*Pure & Appl. Chem.*, Vol.54, No.4, pp.715-754, 1982.
Printed in Great Britain.

0033-4545/82/040715-40$03.00/0
Pergamon Press Ltd
IUPAC

# QUALITY OF ANALYTICAL RESULTS, WITH SPECIAL REFERENCE TO TRACE ANALYSIS AND SOCIOCHEMICAL PROBLEMS*

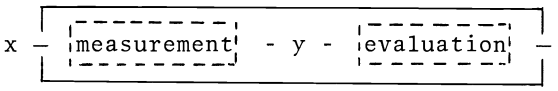
L. A. Currie

Center for Analytical Chemistry, National Bureau of Standards, Washington, D. C. 20234

Abstract - High quality trace analysis is becoming increasingly important for technological development, both with regard to the production and monitoring of high-purity technological materials and processes, and with respect to monitoring and understanding the environmental and societal impacts of industrial waste products. The analytical scientist faces an enormous challenge in meeting these requirements because of the range of concentrations (to less than $10^{-12}$ g/g) and complexity of matrices, as well as the importance of the results to the future of mankind.

The quest for accuracy in trace analysis is best viewed in the framework of a structured *Chemical Measurement Process* (CMP), which may be symbolized as follows:

CHEMICAL MEASUREMENT PROCESS




That is, given a sample containing the analyte at concentration $x$, the analyst carries out a *CMP* which includes the observation of an analytical signal $y$ and the computation of an estimated concentration $\hat{x}$ and its uncertainty $U_{\hat{x}}$. The best means to assure high quality for the overall *CMP* is to (a) perform regular assays of knowns (Certified Reference Materials) and interlaboratory comparison samples, and (b) examine and bring into control each constituent step. Illustrations are presented of assumptions and common pitfalls which are characteristic of each of the *CMP* steps, with special emphasis on those which arise in trace analysis, such as: contamination, losses and interference, calibration and evaluation-model errors, and information loss due to inadequate reporting of results and uncertainties. The question of hypothesis testing and detection limits is given special focus.

A more formidable challenge faces the analytical chemist who would consider the *overall* analytical problem – i.e., treating the broader socio-scientific question involving both the fundamental design of the measurement program (why, what, how to measure) and the interpretation of the resulting observations. Such an augmented *CMP* may be designated as the *Analytical Measurement Process (AMP)*; it is one which calls for the analyst to be intimately involved in an interdisciplinary experimental design effort (prior to the *CMP*) as well as a corresponding evaluation effort (following the CMP). Two examples, based on current programs, are presented: (a) the use of "chemical fingerprints" to identify sources of environmental contamination (Receptor Modeling), and (b) the establishment of a national environmental specimen bank (Archival Sampling).

*Contribution of the National Bureau of Standards. Not subject to copyright.

## INTRODUCTION

Approaches and constraints to technological development are coming to depend increasingly on high quality chemical analysis, especially at trace levels and in both relatively pure materials and complex geological, biological and environmental media. In planning the production of food and energy, for example, we must assess the environmental pathways and further evaluate chemical effects on biological, ecological and climatic systems. Chemical measurements, generally at the trace level, form a critical part of such assessments both for the discovery of changes in the environmental system and for the construction and validation of models used for estimating future system changes.

The analytical scientist has a vital role to play in helping to define the technical or chemical components of the societal problems associated with the development of mankind, and he has primary responsibility for (a) defining the *Analytical Measurement Process* (AMP) which [explicitly] links physicochemical data with the technical information required by decision makers (such as the identification of pollutant sources or the effects of carbon dioxide production on climate), and (b) providing such data through the use of *Chemical Measurement Processes* (CMP) of adequate sensitivity, precision and accuracy. Failure either to correctly link (model) the environmental (or other external) system with chemical observables or to generate high-quality chemical data can result in erroneous societal decisions or dangerous reliance on "theory" (1). Following a brief review of the scope of modern trace analysis and its relation to societal decision-making, we shall examine in the context of the CMP and the AMP the steps which may be taken to generate reliable chemical data and subsequent scientific conclusions.

## SOCIOCHEMICAL PROBLEMS AND THE SCOPE OF MODERN TRACE ANALYSIS

### Chemistry and Society

As the human society develops in number and complexity, its perturbations of the natural environment begin to approach or exceed the natural variations themselves. This is especially evident in the chemical cycles, where for example fossil fuel combustion has produced about a 15% increase in atmospheric $CO_2$ during the past century, and where completely new and persistent substances (chlorofluorocarbons, plutonium, chlorinated polynuclear hydrocarbons) are being added to the environment.

Although such perturbations may have serious consequences, they generally are difficult either to predict reliably or to detect experimentally – especially if they are characterized by long time constants. As a result, man is faced with an entirely new class of problems having complicated societal and scientific aspects, as indicated in Fig. 1 (1-3).

The chief characteristics of these new problems are their diversity and their difficulty. That is, they span disciplines, time, and space; and major uncertainties and difficulties are associated both with the sociopolitical implications and with scientific measurement or forecasting. It is important for the scientist to realize that the sociopolitical problem is by far the more complex – including diverse special interests and perceptions, regional and temporal inequities, and the need to make timely decisions in the face of uncertainty (2-4). It is likewise important to appreciate the facts that many of the decisions are based strictly on scientific theory (models) (5,6), and that errors (especially systematic errors) connected with models or with chemical measurements are frequently assumed trivial or at worst transitory by the general public (1). Experience, of course, belies such blind reliance on the quality of our models and our data. For example: modeling the effects of agricultural chemicals and supersonic aircraft on the stratospheric ozone led to erroneous conclusions because of assumed values for certain chemical rate constants (7); in connection with the "carbon dioxide problem" models support opposing views on the increase of the terrestrial biosphere (8); erroneous clinical laboratory data, with potentially dangerous diagnostic consequences, continue to appear in interlaboratory comparisons (9,10); and inaccurate pollutant measurement procedures have led to overly stringent $NO_x$ controls (11) and misleading assessments of "photochemical oxidant" (ozone) pollution between neighboring regions (12).

## MANKIND'S DEVELOPMENT AND CHEMICAL PERTURBATIONS OF THE NATURAL SYSTEM

**Impacts** • food, climate, health, resources (e.g., $CO_2$, CO, particles, nuclear and toxic wastes)

**Features** • interdisciplinary, intergenerational, global, model-dependent, surprises, decisions under uncertainty

**Limiting Aspects** • sociopolitical > health effects, modeling > chemical measurements

**Roles of the Analyst** • interdisciplinary problem assessment
• archival samples
• quality methods and timely data for monitoring, and model-parameterization and validation

Fig. 1.

In view of these facts, the trace analyst has a number of very important contributions to make. First, because of the interdisciplinary nature of these sociochemical problems, it is vital that he participate in the overall definition and assessment of the problem in order to effectively identify the specific scientific (modeling, measurement) needs and limitations at the outset. The importance of interdisciplinary problem definition is increasingly illustrated by topical workshops attended by a broad class of experts, ranging from economists to biological and physical scientists. Instructive examples include: the assessment of the technological, economic, environmental and institutional incentives and barriers associated with hydrogen as an energy carrier (13); the examination of the scientific and societal issues linking the use of fossil fuel, the carbon cycle, and climate (14); and the establishment of an international environmental monitoring program of biological sampling and specimen banking (15). This last example illustrates a second type of role for the analytical scientist which is assuming increasing importance: the collection and analysis of archival samples. Both natural archives (ice cores, ocean sediment, tree rings, ...) and "anthropogenic" archives (collections of artifacts, meteorites, atmospheric gases and particles, biological specimens, ...) carry chemical information about the past which seldom can be retrieved in any other way. Information on the paleoatmosphere, for example, has been derived from air trapped in the polar ice cores (16-18), and tritium hydrology was established through the analysis of vintage wines (19). From the perspective of trace analysis, archival samples serve as a critical resource to improve the quality of our knowledge as improved methods of analysis become available. In addition, they make possible the use of retrospective monitoring, when new or unsuspected environmental contaminants are discovered. As we shall see later, acquisition and preservation of high-quality chemical archives takes no little skill on the part of the analyst.

Finally, the development of high-quality methods of trace analysis, specially-designed to meet the needs identified in the process of interdisciplinary problem assessment, represents the classical but evermore important responsibility of the analyst. The importance and difficulty of this task will be highlighted in the discussion of the scope of modern trace analysis, which follows, and it is underscored by the enormous reliance on models (as opposed to data) in addressing current public science policy issues. A decision made within just the last few months illustrates the dilemma: "because industry should not be held to a higher standard of purity than nature", the U.S. Environmental Protection Agency is seeking to "model nature's chemistry" in order to infer the contribution of tree emissions (isoprene, α-pinene) to urban photochemical smog (20). Models can serve as indispensible guides, but the results of "theory" are too easily accepted as truth by the lay public. So, it behooves the analytical scientist to perfect methods for obtaining high-quality experimental data to supplant, supplement, or at the very least

to verify the presumed models. Equally important is the provision of realistic estimates of uncertainty for experimental or model results.

The Scope of Trace Analysis

Until the last few years trace analysis measurements were limited to one part in $10^6$ (ppm) or at most one part in $10^9$ (ppb). With the development of new instrumentation and the application of new physical principles, however, trace chemical analysis has seen a revolution leading to a doubling of the exponent ($1:10^{17}$) and even "single-atom detection" (21). (A major factor in this revolution has been the development of multi-stage instruments incorporating extremely sensitive and specific detectors and/or extremely powerful and specific exciters – e.g., mass spectrometers as detectors and laser radiation sources.) One consequence has been the discovery of all sorts of contaminants previously unsuspected and a re-examination of legislation (or regulations) which equated absence to "not detected". In order to set the stage for the subsequent discussion concerning the production of quality trace analytical data, specific examples illustrating the present concentration range are given in Table 1. Before discussing the several entries in the table, two general observations may be made: (a) important contemporary problems may be readily identified corresponding to nearly every concentration decade from $10^{-6}$ to $10^{-17}$ *(Note a)*; and (b) as two of the examples show, our concerns are not limited just to small concentrations or quantities of chemical species, but to small *differential* concentrations. Thus, it is appropriate to consider "differential trace analysis" within our purview. The importance of differential trace analysis, which necessarily implies high precision measurement, is that it applies to "just noticeable differences" in concentration which may signal important biological or environmental changes (induced by man).

