Forecast Verification Glossary

A

accuracy.

> The average degree of correspondence between individual pairs of forecasts and observations. It is frequently measured by the **mean square error**.

association.

> Overall strength of the linear relationship between the forecasts and observations. It is frequently measured by the **linear-correlation coefficient**.

B

base rate.

> The uncertainty in the occurrence of the observations. The base rate is determined by the **marginal distribution** of the observations, p(x).

bias.

> The degree of correspondence between the **mean** forecast (<f>) and the **mean** observation (<x>). This type of bias is also known as overall bias, systematic bias, or unconditional bias. The **mean error** is a measure of the overall forecast bias for **continuous** and **probabilistic forecasts**. A measure of bias for **categorical forecasts** is equal to the total number of event forecasts (**hits** + **false alarms**) divided by the total number of observed events. With respect to the 2x2 verification problem example outlined in the definition of **contingency table**, bias= (A+B)/(A+C).

bivariate histogram.

> A 3-dimensional diagram plotting the density function of the **joint distribution** of two variables. See **histogram**.

box plot.

A diagram plotting the **quantiles** of a probability distribution of a variable. The box contains the values between the upper **quartile** and lower quartile, or 50% of the distribution.

Brier score.

See **probability score**.

C

calibration.

See **reliability (in-the-small)**.

categorical forecast.

A forecast in which particular **predictand** values or events are forecast in an unqualified, categorical manner. Also called nonprobabilistic forecasts.

conditional bias 1.

The degree of correspondence between the conditional **mean** observation (<x>|f) and forecast (f), averaged over all forecasts. See **conditional distribution**.

conditional bias 2.

The degree of correspondence between the conditional **mean** forecast (<f>|x) and observation (x), averaged over all observations. See **conditional distribution**.

conditional distribution.

The probability distribution of a variable, given that a related variable is restricted to a certain value. The conditional distribution of the forecasts, given the observations, p(f|x), is related to

the **discrimination** or **likelihood**. The conditional distribution of the observations, given the forecasts, p(x|f), is related to the **calibration** or **reliability**.

conditional quantile plot.

A diagram plotting specific **quantiles** of a **conditional distribution** of a variable.

contingency table.

A two-dimensional "square" table (with *kxk* entries) that displays the discrete joint distribution of forecasts and observations in terms of frequencies or relative frequencies. 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	B
	No	C	D

The "A" table entry is the number of event forecasts that correspond to event observations, or the number of **hits**; entry "B" is the number of event forecasts that do not correspond to observed events, or the number of **false alarms**; entry "C" is the number of no-event forecasts corresponding to observed events, or the number of **misses**; and entry "D" is the number of no-event forecasts corresponding to no events observed, or the number of **correct rejections**. This 2x2 table will be referenced in the definitions of a number of performance measures formulated for the 2x2 verification problem. These measures include percent correct (**PC**), probability of detection (**POD**), false alarm ratio (**FAR**), success ratio (**SR**), threat score (**TS**) or critical success index (**CSI**), true skill statistic (**TSS**), Gilbert skill score (**GS**), Heidke skill score (**HSS**), and a categorical measure of **bias**.

continuous forecast.

A forecast of a continuous (or semi-continuous) variable. Continuous forecasts at SEC are integer value forecasts of a continuous parameter (e.g., A-Index or 10.7 cm Solar Flux).

correct random forecast.

In a categorical verification problem, a correct forecast (either a hit or correct rejection) due purely to random chance. From the marginal probabilities and with respect to the 2x2 **contingency table**, the probability of a correct forecast by chance, or **hit due to chance** is $(A+B)(A+C)/n^2$, and the probability of a **correct rejection** (correct no forecast) by chance is $(B+D)(C+D)/n^2$, where n= A+B+C+D. The expected number of correct random forecasts (hits and correct rejections) is derived from the marginal sums as $[(A+B)(A+C)+(B+D)(C+D)]/n$

correct rejection.

In a categorical verification problem, a no-event forecast that is associated with no event observed, and is also sometimes called a correct null. In rare event forecasts the number of correct rejections dominates the 2x2 **contingency table**.

covariance.

The sample covariance is a measure of the relationship between the forecasts and observations and is defined as the average of the products of the deviations of each forecast/observation pair from their respective **mean**:

$$s_{fx} = \frac{1}{N} \sum_{i=1}^{N} \left[\left(f_i - \bar{f} \right) \left(x_i - \bar{x} \right) \right]$$

,

where N is the sample size, f refers to forecasts, and x refers to observations.

critical success index (CSI).

