NWS Verification Team meeting 05/05/08

The Ensemble Verification System (EVS): an introduction

James Brown

James.D.Brown@noaa.gov

Goals for today

1. **Introduction to EVS software**

   - **Mechanics of EVS (structure, I/O etc.)**

   - **Brief lecture followed by demo.**

2. **Overview of metrics in EVS**

   - **Which metrics are available in EVS?**

   - **What can they tell us (focus on exercises)?**

3. **Brief introduction to exercises**

1a. Overview of EVS

Scope of EVS

Diagnostic verification

- **Problem-focused: what/where errors & why?**
- **Distinguished from real-time verification**

Diagnostic questions include….

- **Are ensembles reliable?**
- **Prob[flood]=0.9: does it occur 9/10 times?**
- **Operational forc. vs. hindcasts (e.g. MODS)**
- **What are the major sources of uncertainty?**

Design goals of EVS

Verification of continuous time-series

- Temperature, precipitation, streamflow etc.
- \> 1 forecast point, but <u>not</u> spatial products

Forecast products at different scales

- Any lead time (e.g. 1 day – 2 years or longer)
- Any forecast resolution (e.g. hourly, daily)
- Temporal aggregation (e.g. hourly to daily)
- Aggregation across forecast points
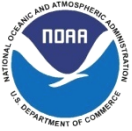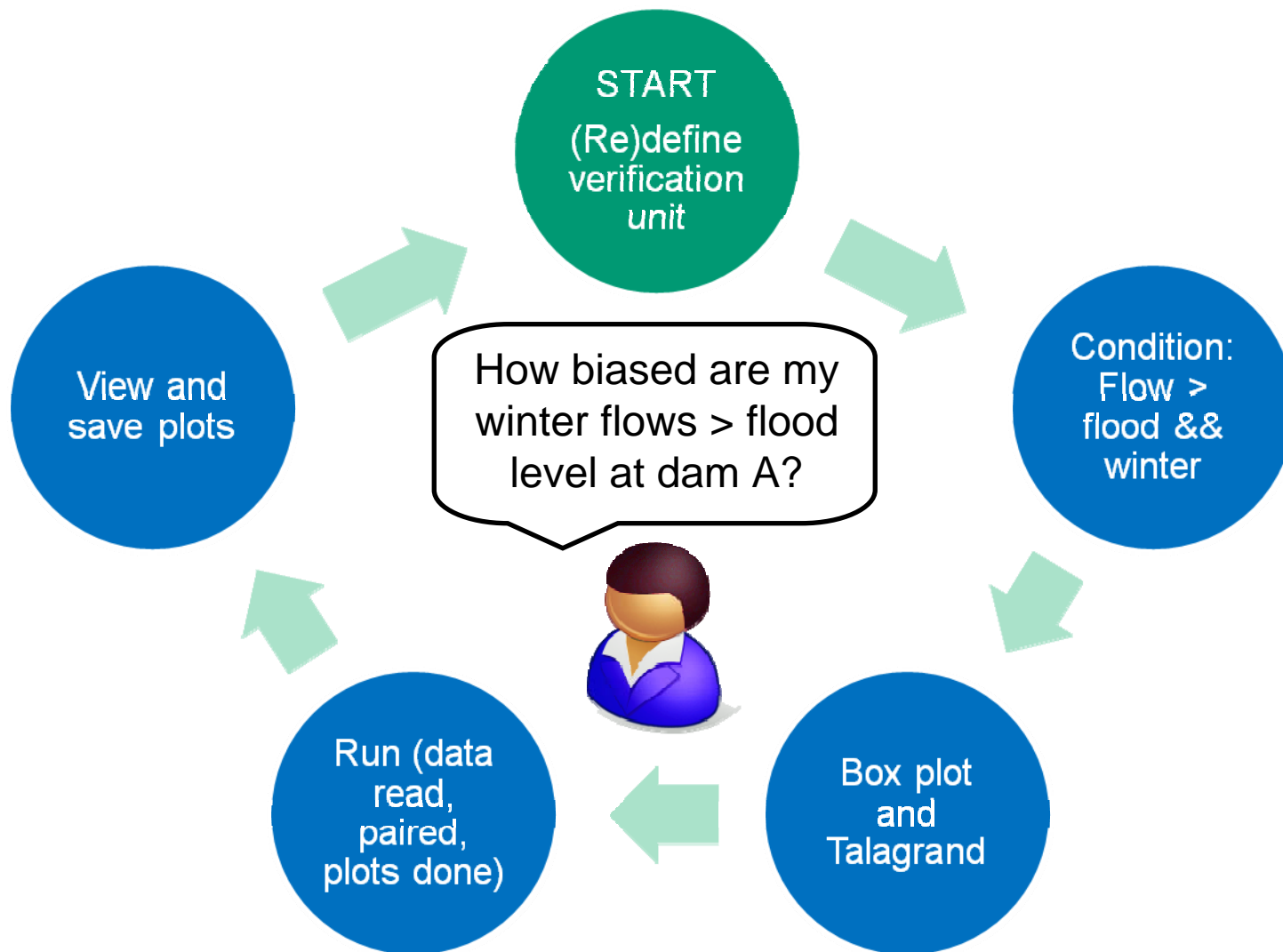
Design goals of EVS

