

# Verification of Probability Forecasts

4/21/04

RFC Ensemble Workshop

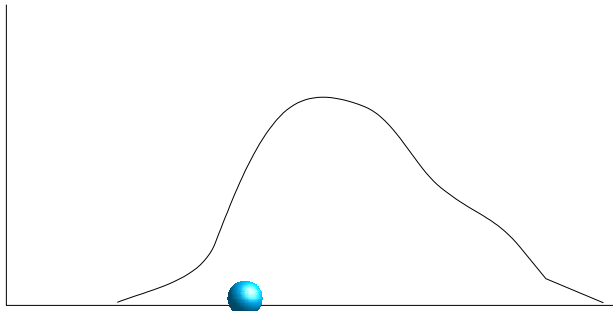
# Reliability and Resolution

## Two Key Characteristics

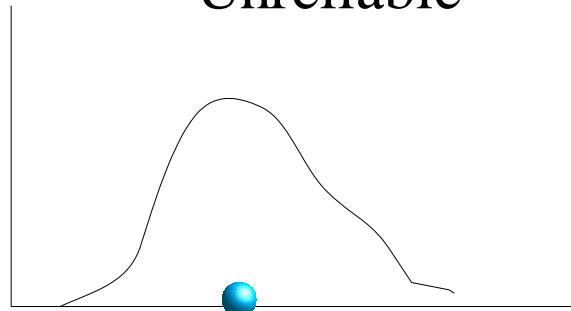
- Reliability
  - Do the forecast probabilities correctly reflect the future uncertainty?
- Resolution
  - Do the forecasts distinguish between upcoming events?

# A Reliability Example

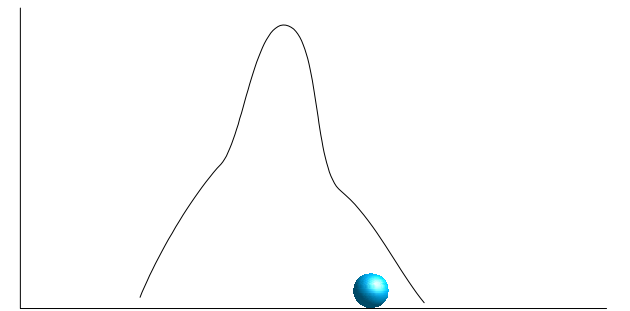
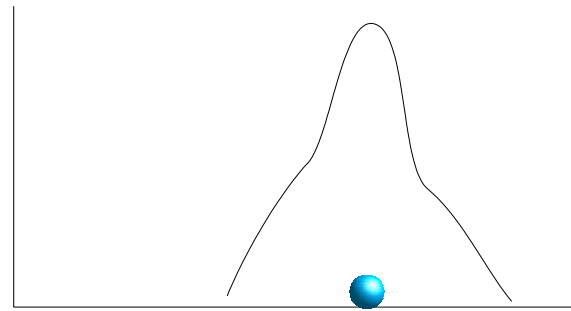
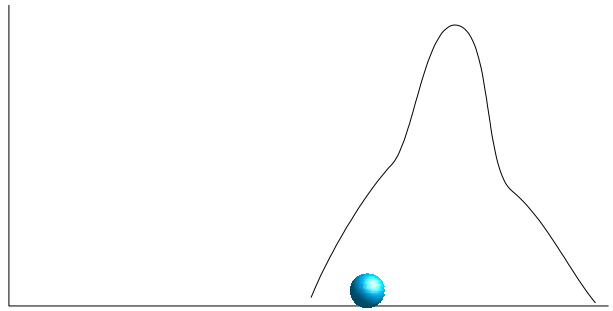
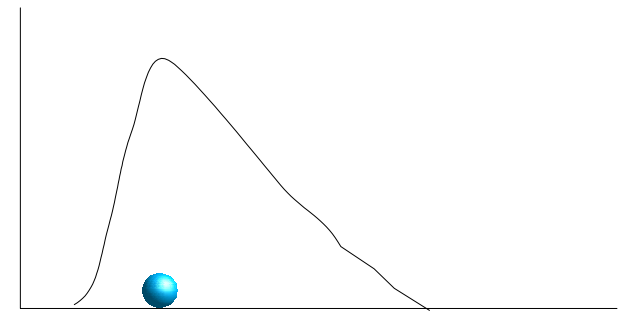
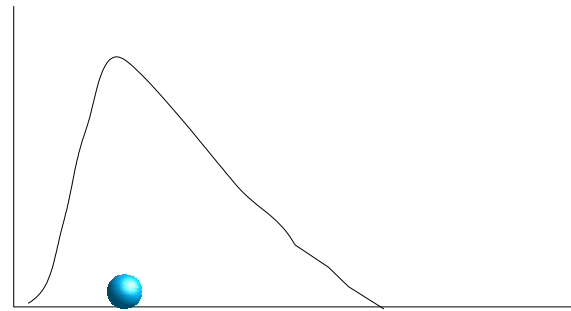
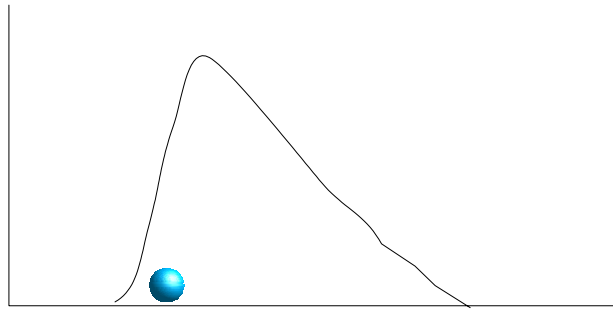
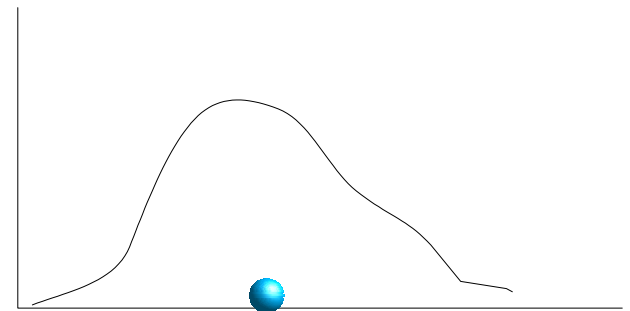
Unreliable



