National DOH Workshop
07/16/08

Ensemble Verification: Status and Plans

Julie Demargne, James Brown, Yuqiong Liu, and Dong-Jun Seo

Contents

- Verification system overview
- Software
  - Ensemble Verification System (EVS)
  - Hydrologic Ensemble Hindcaster
- Science
  - Sampling uncertainty
  - Real-time verification
- Collaborations

1. Verification System Overview

- **Verification System Components:**
  - Logistical Verification to evaluate quality of forecast services
  - Forecast Verification to evaluate quality of forecasts
    - Diagnostic verification and real-time/prognostic verification

- **Forecasts to be verified:**
  - Deterministic and probabilistic (ensemble, water supply)
  - Various space and time domains:
    - point/area vs. grid
    - lead time from 1 hour to several years

1. Verification System Overview

- **Target System Capabilities:** ***Available Tools***

Target System Capabilities	Available Tools
1. Data archiving	➨ *IHFS db, Archive db, Files, WR website*
2. Computing metrics 3. Displaying data & metrics	} *IVP ob8.3, EVS, WR website*
4. Disseminating data & metrics 5. Real-time access to metrics	} *Stats on demand, WR website*
6. Uncertainty analysis	➨ *Studies w/ Hindcaster*

7. **Performance measure tracking**

IVP: Interactive Verification Program (deterministic verification)
EVS: Ensemble Verification System (ensemble verification)
Hindcaster: capability to retroactively generate forecasts using a fixed system