TABLE 1. The Scope of Trace Analysis: Health and the Environment[a]

| Substance | Concentration | Substance | Concentration |
|---|---|---|---|
| $\Delta CO_2$ (seawater) | $10^{-4}$/yr | Pt (liver) | $5 \times 10^{-11}$ |
| Pesticides (food) | $10^{-6}$ | Dioxin (milk) | $10^{-12}$ |
| $\Delta O_2$ (air) | $10^{-7}$/yr | $^{36}Cl$ (water) | $10^{-13}$ |
| Al (ice core) | $10^{-8}$ | $^{14}C$ (atmospheric particles) | $10^{-14}$ |
| Tl (liver) | $10^{-9}$ | All organics (water) | $10^{-15}$ |
| Aflatoxin (food) | $10^{-10}$ | $^{13}CD_4$ (air) | $10^{-17}$ |

[a]References for each substance are given in the text. (See also *Note a.*)

Highlights concerning the entries in Table 1 follow.

$\Delta CO_2$. An important factor in understanding the global $CO_2$ cycle is the rate of uptake in the oceans. The "secular increase in dissolved inorganic carbon ($\Sigma CO_2$)" has been modeled as ∿0.35 percent per decade, but the (annual) change is not yet measurable (8).

Pesticides. Residues of organochlorine and organophosphorus pesticides must be measured at ∿ppm concentrations in foods; overall imprecision (RSD) has been found from collaborative studies to be about 15% (22).

*Note a.* Concentrations throughout the text are expressed in dimensionless units -- e.g., g/g.

$\Delta O_2$. A bold experiment is underway to estimate indirectly the total annual atmospheric production of $CO_2$ by experimentally measuring the decrement in atmospheric oxygen. (Measurement of the atmospheric increment of $CO_2$ yields only the "airborne fraction" of that production.) The best existing measurement *precision* (differential detection capability) is inadequate by about a factor of ten (23).

Al (ice). Trace element impurity patterns in cores taken from the Greenland and Antarctic ice sheets ("natural archives") are of special interest as a source of paleoenvironmental information on discrete events, as well as for pollutant trends and seasonal variations. Aluminum is important in that it serves as an excellent tracer for continental dust (24).

Tl (liver). Thallium was identified as one of the "priority-1" trace elements of special environmental interest by the International Specimen Banking Workshop (15); its concentration in human liver is among the lowest in this priority class (25).

Aflatoxins. These are toxic mold metabolites which are monitored in foods such as peanuts and eggs. At the 0.1 to 1.0 ppb level overall imprecision is about 40% (22).

Pt (liver). A "priority-2" level trace element (15), platinum exhibits one of the lowest elemental concentrations yet measured in biological samples (26).

2,3,7,8-tetrachlorodibenzo-p-dioxin has been characterized as the most toxic small molecule toxin known. Occurrence at the ppt level is already very serious; but collaborative studies have indicated significant difficulty with false positives and false negatives in complex matrices such as milk below about 9 ppt (27).

$^{36}Cl$ (groundwater). This cosmic ray produced nuclide with its $3.01 \times 10^5$ year half-life has special value for the dating of old groundwater, which in turn is important for identifying underground reservoirs that might be suitable for long-term storage of nuclear waste (28). It is noteworthy that the extremely low concentrations of this radionuclide have been measured by *chemical* rather than *radiochemical* means. The measurement process – accelerator mass spectrometry [AMS] (to be further discussed in the following section) – treats the $^{36}Cl$ nuclei exactly like stable nuclei (which of course they are, prior to decay).

$^{14}C$ (atmospheric particles). Naturally-occurring radiocarbon is useful for distinguishing fossil from biogenic pollutant sources, and its measurement takes place by both chemical (AMS) and radiochemical detection schemes (29).

All organics (water). Lamparski, quoting Donaldson (30), notes that "every known organic compound could be detected in water at a level of $10^{-15}$ g/g or higher". This has profound implications for modern trace analysis, in terms of sensitivity *vs.* specificity, for an analyte "at the 10 ppt [$10^{-11}$] concentration level in a sample matrix that is 99.9% pure [could have] interferences from as many as $10^5$ compounds at concentrations $10^3$ times higher" (31).

$^{13}CD_4$ (air). Isotopically-unique "heavy methane" is so rare in the normal atmosphere that a tiny point-source injection can be used successfully for long range ($\sim 10^3$/km) air trajectory model validation. Measurements of this substance, as noted at "The Search for Zero" mass spectrometry symposium [Oct. 1980], represents one of the more extreme challenges to trace analytical chemistry (32,54).

Although in the last decade or so the lower limit for "trace analysis" has been altered by a factor of $10^{11}$ ($10^{-6}$ to $10^{-17}$), it is unlikely that so great a change will occur again. That is, the discipline is already approaching the natural limit set by Poisson noise. If specificity and contamination are not limiting, and if one wishes an imprecision [RSD] of 1% and has an overall recovery-detection efficiency of 1% and initial sample size of 1 mole (host matrix), then the requisite analyte concentration ratio is about $10^{-18}$,

since Poisson counting statistics require $10^4$ detected analyte molecules. Recovery-efficiency changes can alter this by a factor of 100 at most, though non-contaminating enrichment (preconcentration) could significantly extend the range. Noble gases provide the best opportunity in this regard, and in what has been described as one of the most "heroic" experiments in modern radiochemistry, measurements have taken place at the level of $10^{-29}$! [In the Brookhaven Solar Neutrino Experiment, a 100-day exposure leads to the destruction of about thirteen chlorine-37 atoms – determined by the $^{37}Ar$ simultaneously produced – in 400,000 liters of tetrachloroethylene, which is equivalent to more than a megamole of $^{37}Cl$ (33).]

## THE CHEMICAL MEASUREMENT PROCESS (CMP)

The CMP is the key to quality for analytical data. This process, which comprises all the steps which connect the estimated (reported) analyte concentration with its true value, must be rigorously defined, brought under control and exhaustively tested before quality results are posited. Besides internal control and validation procedures, to be discussed below, one of the most effective methods of external validation is the interlaboratory (or inter-method) comparison. While consistency among independent laboratories or independent methods is a necessary condition for accuracy, it may not be sufficient. (Sufficiency may be assumed, however, if *wholly independent* methods can be devised.) In order to make a direct test for accuracy it is desirable to apply the measurement process in question to one or more reference samples whose compositions are known (within acceptable bounds) and similar to those of samples to be measured. With the exception of standards designed specifically for instrument calibration, however, the result obtained from a reference sample must *never* be used to empirically "correct" the overall measurement process. Rather, discrepancies should serve to initiate the search for unanticipated sources of error.

Intercomparisons are nearly always revealing. One example, which illustrates both interlaboratory and standard reference material [SRM] comparison, is given in Fig. 2a (34) *(Note b)*. By plotting results obtained by a given set of laboratories for two different samples in a modified "Youden diagram" (35), we can instantly discern systematic and random error components as well as occasional erratic results (blunders). The line drawn in Fig. 2a has special significance. Its location is totally independent of the points – i.e., it is not "fitted". It has been drawn through the "true value" region as given by the SRM's, and its slope has been fixed at 45°. Deviations along the line thus represent proportionate, interlaboratory (systematic) error and displacements from the line give a measure of intralaboratory precision. As is frequently the case, here the interlaboratory component far outweighed the intralaboratory component. Although the average bias (mean for all of the laboratories) was relatively small in this example, that is not always the case. The laboratory component of error will sometimes arise from a non-random process (such as a mistake in theory) and the resulting Youden diagram will then exhibit clusters of points whose means may be well-displaced from the "truth" [Fig. 2b (36,37)]. The Youden diagram is also helpful for blunder identification. The single result in Fig. 2a obtained without replication (point X), is far removed from the line, as a result of a mistake in measuring either the coal or the fly ash sample. Clearly, replication is one of the most effective safeguards against unanticipated blunders.

---

*Note b.* The abbreviation SRM (Standard Reference Material) is used to denote materials which have been certified by the U.S. National Bureau of Standards. The internationally-accepted generic terminology for such materials is now "Certified Reference Materials".

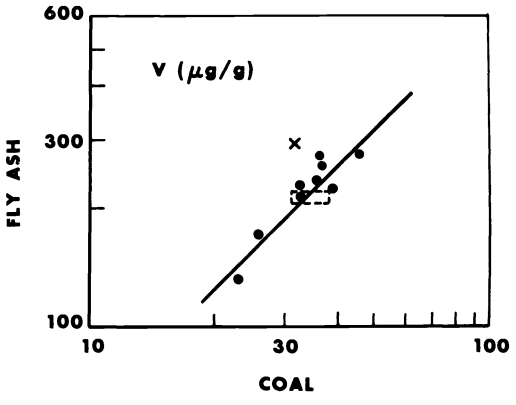



Fig. 2a. Results of an NBS-EPA vanadium intercomparison. The dashed region indicates the certified values for the two Standard Reference Materials, and point-X represents a result which lacks replication. [Ref. 34].

