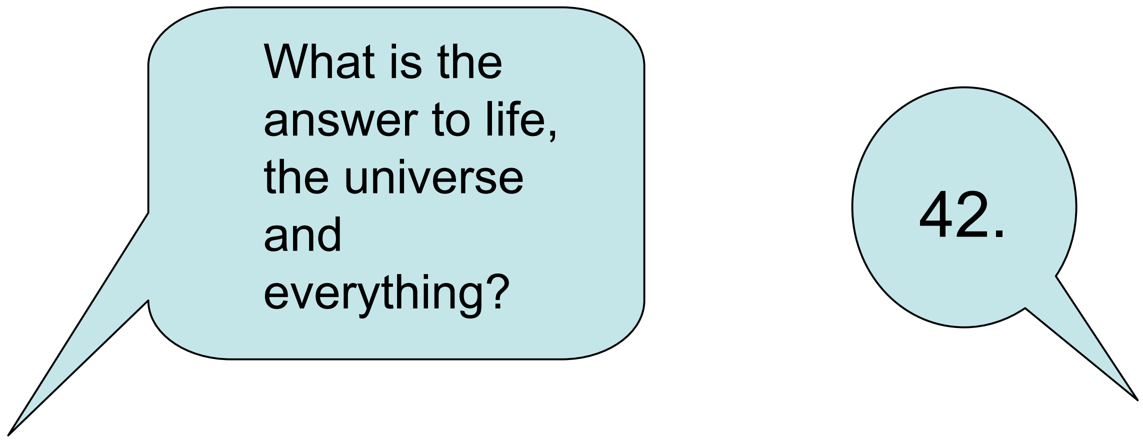
Giving meaning to your forecast verification results

