

INTERNATIONAL UNION OF PURE  
AND APPLIED CHEMISTRY

ANALYTICAL CHEMISTRY DIVISION  
COMMISSION ON ANALYTICAL NOMENCLATURE\*†

**NOMENCLATURE IN EVALUATION OF  
ANALYTICAL METHODS INCLUDING DETECTION  
AND QUANTIFICATION CAPABILITIES**

(IUPAC Recommendations 1995)

*Prepared for publication by*

LLOYD A. CURRIE

Chemical Science and Technology Laboratory  
National Institute of Standards and Technology, Gaithersburg, MD 20899, USA

\*Membership of the Commission during the period (1983–85) when this report was initiated was as follows:

*Chairman:* G. Svehla (UK), *Secretary:* S. P. Perone (USA). *Titular Members:* C. A. M. G. Cramers (Netherlands), R. W. Frei (Netherlands), R. E. van Grieken (Belgium), D. Klockow (FRG). *Associate Members:* L. Currie (USA), L. S. Ettre (USA), A. Fein (USA), H. Freiser (USA), W. Horwitz (USA), M. A. Leonard (UK), D. Leyden (USA), R. F. Martin (USA), B. Schreiber (Switzerland). *National Representatives:* K. Doerffel (GDR), I. Giolito (Brazil), E. Grushka (Israel), W. E. Harris (Canada), H. M. N. H. Irving (S. African Republic), D. Jagner (Sweden), W. Rossett (France), J. Stary (Czechoslovakia).

†The Commission was disbanded in 1989 and this project was transferred to the new COMMISSION ON GENERAL ASPECTS OF ANALYTICAL CHEMISTRY.

---

*Republication of this report is permitted without the need for formal IUPAC permission on condition that an acknowledgement, with full reference together with IUPAC copyright symbol (© 1995 IUPAC), is printed. Publication of a translation into another language is subject to the additional condition of prior approval from the relevant IUPAC National Adhering Organization.*

# Nomenclature in evaluation of analytical methods, including detection and quantification capabilities<sup>1</sup> (IUPAC Recommendations 1995)

## Synopsis

This IUPAC nomenclature document has been prepared to help establish a uniform and meaningful approach to terminology, notation, and formulation for performance characteristics of the Chemical Measurement Process (CMP). Following definition of the CMP and its Performance Characteristics, the document addresses fundamental quantities related to the observed response and calibration, and the complement to the calibration function: the evaluation function. Performance characteristics related to precision and accuracy comprise the heart of the document. These include measures for the means or "expected values" of the relevant chemical quantities, as well as dispersion measures such as variance and standard error. Attention is given also to important issues involving: assumptions, internal and external quality control, estimation, error propagation and uncertainty components, and bounds for systematic error. Special treatment is given to terminology and concepts underlying detection and quantification capabilities in chemical metrology, and the significance of the blank. The document concludes with a note on the important distinction between the Sampled Population and the Target Population, especially in relation to the interlaboratory environment.

## CONTENTS

1. Objective and introduction
  2. Analytical techniques, methods, and the measurement process
  3. Performance characteristics of the chemical measurement process
    - 3.1 Structure of the CMP
    - 3.2 The observed signal; calibration function
    - 3.3 Evaluation function
    - 3.4 Performance characteristics not specifically related to precision and accuracy
    - 3.5 Precision and accuracy – related performance characteristics
    - 3.6 Control and testing of assumptions
    - 3.7 Detection and quantification capabilities
    - 3.8 Estimation
  4. Compound CMPs -- the interlaboratory environment
- References  
Index of terms

## 1. OBJECTIVE AND INTRODUCTION

Effective communication among analytical scientists requires a consistent and uniform system of nomenclature and convention for specifying the Performance Characteristics of the Chemical Measurement Process (CMP) which, following Eisenhart [1], we take to be a fully specified analytical method that has achieved a state of statistical control. This measurement process, which may include substructure such as sample preparation and instrumental sensing, lies between the other two components of the overall analytical system, namely Sampling [2] and the Presentation of Results [3]. Central to all three of these tasks are the issues of precision and accuracy. For this reason we give special attention

---

<sup>1</sup>It is recognized that "Terminology" might be a more appropriate descriptor for the subject of this document, but the term "Nomenclature" is being retained in the title because of the links with the "Orange Book" (*Compendium on Analytical Nomenclature*) and previous documents in this series, that were originated in the IUPAC Commission on Analytical Nomenclature.

to these statistical quantities in our discussion of Performance Characteristics of the CMP [4], in order to help prevent arbitrary and inappropriate usage of terminology among the three areas -- e.g., "uncertainty" vs "inaccuracy" (See section 3.5.11.)

Insofar as possible we shall conform to accepted statistical terminology and notation, even though this may occasionally lead to suggested changes from notation long popular with analytical chemists. A special effort will be made to distinguish between true (or asymptotic) values of parameters and observed or "estimated" values which necessarily exhibit the perturbations of random error. Also, the nature and validity of assumptions (such as normality) will be emphasized; and an effort to minimize information loss will be made by discouraging the use of ambiguous terms or incomplete reporting of data.

**1.1 Measurable Quantity.** The *International Vocabulary of Basic and General Terms in Metrology* defines the measurable quantity as "an attribute of a ... substance which may be distinguished qualitatively and determined quantitatively" [5]. In the context of Analytical Chemistry, the attribute may refer to a physical quantity such as X- or  $\gamma$ -ray energy, or it may refer to a measure of amount such as mass or concentration.

The general expression Qualitative Analysis thus refers to analyses in which substances are identified or classified on the basis of their chemical or physical properties, such as chemical reactivity, solubility, molecular weight, melting point, radiative properties (emission, absorption), mass spectra, nuclear half-life, etc. Quantitative Analysis refers to analyses in which the amount or concentration of an analyte may be determined (estimated) and expressed as a numerical value in appropriate units. Qualitative Analysis may take place without Quantitative Analysis, but Quantitative Analysis requires the identification (qualification) of the analytes for which numerical estimates are given.

**1.2 Analyte; Measurand.** These terms, as well as the analog "determinand" are employed in Analytical Chemistry to indicate the chemical entity involved. The preferred term for Analytical Chemistry is Analyte, defined in the Compendium of Analytical Nomenclature (chapter 10), as "the element [substance] sought or determined in a sample ..." [6]. The term Measurand, as defined in Ref. 5 is more encompassing: "the particular quantity subject to measurement."

## 2. ANALYTICAL TECHNIQUES, METHODS, AND THE MEASUREMENT PROCESS

An excellent classification scheme consisting of a hierarchy of Analytical Techniques, Methods, Procedures, and Protocols has been presented by Parkany [7]. Beginning at the broad level of the Technique, which is closely allied with a basic area of measurement science, each step of the hierarchy becomes increasingly specific; at the Protocol level, one finds "a complete set of definitive directions that must be followed without exception if the analytical results are to be accepted for a given purpose."

Methods of chemical analysis thus can range from rather loosely specified adaptations of basic analytical techniques to explicitly defined test methods that meet the needs of regulatory agencies. Important terminology has been developed, however, to characterize analytical methods from the perspective of precision and accuracy.

**2.1 Definitive Method.** A method of exceptional scientific status which is sufficiently accurate to stand alone in the determination of a given property for the Certification of a Reference Material [8]. Such a method must have a firm theoretical foundation so that systematic error is negligible relative to the intended use. Analyte masses (amounts) or concentrations must be measured directly in terms of the base units of measurements, or indirectly related through sound theoretical equations. Definitive methods, together with Certified Reference Materials, are primary means for transferring accuracy -- i.e., establishing traceability.

Note: Traceability is defined as "the property of a result or measurement whereby it can be related to appropriate standards, generally international or national standards, through an unbroken chain of comparisons" [5].

2.2 **Reference Method.** A method having small, estimated inaccuracies relative to the end use requirement. The accuracy of a reference method must be demonstrated through direct comparison with a Definitive Method or with a primary Reference Material [9].

2.3 **Chemical Measurement Process (CMP).** An analytical method of defined structure that has been brought into a state of statistical control, such that its imprecision and bias are fixed, given the measurement conditions. This is prerequisite for the evaluation of the Performance Characteristics of the method, or the development of meaningful uncertainty statements concerning analytical results.

### 3. PERFORMANCE CHARACTERISTICS OF THE MEASUREMENT PROCESS

#### 3.1 Structure of the CMP

The general structure of the CMP is indicated in Fig. 1. Here, the symbol  $x$  represents the analyte amount (mass or concentration) contained in the sample or test portion [2] taken for analysis. The CMP that operates on  $x$  (solid box in the figure) consists of two primary substructures: sample preparation and instrumental measurement (first dashed-line box) that converts  $x$  to a signal or response  $y$ , and an evaluation unit (second dashed-line box) that transforms  $y$  back into an estimate  $\hat{x}$  of the analyte amount,

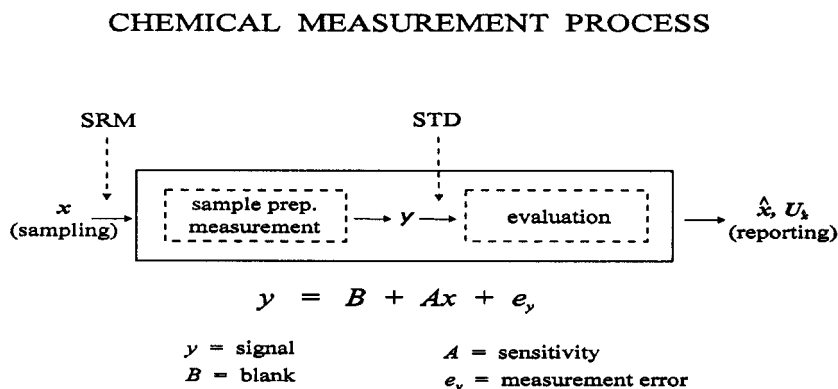


Fig. 1 Schematic diagram of the Chemical Measurement Process. (Adapted from Ref. 10)

together with its uncertainty derived from the detailed structure of the system and/or external calibration (for a non-absolute measurement process) and error estimation. Two important control measures are shown above the diagram: *SRM*, which relates to control of the accuracy of the overall CMP through the use of Certified (or "Standard") Reference Materials; and *STD*, which relates to control of the accuracy of the evaluation (data reduction) step through the use of Standard Test Data. (See section 3.6) Treatment of the output of the CMP -- ie, Presentation of Results -- is the subject of a separate Report of the Commission [3].