2. Software development

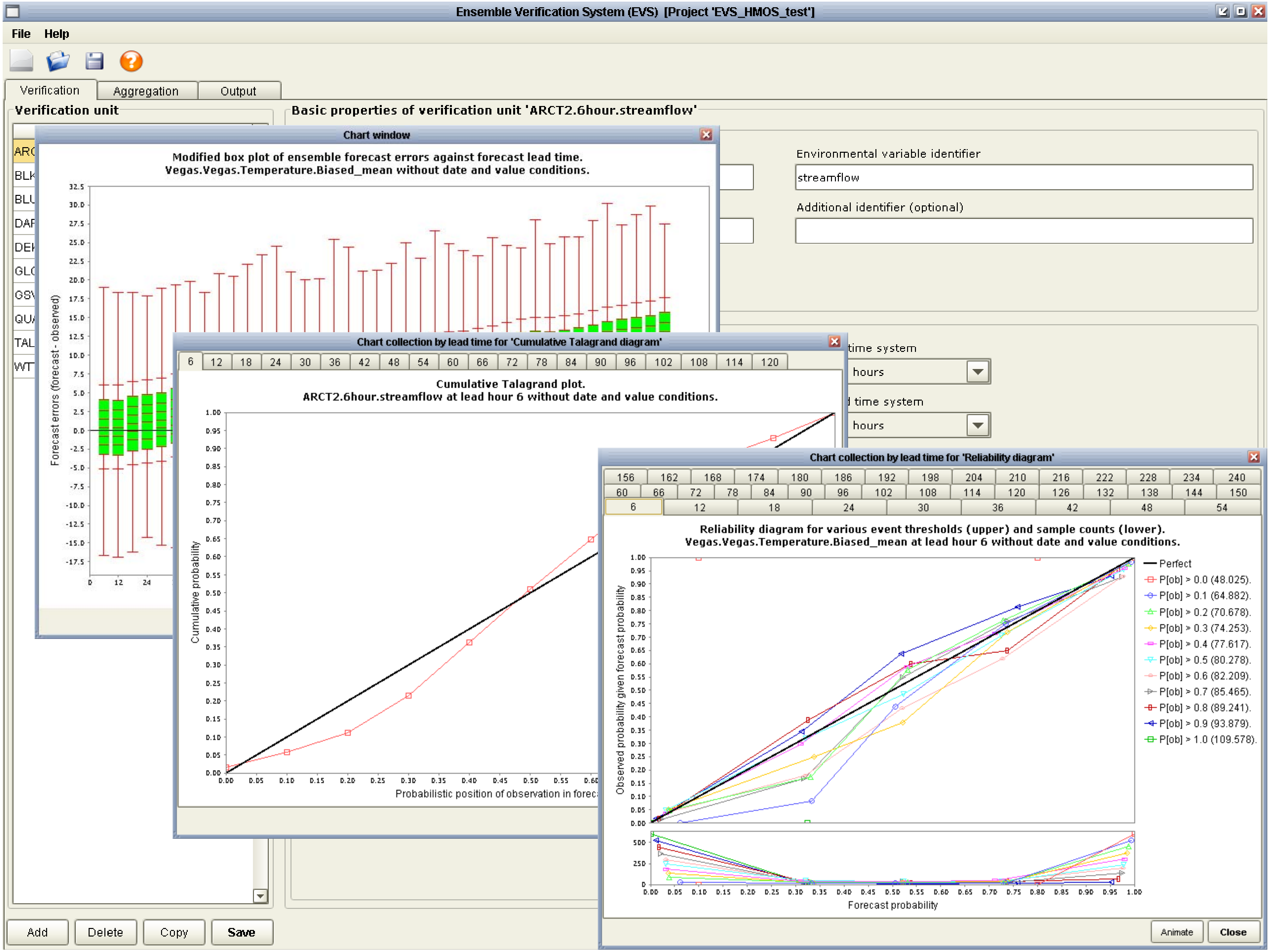
User-Friendly Software

Ensemble Verification System (EVS)

- Java tool with structured GUI
- Verification of numerical time-series
- Flexible “conditional verification”
- Several key metrics, including new ones

Status

- Available to all RFCs (experimental)
- Fully documented and freely available


Ensemble Verification System (EVS) [Project 'EVS_HMOS_test']
File
Help
Verification
Aggregation
Output
Verification unit
Basic properties of verification unit 'ARCT2.6hour.streamflow'
Environmental variable identifier
streamflow
Additional identifier (optional)
time system
hours
time system
hours
Chart window
Modified box plot of ensemble forecast errors against forecast lead time.
Vegas.Vegas.Temperature.Biased_mean without date and value conditions.
Forecast errors (forecast - observed)
Chart collection by lead time for 'Cumulative Talagrand diagram'
Cumulative Talagrand plot.
ARCT2.6hour.streamflow at lead hour 6 without date and value conditions.
Cumulative probability
Probabilistic position of observation in forec
Chart collection by lead time for 'Reliability diagram'
Reliability diagram for various event thresholds (upper) and sample counts (lower).
Vegas.Vegas.Temperature.Biased_mean at lead hour 6 without date and value conditions.
Observed probability given forecast probability
Perfect
P[ob] > 0.0 (48.025).
P[ob] > 0.1 (64.882).
P[ob] > 0.2 (70.678).
P[ob] > 0.3 (74.253).
P[ob] > 0.4 (77.617).
P[ob] > 0.5 (80.278).
P[ob] > 0.6 (82.209).
P[ob] > 0.7 (85.465).
P[ob] > 0.8 (89.241).
P[ob] > 0.9 (93.879).
P[ob] > 1.0 (109.578).
Forecast probability
Animate
Close
Add
Delete
Copy
Save

Verification Software Plans

Enhancements to EVS

- Skill calculations
- Sampling uncertainty
- Separating hydrograph shape/timing errors
- Incorporating feedback from RFCs
- Modify EVS to fit in XEFS, but ultimately…..

