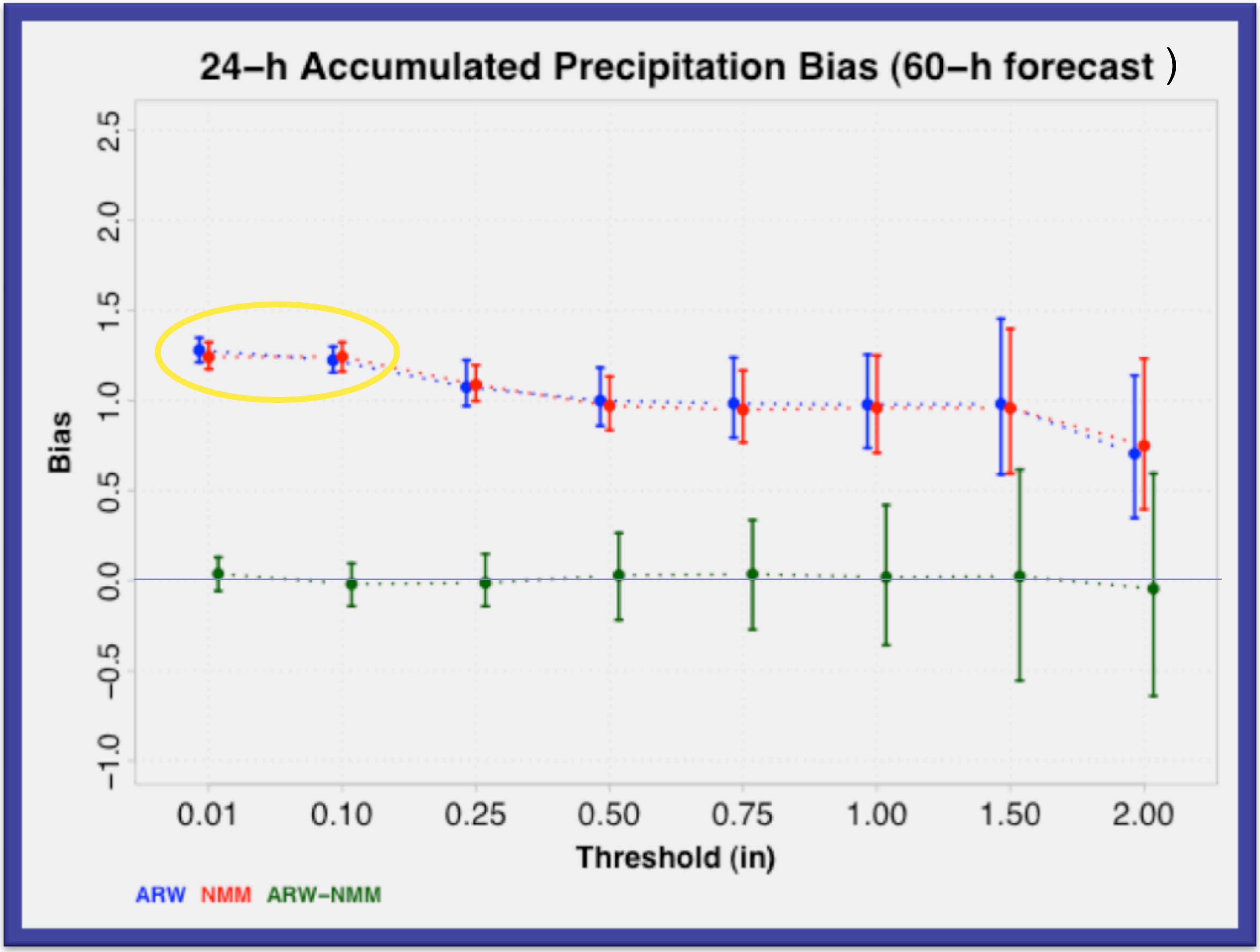
Giving meaning to your forecast verification results

Barbara Brown bgb@ucar.edu

- What does RMSE = 25 mean?
- Is 25 a good value? (What is “good”?)
- Is 25 better than 30?