Uncertainties introduced in both of the principal steps of the CMP become the basis for the statistical performance characteristics. Further detailed specifications for the structure may be necessary to bring these uncertainties within acceptable bounds, and to guarantee an adequate degree of ruggedness. **Ruggedness** means that the precision and accuracy of the method are insensitive to minor changes in environmental and procedural variables, laboratories, personnel, etc. Otherwise, imputed performance characteristics would depend upon a number of uncontrolled factors, and would have limited utility.

The internal structure of the measurement process may include such steps as dissolution, chemical separation and purification, application of a particular instrumental measurement technique, plus the

detailed scheme or algorithm for data reduction. Internal replication, calibration, and blank estimation also would constitute elements of the CMP structure. In certain applications it may be appropriate to augment the structure to include sampling, and in others, to diminish it to focus on instrumental measurement. In any case, all of the structural elements of the specific measurement process ought to be indicated in a complete flow diagram.

### 3.2 The Observed Signal; Calibration Function

The **Calibration Function** is defined as the functional (not statistical) relationship for the CMP, relating the expected value of the observed (gross) signal or response variable  $E(y)$  to the analyte amount  $x$ . The corresponding graphical display for a single analyte is referred to as the calibration curve. When extended to additional variables or analytes which occur in multicomponent analysis, the "curve" becomes a calibration surface or hypersurface. The functional relationship,  $E(y) = F(x)$ , may in general be quite complicated -- and functions assumed (for data reduction) may be wrong, thus comprising a source of systematic error. We consider here only the simplest case, the linear calibration curve, where the observed signal or response  $y$  is given by

$$y = F(x) + e_y \quad (1)$$

with

$$F(x) = B + S = B + Ax \quad (2)$$

where  $S$  denotes the net signal;  $B$ , the blank (or background or baseline, as appropriate);  $x$ , the analyte amount or concentration; and  $A$ , the sensitivity. The error  $e_y$  is taken to be random and normal, with zero mean (no bias) and dispersion parameter  $\sigma$  (standard deviation). The estimated net signal is thus

$$\hat{S} = y - \hat{B} \quad (3)$$

In the more general case of multicomponent analysis Eq. (1) takes the form

$$y = F(x) + e_y \quad (4)$$

where  $y$ ,  $x$ , and  $e_y$  are vectors, and the Calibration Function takes into account the response relations for all analytes and interferences. Under the best of circumstances, Eq. (4) is a linear matrix equation.

**Note:** Symbols used to represent the calibration parameters vary among disciplines. In statistics, for example, it is conventional to use  $\beta_i$  -- e.g., for a quadratic relationship:  $F(x) = \beta_0 + \beta_1 x + \beta_2 x^2$ . In analytical chemistry, identification of  $\beta_0$  and  $\beta_1$  with the blank  $B$  and sensitivity  $A$ , respectively, is valid *only* if the calibration data represent the *entire* CMP and the calibration relation is linear.

**3.2.1 Sensitivity.** In metrology and in analytical chemistry, the sensitivity  $A$  is defined as the slope of the calibration curve [5,6]. (If the curve is in fact a "curve", rather than a straight line, then of course sensitivity will be a function of analyte concentration or amount.) If sensitivity is to be a unique performance characteristic, it must depend only on the CMP, not upon scale factors. For this reason the slope  $dy/dx$  must be defined in absolute terms, such as  $mV/\mu g$ .

**Notes:**

1. Alternative uses for this term in analytical chemistry, such as a qualitative descriptor for *detection* capability, or slope  $A$  divided by  $\sigma$ , etc., are not recommended.
2. It is recognized that the term "sensitivity" has different meanings for different disciplines. In clinical chemistry (diagnostics), for example, sensitivity is defined as "the fraction of all affected subjects in whom the test result is positive: best positivity in the presence of the disease" [11].

3. When a measurement process parameter is estimated by performing an operation on the observed responses, the resulting statistic is called an Estimator; it is designated by a circumflex as shown in equations 3 and 5. Thus,  $\hat{A}$  indicates an estimator such as the least squares estimator for the sensitivity, and its standard deviation  $\sigma(\hat{A})$  determines the random uncertainty component for any particular estimate. If  $E(\hat{A})$ , the expectation or mean value of the  $\hat{A}$  distribution, equals the true value  $A$ , the estimator is said to be unbiased. (See also Section 3.8)

### 3.3 Evaluation Function

The Evaluation Function is the inverse of the calibration function [6, 1st Edit.]. It is generated by applying the Evaluation Operator  $G$  to the calibration function. For a single analyte, this is  $G(F(x))$ , which is equal to  $G(E(y))$ . If the presumed calibration function is the actual calibration function (correct model), then  $G$  will be the inverse of  $F$ , and the operation  $G(F(x))$  will return the actual concentration  $x$ . Application of  $G$  to the observed response  $y$  together with the estimated parameters, leads to  $\hat{x} = G(y)$  for the estimated concentration. Thus, for the simplest (straight line) calibration function we obtain

$$\hat{x} = (y - \hat{B}) / \hat{A} \quad (5)$$

The process is not quite so simple, of course, in terms of possible interference and losses and chemical matrix corrections. Error propagation, particularly if  $e_y$  is non-normal, and for the non-linear parts of the transformation [denominator of Eq. 5], also is not always trivial.

In the more general, multicomponent analytical process, the Evaluation Function is the inverse of the multicomponent calibration function, given by  $G(F(x))$ , where, if the presumed multicomponent calibration model is correct,  $G$  is now the generalized inverse of  $F$ . The estimated concentration vector  $\hat{x}$  is obtained by operating on the observed signal vector  $y$ . Thus,

$$\hat{x} = G(y) \quad (6)$$

Under the best of circumstances,  $G$  will be a linear operator derived from the calibration function, as in linear least squares estimation. Uncertainties in the numbers, identities, and spectra of the component analytes -- as well as multicollinearity (spectrum similarities) -- can lead to severe difficulties in the inversion of Eq. 4. That is, the identity

$$G F = I \quad (7)$$

may not obtain, or the solution may not be numerically stable due to near singularity. For multicomponent analysis, therefore, the Evaluation Function plays a major role in determining the precision and accuracy of the CMP.

### 3.4 Performance Characteristics Not Specifically Related to Precision and Accuracy

A number of terms are necessary to describe the nature of a CMP, such as: analytical technique or method employed, range of "test portion" (sample) sizes to which it may be applied, interference tolerance and saturation effects, instrumentation employed, time of analysis, cost, number and identity of analytes simultaneously measurable, detector type and efficiency, resolution, chemical yield or recovery, etc. All such descriptors deserve attention, and several influence the attainable precision and accuracy, but they lie beyond the scope of the present document.

### 3.5 Precision and Accuracy – Related Performance Characteristics

"Precision" and "accuracy" have thus far been used as general, qualitative descriptors. Such usage is quite common and perhaps even appropriate; however, somewhat different, explicitly defined terms are given below.

3.5.1 **Measurement Result.** The outcome of an analytical measurement (application of the CMP), or "value attributed to a measurand" [5]. This may be the result of direct observation, but more commonly it is given as a statistical estimate  $\hat{x}$  derived from a set of observations. The distribution of such estimates (estimator distribution) characterizes the CMP, in contrast to a particular estimate, which constitutes an experimental result. Additional characteristics become evident if we represent  $\hat{x}$  as follows,

$$\hat{x} = \tau + e = \tau + \underbrace{\Delta}_{\mu} + \delta = \mu + \delta \quad (8)$$

where:

3.5.2 **True Value ( $\tau$ ).** The value  $x$  that would result if the CMP were error-free.

3.5.3 **Error ( $e$ ).** The difference between an observed (estimated) value and the true value; i.e.,  $e = \hat{x} - \tau$  (signed quantity). The total error generally has two components -- bias ( $\Delta$ ) and random error ( $\delta$ ), as indicated above.

3.5.4 **Limiting Mean ( $\mu$ ).** The asymptotic value or population mean of the distribution that characterizes the measured quantity; the value that is approached as the number of observations approaches infinity. Modern statistical terminology labels this quantity the expectation or expected value,  $E(\hat{x})$ .

3.5.5 **Bias ( $\Delta$ ).** The difference between the limiting mean and the true value; i.e.,  $\Delta = \mu - \tau$  (signed quantity).

3.5.6 **Random Error ( $\delta$ ).** The difference between an observed value and the limiting mean; i.e.,  $\delta = \hat{x} - \mu$  (signed quantity). The random error is governed by the cumulative distribution function (*cdf*), which in turn may be described by a specific mathematical function involving one or more parameters. (An example of a one parameter *cdf* is the Poisson distribution, which figures importantly in counting experiments.) Most commonly assumed is the normal or "Gaussian" distribution; this has two parameters: the mean  $\mu$ , and the standard deviation  $\sigma$ . The random error is given by  $\delta = z\sigma$ , where  $z$  is the value of the standard normal variate.

3.5.7 **Standard Deviation ( $\sigma$ ).** Dispersion parameter for the distribution. That is,  $\sigma$  is the performance characteristic that reflects the root mean square random deviation of the observations (results) about the limiting mean; positive square root of the variance.

3.5.8 **Variance ( $V = \sigma^2$ ).** More directly the *cdf* dispersion parameter is the variance, which is defined as the second moment about the mean. For certain non-normal distributions, higher moments may be given.

3.5.9 **Systematic Error.** Taking Systematic Error to be all error components that are not random, we thus far would equate systematic error with the fixed bias of the CMP. Real CMPs, however, should be described by at least two additional quantities:

- Blunders** ( $b$ ) -- which we take as outright mistakes, and
- Lack of Control** ( $f(t)$ ) -- drifts, fluctuations, etc.

Systematic error is defined in the *International Vocabulary of Basic and General Terms in Metrology* as "a component of the error of measurement which, in the course of a number of measurements of the same measurand, remains constant or varies in a predictable way" [5]. A somewhat different perspective on measurement error, advanced by BIPM [12], is presented in the "ISO Guide to the Expression of Uncertainty in Measurement" [13]. This alternative view differs from the classical treatment of random and systematic sources of measurement uncertainty, and assigns "standard deviations" to all error

components. According to the Guide, "it is assumed that, after correction, ... the expected value of the error arising from a systematic effect is zero." A new term, "standard uncertainty" is defined as the "uncertainty of the result of a measurement expressed as a standard deviation." Further, uncertainty components are classified as "type A" and "type B," reflecting those that may be evaluated by statistical methods, and those that are evaluated by other means. Note that the ISO Guide treats *uncertainties of measurement results*, whereas this IUPAC document is concerned with *performance characteristics of measurement processes*.