National Baseline Verification System

- Integrate capabilities of EVS and IVP

Hindcaster: Goal

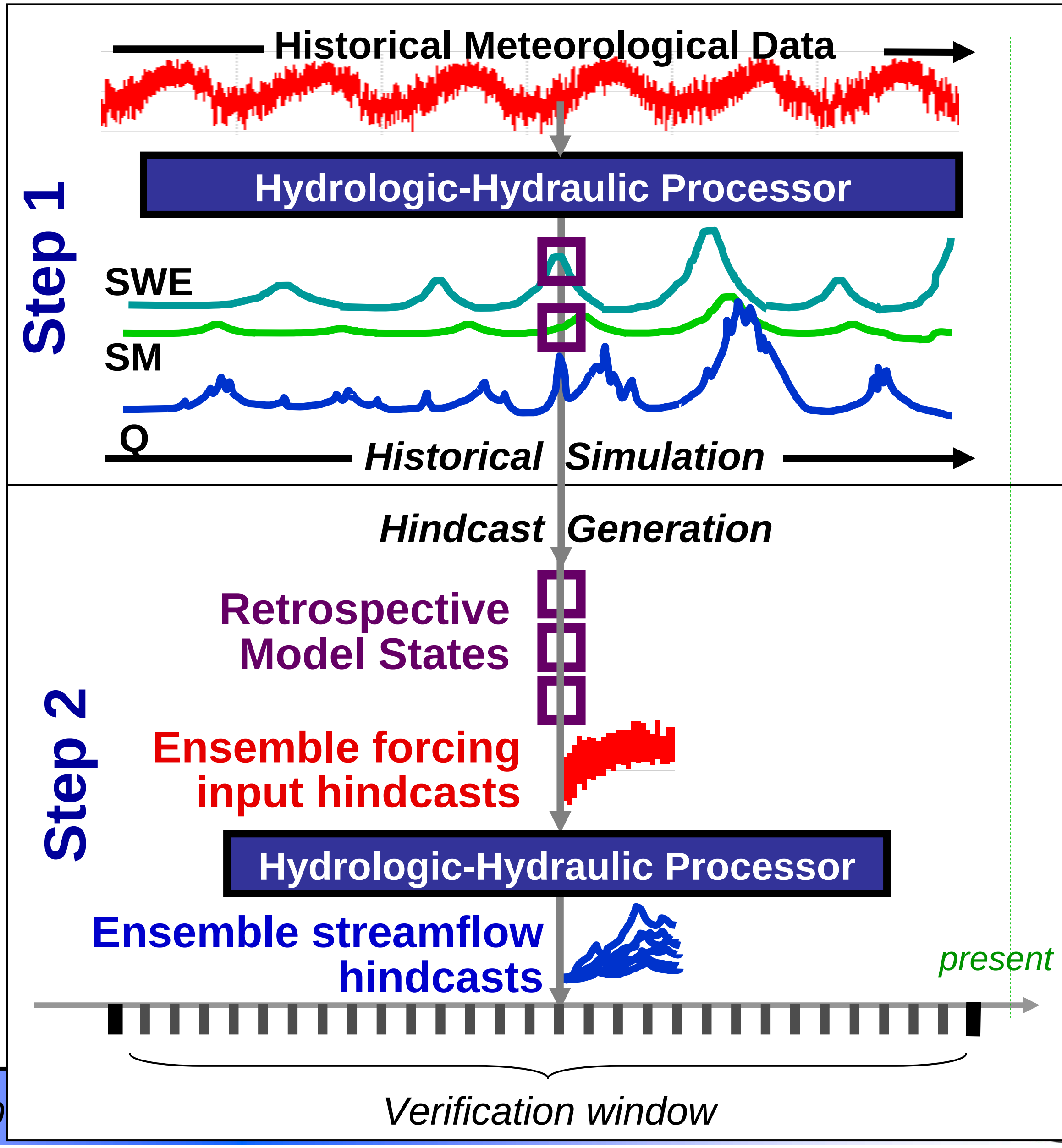
- Goal: systematic hindcasting/re-forecasting for all processes in operational/experimental forecasting system to support verification

- Benefits:
  - validate ensemble science from large samples for fixed forecasting scenarios
  - serve RFC's operational need for calibration and validation
  - quantify uncertainty sources using various hindcasting scenarios

- Verify with various references to quantify error sources:
  - forecast flow vs. simulated flow from perfect forcing inputs
    → forcing input uncertainty
  - forecast flow vs. observed flow
    → forcing input uncertainty + hydrologic uncertainty

Hindcaster: Processes

- Hindcasting done once for a given forecast scenario (fixed models) and a given verification time period:
  - Step 1: produce retrospective model states
  - Step 2: produce hydrologic hindcasts



Hindcaster: Data

- Precipitation and Temperature:
  - Step 1: continuous record of observations up to present
  - Step 2: ensemble forecasts or hindcasts (e.g., from EPP2)
- Other inputs (MAPE, PTPE, QME, etc.):
  - Steps 1 & 2: continuous record of observations up to present
- Streamflow:
  - Observations up to present for verification

Hindcaster: Status

- Current prototype based on NWSRFS ESP:
  - Modified to use enhanced ESP (DR 18809 for ob9)
    - → produce retrospective model states for correct timing
  - Coupled w/ EPP2 hindcaster
    - → produce flow hindcasts from different EPP2 outputs
    - → analyze impact of input and hydrologic uncertainties
  - Run in pseudo single-valued mode
    - → produce raw model hindcasts
    - → analyze impact of operational MODs
  - To be coupled w/ Ensemble Post-Processor
    - → analyze impact of post-processing
- In the future, hindcaster w/ XEFS-CHPS

3. Verification Science Issues

Outstanding Science Issues

- Are verification results statistically reliable given sampling uncertainty (i.e. can we act on them)?
- How can we verify real-time forecasts?
- Can we develop simple verification metrics for all aspects of forecast quality?
- Can we diagnose particular error sources further (e.g. phase vs. amplitude errors)?
- How can we verify extreme events?
- How can we account for error in observations?
- How can we verify forecasts for multi-scale variables (e.g. flow)?
- How can we verify forecasts if non-stationarity exists (e.g. climate change)?

