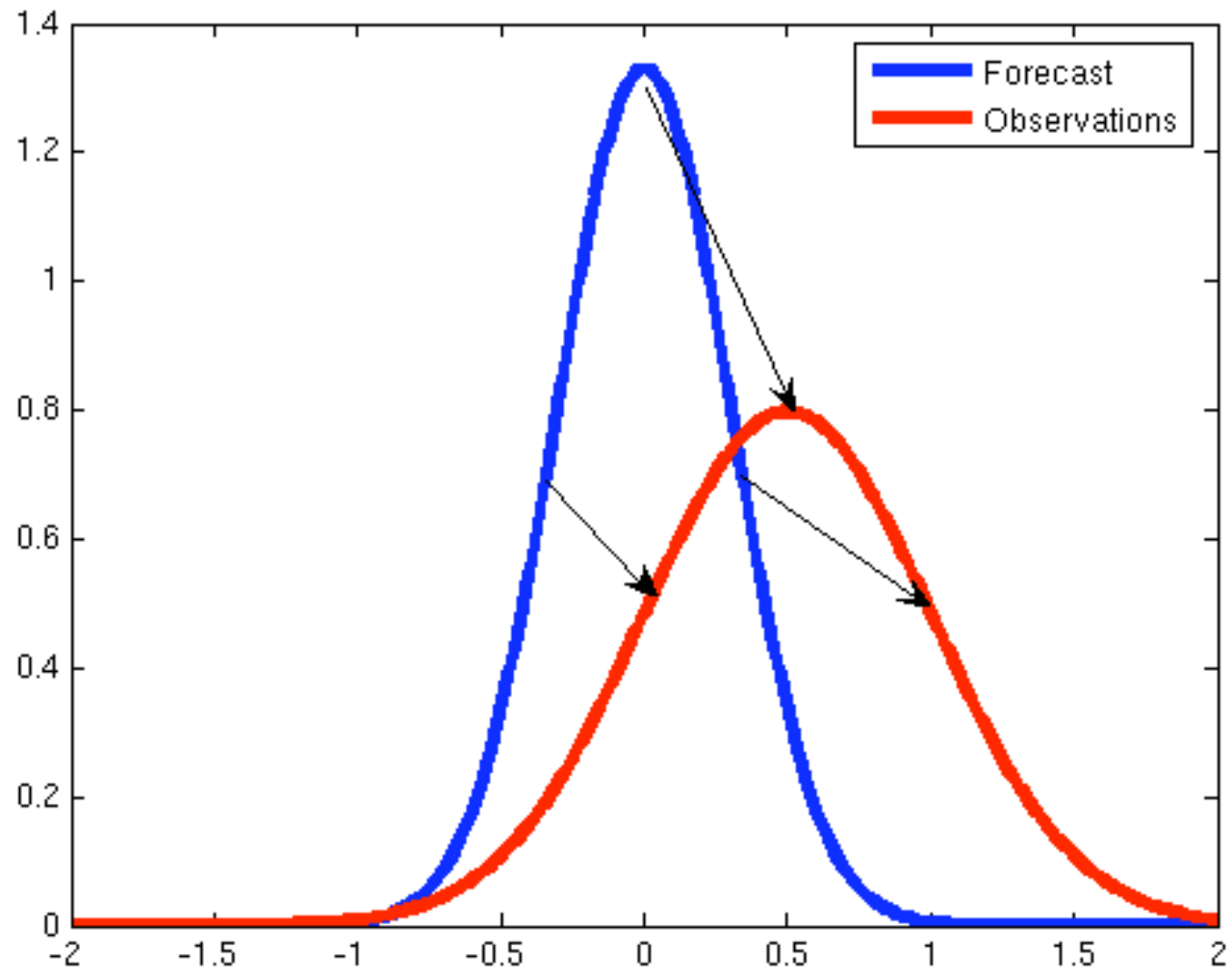
NCAR

- Probabilistic calibration is necessary because the model cannot produce the observed distribution
- This includes correcting both the bias (mean) and the variability (spread)