**3.5.10 Imprecision.** A quantitative term to describe the (lack of) "precision" of a CMP; identical to the Standard Deviation [1, 14].

**3.5.11 Inaccuracy.** A quantitative term to describe the (lack of) accuracy of a CMP; comprises the imprecision and the bias. Inaccuracy *must be* viewed as a 2-component quantity (vector); imprecision and bias should never be combined to give a scalar measure for CMP inaccuracy. (One or the other component may, however, be negligible under certain circumstances.) [1]. Inaccuracy should not be confused with uncertainty. Inaccuracy (imprecision, bias) is characteristic of the Measurement Process, whereas error and uncertainty are characteristics of a Result [3]. (The latter characteristic, of course, derives from the imprecision and bounds for bias of the CMP.)

Note: The resultant bias and imprecision for the overall measurement process generally arise from several individual components, some of which act multiplicatively (eg, sensitivity), and some of which act additively (eg, the blank). (See section 3.8)

## 3.6 Control and Testing of Assumptions

"Quality Metrology concerns itself with the control of measurements and their results which enter into examinations of the quality of materials, devices,...measuring instruments..." [15]. The testing of assumption validity for the Chemical Measurement Process, and thereby its results, necessarily constitutes a fundamental part of Quality Metrology. The control and assessment of imprecision and bias of the CMP -- ie, Quality Assurance [16] -- is accomplished via assumption or Hypothesis Testing, where the null hypothesis is generally taken to be the absence of bias or of an added component of random error.

**3.6.1 Assumption Testing.** The principal concepts involved in the statistical theory of hypothesis testing are presented in section 3.7 with reference to analyte detection. Testing for bias or added imprecision rests upon the same principles. That is, one must postulate null ( $H_0$ ) and alternative ( $H_A$ ) hypotheses, and then define a test statistic and critical value, based upon the acceptable level for the error of the first kind  $\alpha$  -- also known as the significance level of the test. The power of the test, which is described by its operating characteristic [OC curve], is defined as the probability of correctly "accepting" the alternative hypothesis, given  $\alpha$ . The power is thus  $1-\beta$ , where  $\beta$  is the probability of the error of the second kind [17].

Three points deserve emphasis: 1) "Acceptance" of an hypothesis, based on such statistical testing must not be taken literally. More correctly, one simply fails to reject the hypothesis in question. For example, non-detection of an analyte does not prove its absence. Put another way, "acceptance" [non-rejection] may reflect inadequate power [ $1-\beta$ , given  $\alpha$ ] for the test and alternative hypothesis in question. 2) Assumption (hypothesis) testing, itself, rests upon assumptions. The vast majority of statistical tests performed on the CMP and its results, for example, rely upon the assumptions of randomness and normality. Robust estimators and non-parametric or distribution-free tests may be employed when certain common assumptions may not be valid. 3) Assumption tests emerge in many facets of chemical measurement, ranging from analyte detection [section 3.7], to tests of randomness and independence, to tests of means (and bias) using  $z$ - or  $t$ -statistics, and variance (and model) tests using  $\chi^2$  or  $F$  statistics. Such test statistics play a central role in maintaining CMP quality both within and among laboratories; the resultant quality assurance is generally considered from the perspective of internal or external control [16, 18, 19].



## Notes

1. Significance tests may be one-sided or two-sided. Testing for the presence of analyte in excess of the blank (detection test) is one-sided, since the true value of the net analyte concentration cannot be negative. Testing for the presence of bias, on the other hand, is generally two-sided.
2. In many cases, such as testing for the presence of a particular analyte or the presence of systematic error, the null state cannot, *in principle* be attained; nor can the null hypothesis be proved. Recognizing the impossibility of attaining or proving absolute purity or absolute accuracy, it has been suggested that  $H_0$  be displaced from zero to an incremental value consistent with the relevant metrological objectives. In such circumstances, attention would be shifted for example from the Detection Limit to the Discrimination Limit, where the null state would be that characterized by a small, acceptable analyte concentration [20].

3.6.2 Internal Control. Within a given laboratory employing a given method of analysis, control of the CMP can be assessed in part by repeated measurements of samples, such as Reference Materials (RM), having characteristics (composition) similar to the samples of interest. The control in this case is limited to control of the mean (absence of trends, etc.) and control of the variance -- ie, the two quantities that reflect the stability of the CMP. Control Charts are used to maintain a record of such internal control, where critical or control levels are derived from the mean and standard deviation (or ranges) of sets of observations. (At least four observations per set are advisable, to take advantage of the central limit theorem.) When Certified Reference Materials (CRM) or other materials of known composition are available, one may estimate bias as well, within the uncertainty bounds of the CRM. The procedures for accomplishing internal (and external) control, especially from the perspective of the CRM, have been documented by the International Organization for Standardization [9].

3.6.2.1 Repeatability, as measured by the repeatability standard deviation, is an accepted measure of internal variance. Its definition requires that "mutually independent test results [be] obtained with the same method on a test material in the same laboratory with the same equipment by the same operator within a short interval of time" [21]. Thus, repeatability reflects the best achievable internal precision, and realistic uncertainty estimates must take into account possible variations in the constrained factors, as well as possible sources of uncompensated bias. Note that a false level of precision (repeatability) ensues if the observations are not truly mutually independent. Successive readings from an instrument, for example, do not give a valid measure of repeatability for the CMP; rather, they are solely an indication of the instrumental repeatability. (See section 4.1)

3.6.3 External Control. Control may be assessed from without via "blind" replicates (for CMP stability) or "blind" CRMs (for CMP accuracy), submitted without foreknowledge of the measuring laboratory. A common failure of such external control is that the test samples are not totally blind. That is, the appearance or scheduling of the external samples may be sufficient to alert the internal analyst (possibly only subconsciously) to apply extra care, or even lack thereof. Collaborative tests comprise the other form of external control, where a number of (presumably) equivalent laboratories assay test portions from the same homogeneous material. ISO Guide 33 [9] treats CMP assessment via an interlaboratory program; and ISO Guide 35 [8] discusses this approach for the certification of CRMs.

3.6.3.1 Reproducibility, as measured by the reproducibility standard deviation, is the external complement to repeatability. Conditions here are defined such that "test results are obtained with the same method on a test material in different laboratories with different equipment by different operators" [21]. Thus, if the method in question is unbiased, reproducibility meets the objective of varying all factors so that the total error becomes random and thereby experimentally (statistically) estimable. In the International Vocabulary of Basic and General Terms in Metrology [5], the definition appears a little more flexible, in that a list of six types of changing factors is presented (including the method of measurement), accompanied by the notes that a specification of conditions actually subject to change should be indicated, and that the dispersion of results would serve as the quantitative measure of reproducibility.

Control, internal or external, need not be limited to measurement stability and accuracy. Control or assessment of assumed physical (or functional) models as well as random error models (cumulative distribution functions, autocovariance functions) may also be addressed. Both of these elements of modern multivariable and multicomponent measurements are leading to the emergence of a data analogue of Standard (Certified) Reference Materials (SRM), i.e., Standard Test Data (STD) [10, 22, 23]. Such data are supplied, for example, as a regular part of the IAEA Analytical Quality Control Services program (gamma ray spectra) [22]. STD, which represent fully characterized simulations of real analytical signals, have the great merit of providing quality assessment for the evaluation step of the CMP -- the step that is becoming at the same time more common and more complex and more remote from the direct control of the operator, through the advent of sophisticated computational and instrumentation modules. See Fig. 1 for a graphical representation of the SRM and STD control points for the Chemical Measurement Process.

### 3.7 Detection and Quantification Capabilities

Among the most important Performance Characteristics of the Chemical Measurement Process (CMP) are those that can serve as measures of the underlying *detection* and *quantification* capabilities. These are essential for applications in research, international commerce, health, and safety. Such measures are important for *planning* measurements, and for selecting or developing CMPs that can meet specified needs, such as the detection or quantification of a dangerous or regulated level of a toxic substance.

Equations 1-6 provide the basis for our considering the meaning of minimum detectable and minimum quantifiable amounts (signals, concentrations) in Analytical Chemistry [24]. In each case, the determining factor is the distribution function of the estimated quantity (estimated net signal  $\hat{S}$ , concentration or amount  $\hat{x}$ ). If normality can be assumed, it is sufficient to know the standard deviation of the estimated quantity as a function of  $S$  (or  $x$ ). Detection limits (minimum detectable amounts) are based on the theory of hypothesis testing and the probabilities of false positives  $\alpha$ , and false negatives  $\beta$ . Quantification limits are defined in terms of a specified value for the relative standard deviation. It is important to emphasize that both types of limits are CMP Performance Characteristics, associated with underlying *true values* of the quantity of interest; they are *not* associated with any particular outcome or result. The detection decision, on the other hand, is result-specific; it is made by comparing the experimental result with the Critical Value, which is the minimum significant *estimated value* of the quantity of interest.

**3.7.1 Terminology.** Unfortunately, a host of terms have been used within the chemical community to describe detection and quantification capabilities. Perhaps the most widely used is "detection limit" (or "limit of detection") as an indicator of the minimum detectable analyte net signal, amount, or concentration. However, because the distinction between the *minimum significant estimated concentration* and the *minimum detectable true concentration* has not been universally appreciated, the same term and numerical value has been applied by some, perhaps unwittingly, in both contexts. Despite this, the term "Detection Limit" is widely understood and quoted by most chemists as a measure of the inherent detection capability. (For more on the terminological and conceptual history that has beset Detection in Analytical Chemistry, see Currie [20] and Note-3 in Section 3.7.3.2.)

With the goal of harmonizing international terminology in this area, scientists from ISO and IUPAC met in July 1993 [25]. The meeting resulted in full consensus on detection concepts and default parameter choices, and acceptable agreement on terminology. As an outgrowth of that meeting, we recommend the following terms and alternates. (Concepts and formulas will be presented in following sections.) For distinguishing a chemical signal from background noise -- i.e., for making the Detection Decision: the Critical Value ( $L_C$ ) of the appropriate chemical variable (*estimated* net signal, concentration, or amount); alternate: the Critical Level. As the measure of the inherent Detection Capability of a CMP: the Minimum Detectable (true) Value ( $L_D$ ) of the appropriate chemical variable; alternate: the Detection Limit. As the measure of the inherent Quantification Capability of a CMP, the Minimum Quantifiable (true) Value ( $L_Q$ ); alternate: the Quantification Limit. Many other terms such as "Decision Criterion" for  $L_C$ , "Identification Limit" for  $L_D$ , and "Measurement Limit" for  $L_Q$ , appear in the chemical literature.