3(a) Sampling Uncertainty

Sampling Uncertainty In Verification

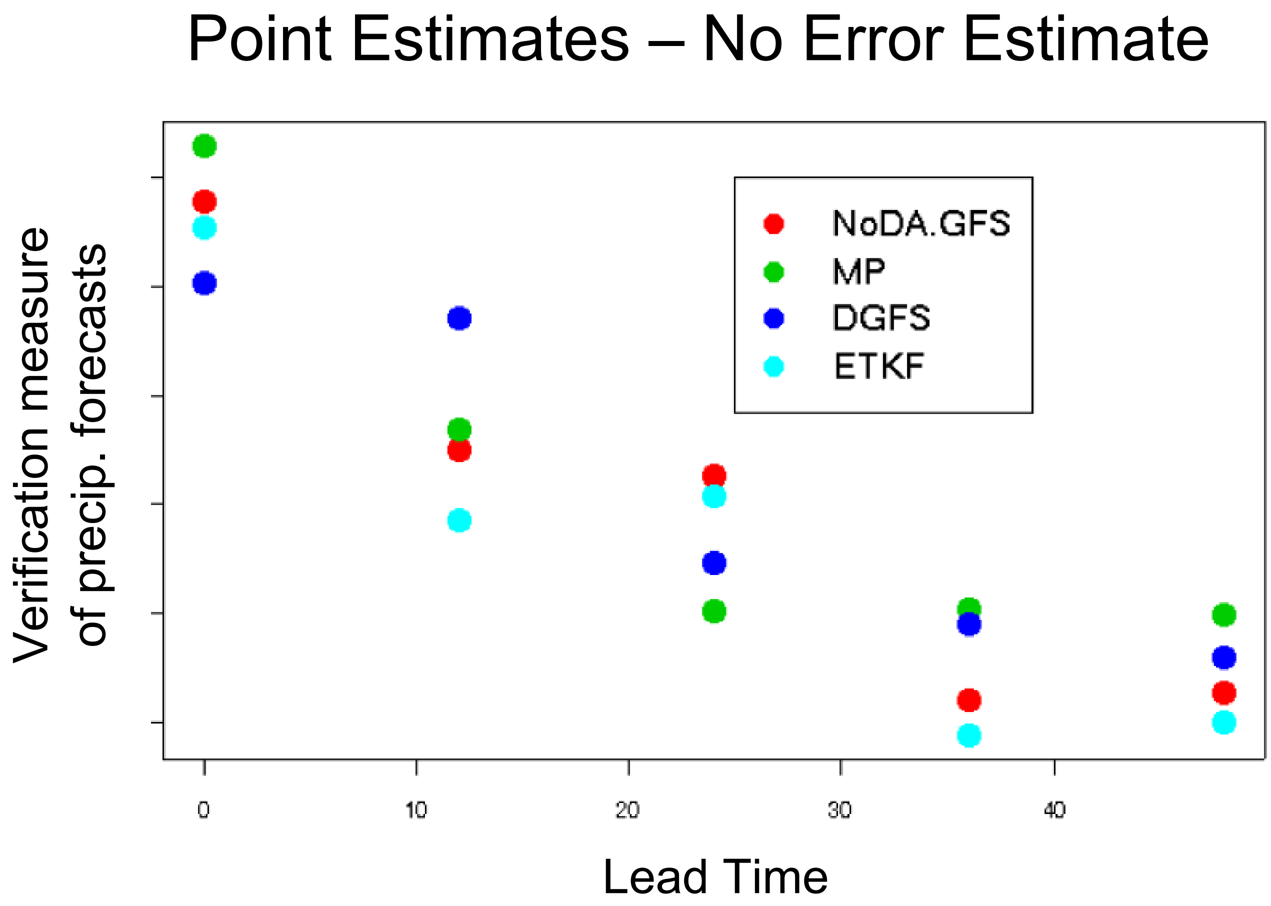
- **Why sampling uncertainty**
  - Verification datasets are finite samples of true underlying population, leading to verification statistics prone to sampling errors
  - Try to answer:

    "Is forecast A significantly different from forecast B?"
- **Reducing sampling uncertainty**
  - Regional pooling to increase effective sample size
  - Using resistant measures
    - E.g., Mean Absolute Error (MAE) is less sensitive to outlier errors than Mean Square Error (MSE)