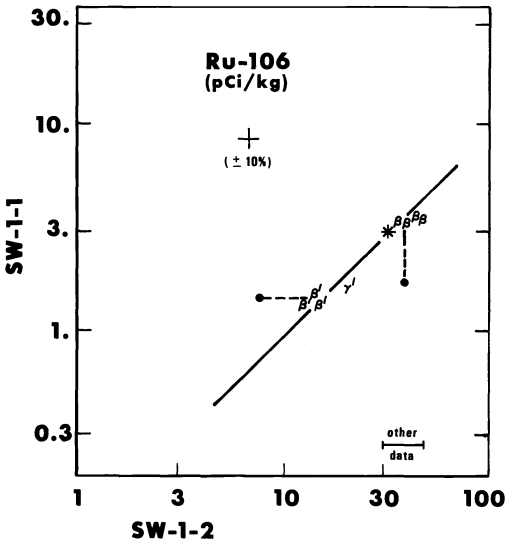



Fig. 2b. Sea water intercalibration plot showing high precision pairs of results for samples SW-1-1 and SW-1-2, [IAEA value indicated by asterisk] and 45° systematic error correlation line. Dashed lines indicate corrections for erratic blunders. Adapted in part from data in reference 36 by permission of the International Atomic Energy Agency [Ref. 37].

CMP Structure and Assumptions

The above example illustrates a useful diagnostic tool for evaluating the nature and magnitude of errors which may be encountered in the trace analysis of complex materials. It does not, however, speak to our main problem: the identification and control of error sources in order to assure quality. That is best addressed in the context of the CMP. Fig. 3 gives a generalized representation of the structure and principal characteristics of the CMP, especially as related to trace analysis. As seen in the figure, the primary steps of sampling, sample preparation (including concentration, purification, mounting for instrumental measurement) and measurement lead to a "signal" $y$. Subsequent steps (evaluation of the signal or data reduction, and reporting) then yield the estimated result $\hat{x}$ and its uncertainty $U_{\hat{x}}$. Each of these steps permits the injection of error; and it is only through the detailed theoretical and experimental assessment of the assumptions and error structure of each step – as well as the formal linkage of the entire sequence of steps – that one can produce a realistic estimate of the "quality" (reliability, range of applicability, robustness, sensitivity, specificity, accuracy, ...) of the overall CMP (37) (see *Note c*). Using the structure as indicated in Fig. 3, we shall first offer some general comments on assumptions and error. Other matters having special relevance to trace analysis will be treated in the following subsections.

The "bottom line" in any evaluation of error for an analytical result $\hat{x}$, is a statement of its overall uncertainty $U_{\hat{x}}$. The error $e_{\hat{x}}$, of course, is generally unknown, but its estimated bounds as given by $U_{\hat{x}}$ must be presented for the result to be useful. $U_{\hat{x}}$, which is not necessarily symmetric, is really a vector comprised of random and systematic components. Such components may be separately estimated by propagation of contributions from each of the CMP steps, and they should always be individually stated (40). A very useful means to assure quality in the estimation of systematic and random components is to require consistency between internal and external estimates. An internal estimate, in this context, is given by propagation through the defined structure of the CMP (37,41). An external estimate for systematic

*Note c.* Descriptors given here for "quality" are intended simply to indicate the general features considered important. A detailed treatment requires explicit definitions of quantitative measures of performance, such as given in references 37 through 39.

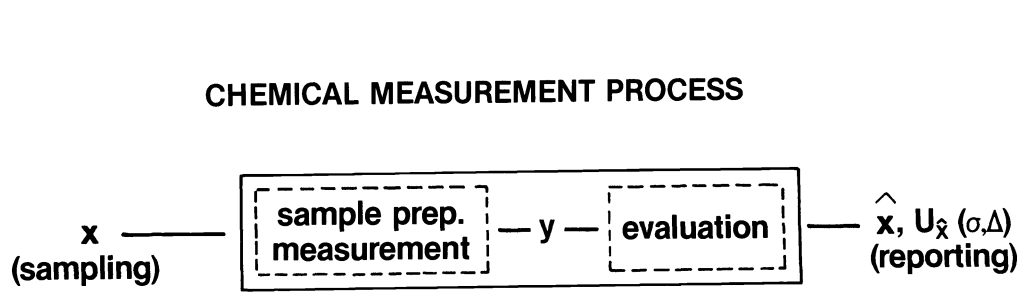



Fig. 3. The Chemical Measurement Process. (See text for explanation of symbols.)

error obtains from the analysis of an SRM; for random error, from replication of the *entire* CMP (42). Finally, as indicated by the second equation below the CMP-diagram, the non-random part of the error may itself comprise two components: a constant bias $\Delta$, and a lack-of-control (e.g., drift or erratic variation with time) term $h(t)$. Most significantly, the first (random) component in the equation diminishes with the number of replications $n$, and therefore uncertainty intervals based on random error only tend to seriously underestimate the true (over-all) error when several replications are involved. ($z$ equals the standard normal variate.)

Extraction of the estimated analyte concentration $\hat{x}$ (which is a vector in the case of multicomponent analysis) from the observed signal $y$, requires an assumed model or functional relationship as symbolized in the first equation in Fig. 3. In general $y$ is given by some unknown functional relationship $f$ of the blank $B$ and the analyte components $x_1 \ldots x_p$, as well as independent variables such as energy $E$ (or wavelength) and time t. When the true model approaches a linear expression having independent terms for each of the components (far right side of the equation for $y$), provided that the number and identities of the components are known, model error or bias will be negligible. More generally, imperfect model assumptions can be a leading cause of error in trace analysis, particularly in complex matrices. An explicit treatment of this question will be given below.

Before leaving the general issue of systematic ($\Delta$) and random ($\delta$) error, it should be observed that these components may be interconvertible. That is, re-design of the CMP may allow the conversion $\Delta \rightarrow \delta$, or the reverse $\delta \rightarrow \Delta$. The former conversion occurs when there is "real" replication, and it yields error reduction as $\sim n^{-1/2}$; the latter conversion should be avoided, as it implies added, non-reducible bias which is often difficult to estimate.

Error Distributions and the Reporting of Trace Analytical Data Distributions. Ultimately, the sensitivity and precision with which trace analytes may be determined is limited by random measurement error. That is, once blunders have been exposed, and bias eliminated by comparative techniques – such as isotope dilution and relative activation analysis, or by conversion to a random error component – the best that one can do with a given *CMP* is given by the measurement (instrumental) standard deviation $\sigma_m$.

Knowledge of $\sigma_m$ is important not just for establishing detection limits and confidence intervals, but it can be a powerful asset in discovering other,

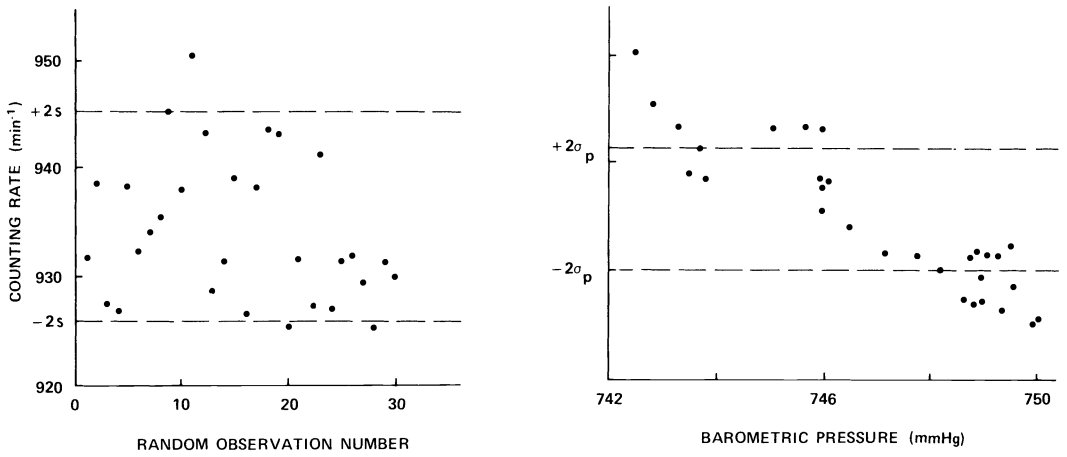



Fig. 4a. Observed meson rates (cosmic-ray guard counter): left, random sequence; right, versus barometric pressure. (Ref. 37).

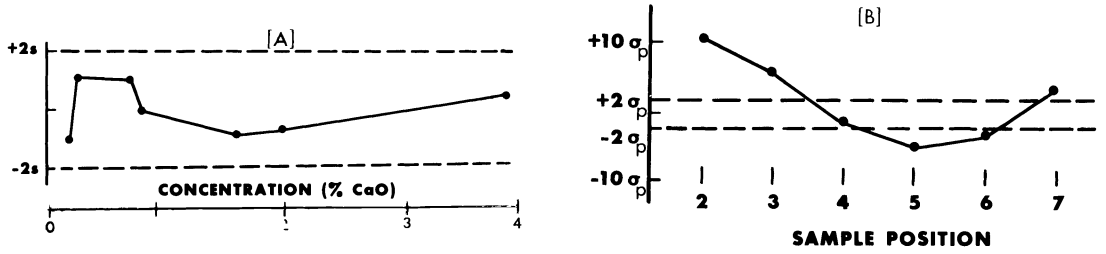



Fig. 4b. Residual diagnosis of systematic error in CaO-X-ray fluorescence analysis. [A] Residual pattern ($y-\hat{y}$) following least squares fitting of a linear calibration curve, [B] normalized Poisson residuals ($A-A_m$) recalculated following deletion of the first point using sample position as the independent variable. (Ref. 34).

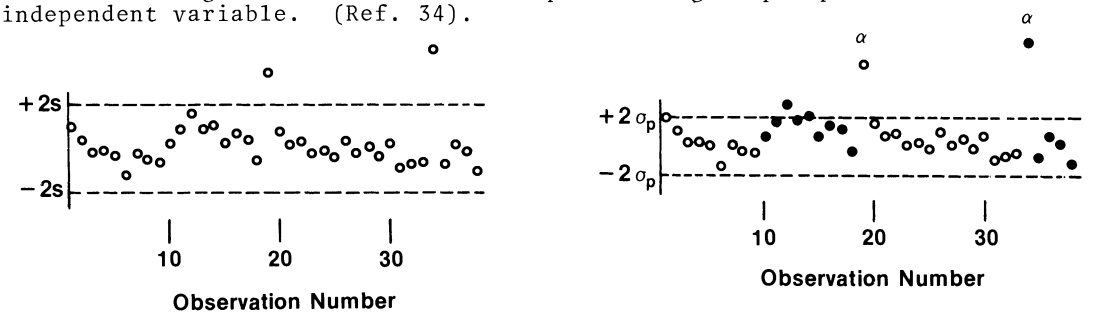



Fig. 4c. $^{14}C$ background observations, effect of impure (ppm) counting gas [points-α]. (Alternating open and solid circles indicate independent measurement sequences – i.e., counter fillings.)