Answer: It depends!!


24–h Accumulated Precipitation Bias (60–h forecast)
Bias
2.5
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
0.01
0.10
0.25
0.50
0.75
1.00
1.50
2.00
Threshold (in)
ARW NMM ARW–NMM

Accounting for Uncertainty

- Observational
- Model
  - Model parameters
  - Physics
  - Verification scores
- Sampling
  - Verification statistic is a realization of a random process
  - What if the experiment were re-run under identical conditions?

Hypothesis Testing and Confidence Intervals

- Hypothesis testing
  - Given a null hypothesis (e.g., "*Model forecast is un-biased*"), is there enough evidence to reject it?
  - Can be *One-* or *two-sided*
  - Test is against a *single null hypothesis*.
- Confidence intervals
  - Related to hypothesis tests, but more useful.
  - How confident are we that the true value of the statistic (e.g., bias) is different from a particular value?

Hypothesis Testing and Confidence Intervals

Example: The difference in bias between two models is 0.01.

Hypothesis test: Is this different from zero?

Confidence interval: Does zero fall within the interval? Does 0.5 fall within the interval?

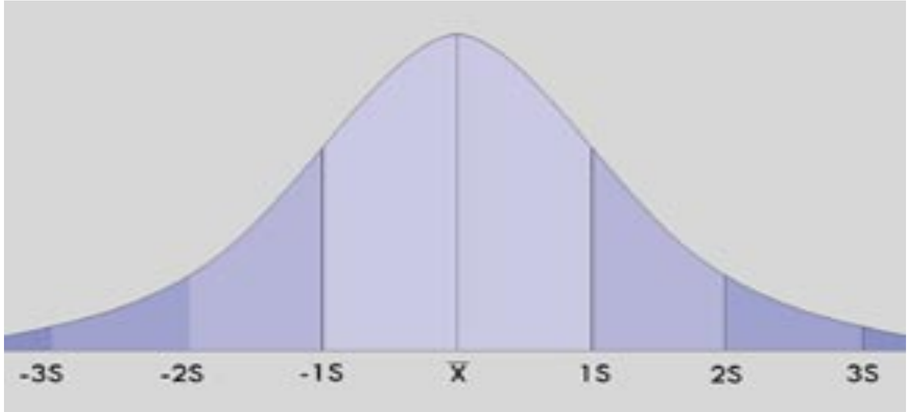
Confidence Intervals (Cis)

"If we re-run the experiment 100 times, and create 100 (1-α)100% CI's, then *we expect the true value of the parameter to fall inside (1-α)100 of the intervals."*

Confidence intervals can be *parametric* or *non-parametric*…

Confidence Intervals (CI's)

- Parametric
  - Assume the observed sample is a realization from a known *population* distribution with possibly unknown parameters (e.g., normal).
  - Normal approximation CI's are most common.
  - Quick and easy.



Confidence Intervals (CI’s)

- Nonparametric
  - Assume the distribution of the observed sample is representative of the *population* distribution.
  - Bootstrap CI’s are most common.
  - Can be computationally intensive, but easy enough.