In the interest of *uniform international nomenclature*, however, only the terms and alternates defined above are recommended.

Note: For presentation of the defining relations, we use  $L$  as the generic symbol for the quantity of interest. This is replaced by  $S$  when treating net analyte signals, and  $x$ , when treating analyte concentrations or amounts. Thus,  $L_O$ ,  $L_D$  and  $L_Q$  may represent  $S_O$ ,  $S_D$  and  $S_Q$ , or  $x_O$ ,  $x_D$  and  $x_Q$ , as appropriate.

**3.7.2 Specification of the Measurement Process.** Just as with other Performance Characteristics,  $L_D$  and  $L_Q$  cannot be specified in the absence of a fully defined measurement process, including such matters as types and levels of interference *as well as* the data reduction algorithm. "Interference free detection limits" and "Instrument detection limits", for example, are perfectly valid within their respective domains; but if detection or quantification characteristics are sought for a more complex chemical measurement process, involving for example sampling, analyte separation and purification, and interference and matrix effects, then it is mandatory that all these factors be considered in deriving values for  $L_D$  and  $L_Q$  for that process. Otherwise the actual performance of the CMP (detection, quantification capabilities) may fall far short of the requisite performance.

**3.7.3 Detection -- Fundamental Relations.** The statistical theory of Hypothesis Testing, introduced in Section 3.6.1, serves as the framework for our treatment of Detection in Analytical Chemistry. Following this theory we consider two kinds of errors (really erroneous decisions): the error of the first kind ("type I," false positive), accepting the "alternative hypothesis" (analyte present) when that is wrong; and the error of the second kind ("type II," false negative), accepting the "null hypothesis" (analyte absent) when that is wrong. The probability of the type I error is indicated by  $\alpha$ ; the probability for the type II error, by  $\beta$ . Default values recommended by IUPAC for  $\alpha$  and  $\beta$  are 0.05, each. These probabilities are directly linked with the one-sided tails of the distributions of the estimated quantities ( $\hat{S}$ ,  $\hat{x}$ ).

A graphical representation of these concepts is given in Fig. 2, where the "driving force" in this hypothetical example is the ability to detect the release of specific chemical precursors of earthquakes (e.g., radon) at levels corresponding to earthquakes of magnitude  $L_R$  and above. Thus  $L_R$  is the "requisite limit" or maximum acceptable limit for undetected earthquakes; this is driven, in turn, by a maximum acceptable loss to society. (Derivation of  $L_R$  values for sociotechnical problems, of course, is far more complex than the subject of this report!) The lower part of the figure shows the minimum detectable value for the chemical precursor  $L_D$ , that must not exceed  $L_R$ , and its relation to the probability density functions (pdf) at  $L = 0$  and at  $L = L_D$  together with  $\alpha$  and  $\beta$ , and the decision point (Critical Value)  $L_C$ . The figure has been purposely constructed to illustrate heteroscedasticity -- in this case, variance increasing with signal level, and unequal  $\alpha$  and  $\beta$ . The point of the latter construct is that, although 0.05 is the recommended default value for these parameters, particular circumstances may dictate more stringent control of the one or the other. Instructive implicit issues in this example are that

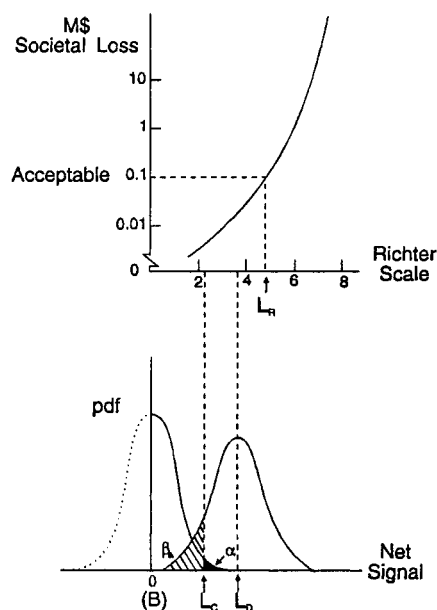


Fig. 2 Detection: needs and capabilities. Top portion shows the requisite limit  $L_R$ , bottom shows detection capability  $L_D$ . (Adapted from Ref. 20)

(1) a major factor governing the detection capability could be the natural variation of the radon background (blank variance) in the environment sampled, and (2) a calibration factor or function is needed in order to couple the two abscissae in the diagram. In principle, the response of a sensing instrument could be calibrated directly to the Richter scale (earthquake magnitude); alternatively, there could be a two-stage calibration: response–radon concentration, and concentration–Richter scale.

A final point is that, with the exception of certain "distribution-free" techniques, Detection Limits *cannot* be derived in the absence of known (or assumed) distributions. As with all Performance Characteristics, the parameters used to compute  $L_C$  and  $L_D$  should be estimated from measurements in the region of interest -- in this case in the range between the blank and the detection limit. Similarly, experimental verification of computed detection limits is highly recommended.

Note: The single, most important application of the detection limit is for *planning* (CMP design or selection). It allows one to judge whether the CMP under consideration is adequate for the detection requirements. This is in sharp contrast to the application of the critical value for decision making, given the *result* of a measurement. The most serious pitfall is inadequate attention to the magnitude and variability of the *overall* blank, which may lead to severe underestimation of the detection limit.

3.7.3.1 Detection Decisions ( $L_C$ ). The decision "detected" or "not detected" is made by comparison of the estimated quantity ( $\hat{L}$ ) with the Critical Value ( $L_C$ ) of the respective distribution, such that the probability of exceeding  $L_C$  is no greater than  $\alpha$  if analyte is absent ( $L = 0$ , null hypothesis). The Critical Value is thus the minimum significant value of an estimated net signal or concentration, applied as a discriminator against background noise. This corresponds to a 1-sided significance test. The above definition of  $L_C$  can be expressed as follows,

$$\Pr(\hat{L} > L_C \mid L=0) \leq \alpha \quad (9)$$

Generally the equation is stated as an equality, but the inequality is given to accommodate discrete distributions, such as the Poisson, where not all values of  $\alpha$  are possible. If  $\hat{L}$  is normally distributed with known variance, Eq. 9 reduces to the following simple expression,

$$L_C = z_{1-\alpha} \sigma_o \quad (10)$$

where  $z_{1-\alpha}$  (or  $z_p$ ) represents the  $(1-\alpha)$ th percentage point or critical value of the standard normal variable, and  $\sigma_o$  is the standard deviation of the estimated quantity (net signal or concentration) under the null hypothesis (true value = 0). Taking the default value for  $\alpha$  (0.05),  $L_C = 1.645 \sigma_o$ .

Note that Eq. 9, *not* Eq. 10, is the defining equation for  $L_C$ , and the result ( $1.645 \sigma_o$ ) applies *only* if the data are normal with known variance and  $\alpha$  is set equal to its default value. If  $\sigma_o$  is estimated by  $s_o$ , based on  $\nu$  degrees of freedom,  $z_{1-\alpha}$  must be replaced by Student's- $t$ . That is,

$$L_C = t_{1-\alpha, \nu} s_o \quad (11)$$

For  $\alpha = 0.05$  and 4 degrees of freedom, for example,  $L_C$  would be equal to  $2.132 s_o$ .

Notes:

1. Some measurement systems impose an artificial hardware or software threshold (*de facto*  $L_C$ ) to discriminate against small signals. In such cases statistical significance is problematic --  $\alpha$  may be quite small and perhaps unknown, but equations 12 and 13 below can still be applied to compute  $L_D$ , given  $L_C$  and  $\beta$ . The impact of such a threshold can be profound, severely eroding the inherent detection capability of the system [26].
2. A result falling below  $L_C$ , triggering the decision "not detected" should not be construed as demonstrating analyte absence. (See section 3.6.1.) Reporting such a result as "zero"

or as " $\langle L_D \rangle$ " is *not* recommended; the estimated value (net signal, concentration) and its uncertainty should *always* be reported.

**3.7.3.2 Minimum Detectable Value; Detection Limit ( $L_D$ ).** The Minimum Detectable Value of the net signal (or concentration) is that value ( $L_D$ ) for which the false negative error is  $\beta$ , given  $L_C$  (or  $\alpha$ ). It is the true net signal (or concentration) for which the probability that the estimated value  $\hat{L}$  does not exceed  $L_C$  is  $\beta$ . The definition of  $L_D$  can thus be expressed as

$$\Pr(\hat{L} \leq L_C \mid L = L_D) = \beta \quad (12)$$

For normal data having known variance structure, this yields,

$$L_D = L_C + z_{1-\beta} \sigma_D \quad (13)$$

For the special situation where the variance is constant between  $L = 0$  and  $L = L_D$ , the right side of Eq. 13 reduces to  $(z_{1-\alpha} + z_{1-\beta})\sigma_o$ ; if in addition  $\alpha$  and  $\beta$  are equal, this gives  $2z_{1-\alpha}\sigma_o$  which equals  $2L_C$ . Taking the default values for  $\alpha$  and  $\beta$  (0.05), this equals  $3.29 \sigma_o$ . If  $L_C$  employs an estimate  $s_o$  based on  $\nu$  degrees of freedom (Eq. 11), then  $(z_{1-\alpha} + z_{1-\beta})$  must be replaced by  $\delta_{\alpha,\beta,\nu}$ , the non-centrality parameter of the non-central- $t$  distribution. For  $\alpha = \beta$ , this parameter is approximately equal to  $2t$  and the appropriate expression (for constant variance) is,

$$L_D = \delta_{\alpha,\beta,\nu} \sigma_o \approx 2t_{1-\alpha,\nu} \sigma_o \quad (14)$$

For 4 degrees of freedom, for example, the use of  $2t$  would give  $L_D = 4.26 \sigma_o$ . (The actual value for  $\delta$  in this case is 4.067.) Note that  $\sigma_o$  *must be used* in Eq. 14. If only an estimate  $s_o$  is available, that means that the minimum detectable value is uncertain by the ratio ( $\sigma/s$ ). Using the techniques of section 3.8.6, confidence limits may then be calculated for  $L_D$ . (A 95% upper limit for  $L_D$ , based on an observed  $s_o$  with 4 degrees of freedom, would be  $\{4.07/(\sqrt{0.178})\} s_o$  or  $9.65 s_o$ .)