Estimating Sampling Uncertainty

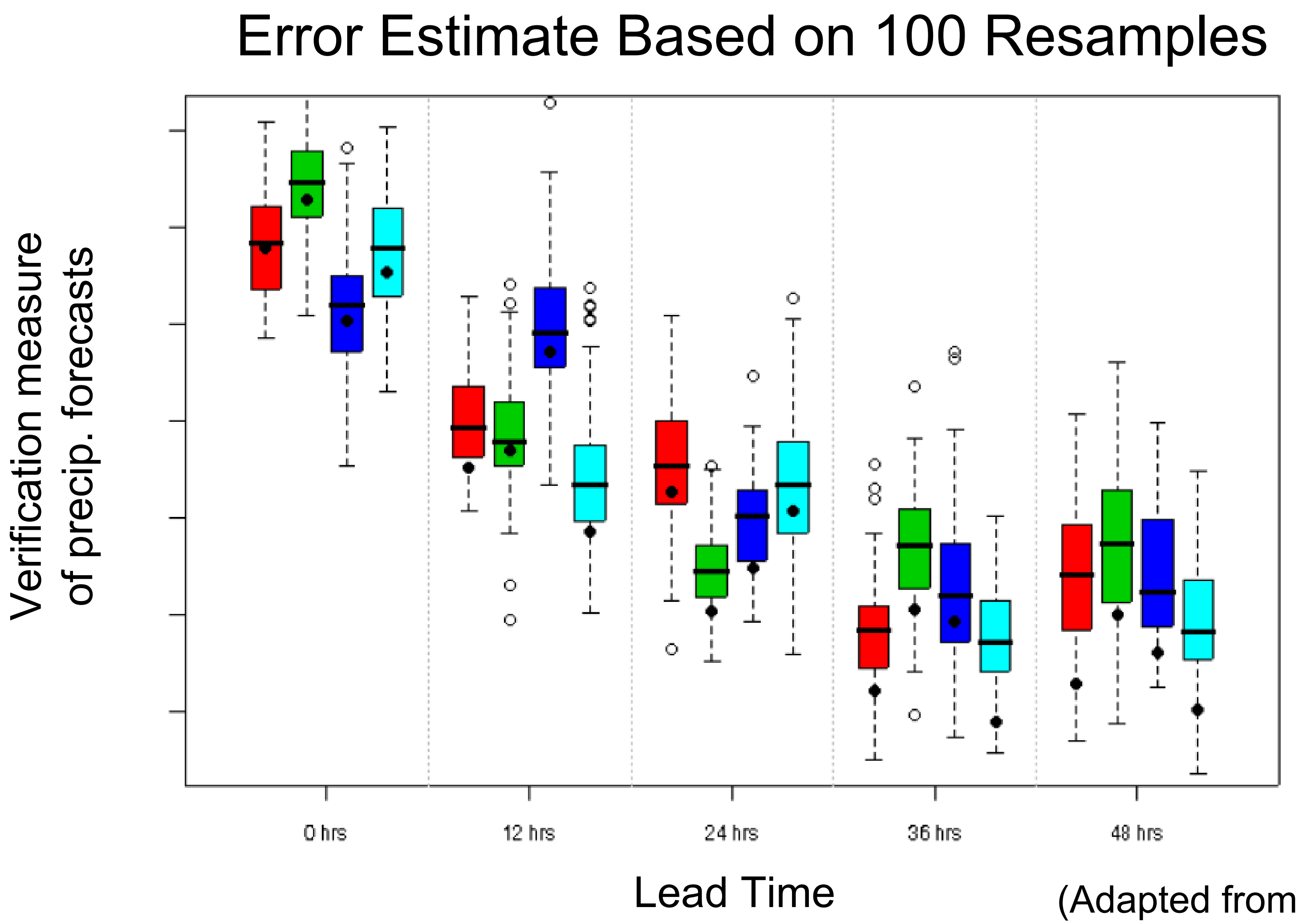
- **Point estimation**
  - ignore uncertainty
- **Standard error estimation**
  - Envelops (error bounds) around nominal values
- **Interval estimation**
  - Confidence intervals
    - random intervals with a specified level of confidence (e.g. 95%, 99%) of including a given a sample value of a measure (statistic)
  - Other intervals
    - Prediction interval, Bayes interval, …

Sampling Uncertainty: Example



(Adapted from Pocernich 2008)