Normal Approximation CI's

Estimate

Standard normal variate

Population ("true") parameter

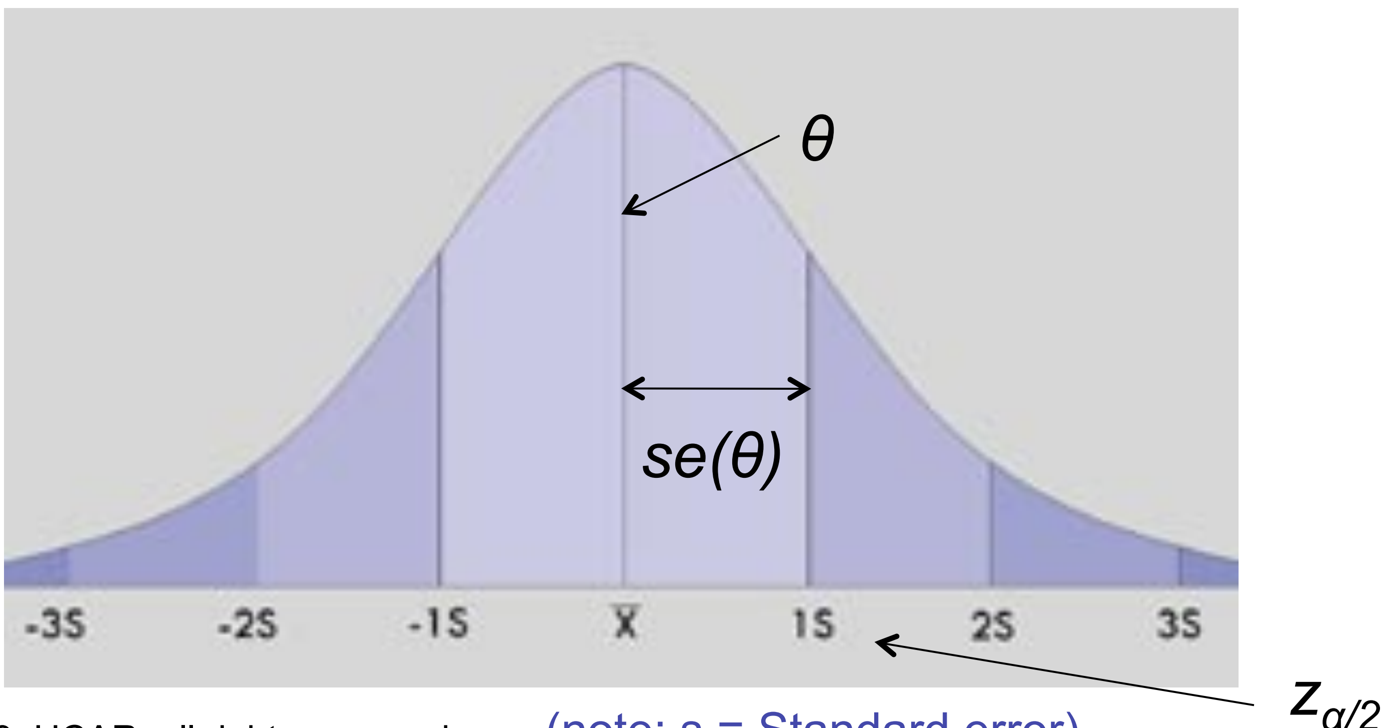
$$\hat{\theta} \pm z_{\alpha/2}\, se(\theta)$$

Is a (1-α)100% Normal CI for ϴ, where

- ϴ is the statistic of interest (e.g., the forecast mean)
- se(ϴ) is the standard error for the statistic
- z_v is the v–th quantile of the standard normal distribution.

Normal Approximation CI's

$$\hat{\theta} \pm z_{\alpha/2}\, se(\theta)$$



(note: s = Standard error)

Normal Approximation CI's

Example: Let $X_1, \ldots, X_n$ be an independent and identically distributed (iid) sample from a normal distribution with variance σ_X^2 .