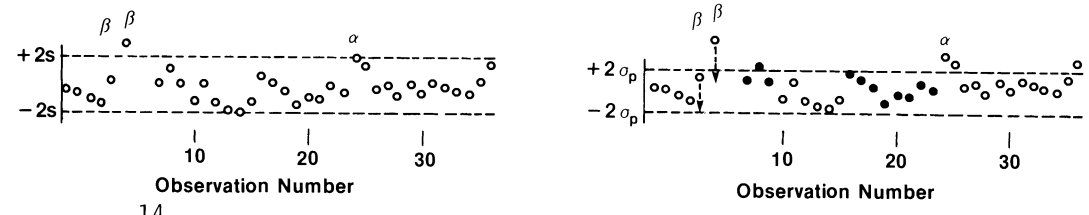



Fig. 4d. $^{14}C$ standard observations, effect of blunders (recording mistakes, points-β). (Alternating open and solid circles indicate independent measurement sequences – i.e., counter fillings.)

Fig. 4. Residual analysis based on total variability (±2s left-hand control charts) vs. instrumental measurement (Poisson) error plus physical insight ($\pm 2\sigma_p$ right-hand control charts).

unsuspected sources of error in a set of data or a data-reduction model. Comparing the observed with the expected variance ($\sigma_m^2$), one is often able to identify and correct such additional error. This process, which will be further discussed under detection theory, is an example of the Analysis of Variance (ANOVA) which forms the cornerstone of modern statistics. When the analytical signal consists of counts of individual (random) events, one has the good fortune of knowing the (minimum) measurement error ($\sigma_p$) from Poisson counting statistics. We have often found it revealing to examine an observation set and its estimated variance ($s^2$) in terms of the Poisson error. Control charts illustrating the point are shown in Fig. 4. As the sequence of "before-after" plots illustrates, foreknowledge of the expected random error magnitude allowed us to detect, and in these cases control, very significant contributions from non-measurement error. The four non-random sources shown are: (a) the effect of barometric pressure on the cosmic-ray background of a nuclear radiation detector (37); (b) the effect of a physically distorted sample holder on the measurement of CaO by x-ray fluorescence spectroscopy (34); (c) the effect of sample gas impurities on an initial background measurement for a radiocarbon detector (43); and (d) the effect of an operator transcription mistake (sample pressure) on the measurement of a radiocarbon calibration sample (43). In all four cases the known Poisson error signaled lack of control, but transformation/rejection was based on *independent* experimental evidence. (Rejection of an observation simply because it is "outlying" is hazardous, especially with a small data set. At times, an "outlier" may be the only *correct* result!)

Assumed distributions for random error represent another source of difficulty in setting detection limits and confidence intervals. The symmetric 99.7% confidence interval extends $\pm 3\sigma$ from the mean for the normal distribution, but only $\pm 2.3\sigma$ for the triangular distribution, and $\pm 1.7\sigma$ for the uniform distribution (37). For skewed or long-tailed distributions, this interval may be many times larger. Unfortunately, random error distributions are not always what we expect, and their tails (hence large confidence intervals) are very difficult to assess experimentally. For example, suppose that one were to make 100 observations of a chemical blank and wished to set the upper 5% limit. It follows from Poisson statistics that above the true 5% limit, one could expect to obtain anywhere from one to ten of the blanks. An illustration of presumed vs. observed distributions is given in Fig. 5, which shows different characteristic deviations from the uniform distribution of the estimated digit from the use of a balance and of a buret (44).

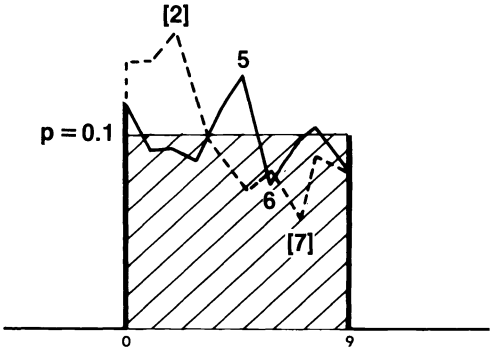



Fig. 5. Random error distributions (last estimated digit) for 1000 student weighings (solid curve; extrema: 5, 6) and 1510 student buret readings (dashed; extrema: [2], [7]). (Adapted from data in Ref. 44.)

Normal distributions are most commonly assumed and most convenient to treat. When one has a linear function (such as signal – blank) of two or more normally distributed variables, the resulting distribution is normal, and the variance is readily propagated (45). Treatment of variance and uncertainty intervals is not so trivial, however, when one encounters non-normal and mixed distributions – as in the case of log-normally distributed concentrations of environmental species convolved with normally distributed measurement error (46). Two useful approaches to these problems are the use of resistant or "distribution-free" statistics when normality cannot be assured, and the use of replication in order to bring about normality. An example of the first approach is the use of the median in place of the arithmetic mean; this has the added advantage of giving protection against isolated outliers (cf. Fig. 4c and 4d). In fact, if one has more than eight observations, it is possible to estimate both the median and its 95% confidence interval from the data alone (no prior $\sigma$-information), with immunity from an outlier of either sign (47). The approach to normality through replication is a result of the

Central Limit Theorem (48) which states that for any distribution as the number of observations (n) increases, the distribution of the arithmetic mean is asymptotically normal. The approach to normality is quite rapid, especially for modest confidence intervals and for symmetric distributions. (The distribution of means from a uniform distribution already looks quite "normal" by the time n=4 (49).) To sum up: the likelihood of quality data and reliable conclusions (detection limits, confidence intervals) can be increased by converting residual systematic errors to random errors by repeating all steps of the CMP, and forcing normality through replication — while using distribution-free techniques when appropriate. These steps make all errors estimable (through $s^2$) and confidence intervals reliable (through normality). Reduction of the standard error (as $n^{-1/2}$) is an added benefit, though not the principal objective.

Reporting. Quality data, poorly reported, leads to needless information loss. This is especially true at the trace level, where results frequently hover about the limit of detection. In particular, reports of upper limits or "not detected" can mask important information, make intercomparison impossible, and even produce bias in an overall data set. An example is given in Fig. 6 which relates to a very difficult radioanalytic problem involving fission products in seawater (36). In this example, only six of the fifteen results could be fully compared and only eight could be used to calculate a mean. Since negative estimates were concealed by "ND" and "<", the mean was necessarily positively biased. (The true value $\tau$ in this exercise was, in fact, essentially zero; and the use of a robust estimator, the median [*m*] does give a consistent estimate.) Although upper limits convey more information than "ND", authors choose conventions ranging from the (possibly negative) estimated mean ($\bar{x}$) plus one standard error to some sort of fixed "detection limit". Such differences are manifest when one finds variable upper limits from one laboratory but constant upper limits from another (50).

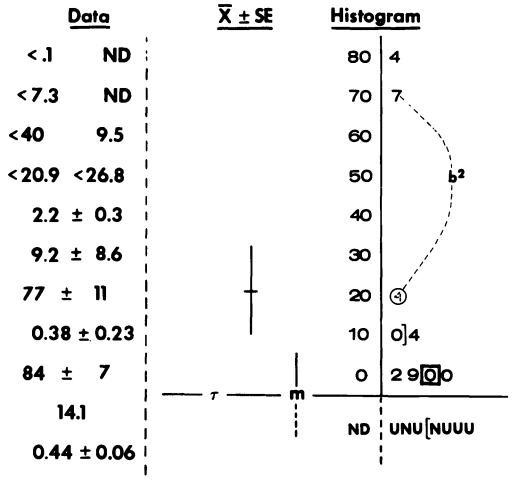



Fig. 6. Reporting deficiencies. International comparison of $^{95}Zr$-$^{95}Nb$ in sample SW-1-1 of seawater (pCi/kg). The symbols have the following meanings: $\tau$ = true value, $\bar{x}$ = arithmetic mean (positive results), m = median (all results), and $b^2$ = a "double blunder" — i.e., inconsistent result 77±11 was originally reported as 24. N and U indicate not detected, and upper limits, respectively. (Data from Ref. 36.)

The solution to the trace analysis reporting dilemma is to provide all relevant information, including *as a minimum:* the number of observations, the estimated value and its standard deviation (see *Note d*), and meaningful bounds for systematic error. More thorough treatments of this issue may be found in Eisenhart (40), and Fennel and West (39).

---

*Note d.* The estimated standard deviation *s* (or the standard error $s/\sqrt{n}$) is preferred to a confidence interval, because *s* carries no assumption of normality. (If he wishes to rely upon this assumption, the user may of course apply Student's — t.)

Sample and Procedure Validity – Certified Reference Materials

Sampling. Sampling forms the first link in the analytical chain. As such, it can maintain or destroy the validity of the entire process. The most likely sources of sampling blunders in trace analysis are contamination, losses, and heterogeneity (non-representative samples). Dangers associated with the first two factors are more or less obvious, i.e., the very fact that one has only a tiny amount of analyte to begin with means that small absolute gains (contamination) or losses can produce large relative errors. Contamination in sampling is often associated with the sampling apparatus, sampling personnel, or sampling medium; and it will be more specifically discussed when we treat the blank and the Environmental Specimen Bank. At extremely low concentrations, significant losses may occur through incomplete extraction of the sample from the host medium, volatilization, and/or chemical reaction (during storage or transport) and adsorption. Much helpful experience comes from radiochemistry and isotopic chemistry, where such problems with losses are countered or effectively monitored through the addition of known amounts of carriers or isotopic tracers at the earliest possible stage. To use such a technique effectively, it is important, of course, to pay attention to one of the basic tenets of isotope dilution methodology: to assure complete isotopic exchange among the possible varied chemical or physical forms of the analyte.