Flexibility to target data of interest

- Two target variables: 1) forecast; 2) observed

- Two conditions: 1) time; 2) variable value

- e.g. observed winter flows > flood stage
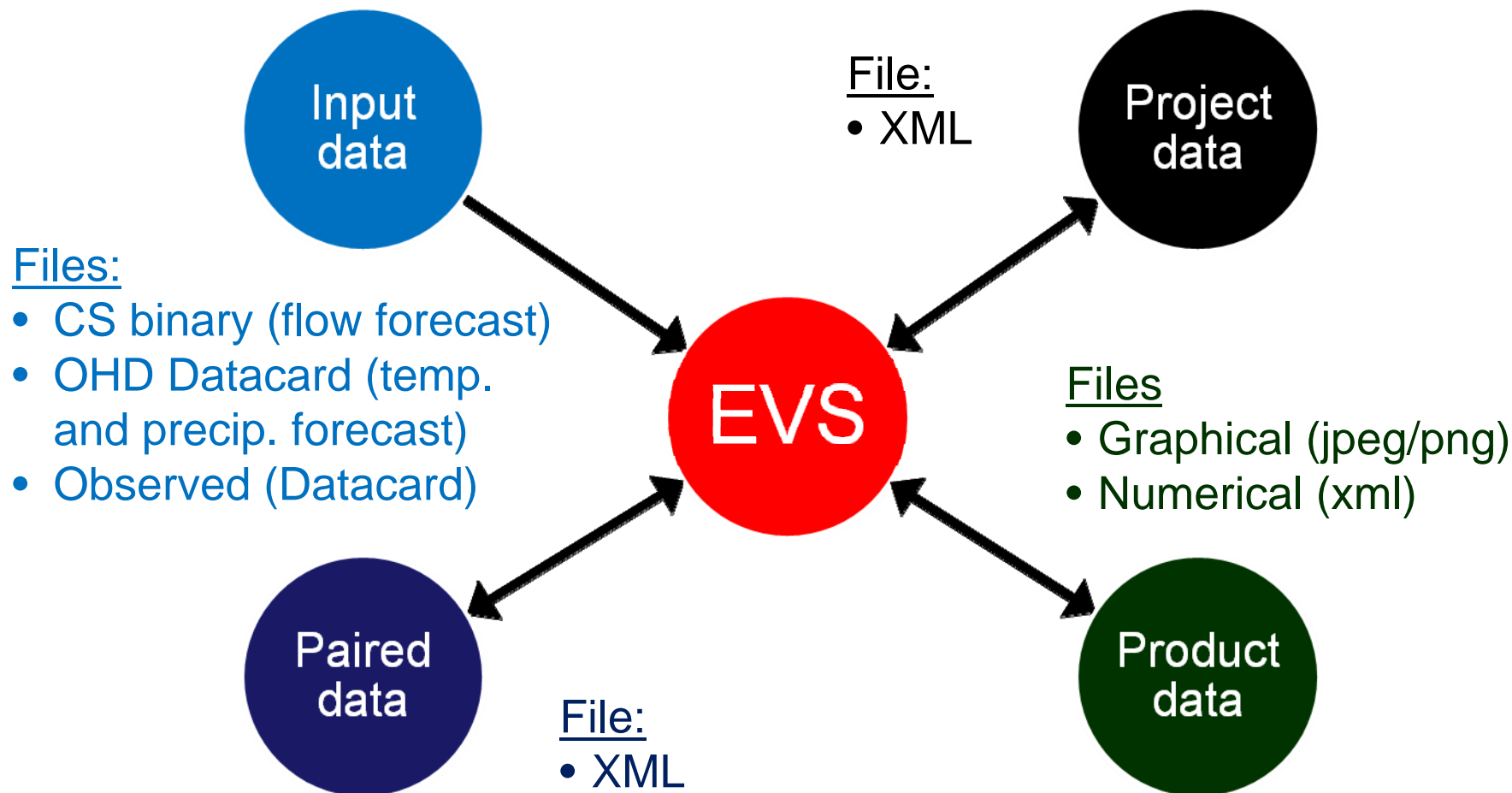
- e.g. ensemble mean temperature < freezing

Carefully selected metrics

- From very detailed to highly summarized

- Documented and explained

Example of workflow

Data I/O and archiving

Input data

File:
• XML

Project data

Files:
• CS binary (flow forecast)
• OHD Datacard (temp. and precip. forecast)
• Observed (Datacard)

EVS

Files
• Graphical (jpeg/png)
• Numerical (xml)

