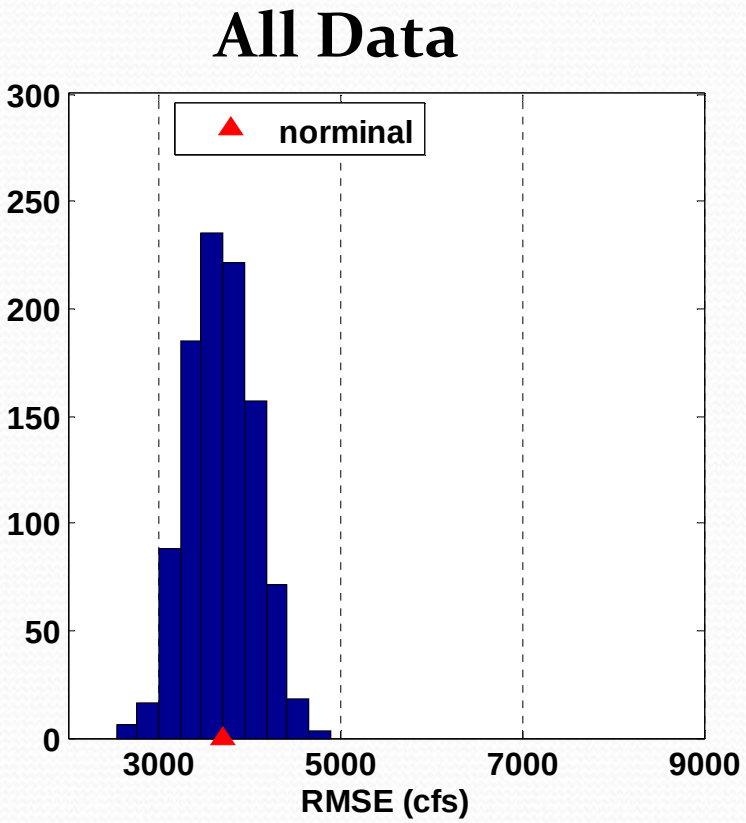
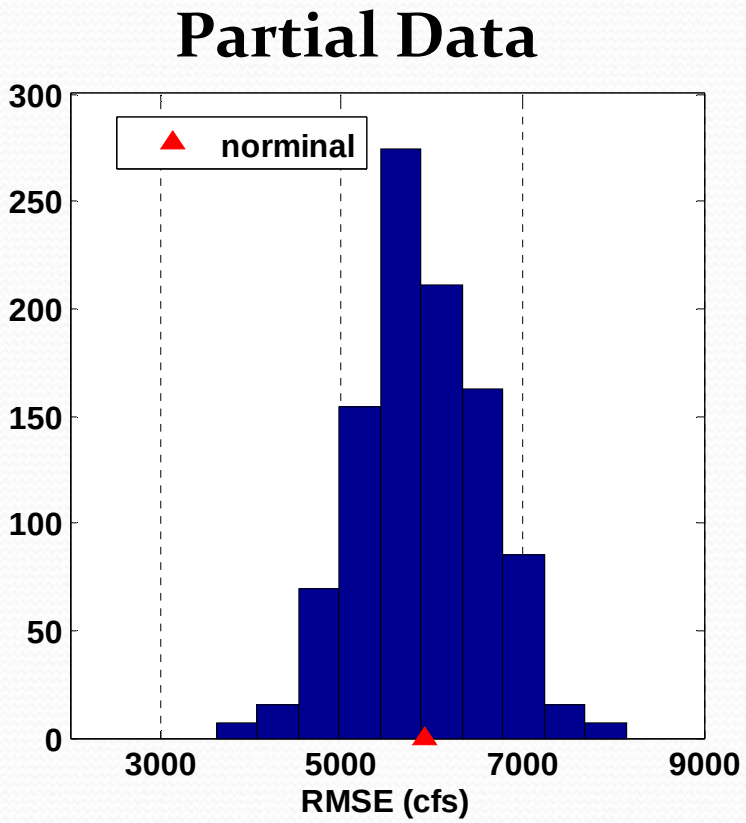
Sampling Uncertainty & Error Decomposition in Hydrologic Verification

Yuqiong Liu, Julie Demargne, James Brown

2nd RFC Verification Workshop
Salt Lake City, UT
November 18-20, 2008

Same Forecast System, Different Statistics





HMOS QUAO2 Ensembles (1999-2008)

Sampling Uncertainty in Verification

- Verification is based on finite samples of forecast-observation pairs
  - Verification statistics are subject to sampling errors
- Try to answer questions like:
  - What is the uncertainty associated with the value of a verification measure?
  - Is forecast ***A*** significantly different from forecast ***B***?

Coping with Sampling Uncertainty

- Reduce sampling uncertainty
  - Regional pooling to increase effective sample size
  - Using resistant measures
    - E.g., Mean Absolute Error (MAE) is less sensitive to sampling uncertainty than Mean Square Error (MSE)
- Quantify sampling uncertainty
  - Full probability distribution
  - Standard errors
  - **Confidence intervals**
    - random intervals with a specified level of confidence (e.g. 95%, 99%) of including a given a sample value of a measure

Approaches to Estimation of Sampling Uncertainty

- Parametric approaches
  - Classical Gaussian approximation
  - Clopper-Person exact interval
  - Many others …
- Non-parametric approaches
  - Analytical methods
    - Bradley at al. 2007 (standard error for BS and BSS)
  - Resampling /numerical methods
    - Permutation
    - **Bootstrap**

Bootstrap

- Recreates the relation between the *population* and the *sample* by resampling (with replacement)
- Assumes independent data
- Outperforms classical normal approximation under i.i.d. condition
- Fails if i.i.d. assumption fails