Sampling Uncertainty: Example



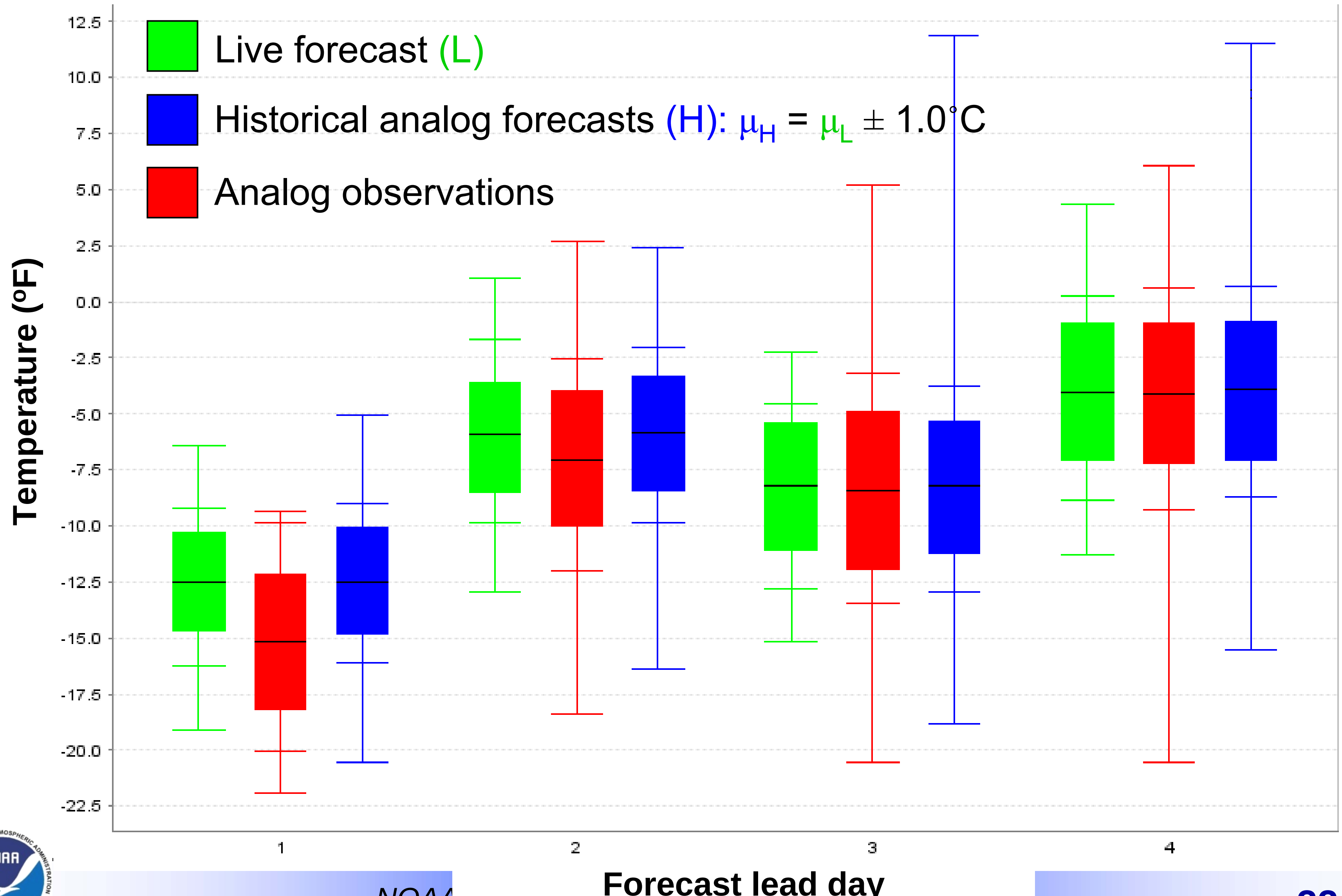
(Adapted from Pocernich 2008)

Ongoing/Future Work on Sampling Uncertainty

- Compute **confidence intervals** for verification measures
  - Analytical approaches
    - Approximate sampling distribution of measures analytically
  - Computational resampling approaches
    - E.g., bootstrap methods
- **Other issues**
  - Observation error
    - So that verification statistics generally appear worse than they really are
  - Spatial and temporal dependence
    - Assumption of data independence often invalid

3(b) Real-time Verification

Informal Example



Formal Approach

"Collect obs. from past, analog, forecasts"

X = observed (unknown for live forecast)