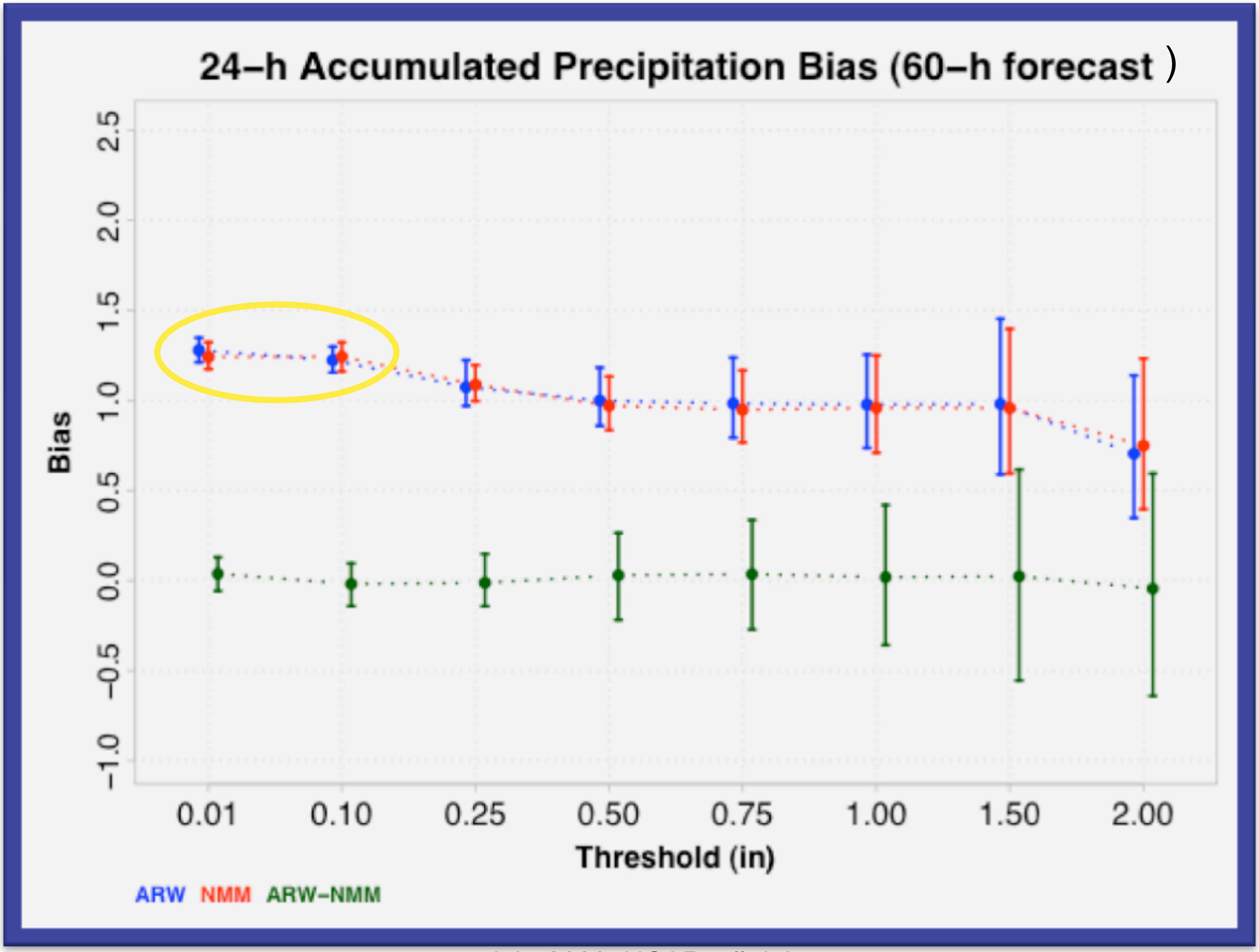


Eric Gilleland

Research Applications Laboratory,

National Center for Atmospheric Research

Boulder, Colorado Email: EricG @ ucar.edu



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., no bias), is there enough evidence to reject it?
  - One- or two-sided, but 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?
  - Interpretation for most *frequentist* intervals is a bit awkward.

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.
- 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

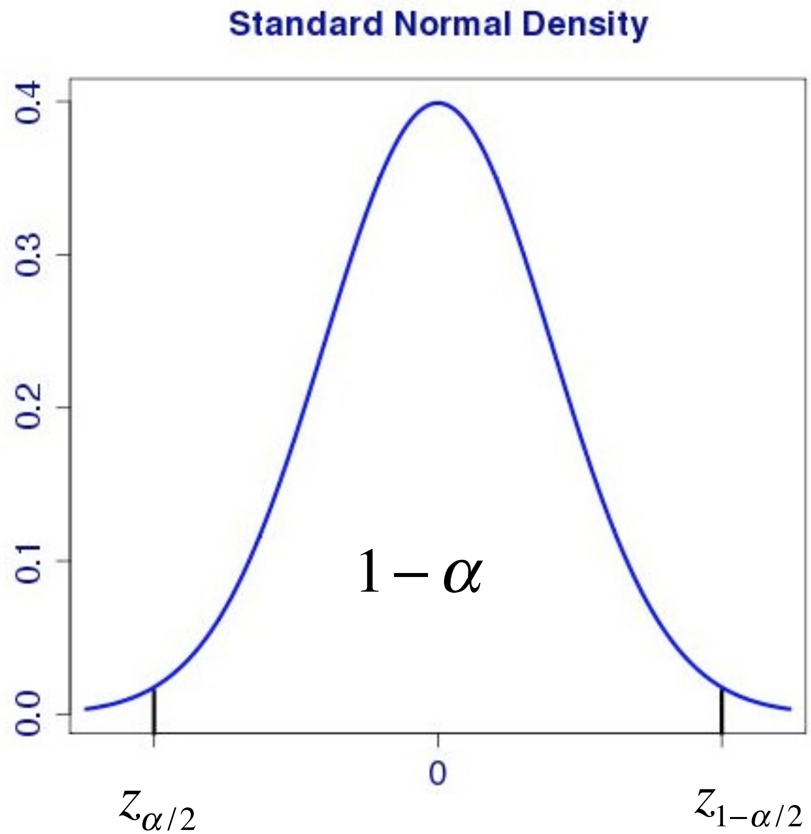
(1-α)100% Normal CI for Θ

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

- Θ is the statistic of interest

  (e.g., the forecast mean),
- $\hat{se}(\theta)$ is the (estimated) standard error for the statistic, Θ, and
- z_v is the v-th quantile of the standard normal distribution.