Example on Sampling Uncertainty

- Dataset: HMOS for QUAO2
  - Entire dataset: 31856 obs-fcst pairs in total
  - Partial dataset: the last 10620 pairs
- Approach: bootstrapping
  - 1000 bootstraps
    - resample dataset 1000 times and call EVS metric functions to calculate the metrics for each sample
  - Percentile intervals
    - 1-2α confidence interval defined by $[\theta^{\alpha} - \theta^{(1-\alpha)}]$,

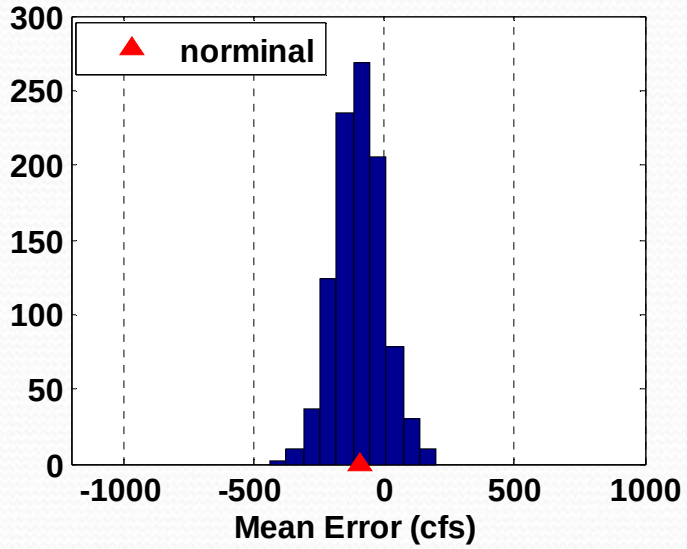
      e.g., 90% CI is [5th percentile, 95th percentile]

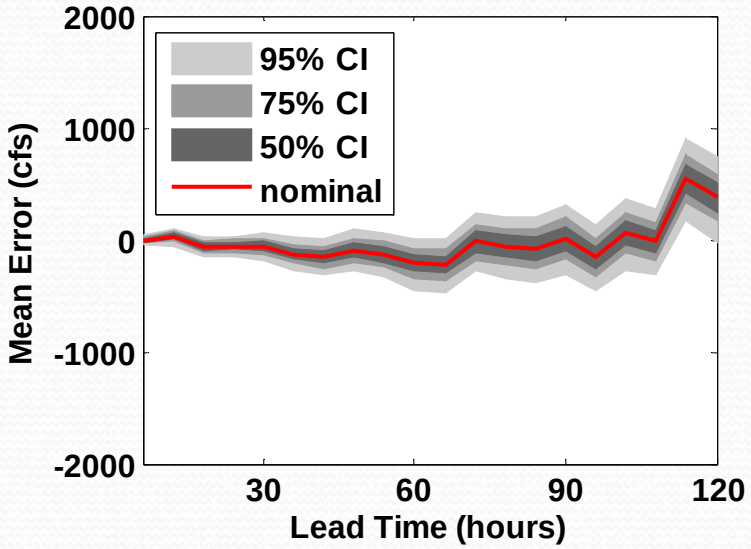
Mean Error

Lead Time 8 (48 hours)