Paired data

File:
• XML

Product data

1b. Demonstration of EVS

2. Verification metrics

Metrics for probabilities

Many ways to classify metrics

1. Tests for single-valued property (e.g. mean)

2. Tests of broader forecast distribution

- Both may involve reference forecasts ("skill")

Caveats in testing probabilities

- Observed probabilities require many events

- <u>Big assumption 1:</u> we can 'pool' events

- <u>Big assumption 2:</u> observations are 'good'

Continuous prob. forecasts

Discrete/categorical forecasts

- **Many metrics rely on discrete forecasts**
- **e.g. will it rain? {yes/no} (rain > 0.01)**
- **e.g. will it flood? {yes/no} (stage > flood level)**

What about continuous forecasts?

- **An infinite number of events**
- **Arbitrary event thresholds (i.e. 'bins')?**
- **Typically, yes (and choice will affect results)**

Metrics vary by design

Observation-centered metrics (discrim.)

- "What do forecasts do when observed do X"?
- i.e. "binning" in terms of observed
- e.g. Relative Operating Characteristic

Forecast-centered metrics (reliability)

- "What do observed do when forecasts do Y"?
- i.e. "binning" in terms of forecasts
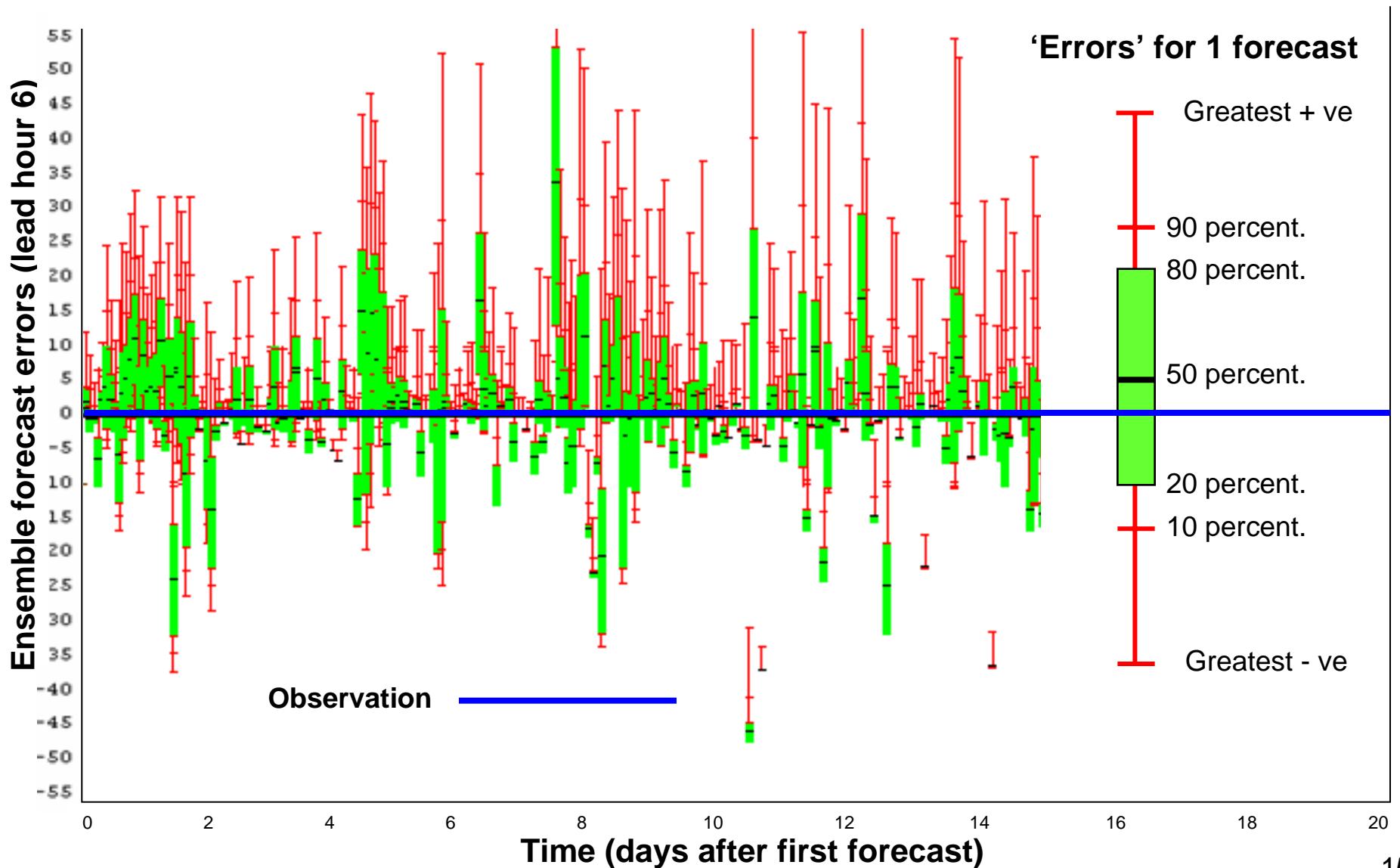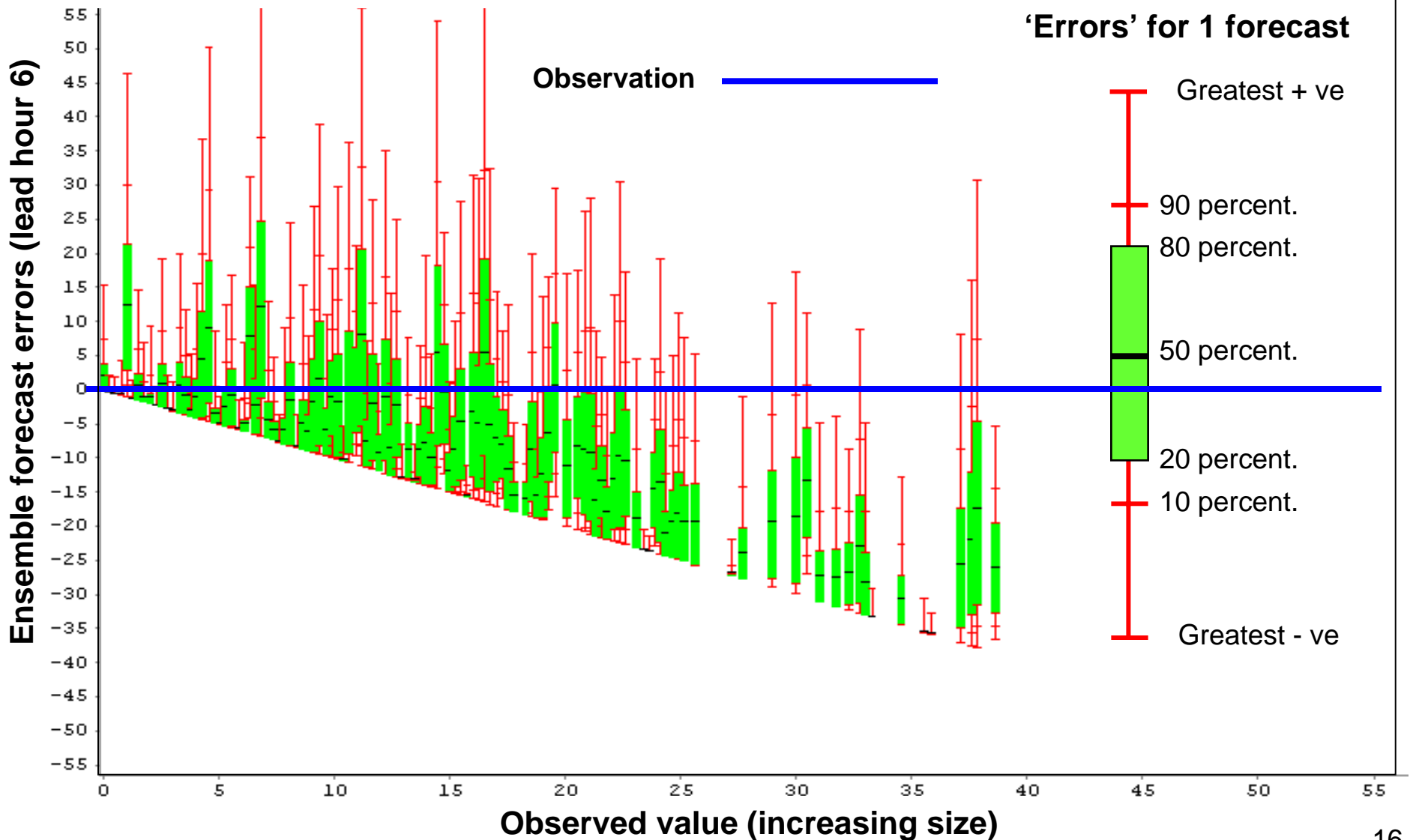- e.g. Reliability Diagram

Metrics vary in detail

Detail varies with verification question

- e.g. inspection of 'blown' forecasts (detailed)

- e.g. avg. reliability of flood forecast (< detail)

- e.g. rapid screening of forecasts (<< detail)

Most detailed (box plot)