Unreliable



Reliable



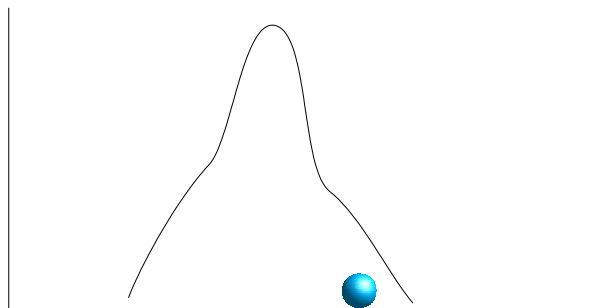
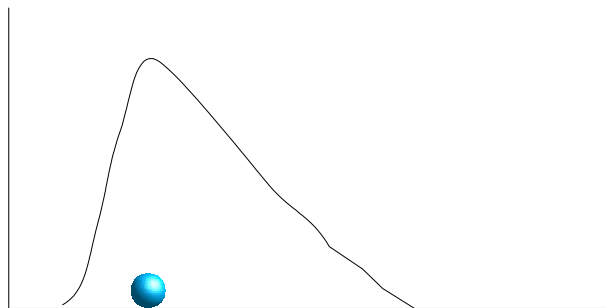
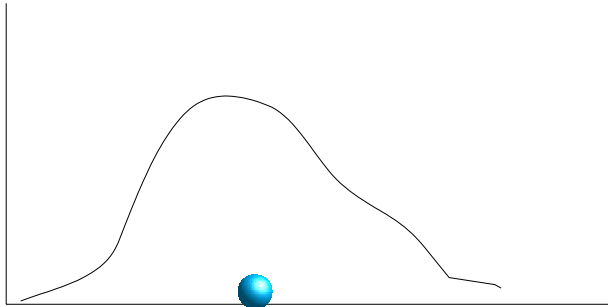
Stage

Stage

Stage

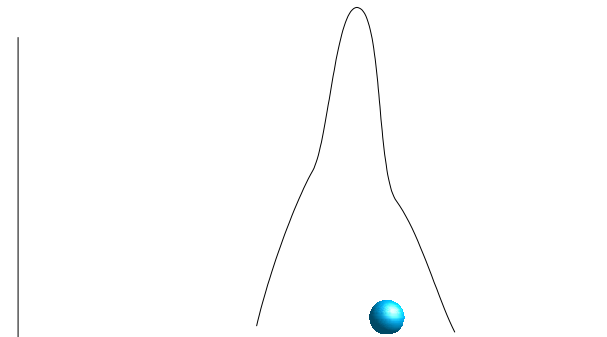
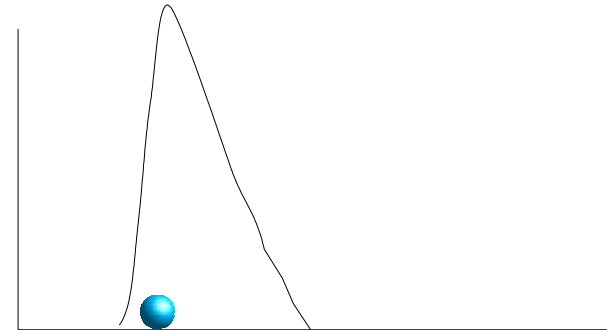
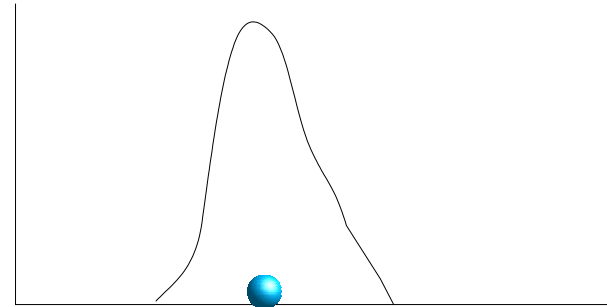
# And Now with Resolution

