Second Verification Workshop
CBRFC, 11/18/08

Ensemble verification refresher

James Brown

james.d.brown@noaa.gov

Aim: reduce forecast bias

Many types of bias. For example:

- **Over- or under-forecasting (e.g. ensemble mean consistently too low or high).**
- **Too little spread in an ensemble forecast to capture observations (“underspread”).**
- **Bias that increases under specific conditions, (“conditional bias”) e.g. under flood flows.**
- **Bias resulting from poor model assumptions (“unreliable”) or resolution (“indiscriminate”).**

1. Types of ensemble verification metric

Types of ensemble metric

Many types of metrics

- Reflects many different types of bias
- Four-dimensions reviewed here

1. Treat ensemble as deterministic

- Convert ensemble forecast to single-valued forecast by choosing “best guess” (mean).
- Apply single-valued metrics (RMSE etc.)
- Easy to understand, but inadequate.

Types of ensemble metrics

2. Simple vs. detailed ensemble metrics

- From summary “scores” (one number)...
- ...to detailed visualizations of raw data (pairs)
- Somewhat application dependent

3. Absolute quality vs. skill

a) Absolute: metric for one forecast model

b) Relative: *skill* of one model over another

Skill needs a metric and reference

Types of ensemble metric

4. Types or attributes of quality

A) When Y was forecast, what was observed?

"Our forecast predicts a 90% chance of flooding."

RELIABLE if observed 9/10 times issued.

B) When X was observed, what was forecast?

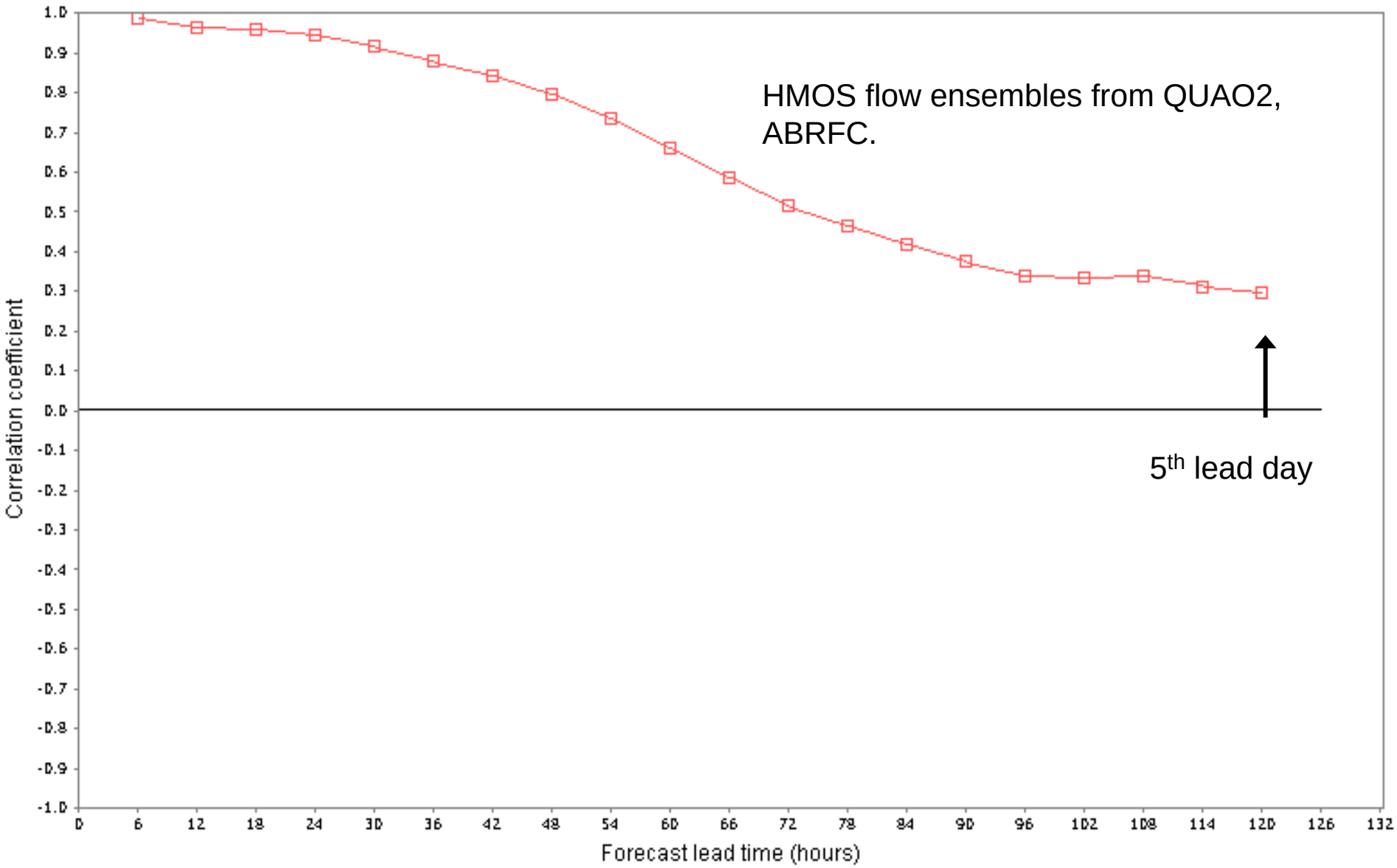
"When we observe Action Stage only, our model predicts a 100% chance of Flood Stage."