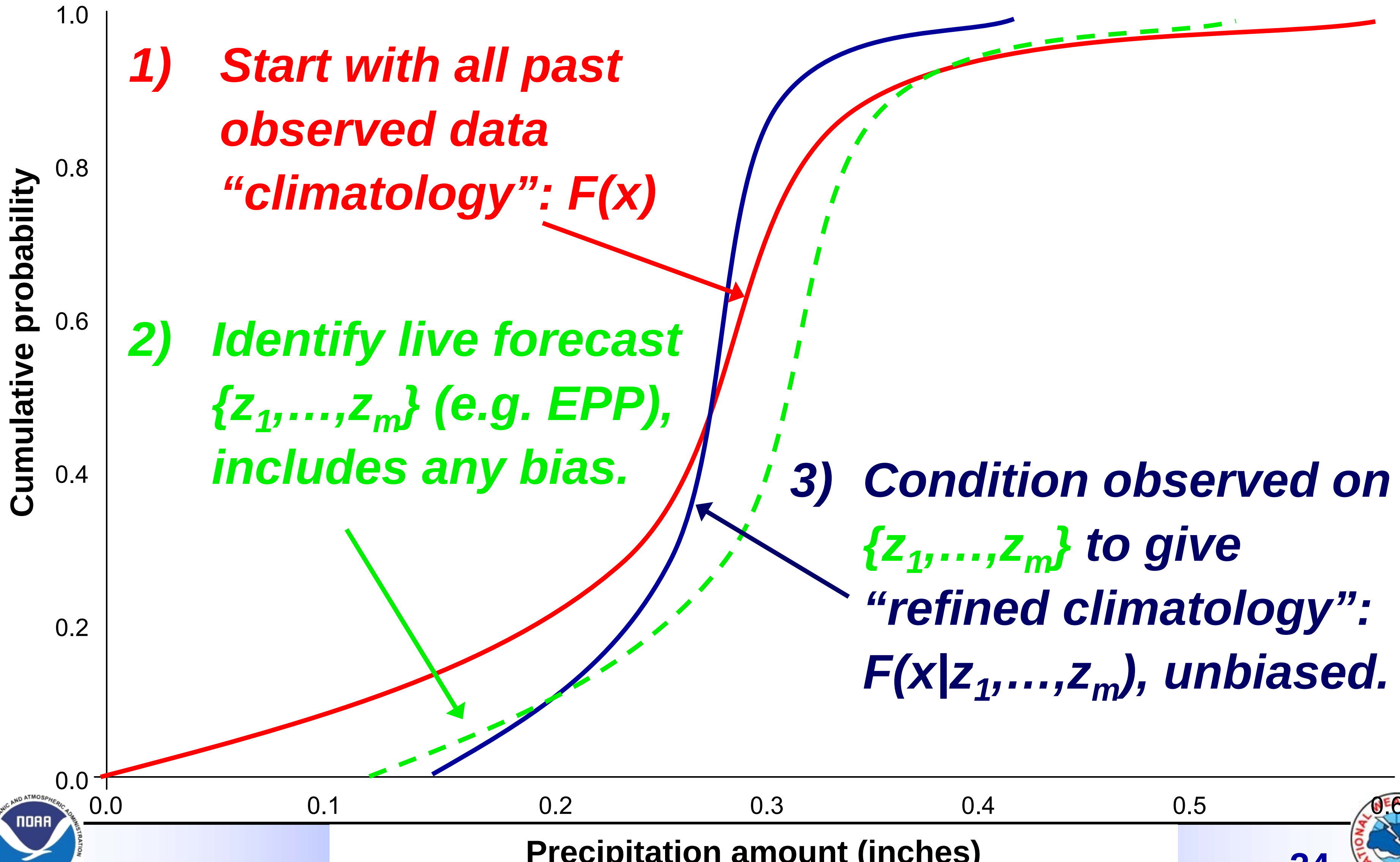
$\boldsymbol{Y} = \{Z_1,\ldots,Z_m\}$, live forecast

The aim is to estimate (from past data):

$F(x|z_1,\ldots,z_m)$

i.e. past observations whose paired forecasts come from parent pop. of *Y*.