Then, $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ is an estimate of the **mean**

of the sample. A (1-α)100% CI for the mean is given by

$$\bar{X} \pm z_{\alpha/2} \frac{\sigma_X}{\sqrt{n}}$$

Note: You can find much more about these ideas in any basic statistics text book

Normal Approximation CI's

- Numerous verification statistics can take this approximation in some form or another
  - Alternative CIs are available for other types of variables
    - Examples: forecast/observation *variance*, linear *correlation*
    - Still rely on the underlying sample being iid normal.
- Contingency table verification scores also have normal approximation CI's (for large enough sample sizes)
  - Examples: POD, FAR

Application of Normal Approximation CI's

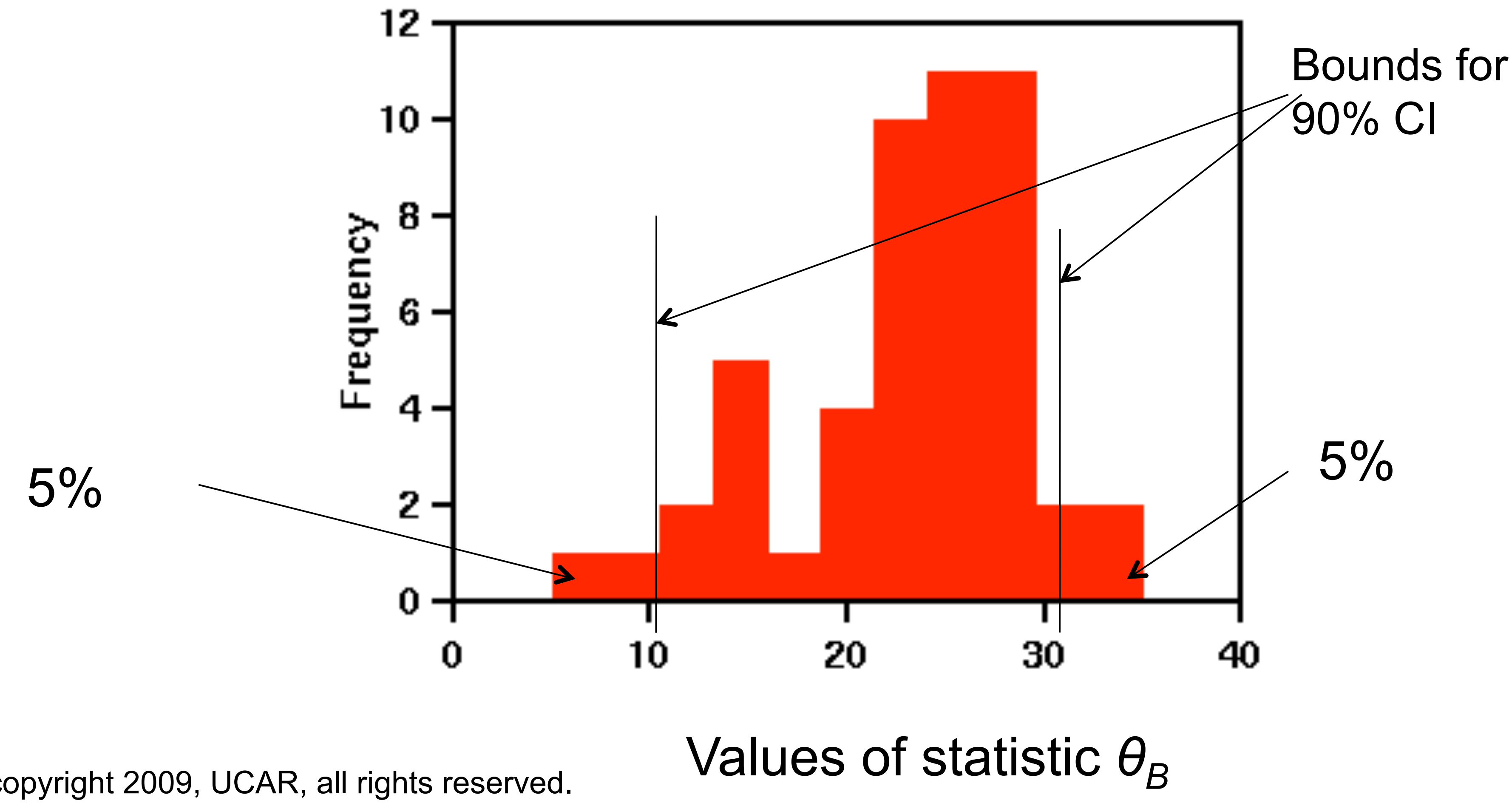
- Independence assumption (i.e., "iid") – temporal and spatial
  - Should check the validity of the independence assumption
  - MET doesn't do this
  - Methods that can take into account dependencies will be added to MET in the future
- Normal distribution assumption
  - Should check validity of the normal distribution (e.g., qq-plots, other methods)
  - MET does not do this – should be done outside of MET
  - ***However***… MET applies appropriate approaches for different verification statistics
- Multiple testing
  - When computing many confidence intervals, the true significance levels are affected (reduced) by the number of tests that are done