Cannot DISCRIMINATE between AS and FS.

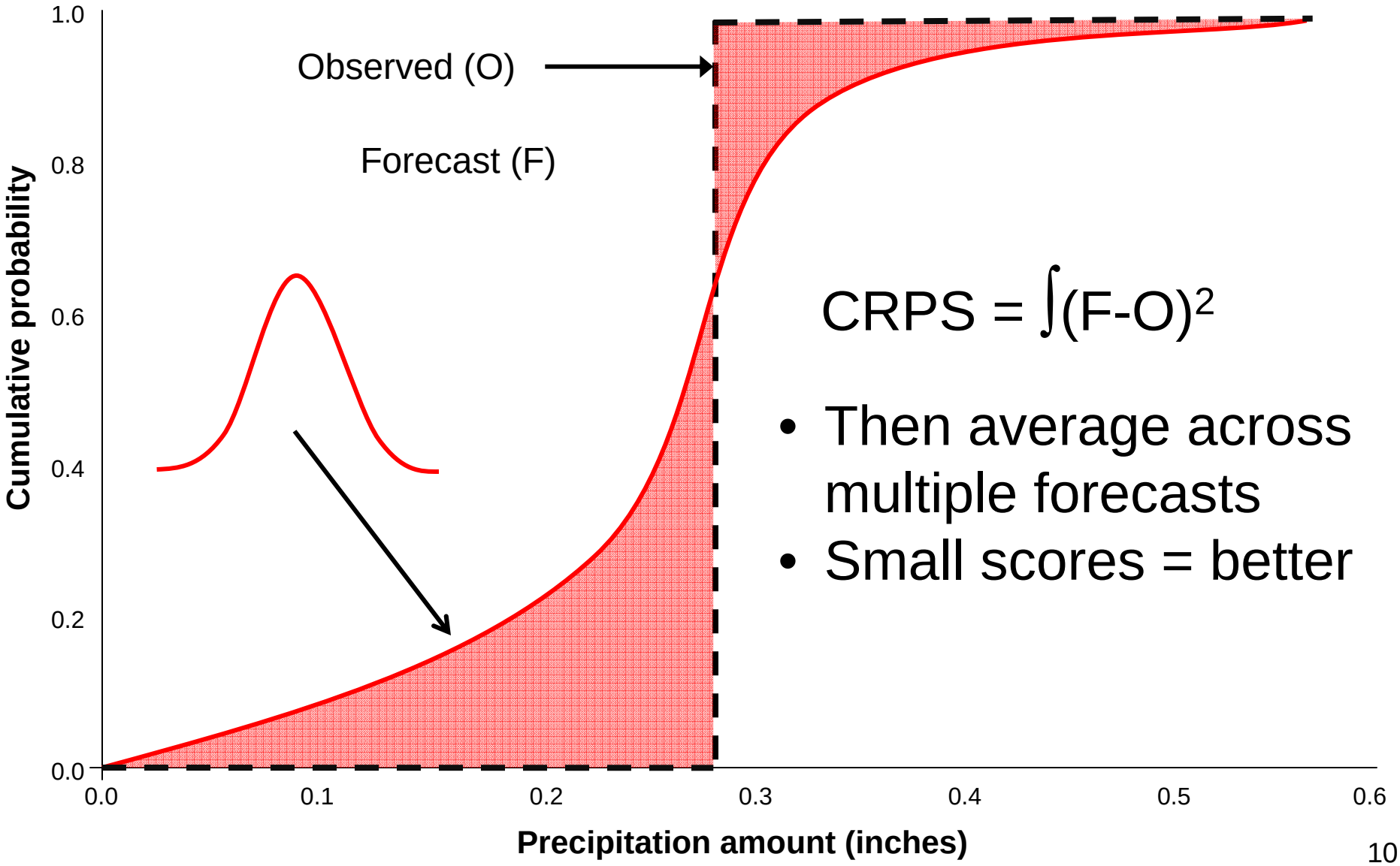
2. Examples of ensemble metrics (available in EVS)