Heterogeneity poses some special problems for trace analysis. First, as the sampling error frequently greatly exceeds the measurement error (51,52), the actual limits of detection or quantitation may be far poorer than would otherwise be expected for the given CMP. Second, if the trace analyte exists primarily in very localized regions of the host medium – such as at interfaces or in individual particles, then random sampling has a good chance of completely missing the analyte fraction (see *Note e*). This is not because of bias – random sampling is always an unbiased procedure – but because of serious statistical inefficiency in an extremely heterogeneous medium. Solutions to this problem include (a) the analysis of the *entire* sample medium, (b) complete mixing to achieve a homogeneous medium, or (c) stratified sampling. While (a) and (b) may not always be feasible, (c) generally is. One limitation to mixing, unless one can produce a true solution free from concentration gradients, is that heterogeneity is apt to remain on a microscale (56). In fact, the possible existence of "microhomogeneity" generally dictates the minimum sample sizes which may be safely used for certified mixtures, such as Standard Reference Materials (57). Also, the mixing operation itself, may lead to difficulties: shaking of particle assemblages having different sizes, compositions and densities is likely to produce nonuniform mixtures. Stratified sampling, which is designed to sample regions of different composition in a representative manner, is often the method of choice (58). (Sampling within each of the strata remains random.) Given a heterogeneous medium, it can be shown that the stratified sample is more efficient (has better precision) than the random sample, and it is less likely to yield an extremely low or extremely high result when one is dealing with ultratrace analysis. Either random or systematic samples may, of course, be combined to form a composite sample before analysis. This operation conserves analytical time, but the cost is information loss on the extent of heterogeneity and its spatial (temporal) patterns.

A classic illustration of the pitfalls of sampling a medium which is heterogeneous with respect to the analyte has been given by Kratochvil (59). This refers to the analysis of aflatoxin in peanuts, where as we saw earlier concentrations of just $10^{-10}$ g/g are already significant. The sampling problem arises because the aflatoxin, which is produced by the fungus *aspergillus flavis*, may be present in only a relatively few peanuts in a several ton storage container. By including representative samples from moist regions where the fungus is most likely to thrive, one is less apt to draw the wrong conclusion about the absence or presence of the toxin in the particular batch or shipment. For a more comprehensive treatment of sampling errors and designs for trace analysis, see references 60 and 61.

Reference Materials (See *Note b*). Perhaps the best possible means to assure CMP validity is through the use of Standard Reference Materials (SRM's). The importance of this means of validation is underscored by the fact that the year 1981 marks the 75th anniversary of the certification of the first

---

*Note e.* Heterogeneity is not limited to physical (spatial) variations. Temporal heterogeneity is well appreciated by those concerned with time-averaged pollutant concentrations, for example (53).

reference materials by the U. S. National Bureau of Standards (NBS) (62). One of the first four SRM's, still in use today, was originally issued in 1906 specifically for technological development – in this case, for use in establishing the accuracy of analytical methods for determining the composition of grey cast irons. Of the 1000 or so SRM's currently available, a major fraction are directly applicable to the establishment and control of quality in industrial processes, while significant numbers have been produced for specific applications to metrology, and environmental and clinical measurements. Some perspective on the importance of SRM's is given by the fact that some 30 clinical SRM's are contributing to the improvement of the reliability of more than $4 \times 10^9$ measurements made annually; and that NBS currently distributes more than 40,000 units annually to approximately 10,000 users world-wide. The certification process itself, in which two or more reliable, independent methods are used to derive estimated analyte concentrations and uncertainties, is necessarily a model for high quality (trace) analysis. To the extent possible, "definitive" methods, having a sound theoretical and proven experimental base, high precision and negligible systematic error are employed in this process (57).

In order to provide a glimpse of the nature of current SRM's, we indicate the major classes of SRM's in Table 2, and we shall briefly discuss three of the more recently-issued materials. The first of these is SRM 909, the first human serum to be certified by NBS for constituents normally found in human serum. This SRM, which is used to assess the accuracy of clinical methods for inorganic and organic constituents, consists of six vials of freeze-dried human serum and six vials of high purity water to reconstitute the serum. A copy of the official announcement for this material is given in Fig. 7. It is worth noting that this SRM is representative of a broad class in which a number of analytes have been certified in their native matrix, and for which there has been a continuing and wide-ranging need for a common reference material for the development and maintenance of reliable analytical methods (63). Concurrent with the development of such SRM's has been the development of the "definitive" methods of analysis, in many cases based on isotope dilution/mass spectrometry (64). In addition to the information concerning SRM 909 given in the announcement, there are some special notes concerning its use, storage and uncertainty limits. Regarding the first of these points, users are cautioned against the possibility of hepatitis transmission. As stated in the SRM certificate: "Although this product was tested with licensed third generation reagents and found nonreactive for the presence of hepatitis B surface antigen ($HB_SAG$), no known test method can offer assurances that products derived from human blood will not transmit hepatitis." With respect to storage, "The freeze-dried serum should be stored in a refrigerator at a temperature between 2 and 8 °C. It should not be frozen nor exposed to sunlight or ultraviolet radiation. Under the recommended storage conditions, this SRM is expected to be stable for at least one year; [but it] is not certified for use after one year from date of purchase." Finally, there is an interesting point concerning uncertainty bounds. That is, the accuracy and even the uncertainty symmetry depends upon how the user reconstitutes the serum. For higher accuracy, and symmetric uncertainty bounds, the reconstitution must be done by mass. Simple transfer of the freeze-dried serum contents of a vial without weighing results in concentration uncertainties which are sometimes larger by a factor of ten, and which are quite asymmetric. This follows from the observed non-Gaussian distribution of masses obtained in filling the vials, which were skewed in the direction of high mass.

The problem of analyte instability, overcome by the reconstitution approach with SRM 909, was solved in an interesting manner for a recent environmental reference material, SRM 1644 – Polynuclear Aromatic Hydrocarbons in Water. In this case nominally $10^{-9}$ g/g concentrations of anthracene, benz(a)-anthracene and benzo(a)pyrene are made available through the use of "generator columns" which are packed with glass beads coated with the compound of interest. The thermodynamic, temperature-dependent solubility serves as the basis for known concentrations, as passage of water through the columns produces saturated aqueous solutions. By controlling the temperature between 10 °C and 30 °C, for example, one can prepare known concentrations of benzo(a)pyrene ranging from 0.6 to $2.3 \times 10^{-9}$ g/g (65).

TABLE 2. Classes of NBS Standard Reference Materials[a,b]

| Metals and Ores | Environmental and Clinical | Metrology, Engineering, and Physical Properties | Radioactivity and Environmental | Spectrophotometry and Primary Chemicals |
|---|---|---|---|---|
| Steels | Clinicals | Electron Probe Microanalytical | Nuclear Materials | Primary Chemicals |
| Steelmaking Alloys | Biologicals | Glasses | Radioactivity | Metallo-organic Compounds |
| Cast Irons | Botanicals | Cements | Isotopics | Optical-Spectrophotometry |
| Cast Steels | Environmentals | X-ray Diffraction | Reference Fuels | |
| Nonferrous Alloys | Industrial Hygiene | Ion Activity | | |
| Gases in Metals | Fertilizers | Mechanical | | |
| High-Purity Metals | Trace Elements | Superconducting | | |
| Ores | | Freezing Points | | |
| Minerals | | Melting Points | | |
| Refractories | | Calorimetric | | |
| Carbides | | Vapor Pressure | | |
| | | Thermal Conductivity | | |
| | | Thermal Expansion | | |
| | | Thermal Resistance | | |
| | | Thermocouple Materials | | |
| | | Magnetic | | |
| | | Optical | | |
| | | Gas Transmission | | |
| | | Permittivity | | |
| | | Resistivity | | |
| | | Rubber Materials | | |
| | | Computer Tapes | | |
| | | Sizing Standards | | |
| | | Color | | |
| | | Photographic | | |
| | | Surface Flammability | | |
| | | Smoke Density | | |
| | | Water Vapor Permeance | | |
| | | Polymers | | |

[a]Further information may be obtained from the Office of Standard Reference Materials, National Bureau of Standards, Washington, D.C. 20234.

[b]Reproduced from Ref. 62.


U.S. DEPARTMENT OF COMMERCE
National Bureau of Standards

# Standard Reference Material 909

## Human Serum

Fall 1980

The Office of Standard Reference Materials announces the availability of a Human Serum Standard Reference Material. SRM 909 was developed for use in assessing the accuracy of clinical methods for specified constituents in human serum, calibrating instrumentation used in these analyses, and validating in-house or commercially produced quality control materials.

The SRM consists of six vials of freeze-dried human serum and six vials of high-purity water to reconstitute the serum. The Certificate of Analysis for the reconstituted serum lists certified concentrations of calcium, cholesterol, chloride, glucose, lithium, potassium, and uric acid.

The constituents were determined at NBS by definitive methods, which are highly accurate analytical methods. Isotope dilution mass spectrometric procedures were used for determining the constituents. Constituents, such as sodium and urea, are also being determined for certification by definitive methods while enzymes are being determined cooperatively by "best available" methods. The Certificate of Analysis will be revised periodically to include new information as it becomes available.

SRM 909 is the first human serum certified by NBS for constituents found normally in human serum. A recently issued human serum matrix, SRM 900, is certified for four antiepilepsy drugs at subtherapeutic, therapeutic, and toxic concentrations. In addition to these matrix SRM's, a number of other clinical SRM's are available. These are described in a brochure on clinical chemistry SRM's as well as in a catalog of Standard Reference Materials which are available on request from the Office of Standard Reference Materials.