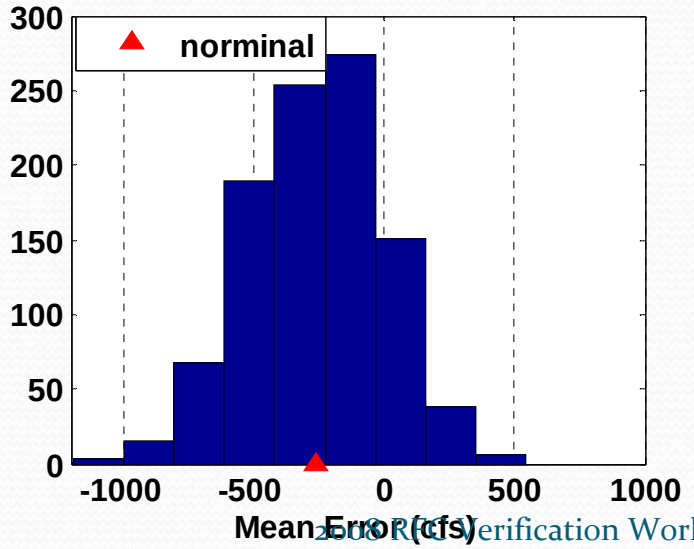
All Lead Times

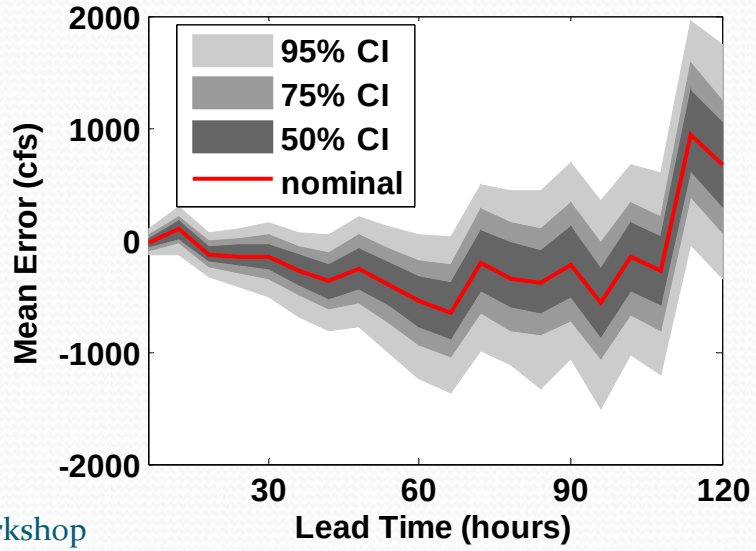
All Data





Partial Data



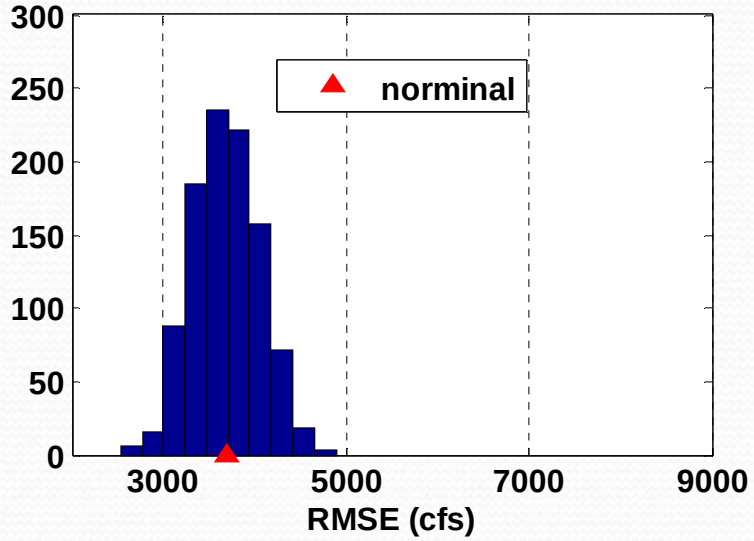


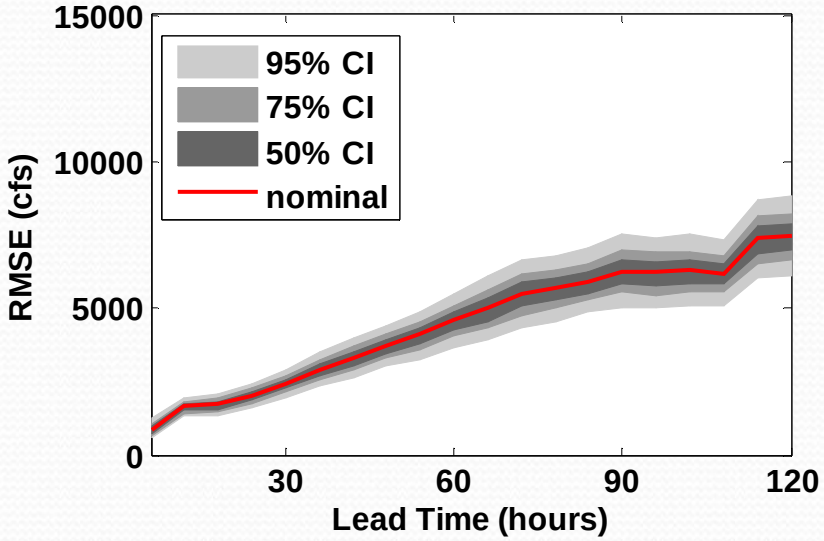
RMSE

Lead Time 8 (48 hours)