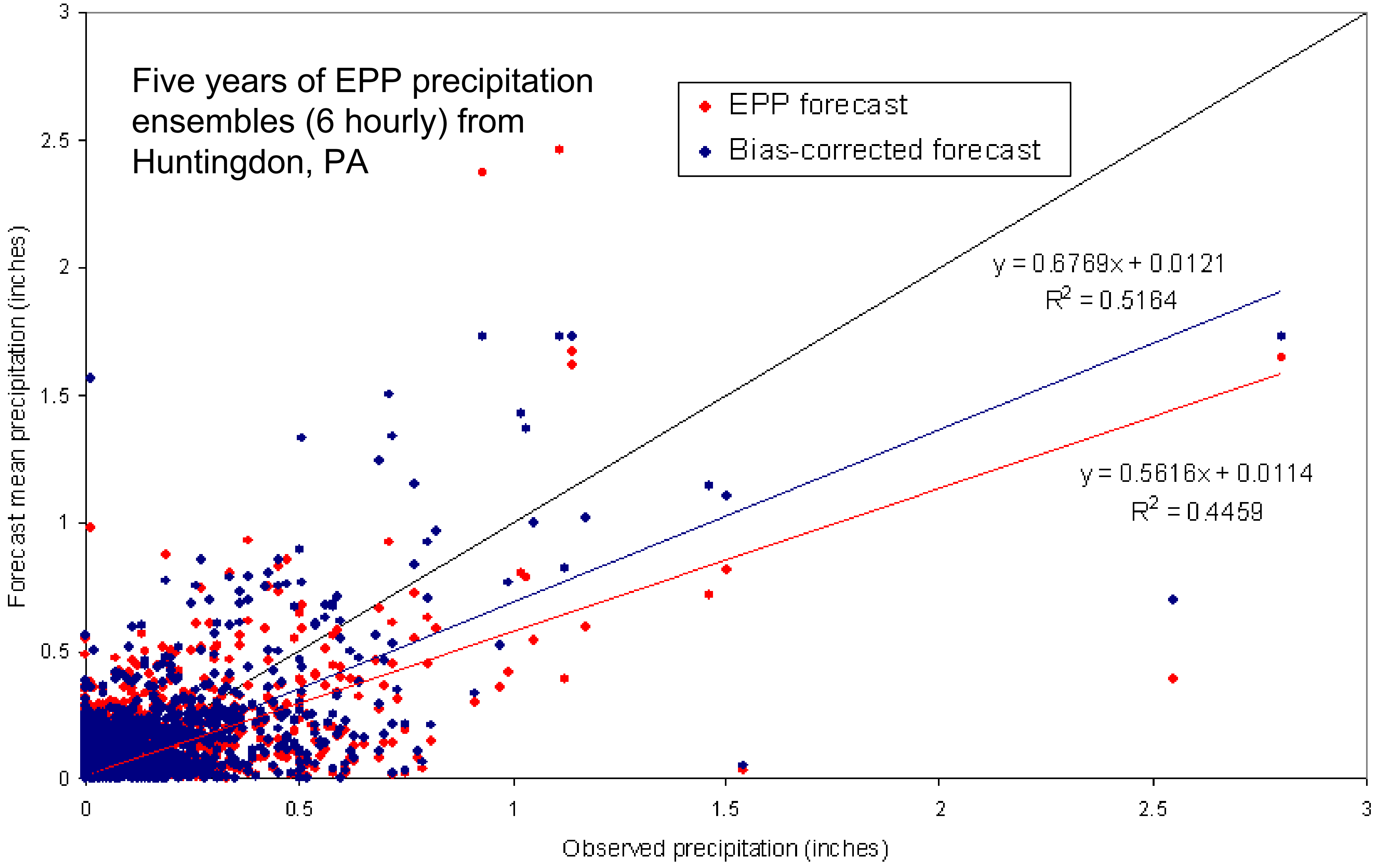
Formal Approach



How to Estimate?

- No single 'parametric' model for all forecast types (e.g. Normal).
- "Indicator regression". An estimate of $\mathrm{Prob}[X \le c_i | Z_j]$ $_{j=1,\ldots,m}$ for several "cutoffs", $_{i=1,\ldots,p}$.
- For each c_i, estimate the average number of times x is below c_i given the z_j's are above or below c_i: multiple regression of 1's and 0's (indicators).

Example of Results



4. Collaborations

RFC Collaborations

NWS Hydro. Forecast Verification team

- **RFC verification workshop** in Aug. 07
- **Exercises** with IVP and EVS
- **RFC verification case studies** with IVP and EVS
- **2nd RFC verification workshop** on Nov. 18-20, 2008
- **Final team report in 2009** to propose standardized verification strategies for identified users and dissemination plan (with performance tracking measures)

http://www.nws.noaa.gov/oh/rfcdev/projects/rfcHVT_chart.html

Other Collaborations

Some key collaborators

- **Iowa State University and University of Iowa**
- **University of California, Irvine**
- **HEPEX**

THORPEX-HYDRO project

- **Verification of met. and hydro. ensembles**

COMET training

- **Online verification module now available!!**

Thank you!

Any questions?