Reliable



Stage

Reliable with Resolution

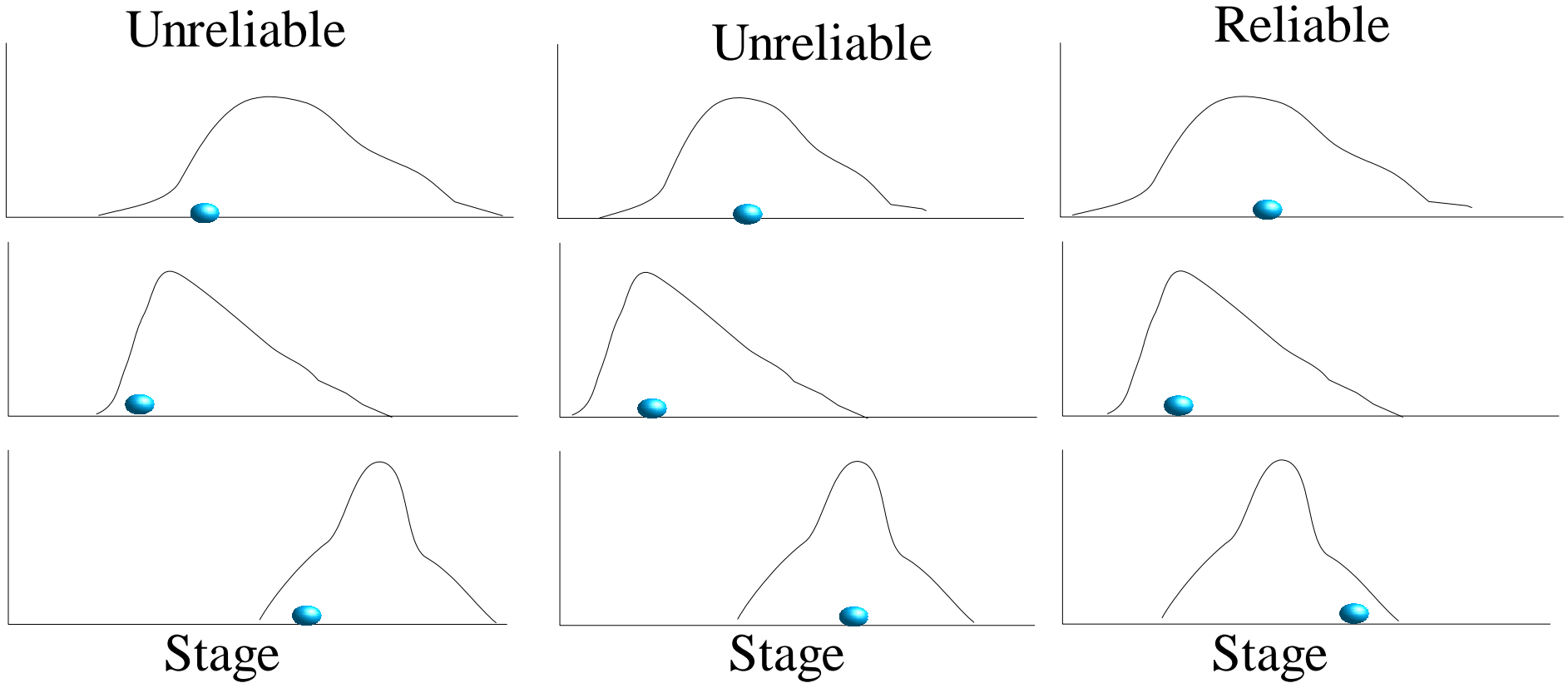
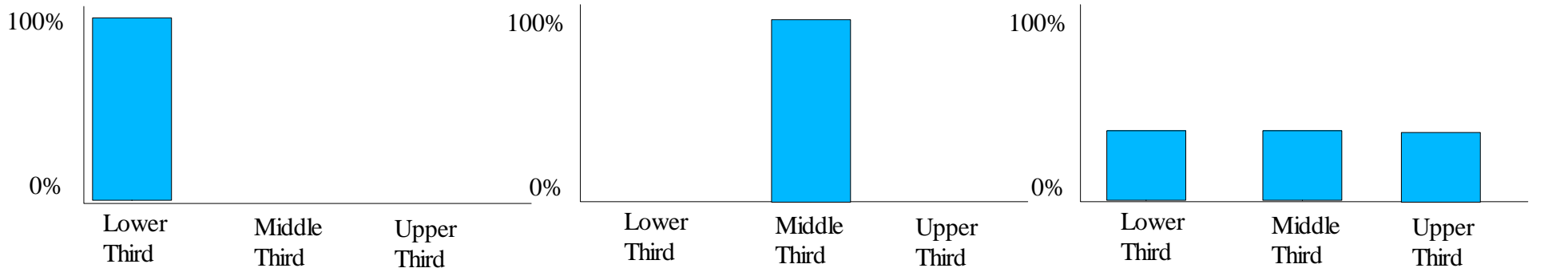