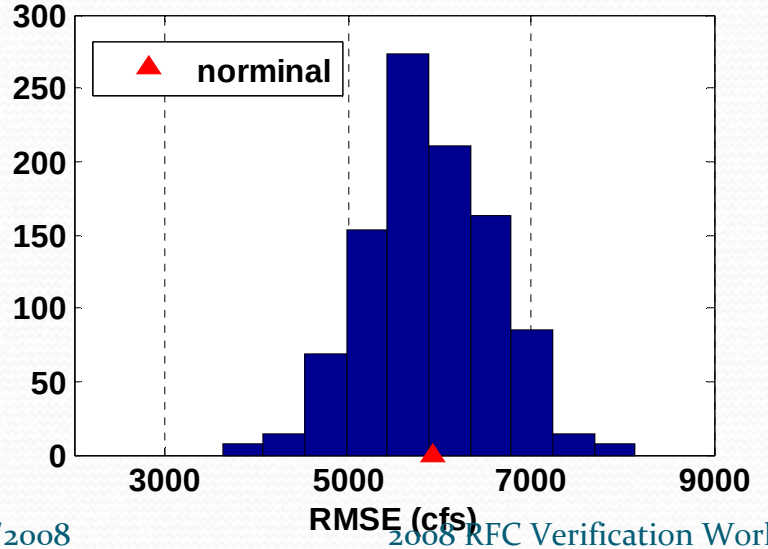
All Lead Times

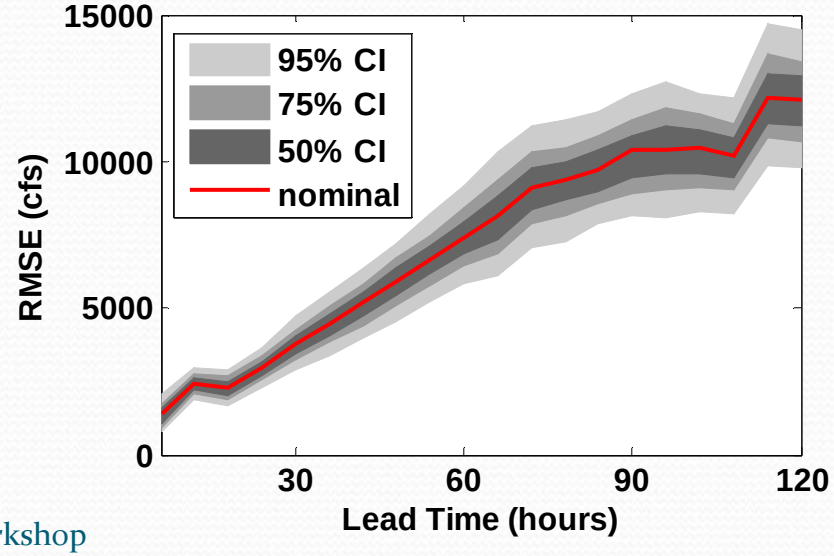
All Data



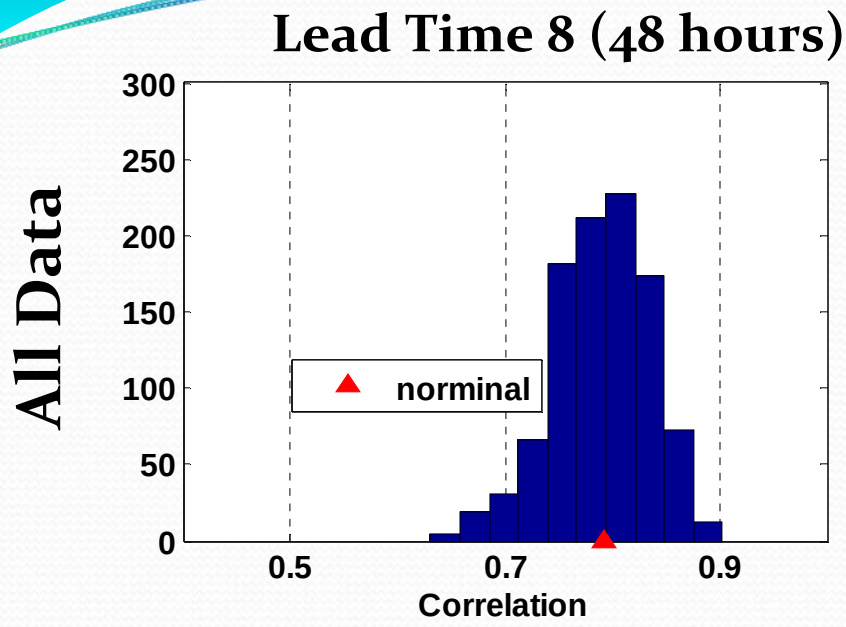


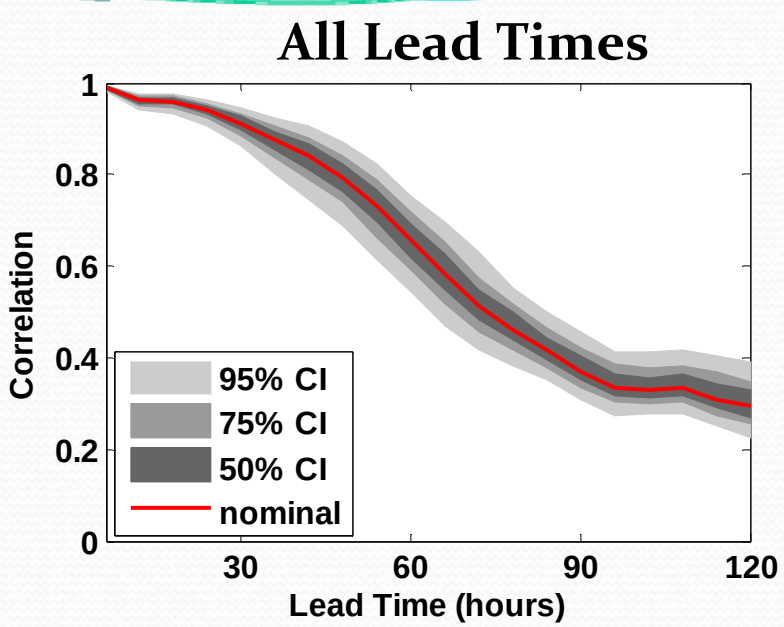
Partial Data