Summarized vs. detailed

Correlation of ensemble mean



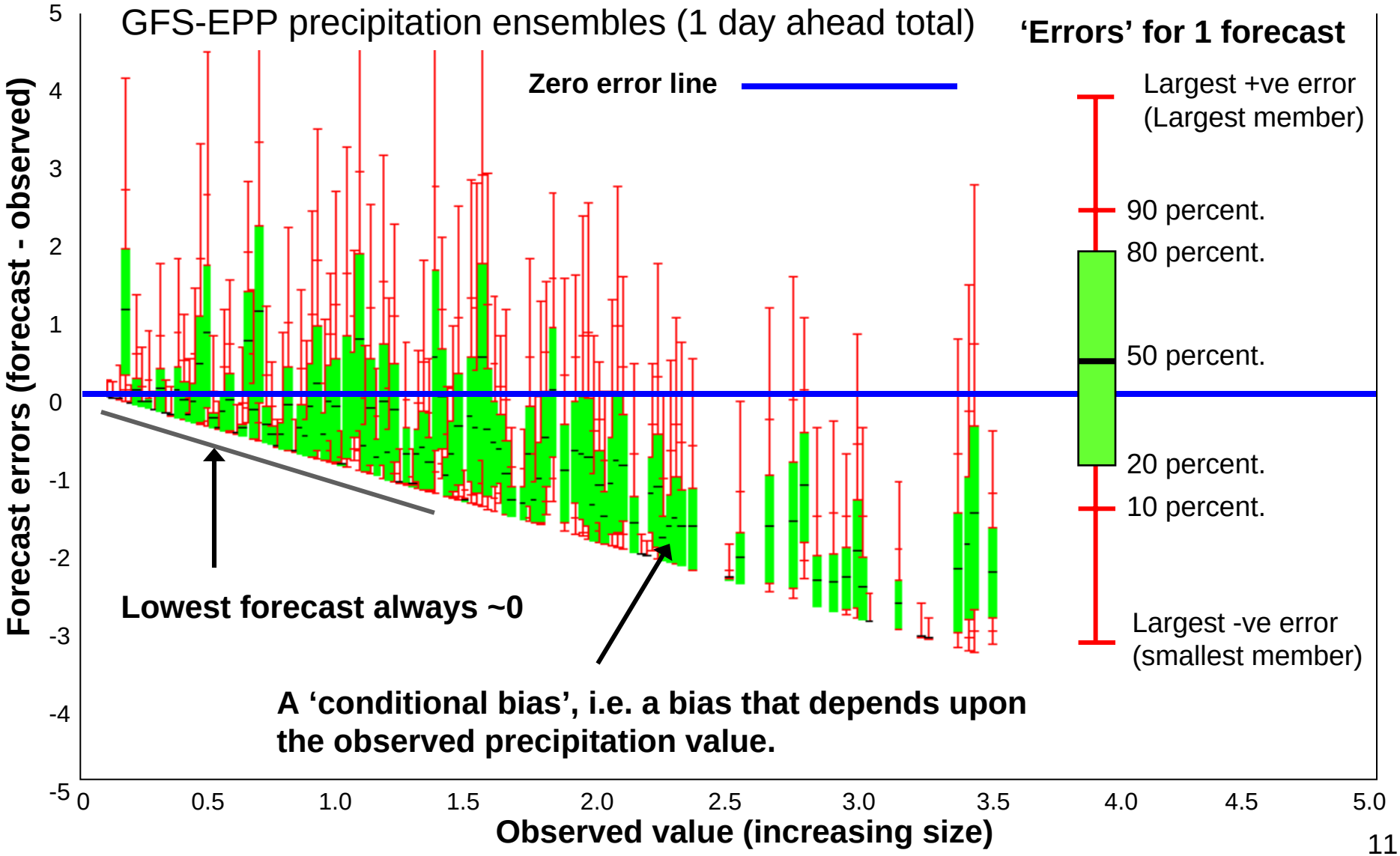
CRPS (simple, ensemble)