Notes:

1. When  $\nu$  is large,  $2t$  is an excellent approximation for  $\delta$ . For  $\nu \geq 25$ , with  $\alpha = \beta = 0.05$ , the difference is no more than 1 %. For fewer degrees of freedom, a very simple correction factor for  $2t$ ,  $4\nu/(4\nu+1)$ , which takes into account the bias in  $s$ , gives values that are within 1 % of  $\delta$  for  $\nu \geq 5$ . For the above example where  $\nu = 4$ ,  $\delta$  would be approximated as  $2(2.132)(16/17)$  which equals 4.013.
2.  $L_D$  is defined by Eq. 12 in terms of the distribution of  $\hat{L}$  when  $L = L_D$ , the probability of the type-II error  $\beta$ , and  $L_C$ , with  $L_C$  being defined (Eq. 9) in terms of the distribution of  $\hat{L}$  when  $L = 0$ , and the probability of the type-I error  $\alpha$ . When certain conditions are satisfied,  $L_D$  can be expressed as the product of a specific coefficient and the standard deviation of the blank, such as  $3.29 \sigma_B$ , when the uncertainty in the mean (expected) value of the blank is negligible,  $\alpha$  and  $\beta$  each equal 0.05, and  $\hat{L}$  is normally distributed with known, constant variance.  $L_D$  is *not defined*, however, simply as a fixed coefficient (2, 3, 6, etc.) times the standard deviation of a pure solution background. To do so can be extremely misleading. The correct expression must be derived from the proper defining equations (Eq. 9 and 12), and it must take into account degrees of freedom,  $\alpha$  and  $\beta$ , and the distribution of  $\hat{L}$  as influenced by such factors as analyte concentration, matrix effects, and interference. (See also section 3.7.2.)
3. The question of detection has been treated extensively by H. Kaiser for spectrochemical analysis. In the earlier editions of the "Orange Book" [6] and related publications of Kaiser on spectrochemical analysis, the use of  $3s_B$  is recommended as the "limit of detection" [translation of *Nachweisgrenze*]. Although originally intended to serve as a measure of the detection capability, this quantity was then used as the "decision criterion" to distinguish an estimated signal from the background noise. Such a definition, which

in effect sets  $L_C$  and  $L_D$  each equal to  $3s$ , corresponds for a normal distribution (large  $v$ ) to a type-I error probability of ca. 0.15 % but a type-II error probability of 50 % ! Kaiser's "limit of guarantee for purity" [ $6s_B$ ; Ref. 30] rectifies the imbalance, but this quantity -- as a measure of detectability, the CMP performance characteristic, is scarcely ever used, and has not appeared in the "Orange Book." Further discussion of the confusion that has resulted from this earlier terminology and *support for identifying the "detection limit" with the CMP performance characteristic  $L_D$*  may be found in several publications and textbooks, including the Standard Practice of ASTM [27] and books by Liteanu and Riča [28, ch. 7], Massart, et al. [31], and Currie [20].

3.7.4 **Signal Domain ( $S_C$   $S_D$ ).** In many cases the smallest signal  $S_D$  that can be reliably distinguished from the blank, given the critical level  $S_C$ , is desired, as in the operation of radiation monitors. Assuming normality and knowledge of  $\sigma$ , simple expressions can be given for the two quantities involved in Signal Detection. Eq. 10 takes the following form for the Critical Value,

$$S_C = z_{1-\alpha} \sigma_o \rightarrow 1.645 \sigma_o \quad (15)$$

where the expression to the right of the arrow results for  $\alpha = 0.05$ . From Eq. 3 the estimated net signal  $\hat{S}$  equals  $y - \hat{B}$ , and its variance is

$$V_{\hat{S}} = V_y + V_{\hat{B}} \rightarrow V_B + V_{\hat{B}} = V_o \quad (16)$$

The quantity to the right of the arrow is  $\sigma_o^2$ , the variance of the estimated net signal when the true value  $S$  is zero. If the variance of  $\hat{B}$  is negligible, then  $\sigma_o \approx \sigma_B$ , the standard deviation of the Blank. If  $B$  is estimated in a "paired" experiment -- i.e.,  $V_{\hat{B}} = V_B$ , then  $\sigma_o = \sigma_B \sqrt{2}$ . Note that  $\sigma_o \approx \sigma_B$ , and  $\sigma_o = \sigma_B \sqrt{2}$ , are limiting cases. More generally,  $\sigma_o = \sigma_B \sqrt{\eta}$ , where  $\eta = 1 + (V_{\hat{B}}/V_B)$ . Thus,  $\eta$  reflects different numbers of replicates, or, for particle or ion counting, different counting times for "sample" vs blank measurements. (See section 3.8.8 for further discussion of the Blank.)

The Minimum Detectable Signal  $S_D$  derives similarly from Eq. 13, that is,

$$S_D = S_C + z_{1-\beta} \sigma_D \quad (17)$$

where  $\sigma_D^2$  represents the variance of  $\hat{S}$  when  $S = S_D$ . For the special case where the *variance is constant* between  $S = 0$  and  $S = S_D$ , and  $\alpha = \beta = 0.05$ , the Minimum Detectable Signal  $S_D$  becomes  $2S_C = 3.29 \sigma_o$ , or  $(3.29\sqrt{2})\sigma_B = 4.65 \sigma_B$  for paired observations. The treatment using an estimated variance,  $s_o^2$  and Student's- $t$  follows that given above in Section 3.7.3.

The above result is not correct for  $S_D$  if the variance depends on the magnitude of the signal. A case in point is the counting of particles or ions in accelerators or mass spectrometers, where the number of counts accumulated follows the Poisson distribution, for which the variance equals the expected number of counts. If the mean value of the background is known precisely, for example,  $\sigma_o^2 = \sigma_B^2$ , which in turn equals the expected number of background counts  $B$ . This leads to approximate expressions of  $1.645 \sqrt{B}$ , and  $2.71 + 3.29 \sqrt{B}$  for  $S_C$  and  $S_D$  (units of counts), respectively, for counting experiments with "well known" blanks [24]. In more complicated cases where net signals are estimated in the presence of chromatographic or spectroscopic baselines, or where they must be deconvolved from overlapping peaks, the limiting standard deviations ( $\sigma_o$  and  $\sigma_D$ ) must be estimated by the *same procedures used to calculate the standard deviation of the estimated (net) signal of interest*. Examples can be found in Ref. 20. (See also sections 3.7.5 and 3.7.6.)

Note: The result for counting data given above is based on the normal approximation for the Poisson distribution. Rigorous treatment of discrete and other non-normal distributions, which is beyond the scope of this document, requires use of the actual distribution together with the defining relations Eq. 9 and Eq. 12.

3.7.5 Concentration Domain ( $x_C$   $x_D$ ). For the special category of "direct reading" instrument systems, the response is given directly in units of concentration (or amount). In this case, the distinction between the signal domain and the concentration domain vanishes, and the treatment in the preceding section applies, with  $y$ ,  $e_y$ ,  $B$ , and  $S$  already expressed in concentration units. As before, the development of particular values for the critical level and detection limit requires distributional assumptions, such as normality, which should be tested. More generally, the transformation to the minimum detectable concentration (amount) involves one or more multiplicative (or divisive) factors, each of which may be subject to error. Thus, one divides the net signal by a theoretically or experimentally determined sensitivity factor or efficiency to convert a gamma ray counting rate into an emission rate; further division by a branching ratio may be needed to determine a radionuclide decay rate; additional correction factors may be needed to treat interference, matrix effects, and chemical recovery; and factors taking into account neutron monitor responses and irradiation and decay times are generally needed in activation analysis. Collectively, these factors comprise the sensitivity  $A$  which relates the net signal to the physical or chemical quantity of interest  $x$ , as indicated in Eq. 5. We consider two cases.

3.7.5.1 Calibration function known ( $e_{\hat{A}}$  negligible or constant). When the uncertainty about the calibration function  $F(x)$  and its parameters is negligible, the minimum detectable concentration  $x_D$  can be calculated as  $F^{-1}(y_D)$ , where  $y_D = B + S_D$ . Problems arise only when the calibration function is not monotonic; and even if it is monotonic, some iteration may be needed if  $F(x)$  is not linear in  $x$ . In the linear case where  $F(x) = B + Ax$ , and the uncertainty of the sensitivity, but not necessarily that of the blank, is negligible, the transformation from the minimum detectable signal to the minimum detectable concentration is simply

$$x_D = S_D/A \quad (18)$$

For normal data with constant, known variance, and  $\alpha = \beta$ , the Minimum Detectable Concentration  $x_D$  is thus  $2S_C/A$ . Taking the default value for  $\alpha$  and  $\beta$ , this becomes  $(3.29 \sigma_o)/A$ , where  $\sigma_o$  is the standard deviation of  $\hat{S}$  when  $S = 0$ . For paired observations this is equivalent to  $(4.65 \sigma_B)/A$ , where  $\sigma_B$  is the standard deviation of the blank. Since only the numerator in Eq. 18 is subject to random error, the detection test will still be made using  $S_C$ . When variance is estimated as  $s^2$ , Student's- $t$  (central and non-central) must be used as shown section 3.7.3.

When the assumed value of the sensitivity  $A$  is fixed, but biased -- as when an independent estimate of the slope from a *single* calibration operation, or a calibration material or a theoretical estimate having non-negligible error, is repeatedly used -- the calculated detection limit will be correspondingly biased. Bounds for the bias in  $A$  can be applied to compute bounds for the true detection limit. Since the biased estimate of  $A$  is fixed, it cannot contribute to the variance of  $\hat{x}$

Note: Repeated use of a fixed estimate for the blank is *not* recommended, unless  $V_B \ll V_B$ , as that may introduce a systematic error comparable to the Detection Limit, itself. This is of fundamental importance in the common situation, especially in trace analysis, where the sensitivity estimate is derived from instrument calibration, but where the blank and its variance depend primarily on non-instrumental parts of the CMP including sample preparation and even sampling. (See section 3.8.8).

3.7.5.2 Calibration function estimated ( $e_{\hat{A}}$  random). When the error in  $\hat{A}$  is random, then its effect on the distribution of  $\hat{x}$  must be taken into account, along with random error in  $y$  and  $\hat{B}$ . This is the case, for example, where sensitivity (slope) estimation is repeated with each application of the measurement process. For the common situation where  $x$  is estimated by  $(y - \hat{B})/\hat{A}$  [Eq. 5], the minimum detectable concentration may be calculated from the defining equations 9 and 12, and their normal equivalents, using the Taylor expansion for the variance of  $\hat{x}$  at the detection limit  $x_D$ :  $V_{\hat{x}}(x=x_D) \approx (V_o + x_D^2 V_{\hat{A}} + 2x_D V_{AB})/A^2$ , with  $V_o$  as given in Eq. 16. For the heteroscedastic case ( $V_y$  varying with concentration),  $V_o$  must be replaced by  $(V_y(x_D) + V_{\hat{B}})$  in the above expression, and weights used [26].

