18.4.3.8 Estimation

Much of the foregoing discussion treats Performance Characteristics as though they were known without error. In fact, apart from definition, this can never obtain. Let us consider the significance (not just statistical) of this limitation for four of the more important CMP characteristics: bias, imprecision (variance, standard deviation), sensitivity, and the blank.

Estimated Bias ($\hat{\Delta}$ or $\tilde{\Delta}$). Estimation of CMP characteristics such as bias and imprecision carries two dichotomies: (1) statistical estimation [circumflex] vs "scientific" (judgment) estimation [tilde]; and (2) "internal" estimation, via propagated contributions of each constituent step of the CMP vs "external" estimation via intercomparison of the overall CMP with an appropriate external standard (or laboratory, or definitive method). In the case of bias, it would seem unlikely that the CMP would be even considered for use if the internally estimated bias were non-negligible. An external bias estimate (statistical) could be formed *ex post facto*, however, during the evaluation of a CMP in comparison to a known standard. A statistically and practically significant bias estimate generally would lead either to rejection of the CMP altogether, or exposure and correction of the source(s) of bias.

Two matters concerning CMP bias are worth noting: (1) The detection limit for bias is intimately tied to the imprecision of the measurement process; bias much smaller than the repeatability-σ is quite difficult to detect. (2) "Correction" or adjustment of bias of a complex CMP based on an observed discrepancy with a natural matrix CRM can be a very tenuous process, unless or until the *cause* of the discrepancy is thoroughly understood.

Bias Uncertainty; Bias Bounds (Δ_M). More commonly, our concern is with the maximum (absolute value) uncorrected bias. Such a quantity is derived from the (scientifically or statistically) estimated bias together with the uncertainty of that estimate. If a statistical estimate is involved, and if one knows the *cdf* and its parameters(s), one can form a confidence interval and upper limit, just as in the case of analytical results.

Estimated Variance ($s^2 = \hat{\sigma}^2 = \hat{V}$). Variance is estimated by the sum of squares of the residuals (deviations of the observed from the estimated or "fitted" values) divided by the *number of degrees of freedom* ν, which equals the number of observations n minus the number of estimated parameters. Thus, for a simple set of observations

$$s^2 = \Sigma(x_i - \bar{x})^2/(n-1) \qquad (18.4.23)$$

where $\bar{x}$ = the estimated (arithmetic) mean.

For a fitted (straight-line) calibration curve,

$$s^2 = \sum (y_i - \hat{y}_i)^2 / (n-2) \qquad (18.4.24)$$

Note that although the standard deviation equals the square root of the variance σ^2, the square root of the *estimated* variance s^2 yields a biased estimate for the standard deviation. An approximate correction is given by multiplying s by $[1+1/(4\nu)]$.

Propagation of "Error" (Variance). An "internal" estimate for the overall variance of a CMP can be constructed from the variances of the contributing elements or steps of the CMP and the functional manner in which they are linked. If the individual *cdf*s are normal and the links are additive (or subtractive), normality is preserved in the overall process. An illustration is the subtraction of an estimated blank from the observed response to get the net signal; in this case variances add. If the parameters for the individual steps are linked multiplicatively, as in the correction of the net signal for the estimated chemical yield, relative variances add. (In this case, normality is only asymptotic, as the *relative* variances become sufficiently small.)

More complicated relations can be treated with the Taylor expansion, suitably adapted for variances:

$$\sigma_f^2 = \Sigma(\partial f/\partial x_j)^2 \, \sigma_{xj}^2 \qquad (18.4.25)$$

where f is the function whose variance is to be determined, and
the x_j are the individual parameters whose variances are known.

Estimated Poisson Variance ("counting statistics"). For counting experiments, if there are no extraneous sources of variance, the distribution of counts is Poisson; hence the variance σ^2 equals the mean μ. Except for the case of relatively few counts, using the observed number of counts as an estimate of the variance is quite adequate.