(Nonparametric) Bootstrap CI's

IID Bootstrap Algorithm

1. Resample *with replacement* from the sample, $X_1, \ldots, X_n$.
2. Calculate the verification statistic(s) of interest from the resample in step 1.
3. Repeat steps 1 and 2 many times, say B times, to obtain a sample of the verification statistic(s) θ_B .
4. Estimate (1-α)100% CI's from the sample in step 3.

Empirical Distribution (Histogram) of statistic calculated on repeated samples



Bootstrap CI’s

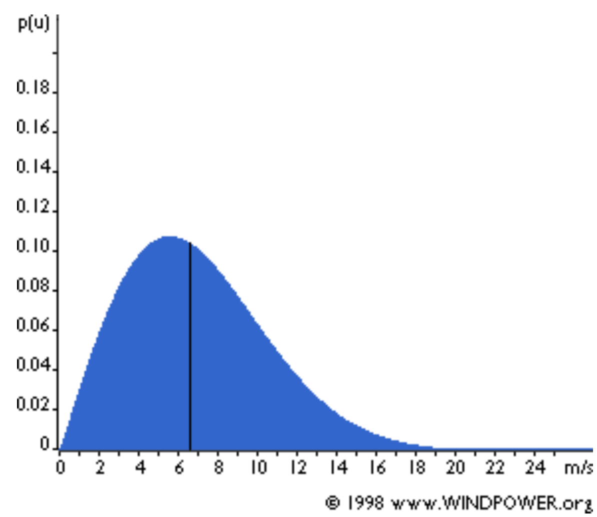
IID Bootstrap Algorithm: Types of CI’s

1. Percentile Method CI’s
2. Bias-corrected and adjusted (BCa)
3. ABC
4. Basic bootstrap CI’s
5. Normal approximation
6. Bootstrap-t

} Available in MET

Bootstrap CI's

- **Sample size** is a configurable parameter in MET
- **Typical approach**: Use same sample size as the original sample
  - Sometimes better to take smaller samples (e.g., heavy-tailed distributions; see Gilleland, 2008).



Practical Considerations

- Availability of CI methods
  - Normal CI is only available for appropriate statistics
  - Bootstrap is available for all/most statistical measures
- Bootstrap can be disabled in MET
- Number of points impacts speed of bootstrap
  - Grid-stat typically uses more points than Point-stat
  - <u>THUS</u>: Bootstrap is quicker with Point-stat.
- May be computationally inefficient to bootstrap over an entire field (e.g., several thousand points)
  - Alternative: Bootstrap the statistics for each field over time.
  - Measures (between-field) uncertainty of the estimates over time, rather than the within field uncertainty.

Practical Considerations

- Normal approximation intervals are quick, and generally accurate
  - Only valid for certain measures
- MET assumes that the samples are independent (probably not totally valid)

Thank you. Questions?

For more information, see:

Developmental Testbed Center, 2009. Model Evaluation Tools User's Guide. Available at: http://www.dtcenter.org/met/

Gilleland E, 2008. Confidence intervals for forecast verification. *Submitted* as an NCAR Technical Note. Available at: http://www.ral.ucar.edu/~ericg/Gilleland2008.pdf

Also look at basic statistical methods textbooks for information on confidence interval methods