# Calibration



NCAR

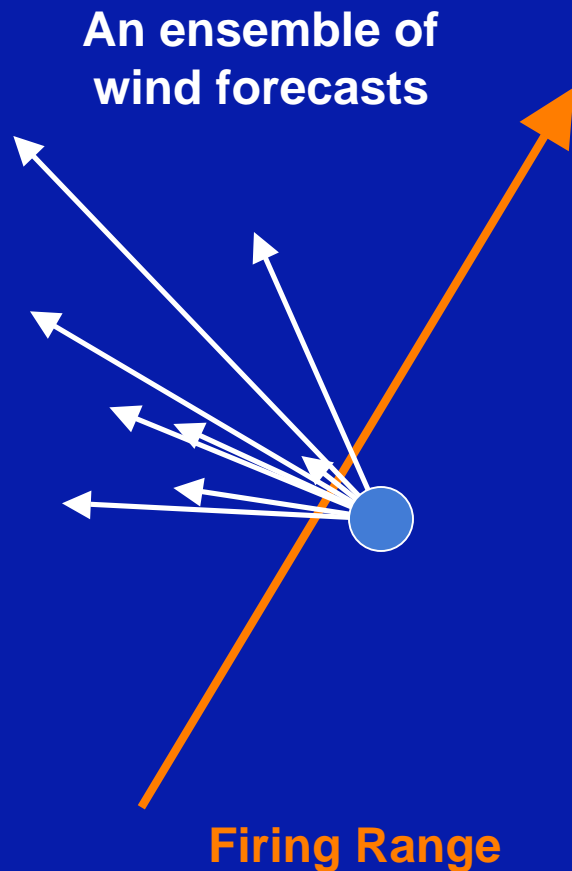


# Test Environment

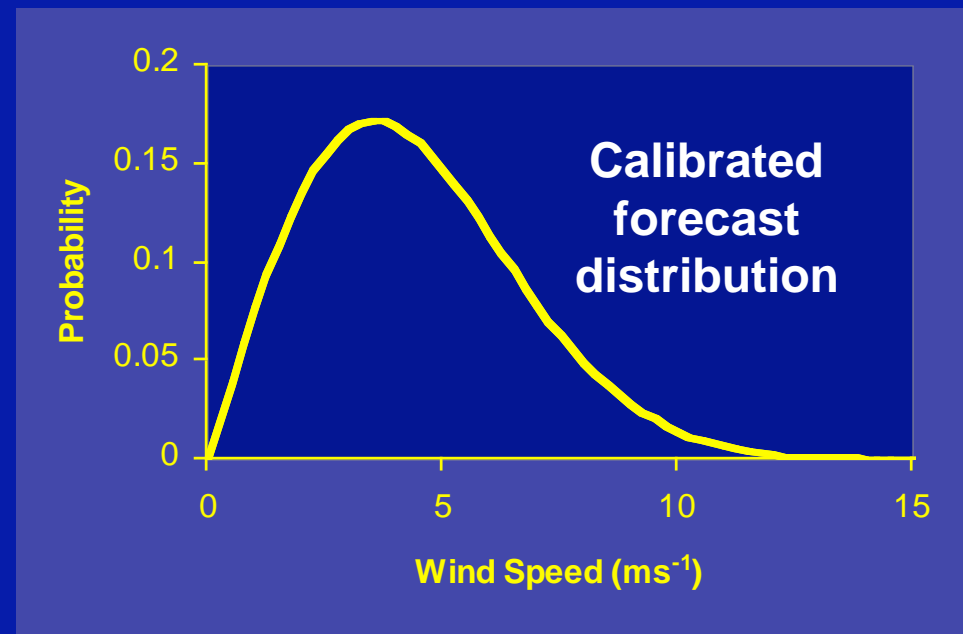


- Lot-acceptance ammunition testing
- Planning and test completion thresholds of 5 and 7 m/s crosswinds
- Peak winds (gust) on-site decisions