Stage

# Measuring Reliability

- Rank Histograms
  - Evaluate the frequency the observed falls into probability intervals.
  - We usually compute a Rank Histogram using the Cumulative Distribution Function
  - Call it a “Cumulative Rank Histogram”.
- Reliability Diagrams
  - Evaluate the frequency the observed falls into probability intervals given a forecast for a specific event.

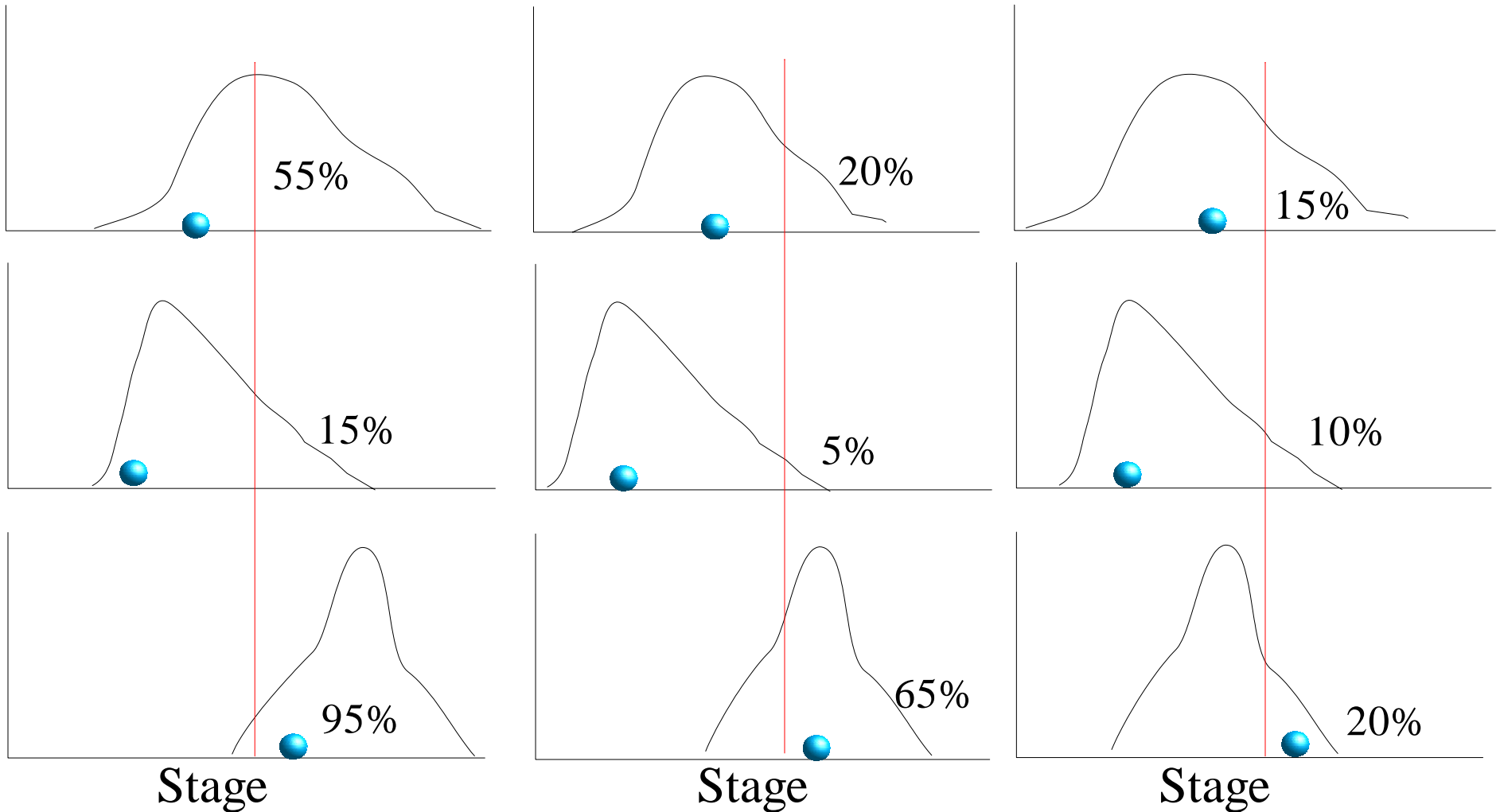