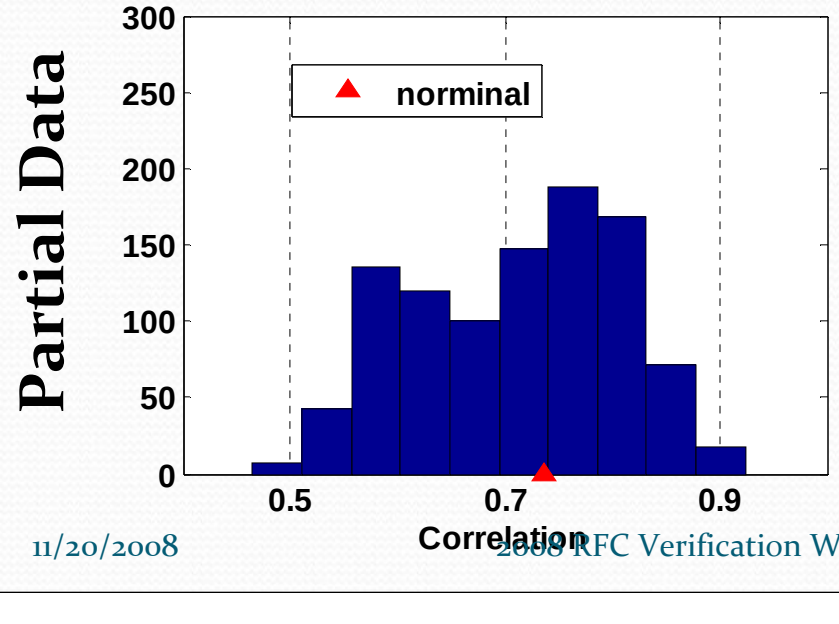


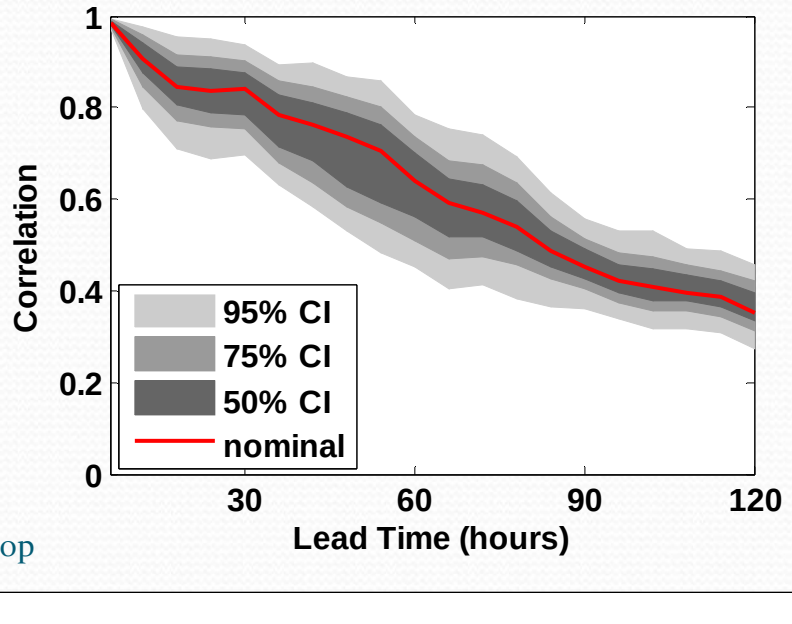


Correlation







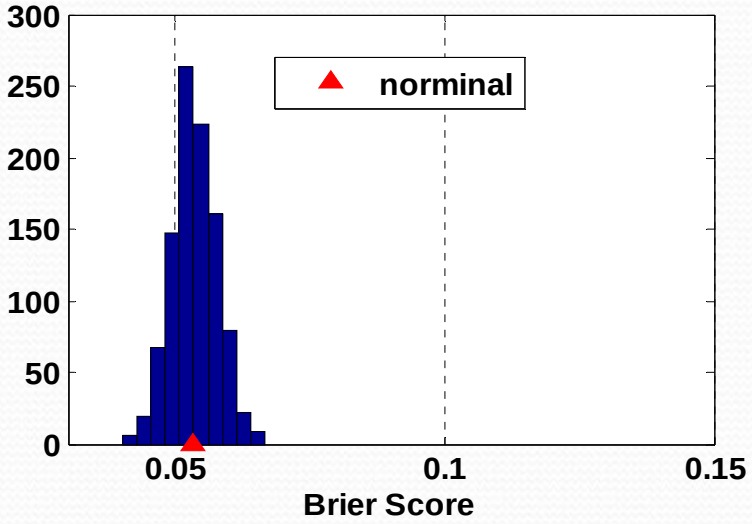


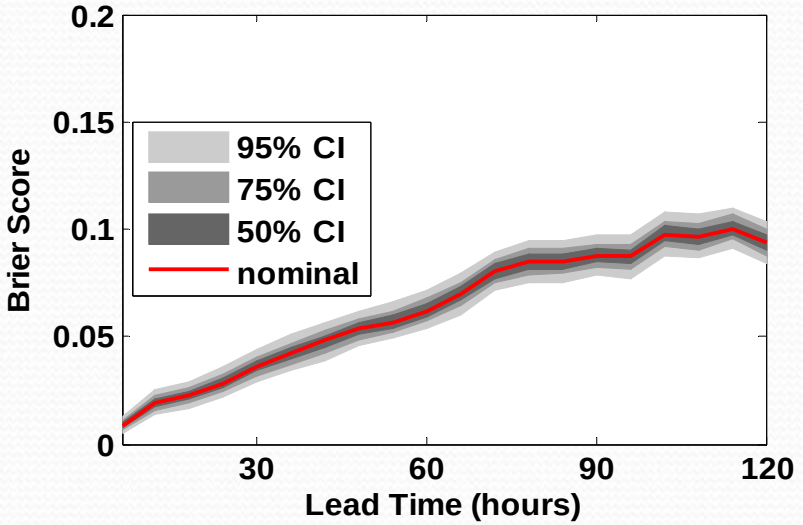
Brier Score (Prob>0.8)

Lead Time 8 (48 hours)