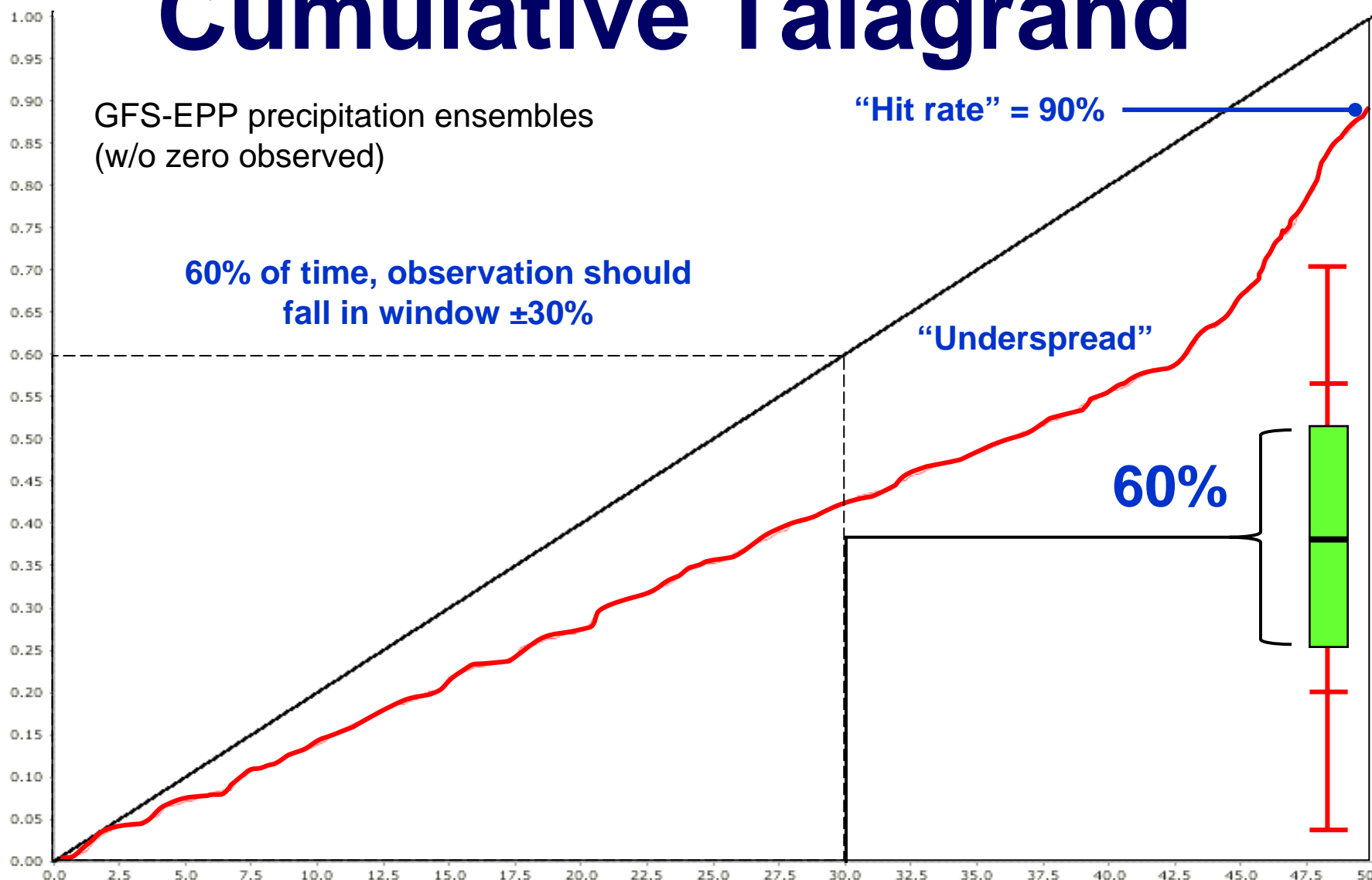# Most detailed (box plot)

Cumulative Talagrand

GFS-EPP precipitation ensembles
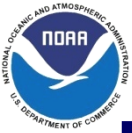(w/o zero observed)