SRM 909 Human Serum may be purchased from the Office of Standard Reference Materials, National Bureau of Standards, Washington, D.C. 20234. (Telephone: 301/921-2045) for $149 per unit of six vials of freeze-dried serum and six vials of high-purity diluent water.

Fig.7. Announcement for Standard Reference Material-909, Human Serum.

A final illustration of a trace analysis reference material is given by SRM 1580, Trace Organics in Shale Oil. This material has been certified in connection with the need to reliably assess the environmental effects of developing new fuel sources, and it serves also as a useful material for evaluating generally the reliability of methods of trace organic analysis in an oil matrix. The need for this certified reference material was demonstrated by an interlaboratory comparison of another oil shale sample for EPA priority pollutants. According to the official announcement for SRM 1580 (Spring, 1980), "in the analysis of a similar shale oil sample [not SRM 1580], nine laboratories showed disagreement ... the determination of fluoranthene ranged from 61 to 220 μg/g [and] phenol from 180 to 399 μg/g". (In SRM 1580, the certified concentrations for these two compounds are 55 ± 5 μg and 407 ± 50 μg/g, respectively. The estimated uncertainties are intended to correspond to approximately 95% confidence limits.) Clearly "certification by concensus" can be a dangerous path (66,67).

Specificity and the Blank

Specificity. There is no greater aid to high-quality trace analysis than highly specific instrumental or chemical separation techniques; and there is no greater limitation than the magnitude and variation of the blank. Without effective chemical or instrumental resolution of interfering species, there would be little hope of determining analytes at the levels shown in Table 1.

That is, even in relatively pure matrices, for analytes sought at concentrations below $\sim 10^{-9}$, there will be overwhelming quantities of potentially interfering species in enormously greater concentrations.

Specificity may be considered in a general sense in terms of two continuing products $\underline{P}$ and $\underline{Q}$, where P represents the probability that the analyte will survive an entire sequence of [n] chemical and/or instrumental discrimination steps; and $Q^{(k)}$, the probability that an interfering substance (k) will survive those same steps. Optimal specificity is thus given by

$$\max P = \max \prod_{i}^{n} p_i \tag{1a}$$

$$\min Q^{(k)} = \min \prod_{i}^{n} q_i^{(k)} \tag{1b}$$

The quantities $p_i$ and $q_i^{(k)}$ refer to the loss of analyte and interfering species-k as a result of the $i^{th}$ discrimination step, and therefore maximizing P gives the optimal recovery or efficiency for the analyte, and minimizing $Q^{(k)}$ gives minimal interference-k or maximum decontamination. The ideal (P=1, Q=0), which is unattainable, would correspond to *unique analytical identification* (if the analyte signal is non-zero). Designing an optimal CMP for detecting ultratrace species in complex matrices thus reduces to striking the best possible balance between Equations (1a) and (1b). The balance required of course depends on the initial amounts of the analyte and interfering substances, and imperfect specificity (P<1, Q>0) can be compensated for when analyte and contaminant have different response patterns. That is, "overlapping" responses, such as chromatograms or spectra may be mathematically resolved – given sufficient measurement precision and response pattern differences.

In order to illustrate specificity at its best, two current examples will be discussed in parallel, one consisting primarily of *chemical* resolution steps and the other, of *instrumental* resolution steps. These are shown in Fig. 8. The left portion of the figure shows highlights from the scheme for uniquely determining 2,3,7,8-tetrachlorodibenzo-p-dioxin (2378-TCDD) at a concentration of $\sim 10^{-12}$ in the presence of (pesticide) contaminants having a million-fold higher concentration (31). The right-hand side of the figure outlines the scheme for determining $^{14}C$ at a concentration of $\sim 10^{-14}$, in which a decontamination factor of $10^{17}$ has been achieved through purely instrumental means (68). In each case the recovery (P) was monitored by means of an internal isotopic standard, which is almost a mandatory procedure when carrying out such extreme separations at ultratrace levels. The recovery for 2378-TCDD was a respectable 40%, while that for $^{14}C$ was just 0.1%. This latter value, while low, was nevertheless adequately normalized by means of the (naturally present) stable isotope, and it was sufficient to yield an acceptable signal. As indicated in the figure, following very substantial initial clean-up measures, the primary steps to resolve 2378-TCDD from the 21 other TCDD isomers consisted of three sequential chromatographic separations, *whose mechanisms depended on substantially different molecular parameters.* It is significant that the isomer-specific chromatographic separations combined with low-resolution mass spectrometry have succeeded in providing reliable results for heavily contaminated sample matrices, whereas less complete chemical separation combined with capillary column gas chromatography-high resolution mass spectrometry has failed (31). Purely instrumental resolution of natural levels of $^{14}C$ from $^{14}N$ and $^{12}C$-molecular fragments more abundant by a factor of $\sim 10^{13}$ has been achieved by major discrimination steps involving enormous differences in lifetimes of negative carbon and nitrogen ions, acceleration to energies yielding complete molecular fragmentation, and finally measurement by particle-specific nuclear detectors (68).

The Blank. With the existing capabilities of "single-atom-detection" (21) and the enormous specificity available through multistep chromatography and accelerator mass spectrometry it might seem that there is no limit to ultrasensitive trace analysis – until one considers the blank (See Fig. 9). Unfortunately, there is no alternative to extreme vigilence when treating the limitations imposed by the blank. In the best of circumstances the mean value

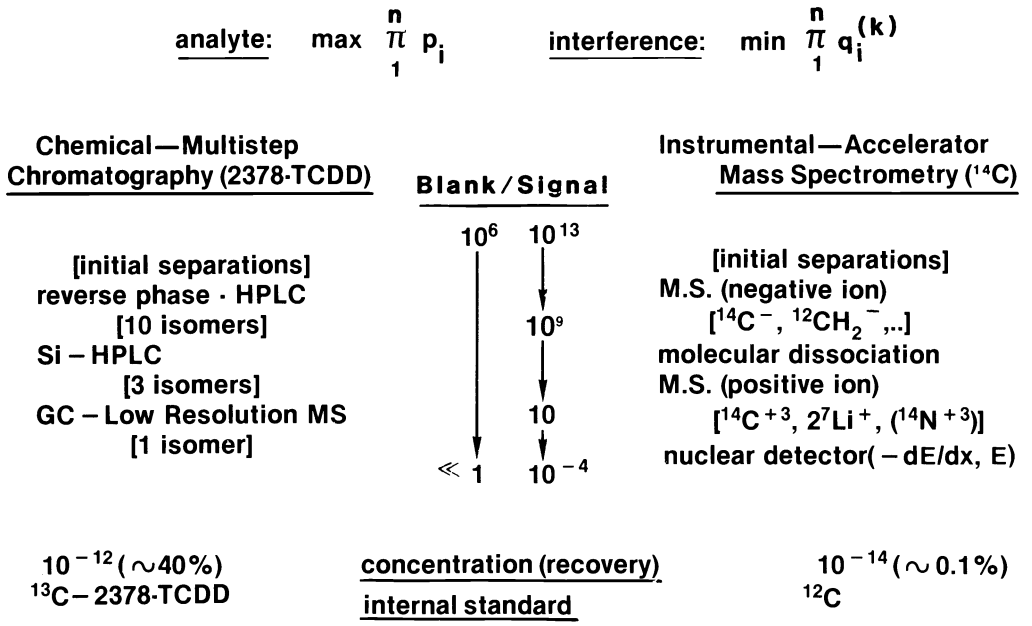



Fig. 8. Specificity. Multistep chemical decontamination (2378, TCDD-Ref. 31) and multistep instrumental resolution ($^{14}C$-Ref. 68).

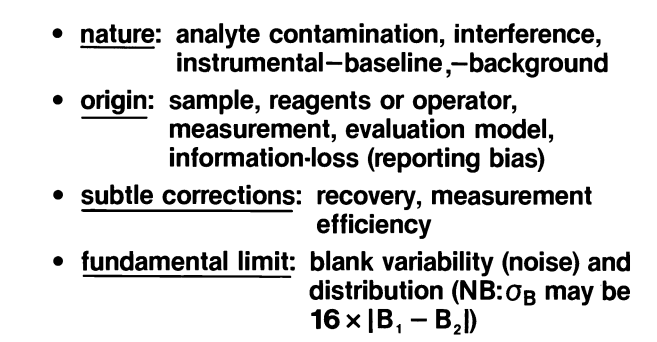



Fig. 9.

of the blank might be expected to be constant and its fluctuations ("noise") normally distributed. Given an adequate number of observations, one could estimate the standard deviation of this noise and therefore set detection limits and precisions for trace signals. In situations where the chemical (analyte) blank remains small compared to the instrumental noise blank this procedure may be valid, as in many low-level counting experiments. Even here, however, to assume that the noise is normally or Poisson distributed, or to

estimate the background from one or two observations is to invite deception. As indicated in Fig. 9, there is a significant chance (5% for normally-distributed blanks) that the expected value of the noise (blank standard deviation) will exceed the observed difference between two blanks by a factor of 16! Subtle perturbations arise even in the instrumental blank situation. For example, if the analyte detection efficiency changes discretely or even fluctuates, it is quite possible that the instrumental blank will suffer a disproportionate change (69). [In order to provide some degree of control over the interpretation of instrumental background, we recommend that it be reported as "Background Equivalent Concentration" (50).]

Certain special cases occur where the blank can be reliably estimated, and therefore adjusted, indirectly. This is the situation: for "on-line" coincidence cancellation of the cosmic-ray mu-meson component of the background in low-level radioactivity measurement (where there is not even a stochastic residue from the adjustment process); for the adjustment of the baseline (due generally to multiple interfering processes) in the fitting of spectra or chromatograms; and for correction for isonuclidic contamination (due to interfering nuclear reactions) in high sensitivity nuclear activation analysis. These last two cases will be treated later in the text.