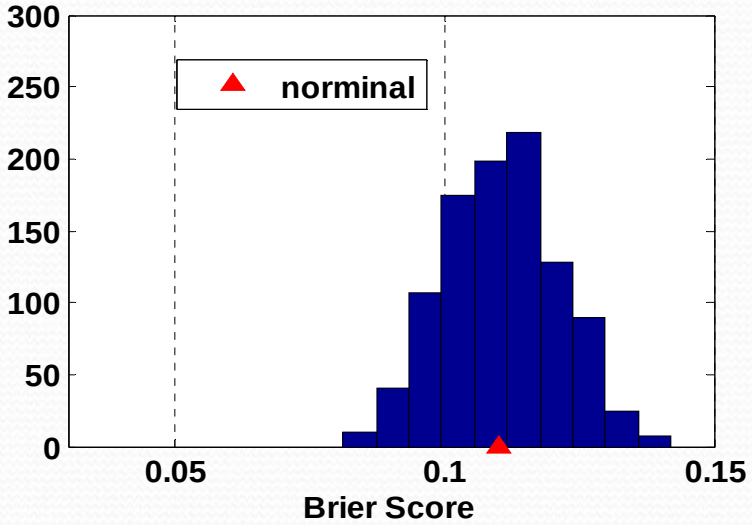
All Lead Times

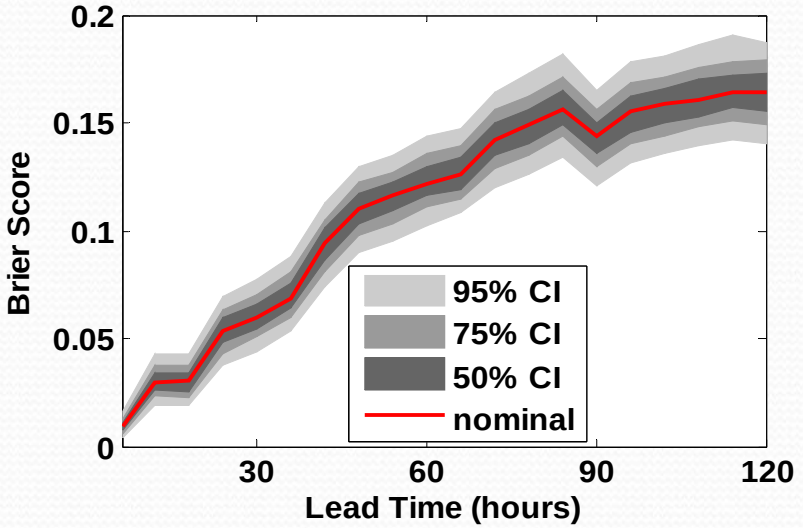
All Data





Partial Data



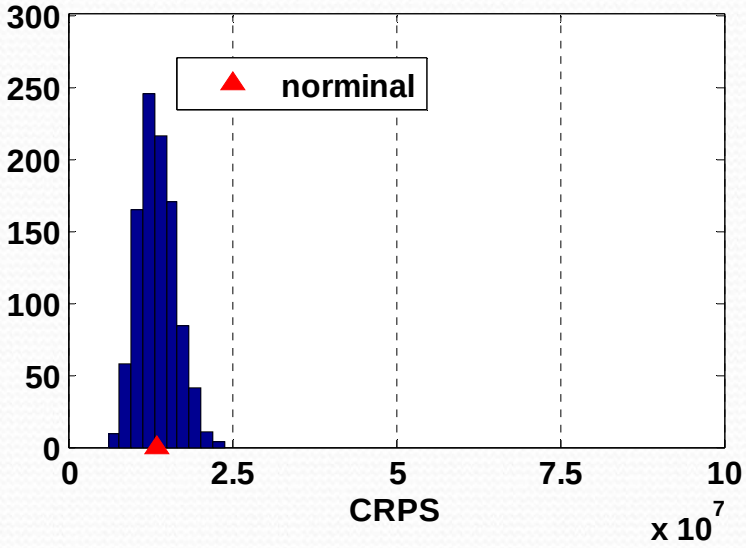


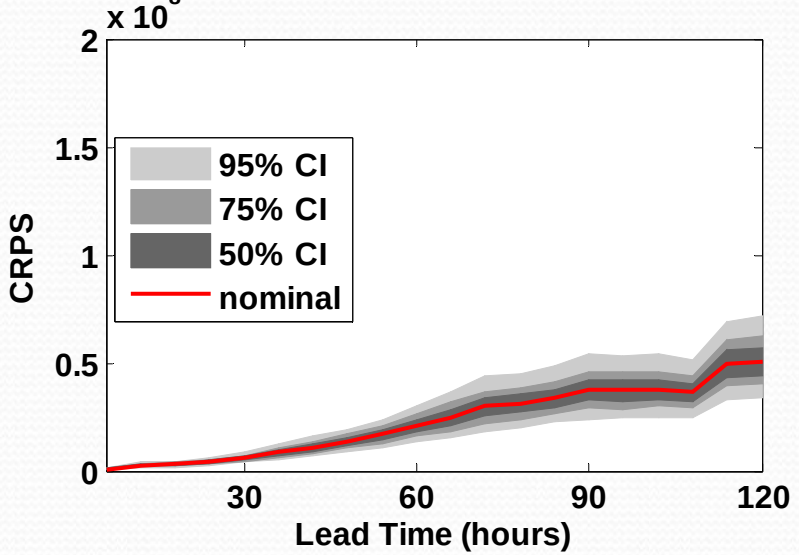
CRPS

Lead Time 8 (48 hours)

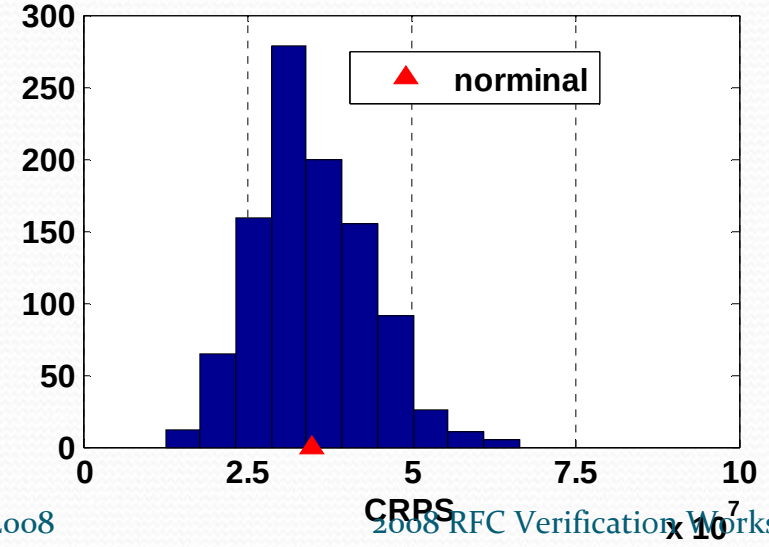
All Lead Times

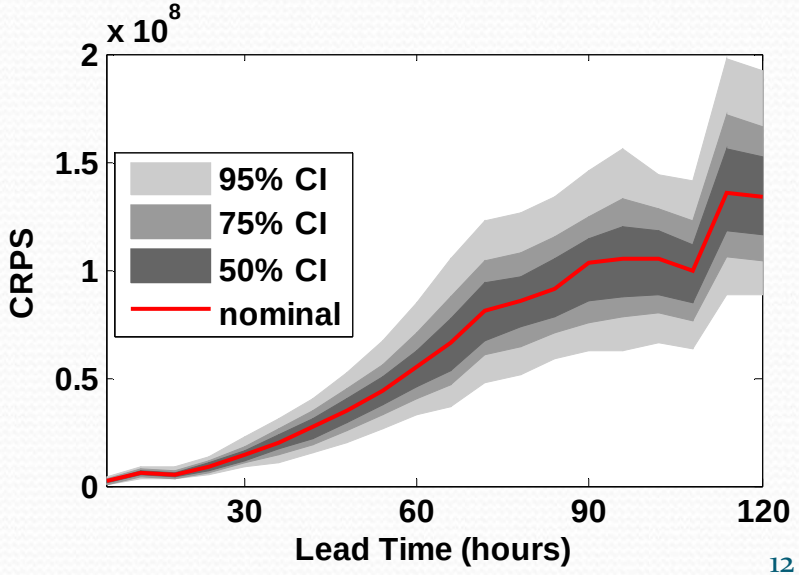
All Data





Partial Data





Observations

- Sampling uncertainty increases with lead time
- Sampling uncertainty increases with decreasing sample size
- Distribution of sampling uncertainty may change with sample size and not always normal
  - Sampling uncertainty of verification statistics needs to be considered!