The results for constant  $V_y$  and  $\alpha = \beta$ , presented in a slightly different form in Ref. 3, are

$$S_C = t_{1-\alpha, \nu} s_o \quad (19)$$

$$x_D = (\delta_{\alpha, \beta, \nu} \sigma_o / A)(K/I) = (2t_{1-\alpha, \nu} \sigma_o / A)(K/I) \quad (20)$$

where:  $K = 1 + r(B, A)(\sigma_B^2 / \sigma_o^2)[t_{1-\alpha, \nu}(\sigma_A^2 / A)]$

$$I = 1 - [t_{1-\alpha, \nu}(\sigma_A^2 / A)]^2$$

When  $B$  and  $A$  are estimated from the *same* calibration data set, the estimates will be negatively correlated with  $r(B, A) = -\bar{x}/x_q$ ,  $x_q$  being the quadratic mean [3]. The ratio  $K/I$  may then range from slightly less than one to very much greater, depending on the calibration design and the magnitude of  $\sigma_y$ . The effect of the factor  $I$  in particular, can cause  $x_D$  to differ substantially from  $2t_{1-\alpha, \nu} \sigma_o / A$ . The extreme occurs when the relative standard deviation of  $\hat{A}$  approaches  $1/t_{1-\alpha, \nu}$ ; then  $x_D$  is unbounded. When  $B$  and  $A$  are estimated independently, then  $r(B, A) = 0$ , and  $K = 1$ . If the relative standard deviation of  $\hat{A}$  is negligible compared to  $1/t_{1-\alpha, \nu}$ , then  $K$  and  $I$  both approach unity, and  $x_D$  reduces to the form given in Eq. 18.

A note of caution: If the parameters used in equations 19 and 20 derive from a calibration operation that fails to encompass the *entire* CMP, the resulting values for  $S_C$  and  $x_D$  are likely to be much too small. Such would be the case, for example, if the response variance and that of the estimated intercept based on instrument calibration data alone were taken as representative of the total CMP, which may have major response and blank variations associated with reagents and the sample preparation environment.

#### Notes:

1. When an estimated value  $\hat{A}$  is used in Eq. 20, it gives a rigorous expression for the maximum (non-detection) upper limit for a particular realization of the calibration curve. (This result derives from the distribution of  $y - \hat{B} - \hat{A}x$  which is normal with mean zero and variance  $V_y + V_{\hat{B}} + x^2 V_{\hat{A}} + 2xV_{AB}$ .) Since  $\hat{A}$  is a random variable, this means for the measurement process as a whole that there is a *distribution of limits*  $x_D(\hat{A})$  corresponding to the distribution of  $\hat{A}$ 's. When  $A$  is used in Eq. 20, the resulting  $x_D$  can be shown to be approximately equal to the median value of the distribution of the maximum upper limits.
2. An alternative approach for establishing  $x_D$ , developed by Liteanu and Riča [28], is based on empirical frequencies for the type II error as a function  $x$ . Using a regression-interpolation technique these authors obtain a direct interval estimate for  $x_D$  corresponding to  $\beta = 0.05$ , given  $x_C$ . This "frequentometric" technique for estimating detection limits is sample intensive, but it has the advantage that, since it operates directly on the empirical  $\hat{x}$  distribution, it can be free from distributional or calibration shape assumptions, apart from monotonicity. For the special case where  $\hat{S}$  and  $\hat{A}$  are distributed normally, a rigorous asymptotic solution to the problem has been developed, based on the theoretical distribution of  $\hat{x}$ , which is not normal [29].

**3.7.6 Multicomponent Detection.** When a sensing instrument responds simultaneously to several analytes, one is faced with the problem of multicomponent detection and analysis. This is a very important situation in chemical analysis, having many facets and a large literature, including such topics as "errors-in-variables-regression" and "multivariate calibration"; but only a brief descriptive outline will be offered here. For the simplest case, where blanks and sensitivities are known and signals additive,  $S$  can be written as the summation of responses of the individual chemical components -- i.e.,  $S_i = \sum S_{ij} = \sum A_{ij} x_j$ , where the summation index- $j$  is the chemical component index, and  $i$ , a time index (chromatography, decay curves), or an energy or mass index (optical, mass spectrometry). In order to obtain a solution,  $S$  must be a vector with at least as many elements  $S_i$  as there are unknown chemical components. Two approaches are common: (1) When the "peak-baseline" situation obtains, as in certain spectroscopies and chromatography, for each non-overlapping peak, the sum  $\sum A x$  can be partitioned into



a one component "peak" and a smooth (constant, linear) baseline composed of all other (interfering) components. This is analogous to Eq. 2, and for each such peak, it can be treated as a pseudo one component problem. (2) In the absence of such a partition, the full matrix equation,  $S = Ax$ , must be treated, with  $x_{kc}$  and  $x_{kd}$  computed for component- $k$ , given the complete sensitivity matrix  $A$  and concentrations of all other (interfering) components. These quantities can be calculated by iteratively computing, from the Weighted Least Squares covariance matrix, the variance of component- $k$  as a function of its concentration, keeping all interfering components constant, and using the defining equations 9 and 12, or their normal forms, equations 10 and 13. Further discussion can be found Ref. 20 and references therein.

An additional topic of some importance for multicomponent analysis is the development of optimal designs and measurement strategies for minimizing multicomponent detection limits. A key element for many of these approaches is the selection of measurement variable values (selected sampling times, optical wavelengths, ion masses, etc.) that produce a sensitivity matrix  $A$  satisfying certain optimality criteria. Pioneering work in this field was done by Kaiser [30]; a review of more recent advances is given by Massart et al [31, ch 8.].

**3.7.7 Minimum Quantifiable Value; Quantification Limit ( $L_Q$ ).** Quantification limits are performance characteristics that mark the ability of a CMP to adequately "quantify" an analyte. Like detection limits, quantification limits are vital for the planning phase of chemical analysis; they serve as benchmarks that indicate whether the CMP can adequately meet the measurement needs.

The ability to quantify is generally expressed in terms of the signal or analyte (true) value that will produce estimates having a specified relative standard deviation (RSD), commonly 10 %. That is,

$$L_Q = k_Q \sigma_Q \quad (21)$$

where  $L_Q$  is the Quantification Limit,  $\sigma_Q$  is the standard deviation at that point, and  $k_Q$  is the multiplier whose reciprocal equals the selected quantifying RSD. The IUPAC default value for  $k_Q$  is 10. As with detection limits, the net signal quantification limit ( $S_Q$ ) and analyte (amount or concentration) quantification limit ( $x_Q$ ) derive from the relations in equations 1-5, and the variance structure of the measurement process. If the sensitivity  $A$  is known, then  $x_Q = S_Q/A$ ; if an estimate  $\hat{A}$  is used computing  $\hat{x}$ , then its variance must be considered in deriving  $x_Q$ . (See Note-1, below.) Just as with the case of  $S_D$  and  $x_D$ , uncertainties in assumed values for  $\sigma$  and  $A$  are reflected in uncertainties in the corresponding Quantification limits.

If  $\sigma$  is known *and constant*, then  $\sigma_Q$  in Eq. 21 can be replaced by  $\sigma_o$ , since the standard deviation of the estimated quantity is independent of concentration. Using the default value for  $k_Q$ , we then have

$$L_Q = 10 \sigma_Q = 10 \sigma_o \quad (22)$$

In this case, the quantification limit is just 3.04 times the detection limit, given normality and  $\alpha = \beta = 0.05$ .

Notes:

1. In analogy with  $x_D$ , the existence of  $x_Q$  is determined by the *RSD* of  $\hat{A}$ . In this case the limiting condition for finite  $x_Q$  is  $RSD(\hat{A}) < 1/k$ . If  $x$  is estimated with Eq. 5, and  $\hat{B}$  and  $\hat{A}$  are independent, and  $\sigma(\hat{S})$  is constant with value  $\sigma_o$ , then  $x_Q = (k\sigma_o/A)/[1-(k\sigma_o/A)^2]^{1/2}$ , where  $(k\sigma_o/A)$  is the limiting result when the random error in  $\hat{A}$  is negligible.]
2. One frequently finds in the chemical literature the term "Determination Limit." Use of this term is *not* recommended, because of ambiguity. It is sometimes employed in the sense of the critical level, for making detection decisions; sometimes, as the detection limit; and still others, as the quantification limit.

3.7.8 **Heteroscedasticity.** If the variance of the estimated quantity is not constant, then its variation must be taken into account in computing detection and quantification limits. The critical level is unaffected, since it reflects the variance of the null signal only. Two types of  $\sigma$  variation are common in chemical and physical metrology: (a)  $\sigma^2$  (variance) proportionate to response, as with shot noise and Poisson counting processes; and (b)  $\sigma$  (standard deviation) increasing in a linear fashion. Detailed treatment of this issue is beyond the scope of this document, but further information may be found in Ref. 20. To illustrate, given normality, negligible uncertainty in the mean value of the blank, and  $\sigma(L)$  increasing with a constant slope ( $d\sigma/dL$ ) of 0.04 -- equivalent to an asymptotic relative standard deviation (RSD) of 4 %, we find the following,

$$L_C = 1.645 \sigma_B \quad (23)$$

$$L_D = L_C + 1.645 \sigma_D, \quad \text{with} \quad \sigma_D = \sigma_B + 0.04 L_D \quad (24)$$

$$L_Q = 10 \sigma_Q, \quad \text{with} \quad \sigma_Q = \sigma_B + 0.04 L_Q \quad (25)$$

Solutions for equations 24 and 25 are  $L_D = 3.52 \sigma_B$ , and  $L_Q = 16.67 \sigma_B$ , respectively. Thus, with a linear relation for  $\sigma(L)$ , with intercept  $\sigma_B$  and a 4 % asymptotic RSD, the ratio of the quantification limit to the detection limit is increased from 3.04 to 4.73.

If  $\sigma$  increases too sharply,  $L_D$  and/or  $L_Q$  may not be attainable for the CMP in question. This problem may be attacked through replication, giving  $\sigma$  reduction by  $1/\sqrt{n}$ , but caution should be exercised since unsuspected systematic error will *not* be correspondingly reduced!