Also called the threat score (TS), is a **verification measure** of **categorical forecast** performance equal to the total number of correct event forecasts (**hits**) divided by the total number of storm forecasts plus the number of misses (**hits** + **false alarms** + **misses**). The CSI is not affected by the number of non-event forecasts that verify (**correct rejections**). However, the CSI is a biased score that is dependent upon the frequency of the event. For an unbiased version of the CSI, see the **Gilbert skill score** (GS). With respect to the 2x2 verification problem example outlined in the definition of **contingency table**, CSI= (A)/(A+B+C).

cumulative forecast.

Probability forecast of an **ordinal predictand** expressed as a cumulative probability distribution corresponding to the original discrete (or continuous) forecast probability distribution. For example, consider the **probability forecast** (treated as a vector) of an N=4 (four event-state) **ordinal predictand**, [0.1, 0.4, 0.49, 0.01]. The corresponding cumulative forecast is [0.1, 0.5, 0.99, 1.0].

cumulative observation.

The cumulative distribution of an **ordinal predictand** corresponding to the original discrete (or continuous) distribution. For example, consider the observation (treated as a vector) of an N=4 (four event-state) **ordinal predictand**, [0, 1, 0, 0]. The corresponding cumulative observation is [0, 1, 1, 1].

D

diagnostic verification.

An approach to **forecast verification** in which attention is focused on determining the basic strengths and weaknesses in the forecasts.

discrimination.

The extent to which the relative frequency of use of the forecasts differs given different observations. It is measured in terms of likelihoods (i.e., p(f|x)) and is the degree to which the conditional mean forecast (<f>|x) differs from the unconditional mean forecast (<f>), averaged over all observations.

discrimination diagram.

A diagram plotting the **conditional distributions** of the forecasts. With respect to 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 event did not occur. Ideally, the two distributions for this example are well separated from one another, becoming two distinct spikes for perfect forecasts.

F

false alarm.

In a categorical verification problem, an event forecast that is associated with no event observed. See **contingency table**.

false alarm ratio (FAR).

A **verification measure** of **categorical forecast** performance equal to the number of **false alarms** divided by the total number of event forecasts. With respect to the 2x2 verification problem example outlined in the definition of **contingency table**, FAR= (B)/(A+B).

forecast verification.

The process of determining the quality of forecasts. The assessment of forecast quality involves the statistical characteristics of the forecasts and matching observations, and the relationships between them. Verification methods are numerous, and the approach used to verify a particular forecast/observation data set is determined by, among other things, the forecast type and the objectives of the particular verification user.

G

Gilbert skill score (GS).

A skill corrected **verification measure** of **categorical forecast** performance similar to the **critical success index** (CSI) but which takes into account the number of hits due to chance. **Hits due to chance** is given as the event frequency multiplied by the number of event forecasts. The GS is equal to the total number of correct event forecasts minus the hits due to chance (**hits** - chance hits) divided by the total number of storm forecasts plus the number of misses minus the hits due to chance (**hits** + **false alarms** + **misses** - **chance hits**). With respect to the 2x2 verification problem example outlined in the definition of **contingency table**, GS= (A-CH)/(A+B+C-CH), where CH= **chance hits**.

H

Heidke skill score (HSS).

> A skill corrected **verification measure** of **categorical forecast** performance similar to the **success ratio** (SR) but which takes into account the number of **correct random forecasts** (**chance hits** + chance correct rejections). The HSS is equal to the total number of correct forecasts minus the correct random forecasts (**hits** + **correct rejections** - **correct random forecasts**) divided by the total number of forecasts minus the correct forecasts due to chance (**hits** + **false alarms** + **misses** + **correct rejections** - **correct random forecasts**). With respect to the 2x2 verification problem example outlined in the definition of **contingency table**, HSS= (A+D-E)/(A+B+C+D-E), where E= **correct random forecasts**. This skill score falls within a (-1, +1) range. No incorrect forecasts give a score of +1, no correct forecasts give a score of -1, and either no events forecast or no events observed give a score of 0. For rare event forecasts, the HSS in the limiting case approaches 2A/(2A+B+C), a simple function of the Critical Success Index (**CSI**).

histogram.