Normal Approximation CI's



$$\hat{\theta} \pm z_{\alpha/2} s\hat{e}(\theta)$$

Note: Normal distribution is symmetric so that

$z_{\alpha/2} = -z_{1-\alpha/2}$,

and therefore normal CI's are symmetric.

Normal Approximation CI's

Example: Let $X_1, \ldots, X_n$ be 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. And a (1-α)100% CI is given by

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

Normal Approximation CI’s

• Numerous verification statistics can take this approximation in some form or another. In other cases (e.g., forecast/observation variance, linear correlation), different parametric CI’s can be used that still rely on the underlying sample’s being iid normal.

• Contingency table verification scores such as probability of detection (POD) and false alarm ratio (FAR) also have normal approximation CI’s (for large enough sample sizes).

Normal Approximation CI's

- Check the validity of the
  - independence assumption.
    - (partial) autocorrelation function,
    - time series plots
    - Effect is to reduce "effective" sample size
      (i.e., increase variability)
  - normal assumption
    - qq-plots,
    - automatic tests exist too.

(cf. Gilleland, 2008)

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).
4. Estimate (1-α)100% CI's from the sample in step 3.

Bootstrap CI's

IID Bootstrap Algorithm

1. Resample *with replacement* from the sample,

$$X_1, \ldots, X_n$$

For example, suppose n=4, so we have $X_1, X_2, X_{3,} X_4$.

One replicate sample might be

$$X_1, X_1, X_4, X_4$$

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.

For example, if we are interested in the mean and variance, then for the replicate sample from the last slide, we would calculate $\bar{X} = \left(X_1 + X_1 + X_4 + X_4\right)/4$ and $\hat{\sigma}_X^2 = \sum_{i=1,1,4,4} (X_i - \bar{X})^2$

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).

B should be small enough for the algorithm to be as fast as possible, and large enough to get an accurate result.

In MET, B can be changed with the "n_boot_rep" argument. Setting n_boot_rep=0 turns bootstrapping off.

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

*1 and 2 are available In MET

Bootstrap CI's

Block Bootstrap Algorithm

- Block bootstrapping is one way to obtain bootstrap CI's for dependent samples.
- Same as IID Bootstrap, but resample blocks of contiguous data points.
- Use percentile method for CI's.
- Block sizes should be substantially larger than the correlation length, but substantially smaller than the sample size. Usually, one takes the greatest integer below the square root of the sample size.

Bootstrap CI's

Block Bootstrap Algorithm: Blocks

Non-overlapping (NBB)

$$Y_1 = \{X_1, \ldots, X_l\}, \ldots, Y_k = \{X_{(k-1)1+1}, \ldots, X_{kl}\}, \ldots, Y_b = \{X_{(b-1)1+1}, \ldots, X_n\}$$

Moving (MBB)

$$Y_i = \{X_i, \ldots, X_{i+l-1}\}$$

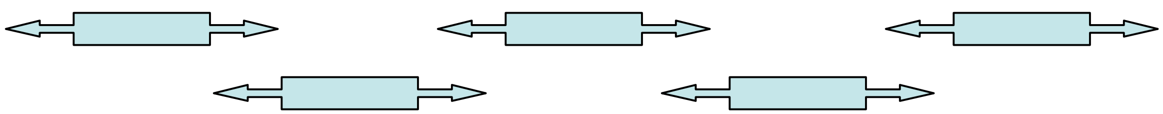
Circular (CBB)