### 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.

3.8.1 **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- $\sigma$  is quite difficult to detect [20]. (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.

3.8.2 **Bias Uncertainty; Bias Bounds ( $\Delta_M$ ).** 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 we know the *cdf* and its parameters(s), we can form a confidence interval and upper limit, just as in the case of analytical results [3]. Scientific, or inferred, bias bounds are much more difficult to generate, for they require skilled and exhaustive scientific evaluation of the entire structure of the CMP. This should *never* be omitted; but, unfortunately it is rarely done.

3.8.3 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  $\nu$ , which equals the number of observations  $n$  minus the number of estimated parameters. Thus, for a simple set of observations

$$s^2 = \Sigma(x_i - \bar{x})^2 / (n-1) \quad (26)$$

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

For a fitted (straight-line) calibration curve,

$$s^2 = \Sigma(y_i - \hat{y}_i)^2 / (n-2) \quad (27)$$

Note that although the standard deviation equals the square root of the variance  $\sigma^2$ , 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  $[1+1/(4\nu)]$  [32].

3.8.4 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 *cdfs* 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:

$$\sigma_f^2 = \Sigma(\partial f / \partial x_i)^2 \sigma_{x_i}^2 \quad (28)$$

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

3.8.5 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  $\sigma^2$  equals the mean  $\mu$ . Except for the case of relatively few counts, using the observed number of counts as an estimate of the variance is quite adequate.

3.8.6 Estimated Variance Bounds. If the observations are distributed normally,  $s^2/\sigma^2$  is distributed as  $\chi^2/\nu$ . An 95% interval estimate for this ratio is therefore given by

$$(\chi^2/\nu)_{.025} < s^2/\sigma^2 < (\chi^2/\nu)_{.975} \quad (29)$$

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

Note: Eq. 29 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. A theoretical analysis of this and other approximations, in comparison to the exact expression which derives from the non-central  $t$  distribution has been given by Vangel [33]. This has special relevance for the Quantification Limit, since the definition of  $L_Q$  is based on a prescribed value for the RSD.

3.8.7 Estimated Sensitivity ( $\hat{A}$ ). The slope (sensitivity) and intercept of the calibration curve are generally estimated using Ordinary Least Squares [3]. 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: Comments on Parameter Estimation (Least Squares Fitting). 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  $F$ ,  $\chi^2$ , 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. (See Natrella [17], for more on statistical vs functional relationships.)

3.8.8 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 (*cdfs*) 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  $\sigma_b$  is not well known. (See section 3.7.3.2)

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 sample preparation 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 [20, ch. 9]. The "external" approach consists of the generation and direct evaluation of a series of ideal or surrogate blanks for the overall measurement process, using 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.

#### 4. COMPOUND CMPS -- THE INTERLABORATORY ENVIRONMENT

When a measurement process consists of two or more segments, it can be properly characterized as a Compound (Chemical) Measurement Process. Specification of the Performance Characteristics of a compound or hierarchical CMP depends upon one's viewing point or position in the hierarchy. That is, at least for the "tree" structure, all segments below the viewing (or "null") node consist of multiple branches or replicates. For the CMP that is in a state of statistical control these replicates yield a crucial measure of random error. (The CMP that is *not* in a state of control is undefined!) Only a single path lies above the null node; this path necessarily fixes the bias of the CMP. By moving up in the hierarchy, one has an opportunity to convert bias into imprecision -- put another way, what is viewed as a fixed (albeit unknown) error at one level of a compound CMP, becomes random at a higher level. This is very important, for random error may be estimated statistically through replication, but bias may not; yet inaccuracy (total error) necessarily comprises both components. Figure 3 presents these concepts schematically [34].

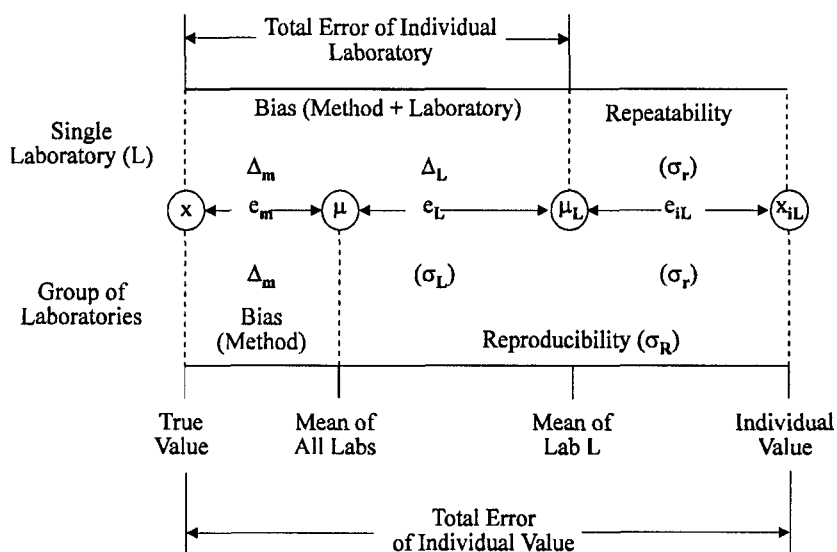


Fig. 3 Partitioning of method, interlaboratory, and intralaboratory error. (Adapted from Ref. 34)

Collaborative or interlaboratory tests, which under the best of circumstances may be found at the uppermost node of the Compound CMP, provide one of the best means for accuracy assessment. In a sense, such intercomparisons epitomize W.J. Youden's recommendation that we vary all possible factors (that might influence analytical results), so that the observed dispersion can give us a direct experimental (statistical) measure of inaccuracy [35]. The basic concept, as indicated in Fig. 3, is that fixed intralaboratory biases are converted into random errors from the interlaboratory perspective. If the overall interlaboratory mean is free from bias, then the observed interlaboratory dispersion is the measure of *both* imprecision and inaccuracy.

4.1 Sampled Population [S] vs Target Population [T]. The above concept has been captured by Natrella [17] by reference to two populations which represent, respectively, the population (of potential measurements) actually sampled, and that which would be sampled in the ideal, bias-free case. The corresponding *S* and *T* populations are shown schematically in Fig. 4, for a two-step measurement

process. When only the  $S$  population is randomly sampled (left side of the figure), the error  $e_1$  from the first step is systematic while  $e_2$  is random. In this case, the estimated uncertainty is likely to be wrong, because a) the apparent imprecision ( $\sigma_s$ ) is too small, and b) an unmeasured bias ( $e_1$ ) has been introduced. Realization of the  $T$ -Population (right side of the figure) requires that all steps of the CMP be random -- ie,  $e_1$  and  $e_2$  in the figure behave as random, independent errors;  $T$  thus represents a Compound Probability Distribution. If the contributing errors combine linearly and are themselves normal, then the  $T$ -distribution also is normal. The concept of the  $S$  and  $T$  populations is absolutely central to all hierarchical measurement processes (Compound CMPs), whether intralaboratory or interlaboratory. Strict attention to the concept is essential if one is to obtain consistent uncertainty estimates for Compound CMPs involving different samples, different instruments, different operators, or even different methods. In the context of (material) sampling, an excellent exposition of the nature and importance of the hierarchical structure has been presented by Horwitz [2].

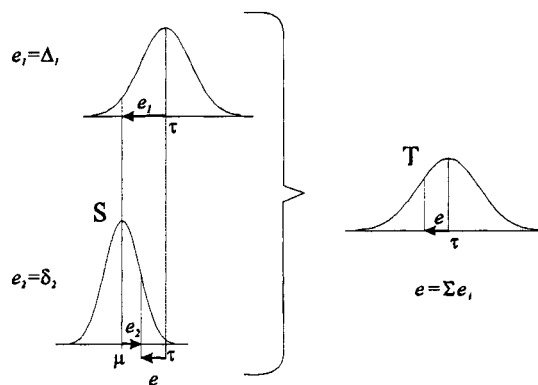


Fig. 4 Sampled [S] and Target [T] Populations.

From the interlaboratory perspective, the first population in Fig. 4 ( $e_1$ ) would represent the distribution of errors among laboratories; the second [S] would reflect intralaboratory variation ("repeatability"); and the third [T], overall variation ("reproducibility") [7,36,37]. If the sample of collaborating laboratories can be taken as unbiased, representative, and homogeneous, then the interlaboratory "process" can be treated as a compound CMP. In this fortunate (doubtless asymptotic) situation, results from individual laboratories are considered random, independent variates from the compound CMP population. For parameter estimation (means, variances) in the interlaboratory environment it may be appropriate to use weights -- for example, when member laboratories employ different numbers of replicates [38].

**4.2 Types of Interlaboratory (Collaborative) Tests.** Interlaboratory studies commonly have one of the following objectives: (1) method evaluation, (2) method comparison, (3) proficiency testing, or (4) establishment of reference values. Each must have its own design and approach to evaluation of the resulting data [36]. Detailed discussion of the various classes of interlaboratory exercises is beyond the scope of this report. The topic is raised because our charge to present nomenclature appropriate for the characterization of the Chemical Measurement Process demands consideration of the Compound CMP, for which the collaborative test is the ultimate manifestation. A separate document, "Nomenclature for Interlaboratory Analytical Studies" [39] provides in-depth treatment of this topic. Also, the reader's attention is directed to excellent documents concerning collaborative testing for the Certification of Reference Materials [8] and for the Validation of Characteristics of a Method of Analysis [36].

## ACKNOWLEDGMENT

Documentation and nomenclature available from national and international organizations, especially ISO, IUPAC, OIML, ASTM, and AOAC, formed the basis for much of the discussion and definitions in this text.

## REFERENCES

1. Eisenhart, C. "Realistic Evaluation of the Precision and Accuracy of Instrument Calibration Systems," *J. Res. NBS*, 67C (1963) 161.
2. Horwitz, W., IUPAC Commission on Analytical Nomenclature, "Nomenclature for Sampling in Analytical Chemistry," *Pure & Appl. Chem.*, 62 (1990) 1193.