"Hit rate" = 90%

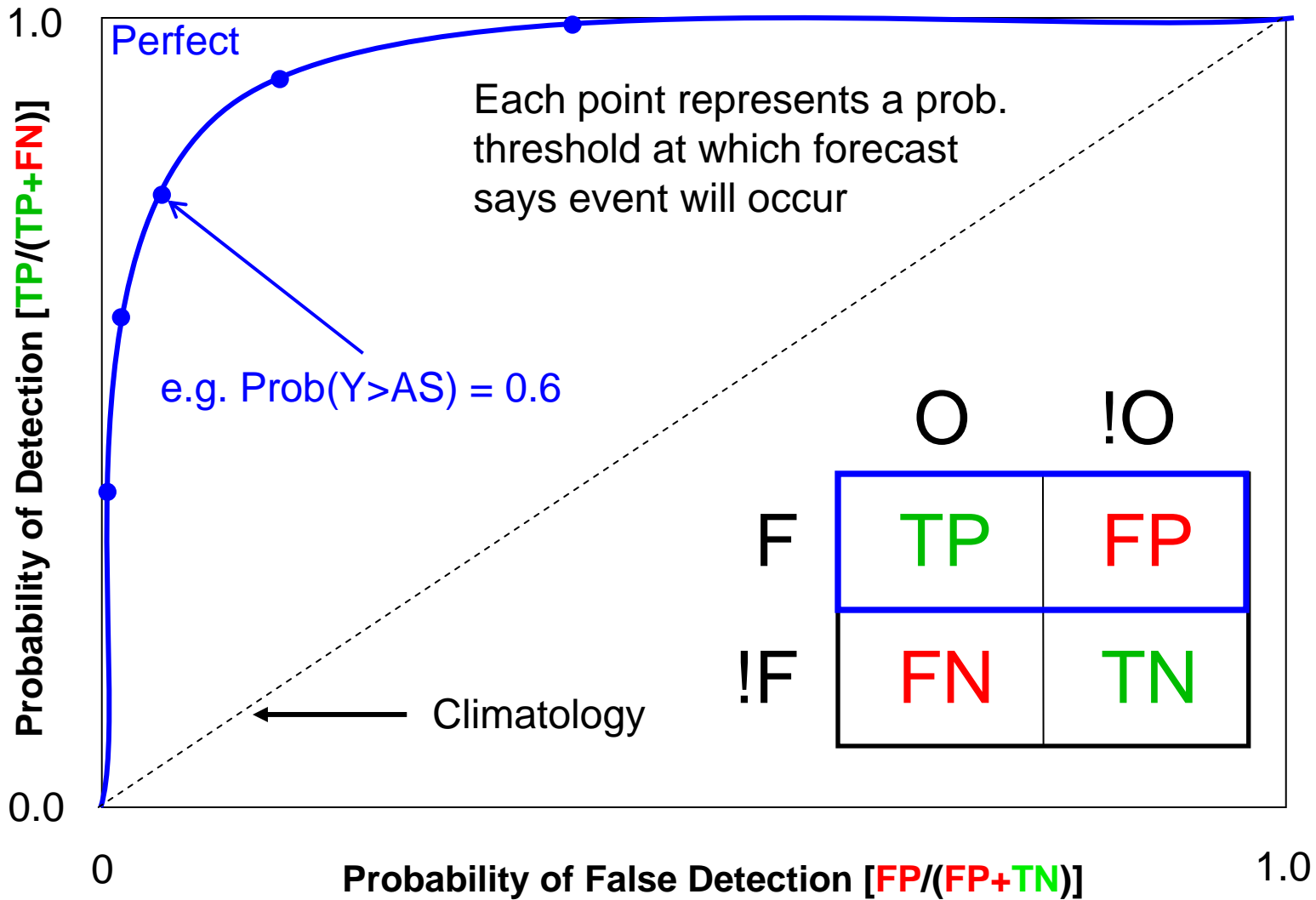**60% of time, observation should
fall in window ±30%**

"Underspread"

60%

Error window (percentile around median)

ROC at Flood Action Stage

Least detailed (a score)

Brier score = $\frac{1}{5} \times \{ (0.8\text{-}1.0)^2 + (0.1\text{-}1.0)^2 + (0.0\text{-}0.0)^2 + (0.95\text{-}1.0)^2 + (1.0\text{-}1.0)^2 \} = 0.8528$

Flood stage   Forecast   ◯ Observation

Least detailed (a score)

Observed (O)

Forecast (F)

Cumulative probability

Precipitation amount (inches)