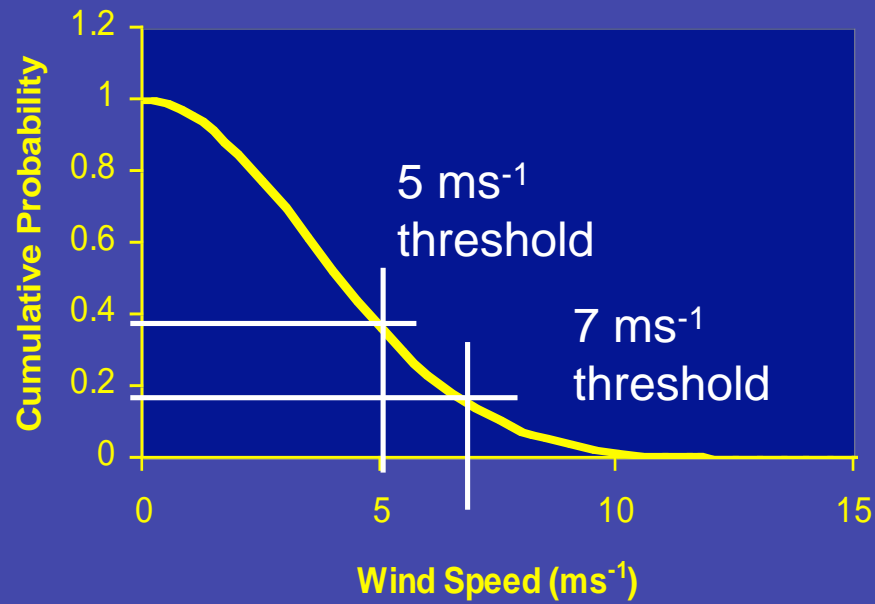
# Probabilistic Forecasts for Direct-Fire Ballistics



Crosswind component:

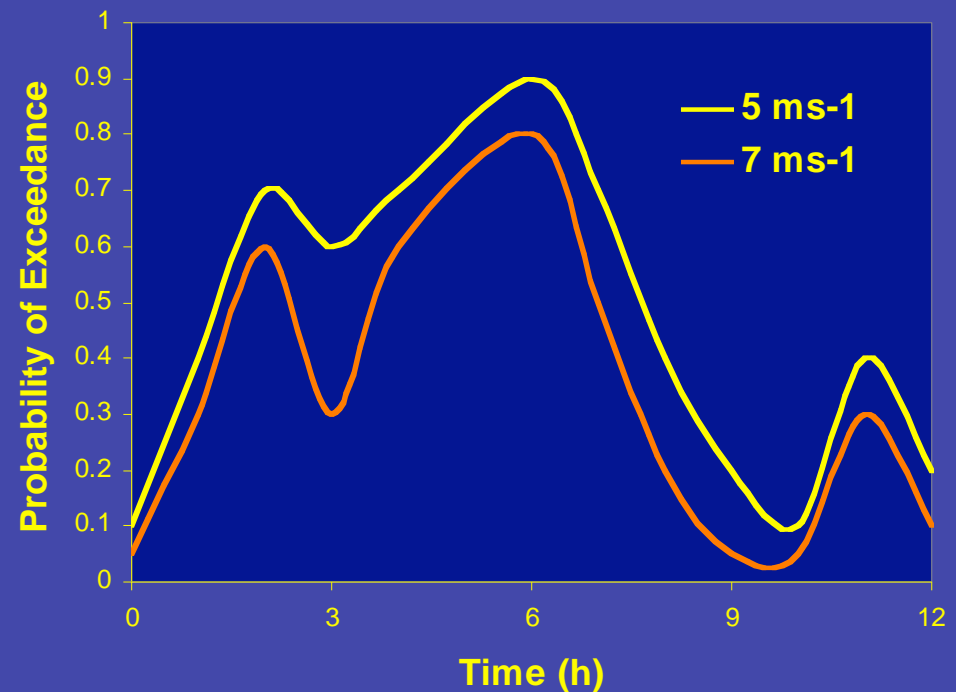


# Probability Forecasts for Direct-Fire Ballistics



Cumulative distribution for a given time

Time series: probability of exceedance



# Goal: Reliable Ensembles for Crosswind Thresholds



NCAR

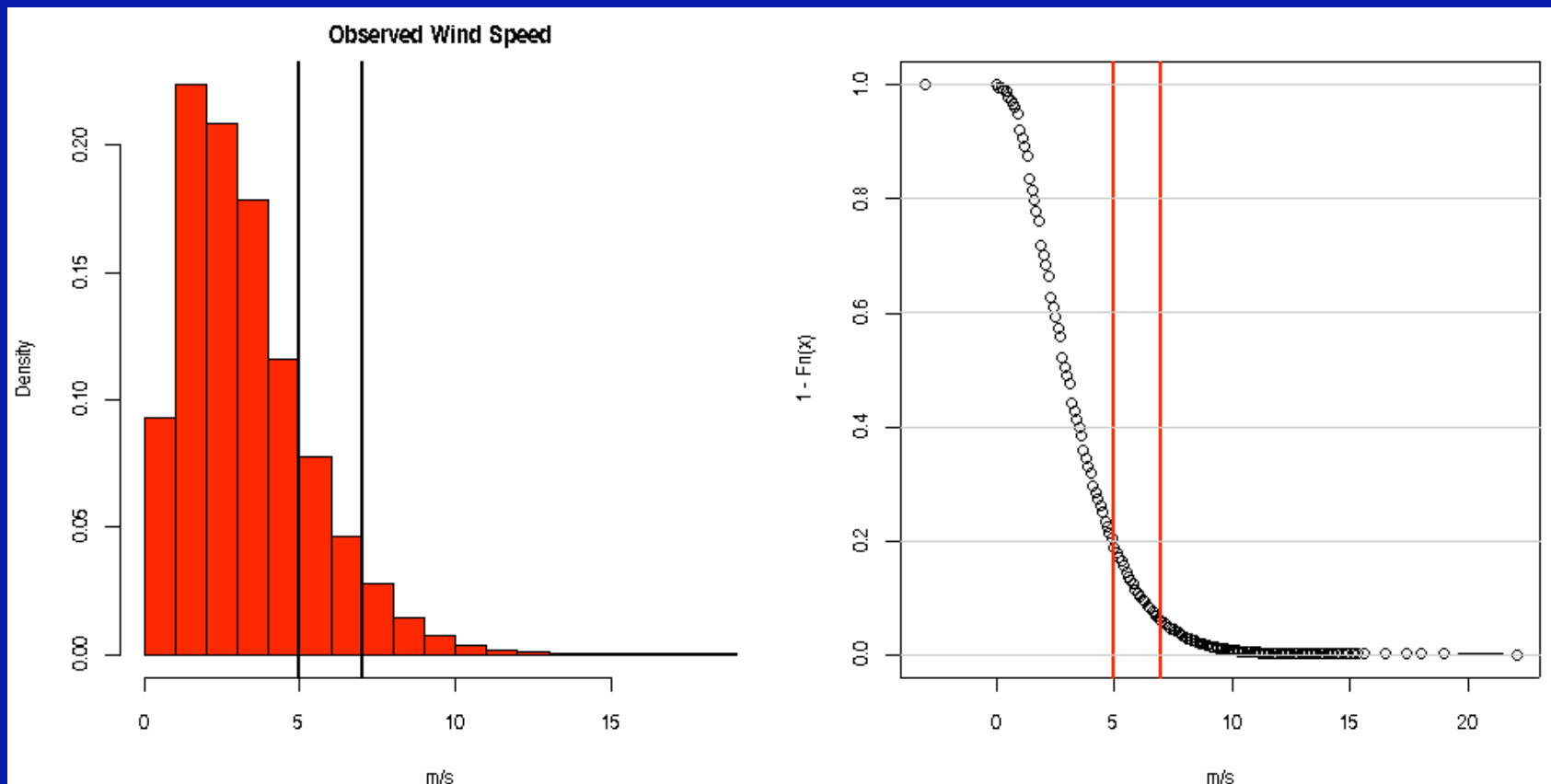
- Over several forecasts, the verification is statistically indistinguishable from the ensemble.
- Model error must be taken into account (calibration).
- Reliability is the first step, later we will consider *resolution*.