Bootstrap CI's

Block Bootstrap Algorithm: Blocks

For example, with n=6, and l=2:

$$X_1, X_2, X_3, X_4, X_5, X_6$$



Bootstrap CI's

Parametric Bootstrap Algorithm:

- Same as IID Bootstrap algorithm, but first model the dependence in the data, then in step 1, take random samples from the model instead of the data.
- Generally preferable to block bootstrap.
- Stronger assumptions about the sample than block bootstrap.
- Use percentile method for CI's.

Bootstrap CI’s

Main MET bootstrap parameters to configure:

1. n_boot_rep

   A value of zero turns off bootstrapping

2. boot_interval

   0 = BCa

   1 = percentile

3. boot_rep_prop

   Default is 1 (i.e., m=n)

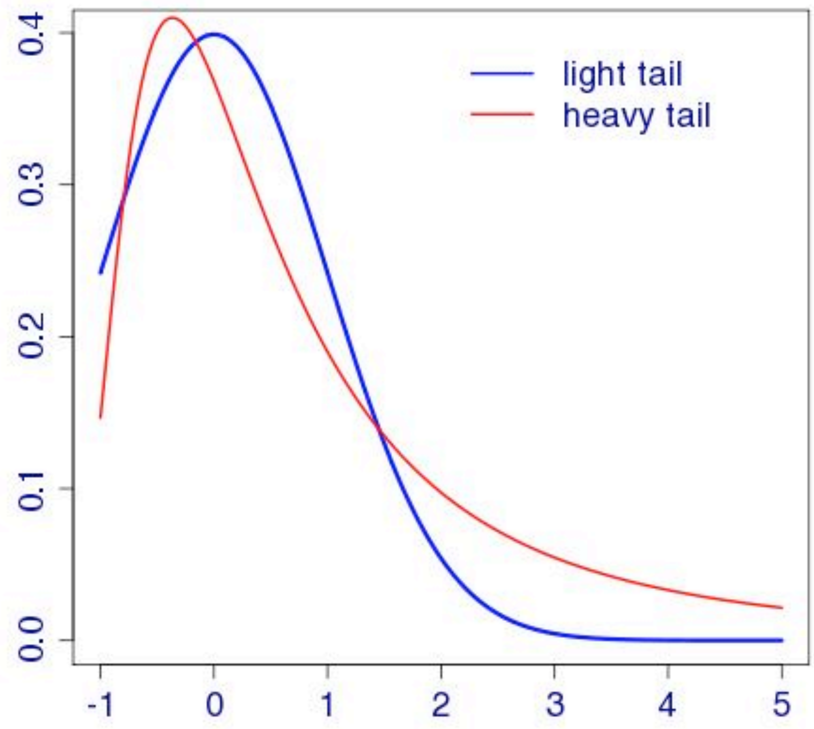
Bootstrap CI's

Main MET bootstrap parameters to configure:

boot_rep_prop

- Let m be the size of each replicate sample.
- Default for boot_rep_prop is 1 (i.e., m=n), which is best for most problems.
- m<n may be appropriate (e.g., if population distribution is heavy-tailed.

Bootstrap CI's



Practical Considerations

- Point-stat is quicker than Grid-stat, so bootstrap is quicker with Point-stat.
- May be prohibitively computationally inefficient to bootstrap over an entire field (i.e., over several thousand points), but can also bootstrap the statistics for each field over time. Measures the (between-field) uncertainty of the estimates over time, rather than the within field uncertainty.
- Normal approximation intervals are quick, and generally accurate. Check the normality assumption!
- Check whether the samples (for either type of interval) are independent or not.

Thank you. Questions?

For more information, see:

Developmental Testbed Center, 2008. Model Evaluation Tools Version 1.1 (METv1.1) 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