# Rank Histograms



# We Need to Look at Forecasts of Something

## Forecasts of Flooding

— Flood Stage

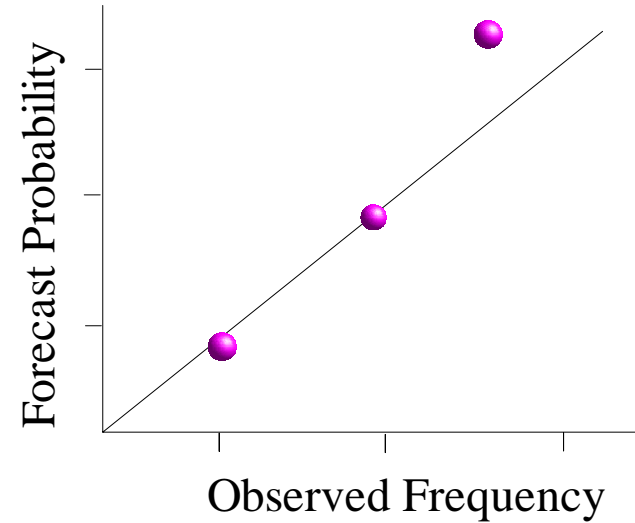


# Reliability Diagrams

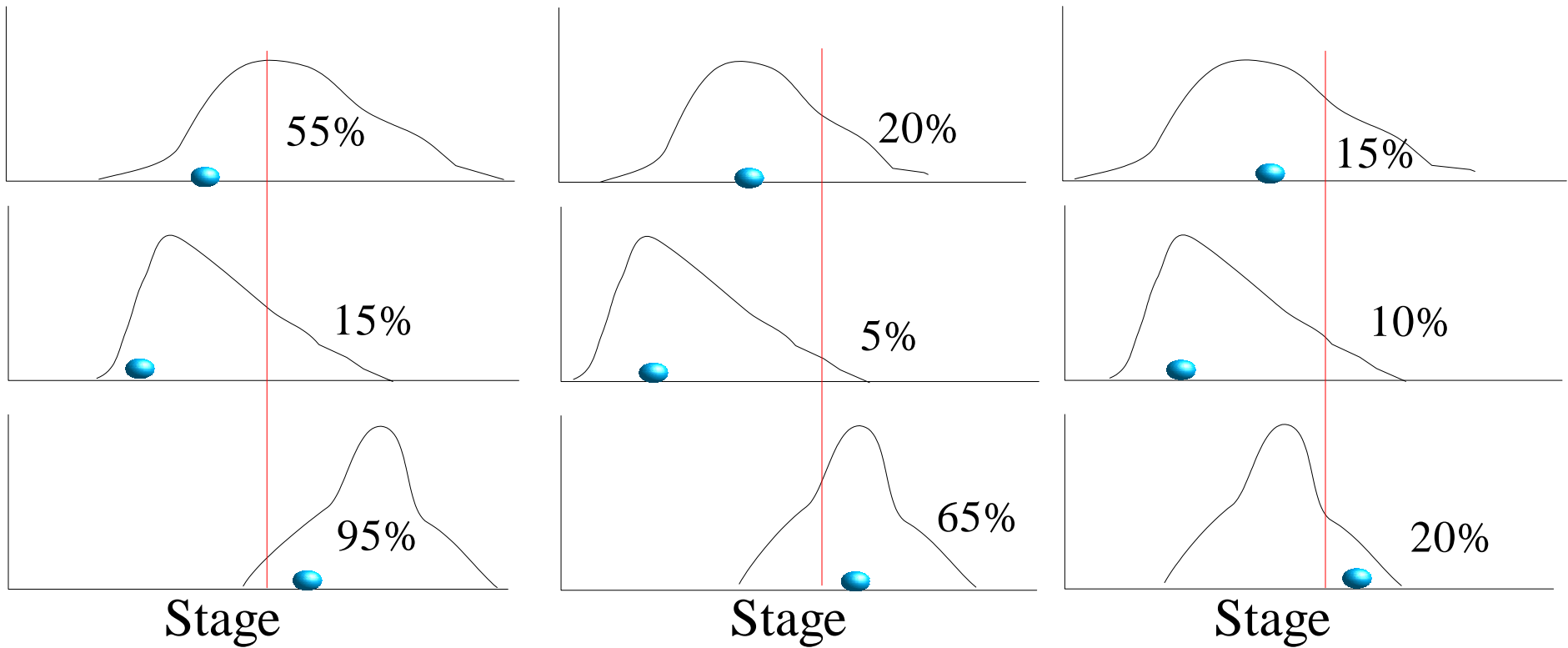
## Forecasts of Flooding

### Forecast chance of a Flood

< 33 % : 6 fcsts, 1 obs  
33% to 66 %: 2 fcsts, 1 obs  
>66%: 1 fcst, 1 obs



— Flood Stage







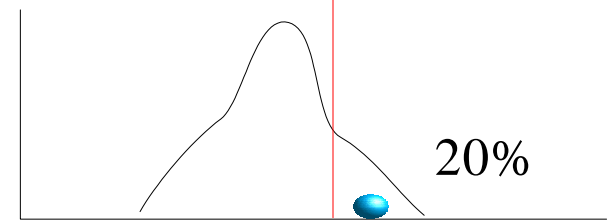