$$CRPS = \int (F-O)^2$$

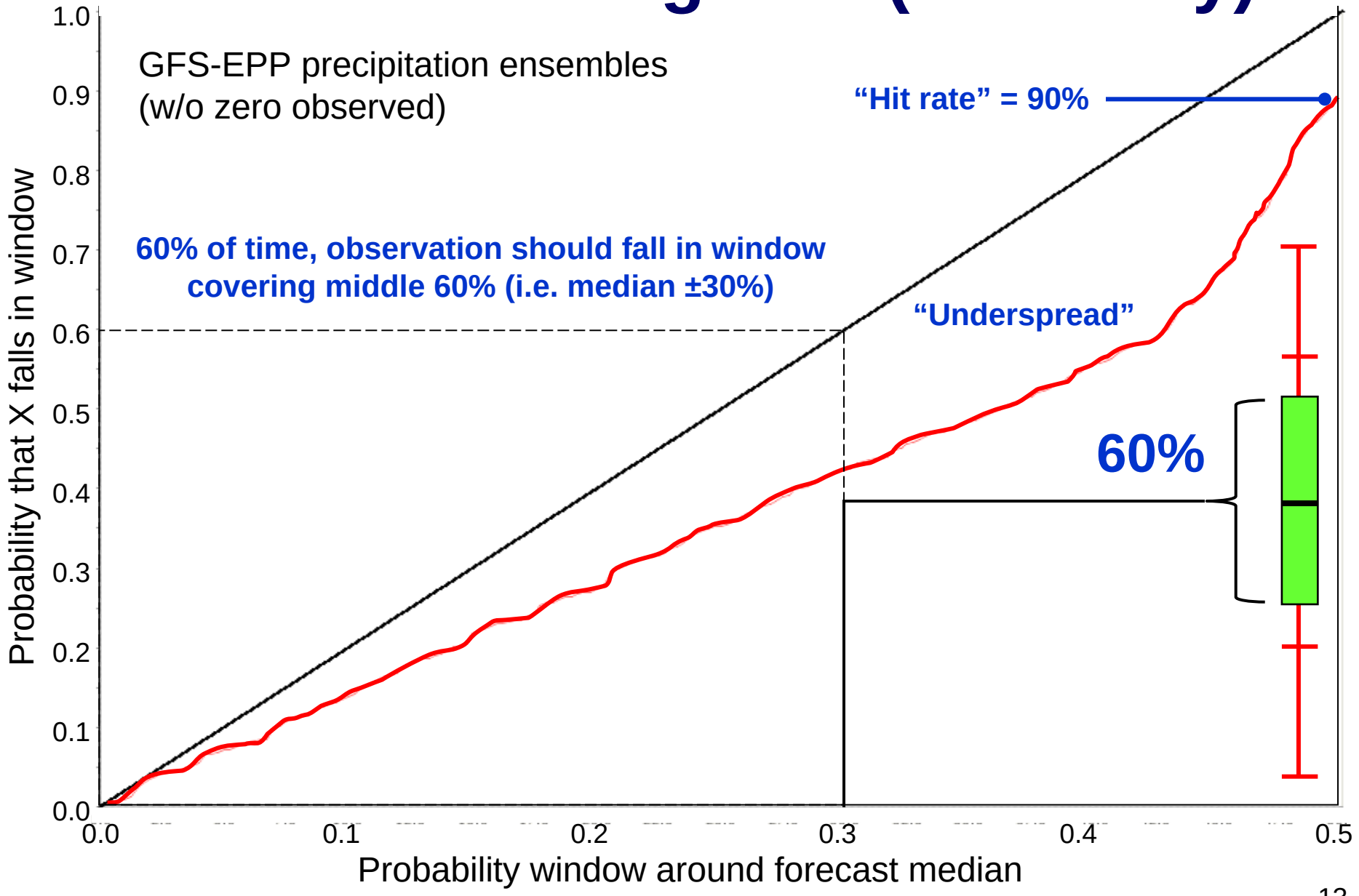
- Then average across multiple forecasts
- Small scores = better