3. Currie, L. A. and Svehla, G., IUPAC Commission on Analytical Nomenclature, "Recommendations for the Presentation of Results of Chemical Analysis," *Pure & Appl. Chem.*, 66 (1994) 595.
4. Wilson, A.L., "Performance Characteristics of Analytical Methods," *Talanta*, 17 (1970) 21, 31; 20( 1973) 725; 21 (1974) 1109.
5. ISO, *International Vocabulary of Basic and General Terms in Metrology*, International Organization for Standardization (1993).
6. IUPAC, "Orange Book," H. Freiser and G. H. Nancollas, Ed., *Compendium of Analytical Nomenclature*, 2nd edition, Blackwell Scientific Publ., Oxford (1987); 1st edition (1978).
7. Parkany, M. "Parameters and Definitions," *Second International Sympos. on the Harmonization of Collaborative Analytical Studies* (1984). Includes 7 annexes (ISO, IEC, OIML, BIPM, IUPAC) treating vocabulary, statistics, analytical nomenclature, precision of test methods, and guidelines for collaborative tests.
8. ISO Guide 35, *Certification of Reference Materials -- General and Statistical Principles* (and references therein), International Organization for Standardization (1989).
9. ISO Guide 33, *Uses of Certified Reference Materials for the Assessment of a Measurement Procedure*, International Organization for Standardization (1989).
10. Currie, L. A., "The Limitations of Models and Measurements as Revealed through Chemometric Intercomparison," *J. Res. NBS*, 90 (1985) 409.
11. Zweig, M. H., "Establishing Clinical Detection Limits of Laboratory Tests," ch. 8 in Currie, Ed., *Detection in Analytical Chemistry: Importance, Theory, and Practice*, ACS Sympos. Series 361, American Chemical Society, Washington (1988).
12. Giacomo, P., "Expression of Experimental Uncertainties," *Metrologia* 17 (1981) 69.
13. ISO, *Guide to the Expression of Uncertainty*, International Organization for Standardization (1993).
14. International Federation of Clinical Chemistry, *Clin. Chim. Acta* 98 (1979) 129; *Pure Appl. Chem.* 51 (1979) 2451.
15. OIML, *Vocabulary of Legal Metrology*, International Organization of Legal Metrology (1978).
16. Taylor, J.K., *Principles of Quality Assurance of Chemical Measurements*, NBSIR 85-3105, National Bureau of Standards, Washington, DC (1985).
17. Natrella, M., *Experimental Statistics*, NBS Handbook 91, U.S. Government Printing Office, Washington (1963).
18. Heydorn, K., *Neutron Activation Analysis for Clinical Trace Element Research*, Chapt 5 (Precision), Chapt 6 (Accuracy), CRC Press, Boca Raton, FL (1985).
19. IAEA, "Quality assurance in biomedical neutron activation analysis," *Anal. Chim. Acta* 165 (1984) 1.
20. Currie, L. A., Ed., *Detection in Analytical Chemistry: Importance, Theory, and Practice*, ch. 1, ACS Sympos. Series 361, American Chemical Society, Washington (1988).
21. ISO Standard 3534-1993, *Statistics - Vocabulary and Symbols*, International Organization for Standardization (1993). See also, ISO Standards Handbook 3, *Statistical Methods* (1989).
22. Parr, R., Houtermans, H., and Schaerf, in *Computers in Activation Analysis and Gamma-ray Spectroscopy*, US Dept of Energy, CONF-780421 (1979) 544.
23. Hopke, P. K. and Massart, D. L., "Reference data sets for chemometrical methods testing," *Chemometrics and Intelligent Laboratory Systems* 19 (1993) 35.
24. Currie, L. A., "Limits for Qualitative Detection and Quantitative Determination," *Analyt. Chem.* 40 (1968) 586.
25. ISO-IUPAC, *Limits of Detection: Nomenclature Harmonization Meeting*, P.-Th. Wilrich (ISO), M. Parkany, chairman (IUPAC), L. A. Currie, Rapporteur, Washington, DC (1993). See also, Hartmann, E., "Erfassungsvermögen von Analysenverfahren," *Fresenius Z. Anal. Chem.*, 335 (1989) 954.
26. Currie, L.A., "The Many Dimensions of Detection in Chemical Analysis, with Special Emphasis on the One-Dimensional Calibration Curve," Chapt. 5 in *Trace Residue Analysis: Chemometric estimations of sampling, amount, and error*, D. Kurtz, Ed., (American Chemical Society), ACS Symposium Series 284 (1985) 49.

27. ASTM, "Standard Practice for Interlaboratory Quality Control Procedures and a Discussion on Reporting Low-Level Data," *ASTM Standard D 4210-89*, American Society for Testing and Materials (1989).
28. Liteanu, C. and Riča, I., *Statistical Theory and Methodology of Trace Analysis*, Wiley (New York, 1980).
29. Currie, L. A., in *International Conf. on Environmetrics and Chemometrics*, Las Vegas, NV (1995). (Proceedings to be published in *Chemometrics and Intelligent Laboratory Systems*.)
30. Kaiser, H., *Z. Anal. Chem.* 216 (1966) 80; 260 (1972) 252.
31. Massart, D. L., Vandeginste, B. G. M., Deming, S. N., Michotte, Y., and Kaufman, L., *Chemometrics: a textbook*, Elsevier, Amsterdam (1988).
32. Dixon, W. J. and F. J. Massey, *Introduction to Statistical Analysis*, 3rd edition, McGraw-Hill, New York (1969).
33. Vangel, M., "Confidence Intervals for a Normal Coefficient of Variation," *American Statistician* (accepted, 1995).
34. Currie, L.A., "Sources of Error and the Approach to Accuracy in Analytical Chemistry," I.M. Kolthoff, P.J. Elving, Ed., *Treatise on Analytical Chemistry*, vol.I, ch. 4 (1978).
35. Youden, W.J., "Uncertainties in Calibration," ch. I-5, in H.H. Ku, Ed., *Precision Measurement and Calibration*, NBS Spec. Publ. 300, U.S. Government Printing Office, Washington, DC (1969).
36. AOAC: "Guidelines for Interlaboratory Collaborative Study Procedure to Validate Characteristics of a Method of Analysis," *J. Assoc. Off. Anal. Chem.*, 67 (1984) 433.
37. ASTM: "Standard Terminology for Statistical Methods," E 456-83a, *1984 Annual Book of Standards*, Sect. 14, p 402; and E 691-79, "Standard Procedure for Conducting an Interlaboratory Test Program to Determine the Precision of Test Methods," *ibid.*, p 556
38. Mandel, J. and R.C. Paule, "Interlaboratory Evaluation of a Material with Unequal Numbers of Replicates," *Analyt. Chem.*, 42 (1970) 1194.
39. Horwitz, W. and Parkany, M., IUPAC Commission on Analytical Nomenclature, "Nomenclature for Interlaboratory Analytical Studies," *Pure & Appl. Chem.*, 66 (1994) 1903.

## INDEX OF TERMS

- |   |   |
|---|---|
| <p>Accuracy - 3.5, 3.5.11<br/>           Analyte - 1.2<br/>           Analytical methods, procedures, protocols, techniques - 2.<br/>           Assumption testing - 3.6.1</p> <p>Bias - 3.5.5<br/>           Bias bounds - 3.8.2<br/>           Bias estimation - 3.8.1<br/>           Blank (background, baseline) - 3.2, 3.8.8<br/>           Blunder - 3.5.9</p> <p>Calibration function - 3.2, 3.7.5<br/>           Chemical Measurement Process - 1, 2.3, 3.1, 3.7.2<br/>           Collaborative and interlaboratory tests - 3.6.3, 4., 4.2<br/>           Confidence ellipse - 3.8.7<br/>           Control - 3.6<br/>           Control, external - 3.6.3<br/>           Control, internal - 3.6.2<br/>           Critical level, value - 3.7.1, 3.7.3.1<br/>           Cumulative distribution function - 3.5.6</p> | <p>Definitive method - 2.2<br/>           Degrees of freedom - 3.7.3, 3.8.3<br/>           Detection and quantification capability - 3.7<br/>           Detection decision - 3.7.1, 3.7.3.1<br/>           Detection limit - 3.7.1, 3.7.3.2<br/>               multicomponent - 3.7.6<br/>           Detection terminology - 3.7.1<br/>           Discrimination Limit - 3.6.1 (note 2)</p> <p>Error - 3.5.3<br/>           Error (variance) propagation - 3.8.4<br/>           Error, random - 3.5.6<br/>           Error, systematic - 3.5.9<br/>           Estimation - 3.2.1, 3.8<br/>           Estimator - 3.2.1, 3.5.1<br/>           Evaluation function - 3.3<br/>           Evaluation operator - 3.3<br/>           Expectation, expected value - 3.2.1, 3.5.4</p> |
|---|---|



- Heteroscedasticity – 3.7.8  
Hypothesis testing – 3.6.1, 3.7.3  
  type-I error, false positive – 3.7.3  
  type-II error, false negative – 3.7.3  
  significance level, power, operating characteristic – 3.6.1
- Imprecision – 3.5.10  
Inaccuracy – 3.5.11
- Limiting mean (expectation) – 3.5.4
- Measurable quantity – 1.1  
Measurand (measured quantity) – 1.2, 3.5.9  
Measured result – 3.5.1  
Minimum detectable concentration – 3.7.5  
Minimum detectable signal – 3.7.4  
Minimum detectable value (detection limit) – 3.7.1, 3.7.3.2  
Minimum quantifiable value (quantification limit) – 3.7.1, 3.7.7  
Minimum significant value (critical level, value) – 3.7.1, 3.7.3.1  
Model – functional, structural (statistical) – 3.8.7
- Parameter estimation – 3.8.7  
Performance characteristics – 3., 3.1, 3.4, 3.5  
Population: sampled, target – 4.1  
Precision – 3.5, 3.5.10
- Qualitative analysis – 1.1  
Quality metrology – 3.6  
Quantification limit – 3.7.1, 3.7.7, 3.8.6 (note)  
Quantitative analysis – 1.1
- Reference material, Certified reference material,  
  Standard reference material – 3.1, 3.6.2, 3.6.3  
Reference method – 2.2  
Relative standard deviation – 3.8.6 (note)  
Repeatability – 3.6.2.1  
Reproducibility – 3.6.3.1  
Response – 3.2  
Ruggedness – 3.1
- Sensitivity – 3.2, 3.7.5, 3.7.6, 3.8.7  
Signal (net; observed) – 3.2  
Significance level, test – 3.6.1, 3.7.3.1  
Standard test data – 3.1, 3.6.3.1  
Standard deviation – 3.5.7, 3.8.3  
Standard uncertainty – 3.5.9
- Threshold – 3.7.3.1  
Traceability – 2.1  
True value – 3.5.2
- Uncertainty – 3.5.9, 3.5.11, 3.8.2
- Variance – 3.5.8, 3.8.3  
Variance bounds – 3.8.6  
Variance estimation – 3.8.3  
Variance, Poisson – 3.8.5
- Weights – 3.7.5.2, 3.8.7, 4.1