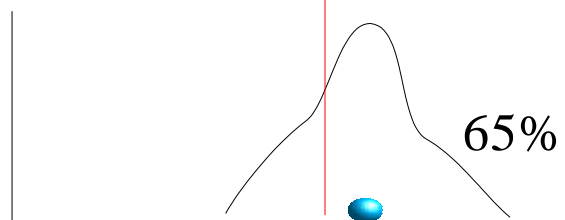
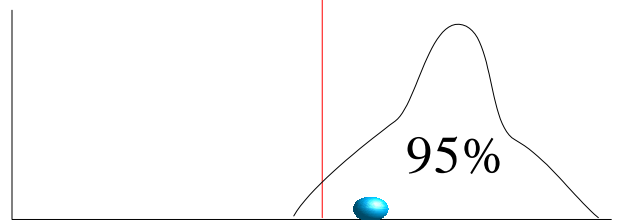
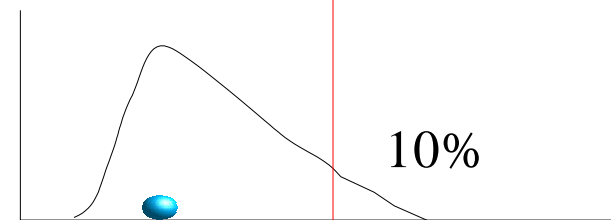
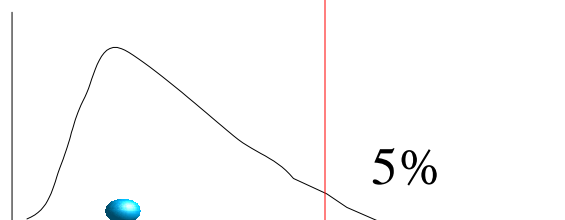
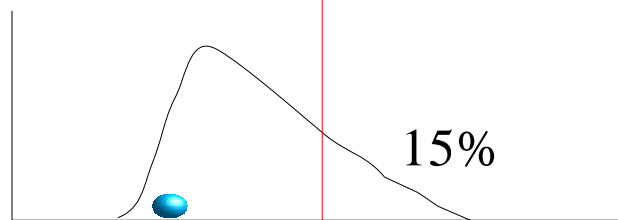
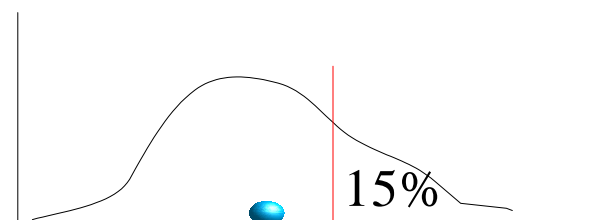
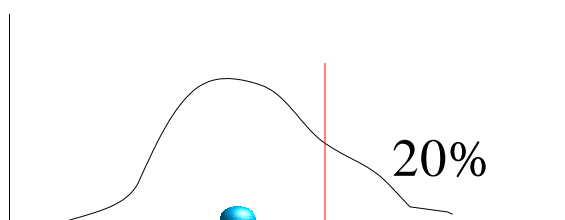
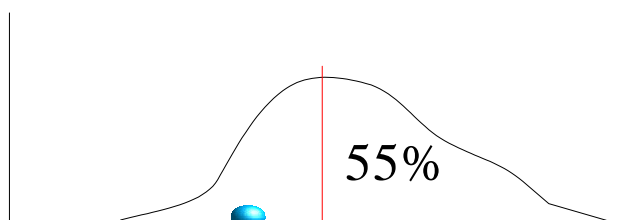
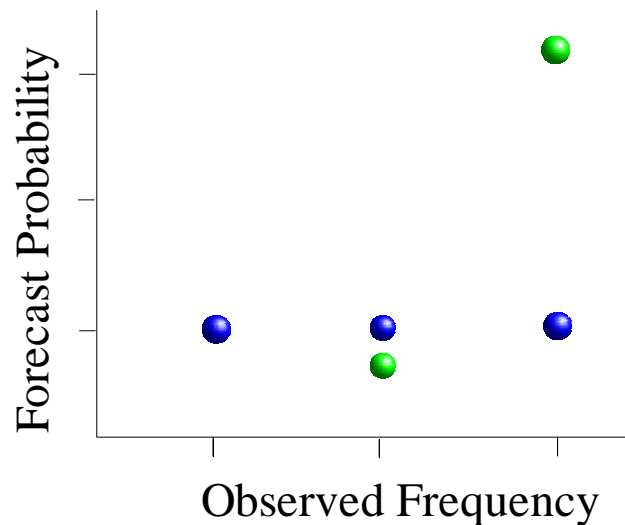
# Measuring Resolution