Very detailed (box plot)



Reliability vs. discrimination

Cumulative Talagrand (reliability)

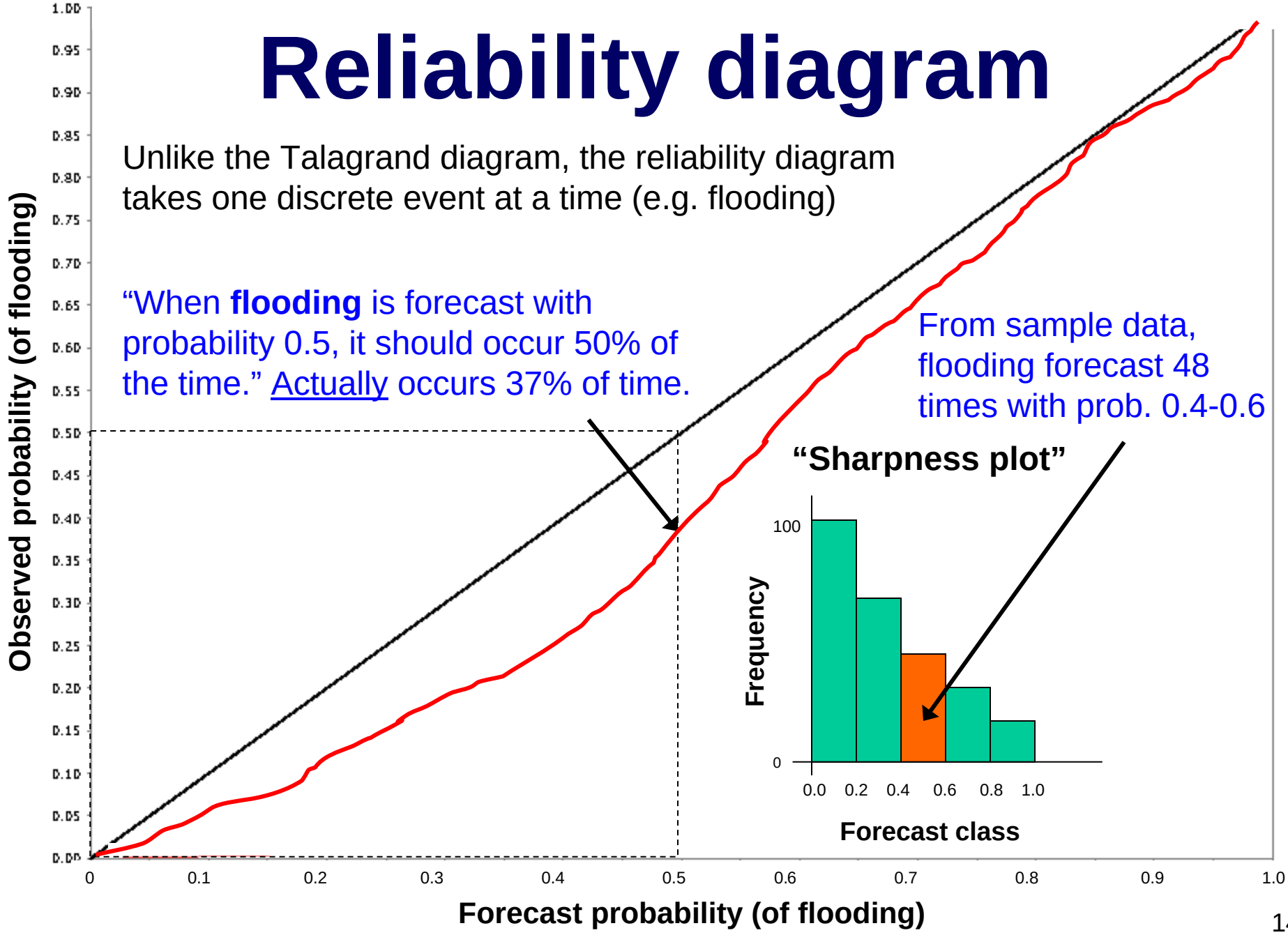


Reliability diagram