Estimated Variance Bounds. If the observations are distributed normally, s^2/σ^2 is distributed as χ^2/ν. A 95% interval estimate for this ratio is therefore given by

$$(\chi^2/\nu)_{.025} < s^2/\sigma^2 < (\chi^2/\nu)_{.975} \qquad (18.4.26)$$

A useful approximation for rapidly estimating the uncertainty in s/σ is $1/\sqrt{2\nu}$. This is roughly equivalent to the standard deviation of the ratio s/σ for large ν. Thus, about 200 degrees of freedom are required before the relative standard uncertainty in σ is decreased to about 5 %.

Note: Eq. 18.4.26 can be used to derive approximate confidence intervals for the relative standard deviation (RSD), given the observed ratio $s/\bar{x}$, without taking into account the uncertainty of $\bar{x}$; the approximation improves with increasing degrees of freedom, and decreasing RSD. This has special relevance for the Quantification Limit, since the definition of L_Q is based on a prescribed value for the RSD.

Estimated Sensitivity ($\hat{A}$). The slope (sensitivity) and intercept of the calibration curve are generally estimated using Ordinary Least Squares. Weighted Least Squares may be justified if at least the *relative* statistical weights are reliably known (or can be assumed), where the weights are taken as inverse variances. Although the intercept of an instrument calibration curve may give some useful information on the magnitude of the blank, for low-level measurements that may be severely affected by contamination, it is advisable to make direct estimates of the components of the blank and their variability.

Note: When a *functional*, as opposed to a *statistical* (structural) relation exists between variables -- as in the case of a calibration curve -- the terms "Regression" and "Correlation" are inappropriate. The quality of the fit should be assessed by appropriate test statistics, such as F, χ^2, the MSSD (Mean Squared Successive Deviation), etc. In some cases, where the individual data are quite precise, such test statistics can show a "fit" to be very poor, even though the linear correlation coefficient is almost unity. A related situation where Correlation is appropriately used is for the (statistical) relation between parameters (slope, intercept) estimated from the same data set. This statistical relation is commonly displayed in the form of a *confidence ellipse*.

The Blank (B). The blank is one of the most crucial quantities in trace analysis, especially in the region of the Detection Limit. In fact, as shown above, the distribution and standard deviation of the blank are intrinsic to calculating the Detection Limit of any CMP. Standard deviations are difficult to estimate with any precision (ca. 50 observations required for 10 % RSD for the Standard Deviation). Distributions (*cdf*s) are harder! It follows that extreme care must be given to the minimization and estimation of realistic blanks for the over-all CMP, and that an adequate number of full scale blanks must be assayed, to generate some confidence in the nature of the blank distribution and some precision in the blank RSD.

Note: An imprecise estimate for the Blank standard deviation is taken into account without difficulty in Detection Decisions, through the use of Student's-t. Detection Limits, however, are themselves rendered imprecise if σ_B is not well known. (See section 18.4.3.7.)

Blanks or null effects may be described by three different terms, depending upon their origin: the *instrumental background* is the null signal (which for certain instruments

may be set to zero, on the average) obtained in the absence of any analyte- or interference-derived signal; the (spectrum or chromatogram) *baseline* comprises the summation of the instrumental background plus signals in the analyte (peak) region of interest due to interfering species; the *chemical (or analyte) blank* is that which arises from contamination from the reagents, sampling procedure, or isolation of the analyte steps which corresponds to the very analyte being sought. Assessment of the blank (and its variability) may be approached by an "external" or "internal" route, in close analogy to the assessment of random and systematic error components. The "external" approach consists of the generation and direct evaluation of a series of ideal or surrogate blanks for the overall measurement process, using test samples which are identical or closely similar to those being taken for analysis -- but containing none of the analyte of interest. The CMP and matrix and interfering species should be unchanged. (The surrogate is simply the best available approximation to the ideal blank -- ie, one having a similar matrix and similar levels of interferants.) The "internal" approach has been described as "Propagation of the Blank." This means that each step of the CMP is examined with respect to contamination and interference contributions, and the whole is then estimated as the sum of its parts -- with due attention to differential recoveries and variance propagation. This is an important point: that the blank introduced at one stage of the CMP will be attenuated by subsequent partial recoveries. Neither the internal nor the external approach to blank assessment is easy, but one or the other is mandatory for accurate low-level measurements; and consistency (internal, external) is requisite for quality. Both approaches require expert chemical knowledge concerning the CMP in question.