- Discrimination Diagrams
  - Like Reliability Diagrams, but sorting by the observation
  - Look at what the forecast was, before a flood.
- Relative Operating Characteristics (ROC) Diagrams
  - Connects to theory to assess value and to deterministic scores.
- Ranked Probability Score and Brier Score
  - Consist mostly of Resolution

# Discrimination Diagrams

## Flood and No Flood Events

 3 Floods	Forecast Prob of Flood		
	High	Med	Low
	1	1	1
 6 No Flood	Forecast Prob of No Flood		
	High	Med	Low
	5	1	0



— Flood Stage Stage

Stage

Stage

# Relative Operating Characteristics (ROC) Diagrams for Flood Forecasts

>15% Flood OBS

F      NF

	F	NF
FCST F	3	4
FCST NF	0	2

>50% Flood OBS

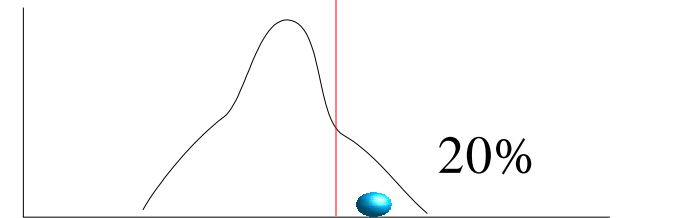
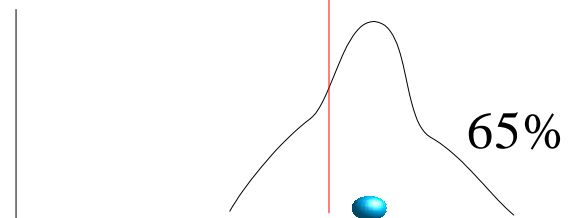
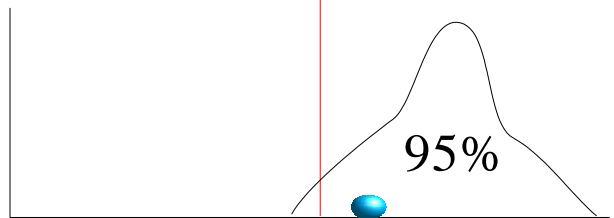
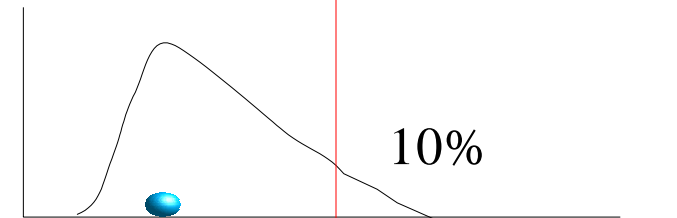