$CRPS = (F-O)^2$

- Then average across multiple forecasts
- Small scores = better
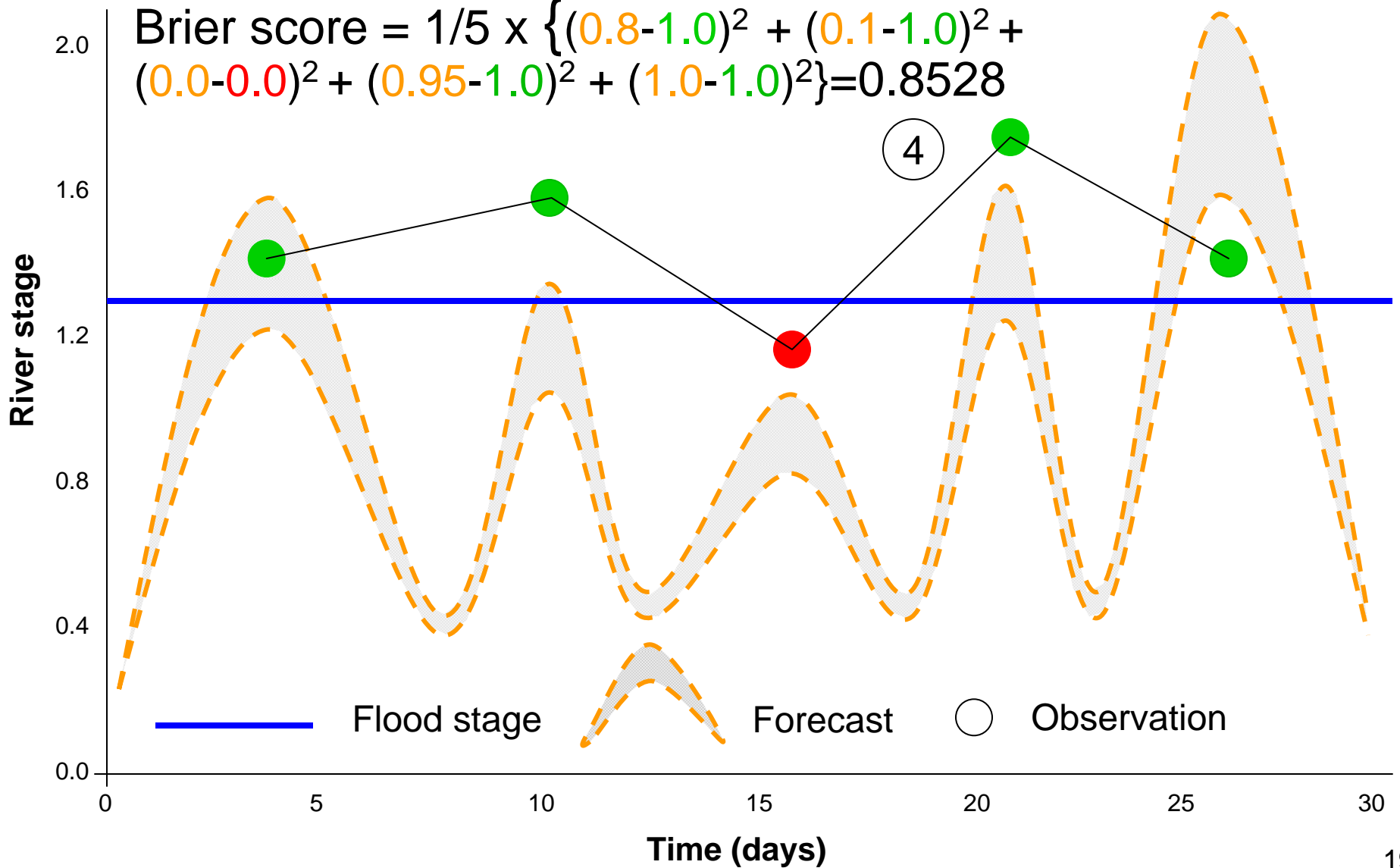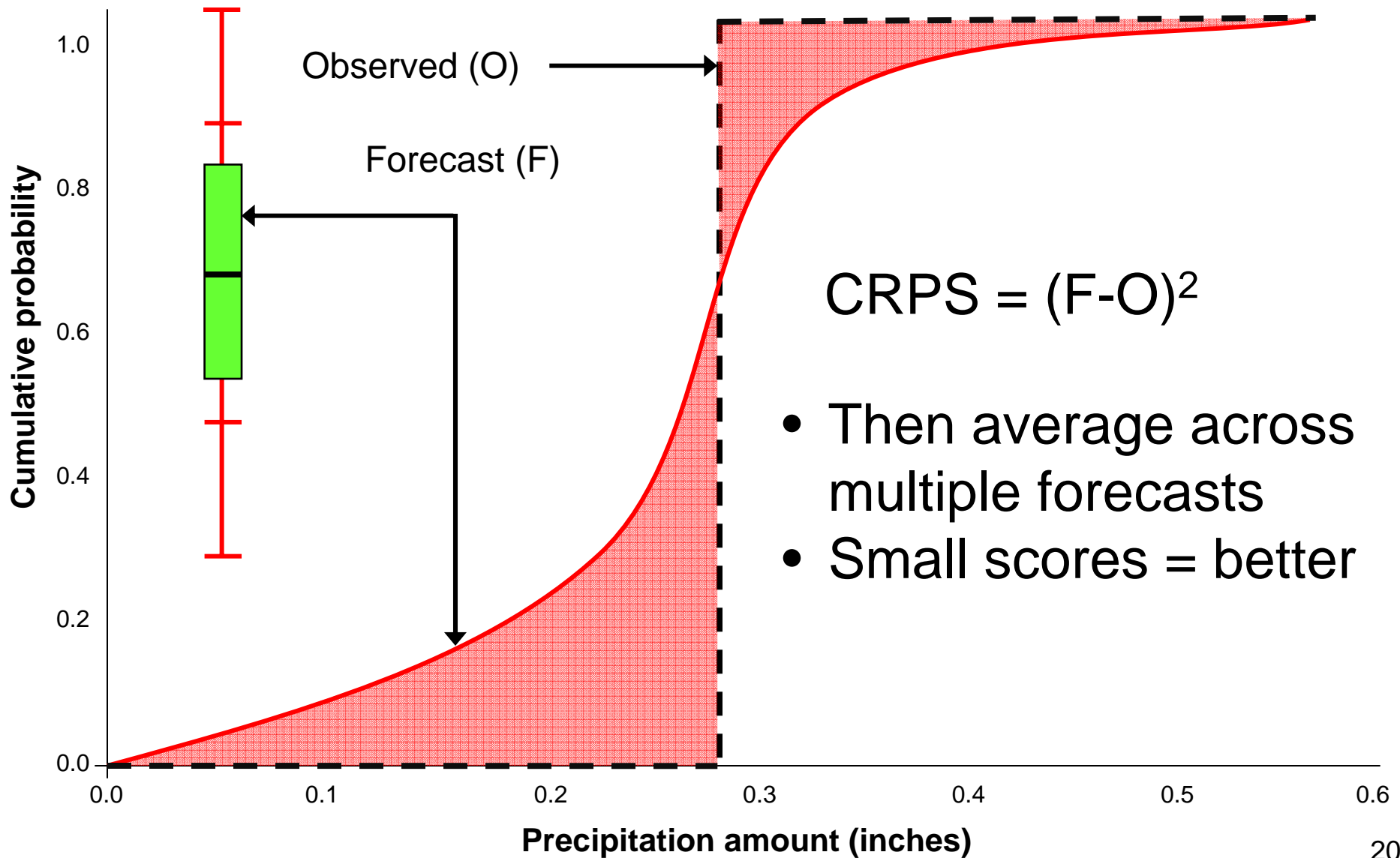
3. Exercises