# Calibration



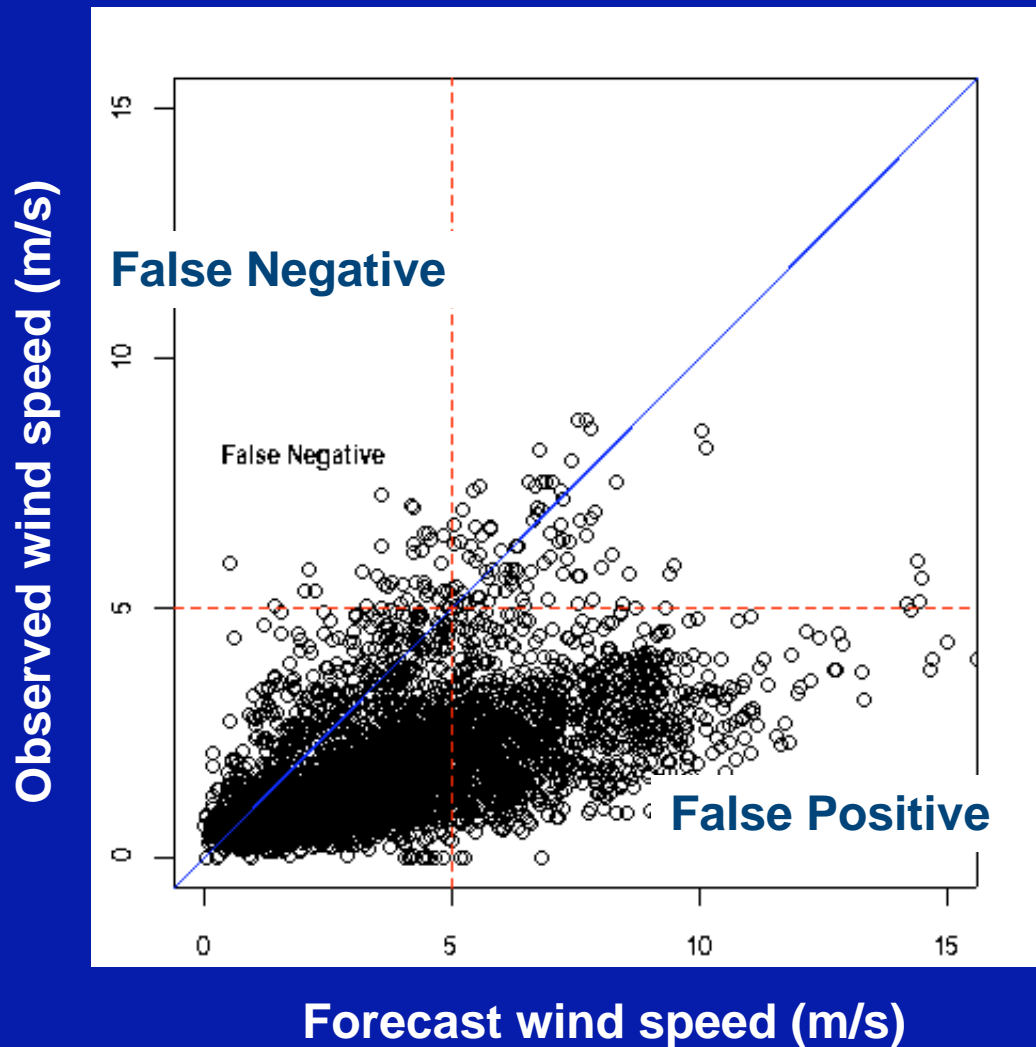
NCAR

We are shooting for this from the model:



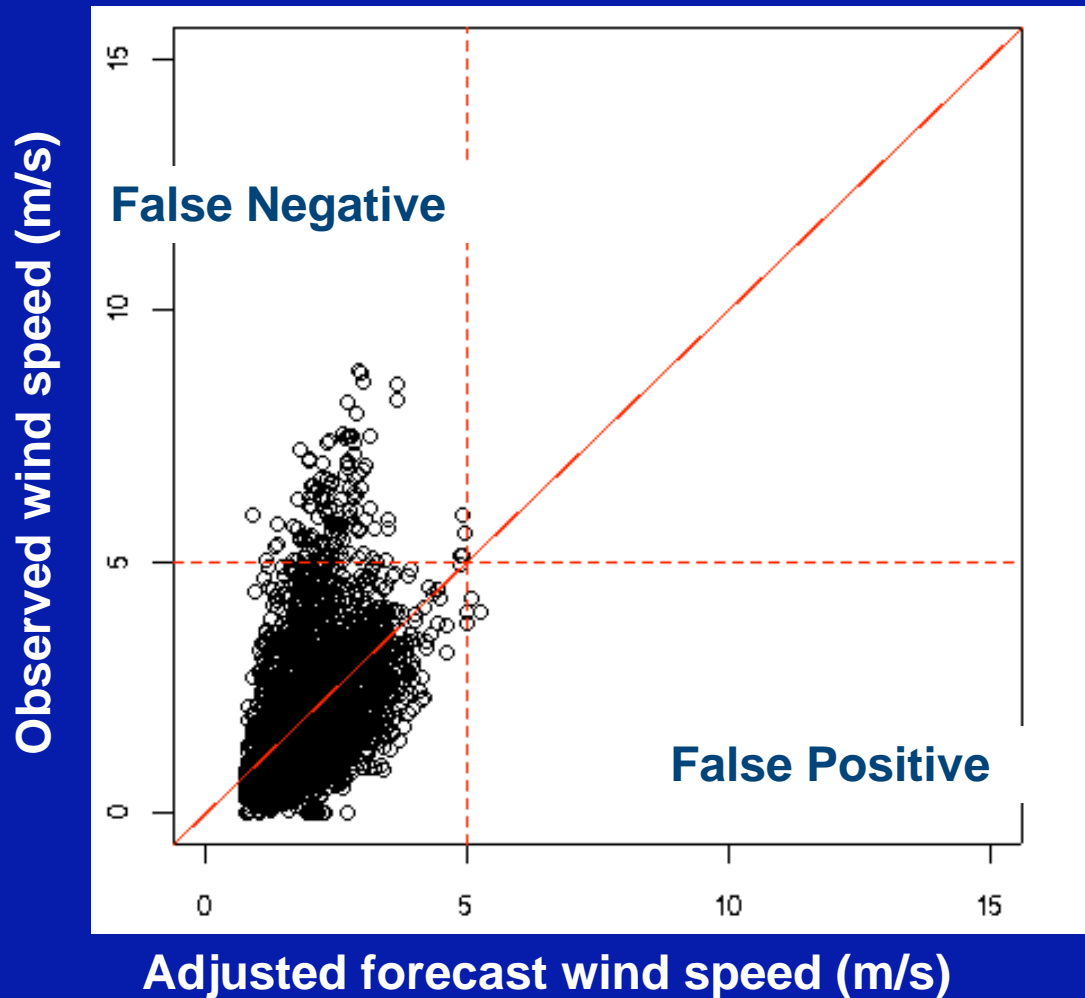
**These distributions are lognormal, and we correct the mean and variance in the same way.**

