Glossary of Forecast Verification Metrics

Bias
The difference between the mean of the forecasts and the mean of the observations. Could be expressed as a percentage of the mean observation. Also known as overall bias, systematic bias, or unconditional bias. For categorical forecasts, bias (also known as frequency bias) is equal to the total number of events forecast divided by the total number of events observed. With the (2x2) **contingency table**, Bias = (a+b)/(a+c). Perfect score: 1.

Brier Score (BS)
The mean square error of probabilistic two-category forecasts where the observations are either 0 (no occurrence) or 1 (occurrence) and forecast probability may be arbitrarily distributed between occurrence and non-occurrence. BS=0 for perfect (deterministic) forecasts. BS=1 for forecasts that are always incorrect.

Brier Skill Score (BSS)
A **Skill Score** based on **BS** values. The most commonly used reference forecasts are persistence and climatology. Perfect score: 1.

Contingency Table
A two-dimensional table that gives the discrete joint distribution of forecasts and observations in terms of cell counts. For dichotomous categorical forecasts, having only two possible outcomes (Yes or No), the following (2x2) contingency table can be defined:

2x2 Contingency Table		Event Observed	
		Yes	No
Event Forecast	Yes	a (hits)	b (false alarms)
	No	c (misses)	d (correct negatives)

Continuous Ranked Probability Score (CRPS)
A measure of the integrated squared difference between the cumulative distribution function of the forecasts and the corresponding cumulative distribution function of the observations. It is an extension of the **Ranked Probability Score (RPS)** for continuous probability forecasts. Perfect score: 0.

Correlation Coefficient
A measure of the linear association between forecasts and observations independent of the mean and variance of the marginal distributions. Pearson Correlation Coefficient and Spearman Rank Correlation are the most widely used ones. Perfect score: 1.

Discrimination Diagram
A diagram plotting the conditional distributions of the forecasts. For binary events, this diagram plots the conditional distribution of the forecasts given that the event occurred, and the conditional distribution of the forecasts given that the even did not occur. Ideally, the two distributions are well separated from one another, becoming two distinct spikes for perfect forecasts.

False Alarm Ratio (FAR)
For categorical forecast, the number of false alarms divided by the total number of events forecast. A measure of reliability. With the (2x2) **contingency table**, FAR = b/(a+b). Not to be confused with the

Probability of False Detection (POFD) (also called **False Alarm Rate**) (which is conditioned on observations rather than forecasts). Range: 0 to 1. Perfect score: 1.

Lead Time of Detection (LTD)
The average lead time of forecasts that correspond to hits in the contingency table.

Mean Absolute Error (MAE)
The average of the absolute differences between forecasts and observations. A more robust measure of forecast accuracy than Mean Square Error that is sensitive to large outlier forecast errors. Perfect score: 0. Note: the overbar denotes the mean.

$$MAE = \overline{(|\ f - o\ |)}$$

Mean Error (ME)
The average difference between forecasts and observations. Note: it is possible to get a perfect score if there are compensating errors. Perfect score: 0.

$$ME = \overline{(f - o)}$$

Probability Of Detection (POD) (or Hit Rate)
For categorical forecast, the number of hits divided by the total number of events observed. A measure of discrimination. For the (2x2) **contingency table**, POD = a/(a+c). Range: 0 to 1. Perfect score: 1.

Probability Of False Detection (POFD) (or False Alarm Rate)
For categorical forecast, the number of false alarms divided by the total number of events observed. A measure of discrimination. For the (2x2) **contingency table**, POFD = b/(b+d). Not to be confused with the **False Alarm Ratio (FAR)** (which is conditioned on forecasts rather than observations). Range: 0 to 1. Perfect score: 0.

Root Mean Square Error (RMSE)
The square root of the average of the squared differences between forecasts and observations. It puts a greater influence on large errors than smaller errors, which may be good if large errors are especially undesirable, but may also encourage conservative forecasting. Perfect score: 0.

$$RMSE = \sqrt{\overline{(f - o)^2}}$$

Ranked Probability Score (RPS)
The mean square error of probabilistic multi-category forecasts where observations are 1 (occurrence) for the observed category and 0 for all other categories and forecast probability may be arbitrarily distributed between all categories. By using cumulative probabilities, it takes into account the ordering of the categories. For two category forecasts, the RPS is the same as **Brier Score**. Perfect score: 0.

Ranked Probability Skill Score (RPSS)
A **Skill Score** based on **RPS** values. The most commonly used reference forecasts are persistence and climatology. Perfect score: 1.

Rank Histogram (or Talagrand Diagram)
A plot of observed frequencies for k non-overlapping bins of equal probability for the forecast distribution. It measures how well the observed probability distribution is represented by the forecasts. For perfect forecasts, the rank histogram is flat since the observation is equally likely to fall between any two members. For U-shaped histogram, the ensemble spread is too small, most observations

falling outside the extremes of the ensemble. For dome-shaped histogram, the ensemble spread is too large, most observations falling near the center of the ensemble. For asymmetric histogram, the model has a bias to one side.

Relative (or Receiver) Operating Characteristic (ROC)
A signal detection curve for binary forecasts obtained by plotting **POD** (y-axis) versus **POFD** (x-axis) using a set of increasing thresholds to make the yes/no decision. The 45 degree diagonal line indicates no skill. It is conditioned on the observations (given that Y occurred, what was the corresponding forecast?). It is a good companion to the **Reliability Diagram**, which is conditioned on the forecasts. Perfect: curve travels from bottom left to top left of the diagram, then across to top right of the diagram.

Reliability Diagram
A diagram in which the frequency of the observations, given the forecast probability, is plotted against the forecast probability, where the range of forecast probabilities is divided in to K bins. The sample size in each bin is often included as a histogram or values beside the data points. Perfectly reliable forecasts have points that lie on the 45 degree diagonal line. The deviation from the diagonal line gives the conditional bias. The Reliability Diagram is called the Attributes Diagram when the no-resolution line and the no-skill line with reference to climatology are included. It is conditioned on the forecasts (given that X was predicted, what was the outcome?). It is a good partner of the **ROC**, which is conditioned on the observations.

Root Mean Square Error Skill Score (SS-RMSE)
A **Skill Score** based on **RMSE** values. The most commonly used reference forecasts are persistence and climatology.

Sample Size
A numeration of the number of forecasts involved in the calculation of a metric appropriate to the type of forecast (e.g., categorical forecasts should numerate forecasts and observations by categories, etc.)

Skill Score
A measure of the relative improvement of the forecast over some (usually 'low-skilled') benchmark forecast. Commonly used reference forecasts include climatology, persistence, or output from an earlier version of the forecasting system. In general, skill scores are the percentage difference between verification scores for two sets of forecasts (e.g., operational forecasts versus climatology). Perfect score: 1.

$$SS = 1 - \frac{Score(\ forecast\_set\_1)}{Score(\ forecast\_set\_2)}$$

Uncertainty
The degree of variability in the observations. Most simply measured by the variance of the observations. Important aspect in the performance of a forecasting system, over which the forecaster has no control.

On-line References

http://www.bom.gov.au/bmrc/wefor/staff/eee/verif/verif_web_page.shtml
http://www.sel.noaa.gov/forecast_verification/Glossary.html
http://www.met.rdg.ac.uk/cag/publications/Glossary.pdf