> A diagram plotting the density function of the distribution of a variable (or variables). A 2-dimensional histogram is a diagram plotting the **marginal distribution** of a variable in terms of its frequency of occurrence. A **bivariate histogram** is a 3-dimensional diagram plotting the density function of the **joint distribution** of two variables. Higher dimensional histograms are similarly generalized.

hit.

> In a categorical verification problem, an event forecast that is associated with an event observed. See **contingency table**.

hit due to chance.

> In a categorical verification problem, a correct yes forecast (hit) due purely to random chance. Derived from the marginal probabilities and with respect to the 2x2 **contingency table**, the probability of a correct forecast (hit) by chance is $(A+B)(A+C)/n^2$, where n is the total sample size (A+B+C+D). The expected number of chance hits (CH) is given by the event frequency multiplied by the number of event forecasts, CH= (A+B)(A+C)/n.

J

joint distribution.

The probability distribution defined over two or more variables. The joint distribution of the forecasts and observations, p(f,x), contains all of the information relevant to the verification problem (except for time relationships). This joint distribution can be decomposed into expressions involving **conditional** and **marginal** distributions:

p(f,x) = p(x|f) p(f), the **calibration** - **refinement** factorization and

p(f,x) = p(f|x) p(x), the **likelihood** - **base rate** factorization.

L

likelihood.

See **discrimination**.

linear-correlation coefficient (r).

A measure of the linear **association** between the forecasts and observations. The sample linear-correlation coefficient is defined as

$$r_{fx} = \frac{s_{fx}}{s_f s_x},$$

where s_{fx} is the sample **covariance**, s_f and s_x are sample **standard deviations**, f refers to forecasts, and x refers to observations.

M

marginal distribution.

A probability distribution of a single variable. The marginal distribution of the forecasts, p(f), is related to the **refinement** or **sharpness**. The marginal distribution of the observations, p(x), is related to the uncertainty or **base rate**.

mean.

The sample mean, arithmetic mean, or average, is defined as

$$\bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i$$,

where x is the variable and N is the sample size. The mean of a variable is also indicated by angled brackets, e.g. <x>.

mean error (ME).

The **mean** of the differences of the forecasts and observations. In this case, the variable in the definition of **mean** is (f-x), where f refers to forecasts and x refers to observations. The mean error is a measure of the overall **bias** of the forecasts. A positive mean error indicates that, on the average, the forecasts were larger than the corresponding observations.

mean square error (MSE).

The **mean** of the square of the differences of the forecasts and observations. For mean square error, the variable in the definition of **mean** is $(f-x)^2$, where f refers to forecasts and x refers to observations. The mean square error is a measure of forecast **accuracy**. The lower the mean square error, the more accurate the forecasts.

median.

The value of a variable that divides a sample in half. There is equal probability that the value of a variable will be greater or smaller than the sample median. The median of variable x is abbreviated as Med(x).

miss.

In a categorical verification problem, a no-event forecast that is associated with an event observed. See **contingency table**.

N

nominal predictand.

A set of two or more mutually exclusive and collectively exhaustive (discrete) events or states. These events may be ordered or unordered but are discrete variables by definition. Note that it is possible (and sometimes desirable) to treat discrete variables as nominal predictands for some purposes and **ordinal predictands** for others.

normal distribution.

The mathematical function that describes the symmetric bell-shaped curve defined by the Gaussian:

$$G_{X, \sigma} = \frac{1}{\sigma\sqrt{2\pi}} e^{(x-X)^2 / 2\sigma^2} ,$$

where X and sigma determine the center and width of the distribution, respectively. Measurements whose limiting distribution is given by this function are said to be *normally distributed*.

O

ordinal predictand.

A set of two or more mutually exclusive and collectively exhaustive ordered events or states. Ordinal predictands can be either continuous or discrete variables. Note that it is possible (and sometimes desirable) to treat discrete variables as ordinal predictands for some purposes and **nominal predictands** for others.

P

percent correct (PC).

A **verification measure** of **categorical forecast** performance equal to the total number of correct forecasts (**hits** + **correct rejections**) divided by the total number of forecasts. Simply stated, it is the percent of correct forecasts. With respect to the 2x2 verification problem example outlined in the definition of **contingency table**, PC= (A+D)/(A+B+C+D).

predictand.

The variable or quantity for which the forecasts are formulated, usually defined in terms of a set of numerical values or a set of two or more events. Predictands can be **nominal** or **ordinal**.

prediction efficiency.