# Forecast vs. Observed



- Forecast has a large positive bias in wind speed
- False positive forecasts for winds  $> 5$  m/s 28% of the time.

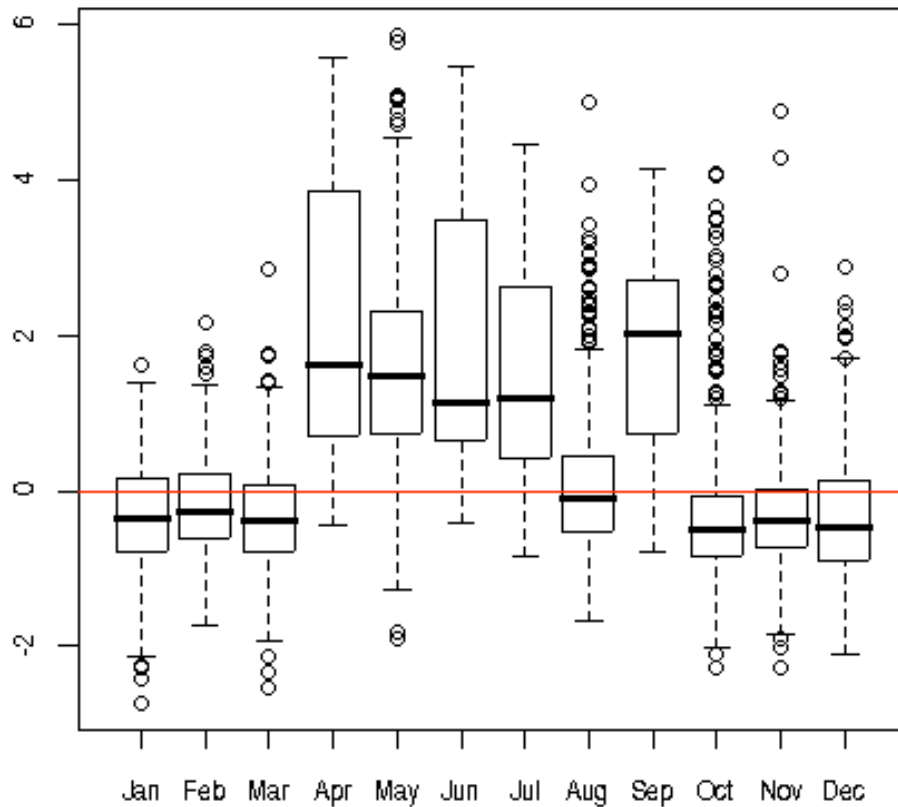
# Simple Solutions Inadequate



- Linear regression to correct
- Removes false positives
- Introduces more false negatives
- Bimodality may be a problem

# Monthly Variability

Linear Model Residuals



Month

- Distributions of regression residuals each month
- Shows that a single calibration for all times is not appropriate

# Summary



NCAR

- We want an estimate pdf useful for decision making (gambling).
- An ensemble forecast can be the basis.
- Calibration is necessary, but can be difficult.