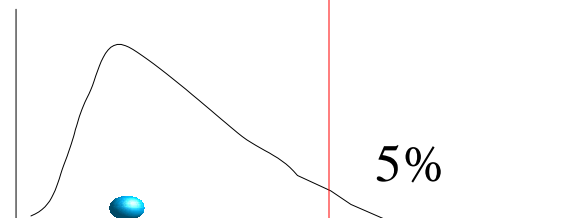
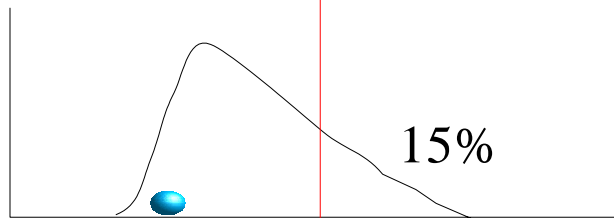
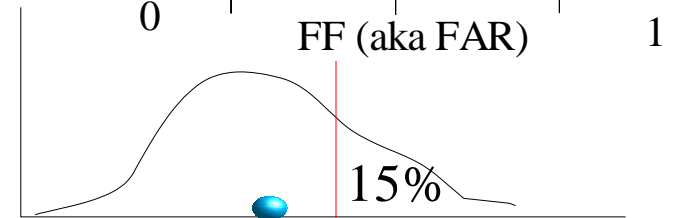
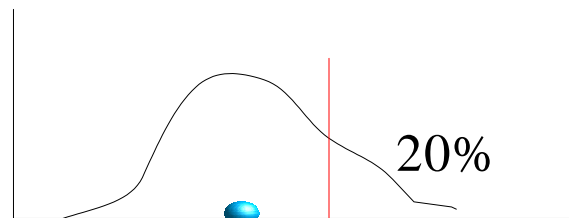
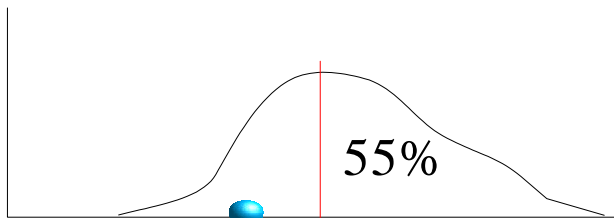
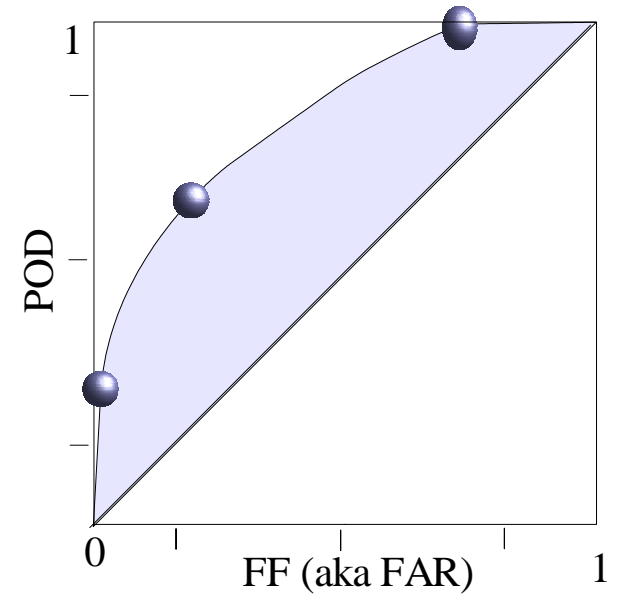
F      NF

	F	NF
FCST F	2	1
FCST NF	1	5

>85% Flood OBS

F      NF

	F	NF
FCST F	1	0
FCST NF	2	6



— Flood Stage      Stage

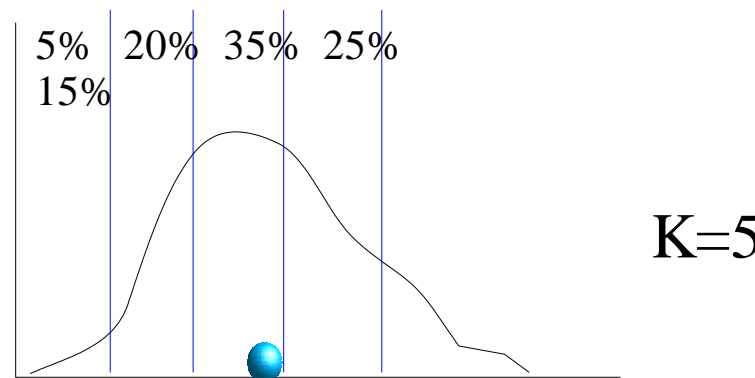
Stage

Stage

# Ranked Probability Score (RPS) for Flood Forecasts

$$\text{RPS} = \sum_k (f_k - o_k)^2$$

Where  $f=0$  if the event did not occur  
and  $f=1$  if the event did occur  
and  $k$  is the category number



$$\begin{aligned} \text{RPS} &= ((0 - 0.05)^2 + (0 - 0.25)^2 + (1 - 0.60)^2 + (1 - 0.85)^2 + (1 - 1.00)^2) \\ &= 0.25 \end{aligned}$$

For multiple forecasts, take the average RPS.

If  $K=2$ ; RPS = The Brier Score