Installation

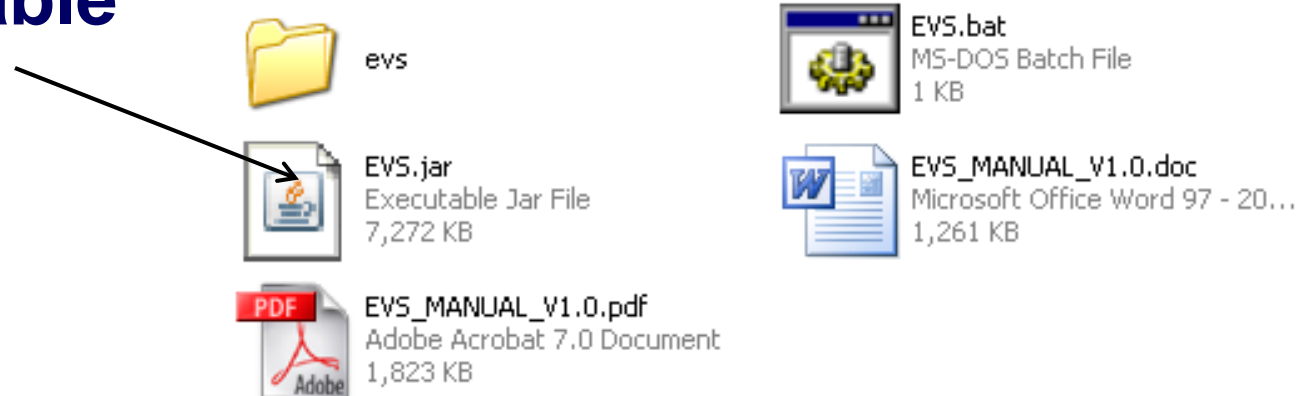## See EVS User's Manual (pp. 6-8)

- **Will run under any OS (tested for Lx/Win.)**

- **Software provided in folder**

- **Recommend JRE version 1.6.0 (1.5.0_12 min.)**

Executable

evs

EVS.jar
Executable Jar File
7,272 KB

EVS_MANUAL_V1.0.pdf
Adobe Acrobat 7.0 Document
1,823 KB

EVS.bat
MS-DOS Batch File
1 KB

EVS_MANUAL_V1.0.doc
Microsoft Office Word 97 - 20…
1,261 KB

Data/instructions

All data/instructions by COB 9th May

- Word document containing exercises

- Folder containing data for each exercise

- Folder containing software

Exercises

Three exercises (increasingly complex)

- **First two exercises deal with synthetic data…**

- **….linear regression model for temperature**

- **Exercise 1: forecasts unbiased**

- **Exercise 2: forecasts biased in mean/spread**

- **Exercise 3: deals with real flow (MARFC)**

- **'Real' biases are less easy to detect!**

- **Need to <u>create</u> plots and <u>analyze</u> them**

Next meeting (06/12)

Go through EVS results

- **What did you learn?**

- **What did you find difficult?**

- **What were the main problems with EVS?**

- **What were the main conceptual problems?**

Use list server for data/software issues!!

- **We will respond to technical/software issues**

- **Conceptual issues addressed in next meeting**

Next meeting (06/12)

Discuss the COMET training module

- **Available in early June**

- **…..E-mail from Matt Kelsch**

- **Feedback from the team**

- **What aspects were easy/difficult?**

Verif-hydro list server for questions

Email: verif-hydro@infolist.nws.noaa.gov

Website: http://infolist.nws.noaa.gov/read/login