When the blank is due to contamination (as opposed to interferences or instrumental background), high quality trace analysis is at its greatest risk. Assumptions of constancy, normality or even randomness are not to be trusted. An apparent analyte signal may be almost entirely due to contamination (70); and blank correction must take into account its point(s) of introduction and subsequent analyte recoveries. The randomness assumption may be inappropriate because the blank may depend upon the specific history of the sample, container or reagents (71). Also when procedures are applied to real sample matrices as opposed to pure solutions blank problems abound, as was observed, for example, in the analysis of Pb (at a concentration of 30 ng/g) in porcine blood in contrast to aqueous solutions (72). (Reference 72 is also commended to the reader for a more complete treatment of the blank in trace analysis.) The most severe test of this sort comes when "blind" blanks together with samples at or near the detection limit, all in actual sample matrices, are submitted for analysis. Horwitz, for example, referring to collaborative tests of "unknowns" for 2378-TCDD in pure solutions, beef fat, and human milk, noted that significant numbers of false negatives began to appear when concentrations were less than $9 \times 10^{-12}$, and that false positives increased from 19% for blank "standards" to over 90% for human milk samples (73)!

Finally, an instructive display of blank variations is given in Fig. 10, which derives from a recent investigation of blanks connected with the analysis of chromium at the ng/g level in biological and botanical SRM's by isotope dilution mass spectrometry (74). Worth noting are the facts that: (a) two discrete and significant steps of reduction took place, with an overall reduction by about a factor of 20; (b) within given regions one observes wide variations (factor of $\sim$2 in region I) and local non-random variations (trend around blank-21, region III); (c) the blanks are always positive, consistent with contamination as the source. The excursions seen in regions I and II imply that uncontrolled variables are significantly affecting the blank and that normality and randomness might be misleading assumptions (see *Note f*). This is supported by the very large changes brought about through control of environmental and procedural variables as indicated in the figure caption. [Sub-boiling distillation of $HNO_3$, for example, reduced total inorganic impurities by a factor of 100, to 2.3 ppb (55)]. *One is led to the conclusion that, unless or until the maximum blank excursions are trivial compared to other sources of measurement error, the best course includes: continual monitoring of the blank, provision of the highest purity reagents and cleanest possible apparatus, and above all isolation of the operator and his environment (atmosphere) from the sample (75).*

---

*Note f.* For protection against occasional outliers without outright rejection and mild non-normality, we are investigating the use of various "robust" statistics (as the median) taking an adequate number of observations. For reliable application of these methods it is desirable to treat the sample in exactly the same manner as the blank (76).

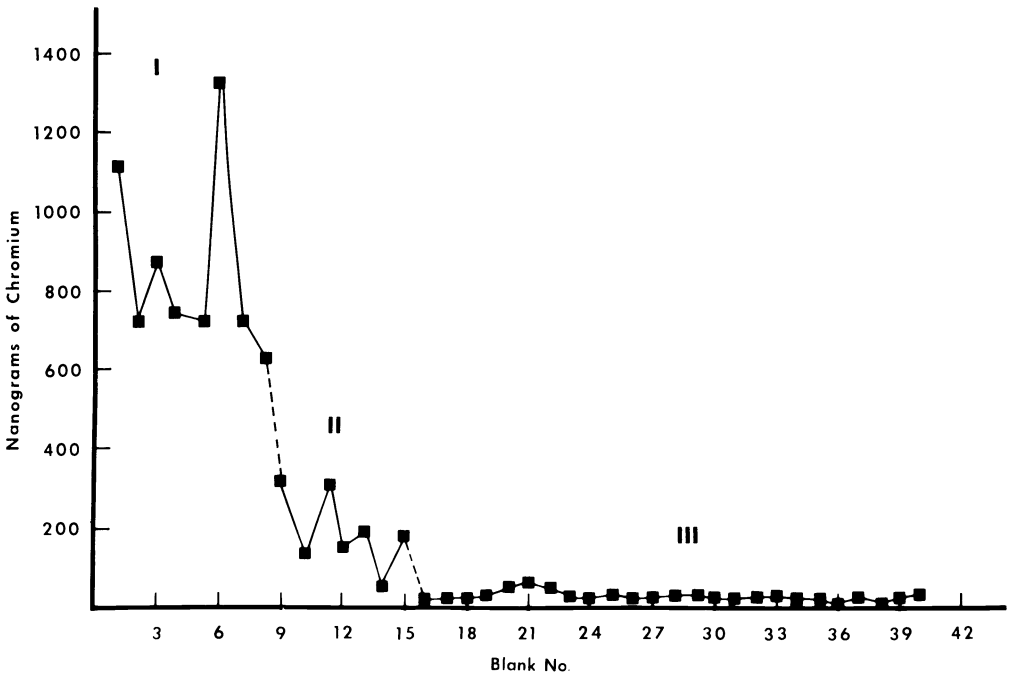



Fig. 10. Chromium Analytical Blanks. Region I: ordinary laboratory and reagents. Region II: purified atmosphere (clean room and clean, laminar-flow bench). Region III: purified atmosphere and purified reagents (Ref. 74,75).

Detection Theory and Hypothesis Testing

Central to all decisions based on the outcomes of stochastic experiments is the statistical theory of Hypothesis Testing (77). The theory has its most obvious and well-utilized application to trace analysis in connection with the Detection Limit (78). Its importance for Analytical Chemical decision-making extends well beyond this particular application, however, and therefore it is worth investing a few paragraphs to expose the assumptions and chemical applications of this theory (see Fig. 11).

## DETECTION · HYPOTHESIS TESTING

- **false-positives, -negatives ($\alpha$,$\beta$)**
- **analyte detection**
  - **— $\sigma_0$ and $\sigma$ (concentration) needed**
  - **— explicit algorithm needed [$\hat{x}$]**
  - **— effects of interference, degrees of freedom, number of decisions, systematic error.**
- **bias detection**
- **blunder (outlier) detection**
- **model error detection (bias matrix)**

Fig. 11.

The basic issue we wish to address is whether one primary hypothesis [the "null hypothesis", $H_0$] describes the state of the system at the point (or time) of sampling or whether one or more "alternative hypotheses" [$H_1$, $H_2$...] describes it. The actual test is one of consistency – i.e., given the experimental sample, are the data consistent with $H_0$, at the specified level of significance, $\alpha$? That is the first question, and if we draw (unknowingly) the wrong conclusion, it is called an error of the first kind. This is equivalent to a false positive in the case of trace analysis – i.e., although the (unknown) true analyte signal S equals zero (state $H_0$), the analyst reports, "detected".

The second question relates to discrimination. That is, given a decision-(or critical-) level $S_C$ used for deciding upon consistency of the experimental sample with $H_0$, what true signal level $S_D$ can be distinguished from $S_C$ at a level of significance $\beta$? If the state of the system corresponds to $H_1$ ($S=S_D$) and we falsely conclude that it is in state $H_0$, that is called an error of the second kind, and it corresponds in trace analysis to a false negative. The probabilities of making correct decisions are therefore $1-\alpha$ (given $H_0$) and $1-\beta$ (given $H_1$); $1-\beta$ is also known as the "power" of the test, and it is fixed by $1-\alpha$ (or $S_C$) and $S_D$. One major objective in selecting a CMP for trace analysis is thus to achieve adequate detection power ($1-\beta$) at the signal level of interest ($S_D$), while minimizing the risk ($\alpha$) of false positives. Given $\alpha$ and $\beta$ (commonly taken to be 5% each), there are clearly two derived quantities of interest: $S_C$ for making the detection decision, and $S_D$ the detection limit.

An assumption underlying the above test procedure is that the estimated net signal $\hat{S}$ is an independent random variable having a known distribution. (This is identical to the prerequisite for specifying confidence intervals.) Thus knowing (or having a statistical estimate for) the standard deviation of the estimated net signal $\hat{S}$, one can calculate $S_C$ and $S_D$, given the form of the distribution and $\alpha$ and $\beta$. If the distribution is normal, and $\alpha=\beta=0.05$, $S_D \tilde{=} 3.29\ \sigma_{\hat{S}}$ and $S_C \tilde{=} S_D/2$. Thus, the relative standard deviation of the estimated net signal equals 30% at the detection limit (78). Incidentally, the theory of differential detection follows exactly that of detection, except that $\Delta S_{JND}$ (the "just noticeable difference") takes the place of $S_D$, and for $H_0$ reference is made to the base level $S_0$ of the analyte rather than the zero level (blank). A small fractional change $(\Delta S/S)_D$ thus requires even smaller imprecision.

Obviously, the smallest detection limits obtain for interference-free measurements and in the absence of systematic error. Allowance for these factors not only increases $S_D$, but (at least in the case of systematic error) distorts the probabilistic setting, just as it does with confidence intervals. Special treatments for these questions and for non-normal distributions are needed, but are beyond the scope of this paper. Not so obvious perhaps is the fact that $S_D$ depends on the specific algorithm selected for data reduction. As with interference effects on $S_D$, this dependence comes about because of the effect on $\sigma_{\hat{S}}$, the standard deviation of the estimated net signal. Some more explicit coverage of these matters appears in reference 37, and some will be given below in the discussion of the IAEA $\gamma$-ray spectrum evaluation intercomparison.

Hypothesis testing is extremely important for other phases of chemical analysis, in addition to the question of analyte detection limits. Through the use of appropriate test statistics, one may test data sets for bias, for unexpected random error components, for outliers, and even for erroneous evaluation (data reduction) models (37). Because of statistical limitations of such tests, especially when there are relatively few degrees of freedom, they are somewhat insensitive (lack power) except for quite large effects. *For this reason it is worth considerable effort on the part of the analyst to*

*construct his CMP so that it is as free from or resistant to bias, blunders, and imperfect models as possible.*

Fig. 12 gives an illustration of the difficulties of detecting both systematic error and excess random error. There we see that just to detect systematic error when it is comparable to the random error ($\sigma$) requires about 15 observations; and to detect an extra random error component having a comparable $\sigma$ requires 47 observations (34). In a simple case involving model error it has been shown that analyte components omitted from a least-squares multicomponent spectrum fitting exercise must be significantly above their detection limits (given the correct model) before mis-fit statistics signal the error (37). This limitation in "statistical power" to prevent significant model error bias, especially in the fitting of multicomponent spectra, is one of the most important reasons for developing multidimensional chemical or instrumental procedures and improved detectors of high specificity or resolution.