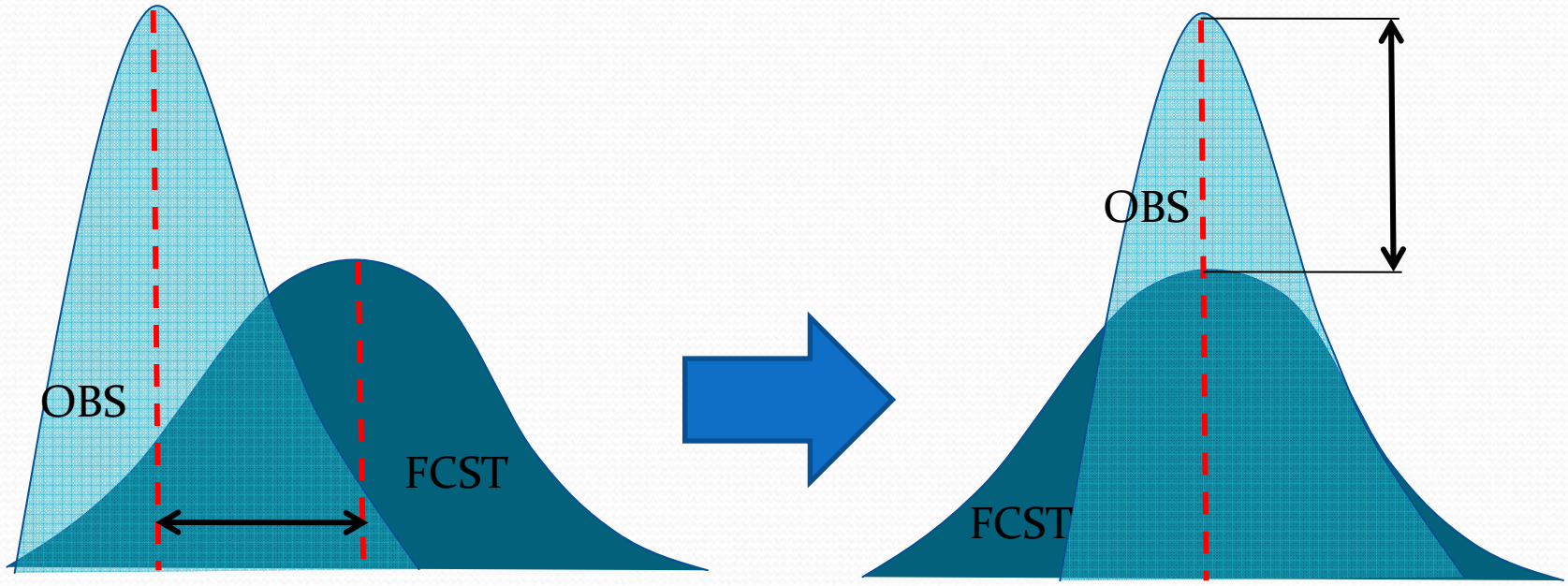
Ongoing/Future Work

- Numerical approaches
  - Bias-corrected and accelerated bootstrap
  - Moving –blocks bootstrap
  - Approximate bootstrap confidence intervals
- Analytical approaches
  - Bradley (2007): Brier score and Brier skill score
- Parametric approaches
  - Classical normal approximation