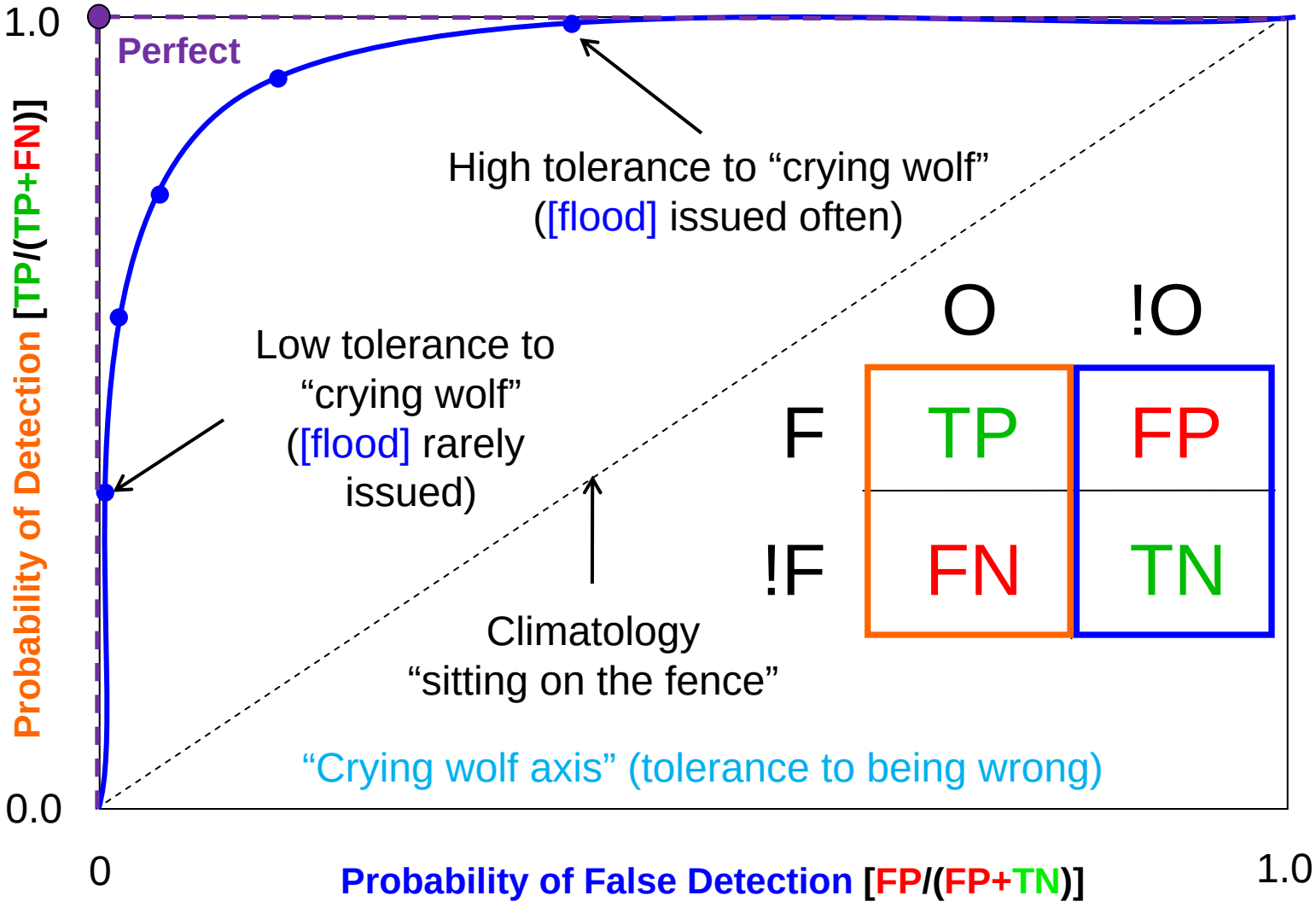
Unlike the Talagrand diagram, the reliability diagram takes one discrete event at a time (e.g. flooding)