The same as the **skill score** with respect to climatology. The prediction efficiency gives a measure of how well on average the forecasts can account for the variance of the observations.

probability forecast.

A forecast that specifies the likelihood of occurrence of a mutually exclusive and collectively exhaustive set of events (see **nominal** and **ordinal predictand**) during a given time frame. Probabilistic forecasts range from 0 (event cannot occur) to 1.0 (event is certain to occur). The probability forecast range at SEC is limited from 0.01 to 0.99.

probability of detection (POD).

A **verification measure** of **categorical forecast** performance equal to the total number of correct event forecasts (**hits**) divided by the total number of events observed. Simply stated, it is the percent of events that are forecast. With respect to the 2x2 verification problem example outlined in the definition of **contingency table**, POD= (A)/(A+C).

probability score.

Or Brier score (BS), is the **mean square error** of **probability forecasts**. The Brier score is defined as

$$\blacksquare,$$

where K is the number of forecasts in the sample, r refers to the forecast vector, and d refers to the observation vector. The range of the probability score is the closed interval [0,2] and has a negative orientation; that is, smaller values are better.

Q

quantile.

The specific value of a variable that divides the distribution into two parts, those values greater than the quantile value and those values that are less. For instance, p percent of the values are less than the pth quantile.

quartile.

A **quantile** that separates one quarter of the values in a distribution from the remaining three quarters.

R

ranked probability score (RPS).

The **mean square error** of **probability forecasts** for **ordinal predictands**. It is the difference between **cumulative forecasts** and **observations**. RPS is defined as

$$RPS = \left(1/K\right)\sum_{k=1}^{K}\left(\mathbf{R}_k - \mathbf{D}_k\right)\left(\mathbf{R}_k - \mathbf{D}_k\right)^{T},$$

where K is the number of forecasts in the sample, and where R and D refer to the **cumulative forecast** and **observation** vectors, respectively. Two useful vector partitions of the RPS are

$$RPS = RPS_{sc} + REL - RES$$
$$= REL + RES'$$
,

where REL is the **reliability** and RES and RES' are the **resolution** as defined by Murphy and Sanders, respectively. The RPS$_{sc}$ is the ranked probability score with respect to the **sample climatology** as a forecast and is defined as

$$RPS_{sc} = S_x^2$$
$$= RES + RES'$$
,

where S$_x^2$ is the [variance](#) of the observations over the sample.

refinement.

Or **sharpness**, only applies to **probability forecasts** and is the extent to which an individual forecast differs from the overall average forecast. It is related to the forecast **marginal distribution**, p(f), and can also be thought of as the degree to which probability forecasts approach categorical forecasts of 0 or 1.

reliability (in-the-small) (REL).

The same as **calibration**, it is the degree of correspondence, over one or more subsamples of verification data involving identical forecasts, between the average observations for the subsamples and the respective forecasts. For J subsamples of K **probability forecasts**, the reliability is defined as

$$REL = (1/K) \sum_{J=1}^{J} K^J \left(\mathbf{r}^J - \overline{\mathbf{d}}^J \right) \left(\mathbf{r}^J - \overline{\mathbf{d}}^J \right)^T$$
,

where the overbars denote the **mean**, and r and d refer to the forecast and observation vectors, respectively. Note that REL > 0 and has a negative orientation; that is, smaller values indicate greater reliability.

reliability diagram.

A diagram in which the **conditional distribution** of the observations, given the forecast probability, is plotted against the forecast probability. The distribution for perfectly reliable forecasts will plot along the 45 degree diagonal.

resolution (RES or RES').

The degree to which the conditional **mean** observation (<x>|f) differs from the unconditional **mean** observation (<x>), averaged over all forecasts. See **conditional distribution**. For **probability forecasts**, resolution relates to the correspondence of occurrence of events in subsamples of observations associated with distinct forecasts. For J subsamples of K **probability forecasts**, a definition of resolution is given by Murphy as

$$RES = (1/K) \sum_{j=1}^{J} K^j \left(\overline{\mathbf{d}}^j - \overline{\mathbf{d}} \right) \left(\overline{\mathbf{d}}^j - \overline{\mathbf{d}} \right)^T$$

,

where the overbars indicate the **mean** and d is the observation vector. Another definition of resolution by Sanders, is given as

$$RES' = (1/K) \sum_{j=1}^{J} K^j \overline{\mathbf{d}}^j \left(1 - \overline{\mathbf{d}}^j \right)^T$$

,

where 1 is the identity vector. Note that both RES and RES' are greater than or equal to zero; however, larger values of RES indicate more resolution whereas smaller values of RES' indicate more resolution.

robust measure.