Detangling Timing, Peak Value, and Shape Errors in Streamflow Hydrograph

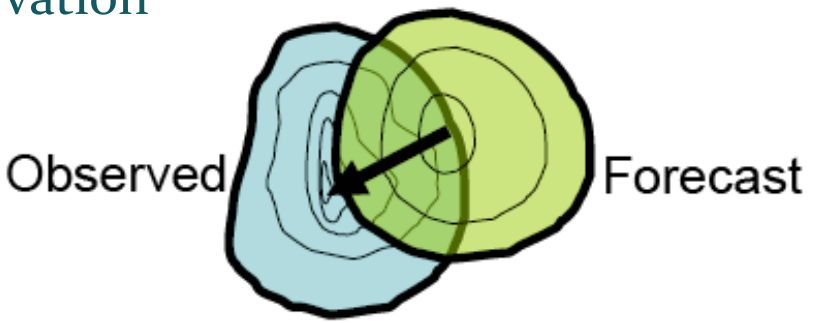
Different Types of Errors

- Peak Timing
- Peak Value
- Shape



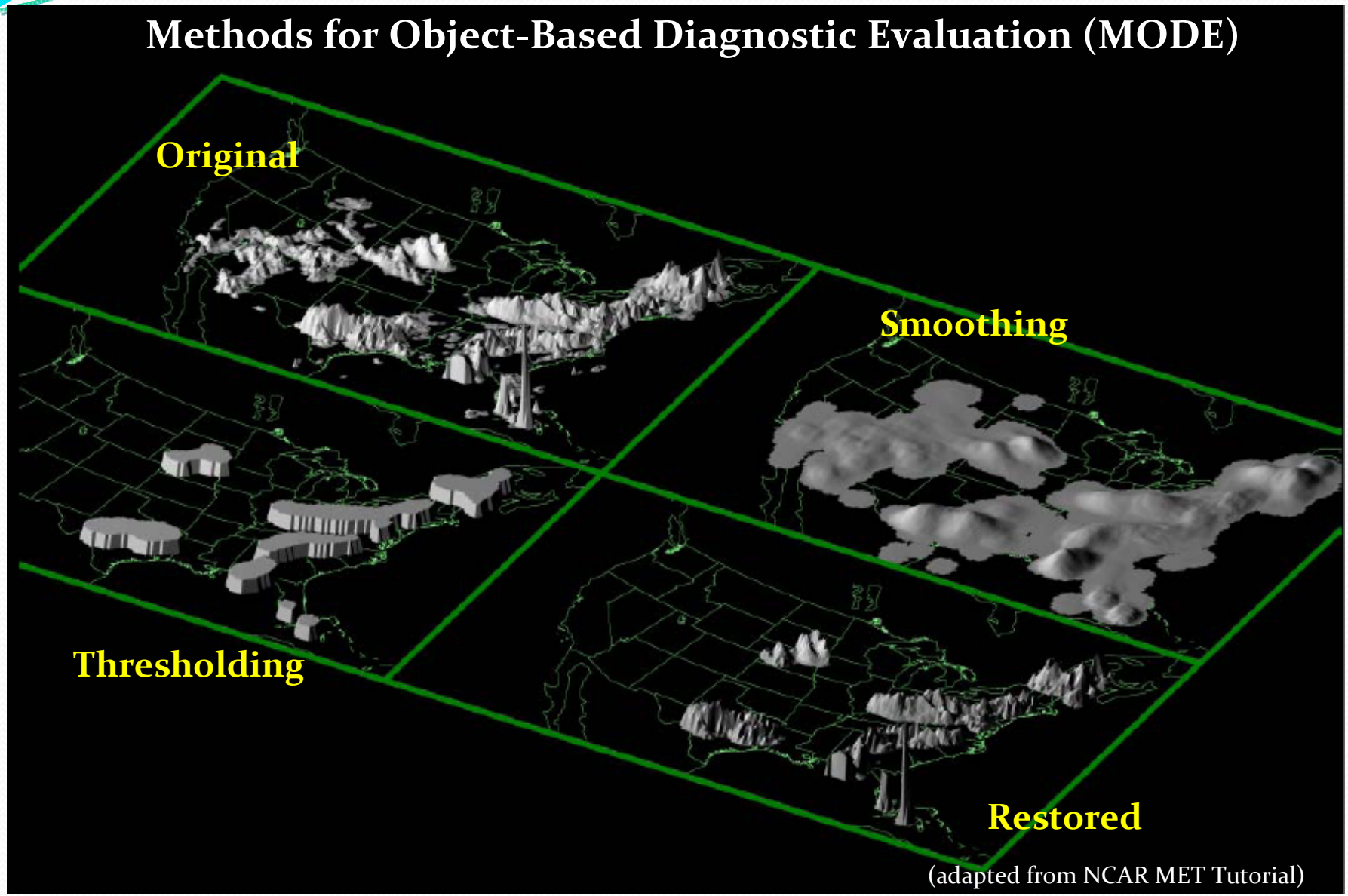
Learn From Spatial Verification

- Object-based methods
  - Detect objects of forecast and observation
    - Smoothing
    - Thresholding
  - Extract attributes of objects
  - Merge objects
  - Match/pair forecast & observation objects
    - Fuzzy logic
  - Compare attributes of paired objects
    - Displacement, intensity, size, location, aspect ration, …



(adapted from NCAR MET Tutorial)

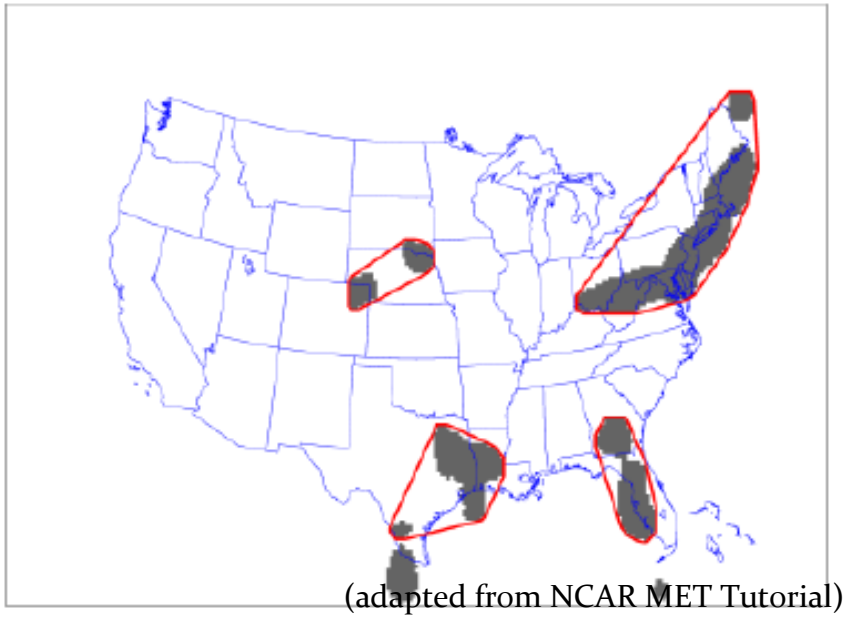
Methods for Object-Based Diagnostic Evaluation (MODE)



(adapted from NCAR MET Tutorial)

Forecast

Observed



(adapted from NCAR MET Tutorial)

MODE Applying to Hydrographs

- Identify events for forecasts & observations
  - Define event time window
  - Smooth hydrographs
  - Define event threshold
  - Apply threshold to hydrograph to produce events (objects)
- Merge events for forecasts & observations using fuzzy logic
- Pair events of forecasts & observations using fuzzy logic
- Compare attributes of pairs
  - Timing, peak value, shape (rising, recession)

Curve Registration

- Measure goodness-of-fit of temporal predictions
- First deformation along x-axis(time)
  - Amount of deformation (timing and shape errors)
- Then scaling along y-axis
  - Amount of scaling (peak error)

Reilly et al. 2004