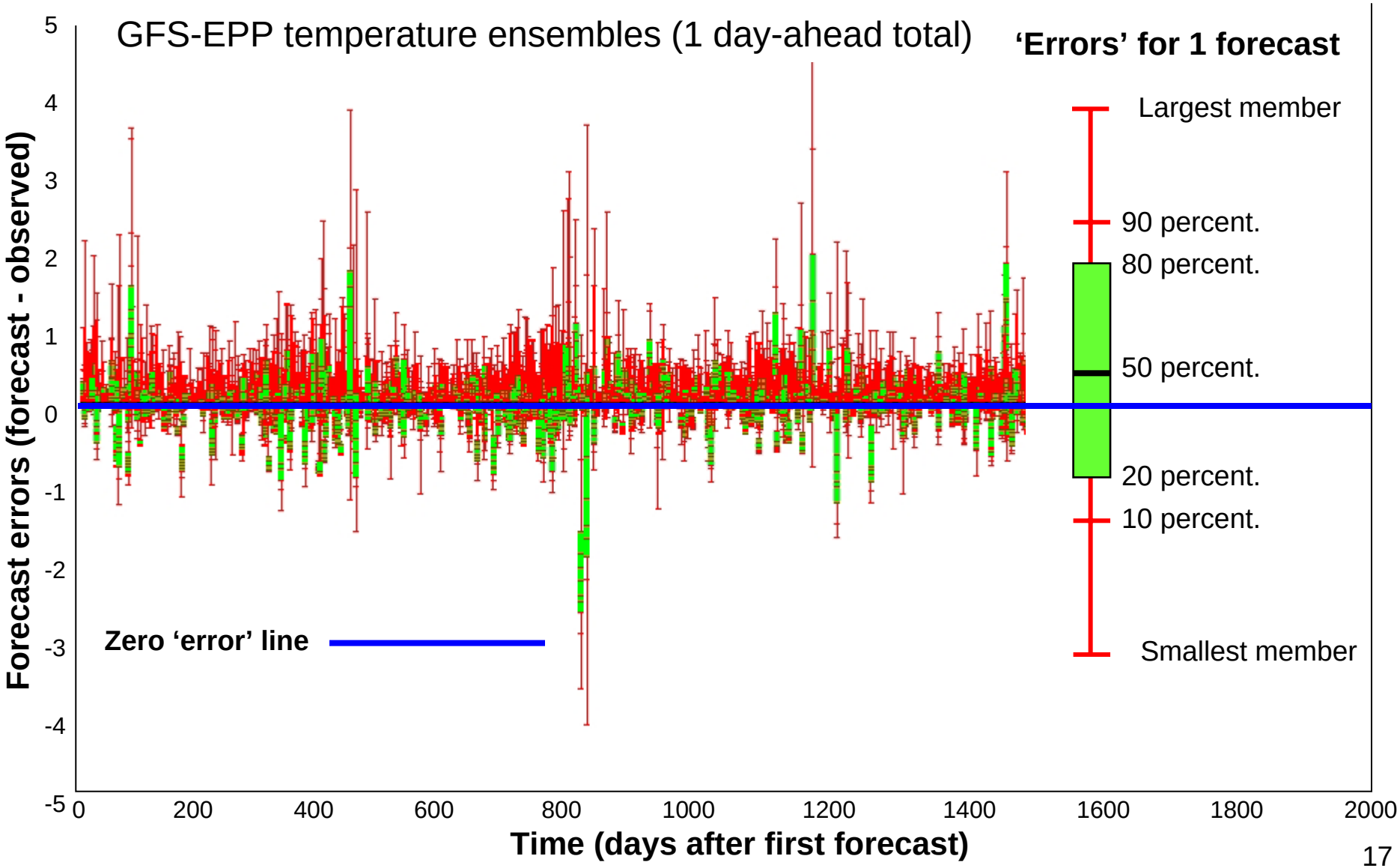
"When **flooding** is forecast with probability 0.5, it should occur 50% of the time." Actually occurs 37% of time.

From sample data, flooding forecast 48 times with prob. 0.4-0.6



ROC (event discrimination)



Questions ???

Very detailed (box plot)