A **verification measure** whose values are not unduly influenced by the fact that the variables on which these values are based are not **normally distributed**.

root mean square error (RMSE).

The square root of the **mean square error**.

S

sample climatology.

An assertion of the prevailing (characteristic) observed conditions over a forecast verification data set. A convenient measure of climatology is the **mean** of the observations in the data set. Sample climatology is commonly used as a reference forecast.

sharpness.

See **refinement**.

skill.

The average **accuracy** of the forecasts in the sample relative to the **accuracy** of forecasts produced by a reference method. Examples of a suitable reference include forecasts of recurrence, persistence, **sample climatology**, or the output of a forecast model. Skill can be measured by any number of so-called skill scores. When based on **mean square error**, the skill score is defined as

$$ SS = 1 - \left(\frac{MSE}{MSE_r} \right), $$

where MSE is the **mean square error** of the sample forecasts and MSE_r is that of the reference forecasts. For instance, the skill score with respect to **sample climatology** (SS_{sc}) uses the **mean** observation over the sample period as the reference forecast. For **probabilistic forecasts**, where the observations are either 0 or 1, the sample climatology skill score can be written as

$$ SS_{sc} = 1 - \left(\frac{MSE}{MSE_{sc}} \right) $$

$$ = 1 - \left(\frac{MSE}{s_x^2} \right), $$

where s_x^2 is the sample **variance** of the observations. This type of skill score is also called the **prediction efficiency**. In addition, skill scores based on the MSE can be decomposed as follows:

$$ SS = r_{fx}^{2} - \left[r_{fx} - \left(\frac{s_f}{s_x} \right) \right]^{2} - \left[\frac{\left(\bar{f} - \bar{x} \right)}{s_x} \right]^{2} , $$

where r_{fx} is the **linear-correlation coefficient**, s is the sample **standard deviation**, f refers to forecasts, and x refers to observations. The second term in this decomposition is related to the conditional **bias** of the forecasts and the third term is related to the unconditional **bias**.

standard deviation.

The square root of the **variance**.

statistical characteristic.

A property of the **joint**, **conditional**, or **marginal distribution** of forecasts and observations or of the correspondence between forecasts and observations. Also called attribute.

success ratio (SR).

A **verification measure** of **categorical forecast** performance equal to the number of **hits** divided by the total number of event forecasts (**hits** + **false alarms**). With respect to the 2x2 verification problem example outlined in the definition of **contingency table**, SR= (A)/(A+B). Also, note that the success ratio is equal to one minus the **false alarm ratio** (SR= 1-FAR).

T

threat score (TS).

Same as **critical success index** (CSI).

true skill statistic (TSS).

Also called the Hanssen_Kuipers discriminant or Kuipers' performance index, is a **verification measure** of **categorical forecast** performance formulated similarly to the Heidke skill score (**HSS**). This skill score incorporates random correct forecasts that are constrained to be unbiased. The random reference forecasts in the denominator have a marginal distribution equal to the sample climatology. With respect to the 2x2 verification problem example outlined in the definition of **contingency table**, TSS= (AD-BC)/(A+C)(B+D). Perfect forecasts receive a score of +1, random forecasts receive a score of 0 and forecasts inferior to random forecasts receive a negative score. The TSS has some desirable characteristics for evaluating rare event forecasts. For instance both random forecasts and constant forecasts receive a TSS score of 0. In addition the contribution to this score for a correct no forecast (**correct rejection**) increases as the event becomes more likely, while the contribution for a correct yes forecast (**hit**) increases as the event becomes less likely. For extremely rare events, as the number of **correct rejections** approaches infinity, the TSS approaches the Probability of Detection (**POD**).

U

uncertainty.

The degree of variability in the observations.

V

variance.

A measure of the dispersion of a data set. The sample variance is defined as

$$s^2 = \frac{1}{N-1}\sum\left(x_i - \bar{x}\right)^2$$

,

where x is the variable and N is the sample size. The sample standard deviation is the square root of the sample variance.

verification measure.

A mathematical function that describes a **statistical characteristic** of forecasts and/or observations.