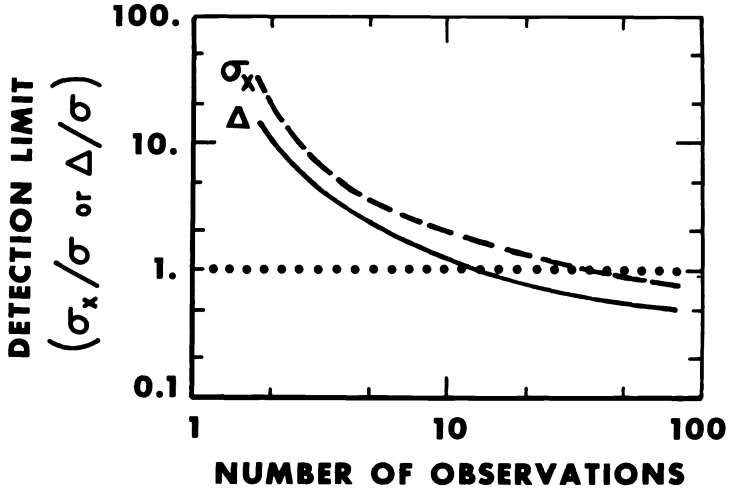



Fig. 12. Detection limits vs. number of observations for extraneous random error ($\sigma_x$, dashed curve) and systematic error ($\Delta$, solid curve). (Ref. 34).

One example will be given to illustrate the successful use of a test statistic ($\chi^2$) to assess the accuracy of a fitting model. This involves an assumed linear calibration curve for the analysis of CaO by x-ray fluorescence analysis. Critical to the success of the statistical test was knowledge of the intrinsic (Poisson) random error standard deviation. Residual patterns before and after using this knowledge are shown in Fig. 4b, the final pattern of deviations being given as a function of sample position (34). In this case $H_0$ (straight line calibration model) was rejected by the extremely large value of $\chi^2$/degree of freedom, about 50 for five degrees of freedom, the critical level ($\alpha$=0.05) being just 2.2.

Data Evaluation – Pitfalls Associated with Peak Detection

The final phase of the CMP consists of data reduction. Just as with each of the other serial steps of the CMP, this final step can seriously compromise the quality of the overall result if an inaccurate evaluation model is employed. To illustrate some of the pitfalls which may arise, the problem of multicomponent peak detection will be examined.

Before examining this issue in detail, let us first consider the various classes of evaluation models. These range from the most general form given in the first equation in Fig. 3, to the simplest, gross signal minus blank, discussed earlier. Taking E to represent a controllable discrete or continuous experimental variable (energy, time, location, ...), we find

full spectrum

$$y(E) = f(B; x_1, \ldots x_p; E) + e(E) \quad (2a)$$

$$y(E) = B(E) + \underline{A}(E)\cdot\underline{x} + e(E) \quad (2b)$$

$$\downarrow$$

peak region(s)

$$y(\Delta E) = B(\Delta E) + \underline{A}(\Delta E)\cdot\underline{x} + e(\Delta E) \quad (2c)$$

$$y(\Delta E) = B(\Delta E) + S(\Delta E) + e(\Delta E) \quad (2d)$$

$$\downarrow$$

point observation

$$y = B + S + e \quad (2e)$$

where: y represents the gross signal; B, the baseline or background or blank (including unresolved interference); $x_1, \ldots x_p$, the concentrations of analytes 1, ...P; $\underline{A}$, the calibration matrix for these P-components; e, the random error vector; $\Delta E$, a restricted "peak region" of the controlled variable, and S=AX, the net signal or "peak area" for a single component.

The transition from the highly-convolved, generally non-analytic, non-linear model (2a) to the extremely simple point observation model (2e) comes about through chemical and instrumental specificity. Equation (2b) represents the situation in which the model is linear; and (2c) and (2d) arise when component responses are restricted to localized regions $\Delta E$. To illustrate the role of evaluation models in trace analysis we shall focus on these two equations which cover a major fraction of the situations in chromatography and spectroscopy. Equation (2c) differs from (2d) in that it includes a term for overlapping peaks $\underline{A}(\Delta E)\cdot\underline{x}$. In all of the equations but the last, all quantities except $\underline{x}$ are functions of the controlled variable. They are generally not simple mathematical functions, however; ignoring this fact is a principal reason for bias in data reduction. The error term e(E) is very special. Reliable knowledge about its magnitude (i.e., $\sigma$), distribution, and independence (or correlation) is mandatory to assure accurate deconvolution. Resolution and detection limits are, in turn, directly controlled by this final term.

From the perspective of equations (2c) and (2d), we find the following information critical for successful trace analytical peak evaluation: (a) the selection of the correct numbers and locations (energy, wavelength, time) of overlapping peaks, (b) establishing the correct shape for the baseline, and (c) the proper application of statistical detection theory for extracting subliminal peaks from the measurement noise. The problem of unresolved peaks is especially severe when one is searching for a trace component in the presence of a large neighbor. Frequently a perfectly adequate fit may be obtained when the trace member is excluded from the model (even though it exceeds its detection limit). At the same time, this omission adds significant bias to the estimate for the major peak (34,37). Both the difficulty of this problem and the importance of non-statistical solutions are immediately evident from Fig. 13. In this figure close lying $\gamma$-ray lines are shown for two radioactive gold nuclei (important in activation analysis) which are essentially non-separable (by chemistry or differential decay). What looks like a doublet with the primordial NaI $\gamma$-ray detector becomes a quartet under the resolution of Ge(Li).

Inaccurate baseline shapes probably account for the second most important difficulty in peak fitting, following the questions of the number and location of peaks. Incorrect assumptions about baseline shapes in fact play the same part in peak detection and estimation as do inaccurate blanks, in simple analyte measurements. Probably the most common assumption used in peak estimation is that the baseline is linear. If non-linearities are small compared to random measurement errors over the width of the peaks, which is commonly the case, then the assumption is acceptable, and little bias (or false positives and negatives) will result (79). Small amounts of simple non-linearity may also be tolerated using somewhat more sophisticated fitting techniques, but assumed (mathematical) analytic shapes may nevertheless fail to adequately represent the contortions of real baselines.

A dual approach to the problem was taken a few years ago, in which the fitted baseline is not required to conform to any particular mathematical function, but rather to be consistent with the following: (a) the shape of the peak being estimated, (b) the observed gross spectrum (peak plus baseline) and its random errors, and (c) the maximum "smoothness" allowed by (a) and (b) (80).

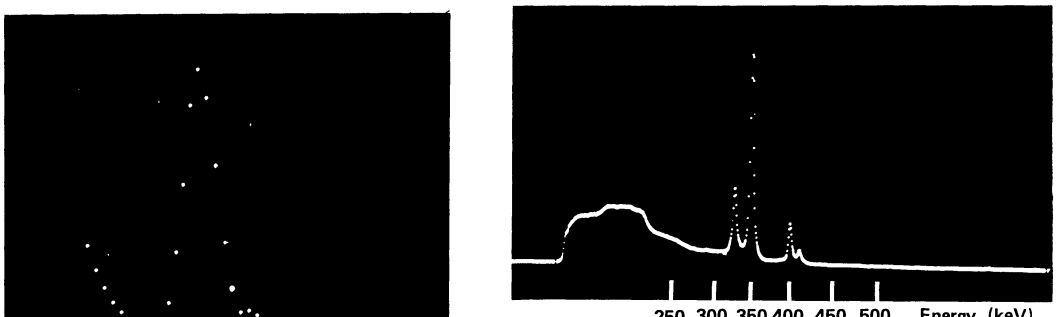



Fig. 13. Gamma-ray spectrum from bremsstrahlung-activated gold. Peaks at 333, 356, and 426 keV arise from $^{196}Au$; the 412-keV peak comes from $^{198}Au$ (Ref. 34). NaI(Tl) detector, left plot; Ge(Li) detector, upper plot.

This method which has been dubbed the Smoothest Consistent Baseline (SCB), is an example of fitting a "partially known model" – peak shape known, baseline shape not known – *in which the partial knowledge is recognized rather than being replaced with an assumed analytic function.* Solution is made possible by the inequality constraint (minimum baseline structure), and the fit is accomplished through the use of the Lagrange undetermined multiplier. The dual part of this approach is to employ physical-chemical knowledge of the possible local excursions in actual baseline structure, in order to estimate limits for peak bias. This follows from a "Bias Matrix" which derives directly from the constrained fit (80).

An illustration of a rather severe deviation from baseline linearity is given in Fig. 14 which shows a γ-ray peak ($^{65}Zn$) above a Compton edge of $^{60}Co$. In Fig. 15 one superposition of the $^{65}Zn$ peak with this non-linear, non-analytic baseline is given (top-most curve), as well as the true baseline and the SCB-derived baseline. This method of peak detection is currently being extended to the resolution of a peak doublet on a baseline of uncertain shape, and its mathematical properties are being further examined (81).

One of the most revealing tests of γ-ray peak evaluation algorithms was undertaken by the International Atomic Energy Agency (IAEA) in 1977. In this exercise, some 200 participants including this author were invited to apply their methods for peak estimation, detection and resolution to a simulated data set constructed by the IAEA. The basis for the data were actual Ge(Li) γ-ray observations made at high precision. Following this, the intercomparison organizers systematically altered peak positions and intensities, added known replicate Poisson random errors, created a set of marginally detectable peaks, and prepared one spectrum comprising nine doublets. The advantage was that the "truth was known" (to the IAEA), *so the exercise provided an authentic test of precision and accuracy of the crucial evaluation step of the CMP.*

Standard, doublet and peak detection spectra are shown in Fig. 16 and 17; and Fig. 18 summarizes the results (82,83). While most participants were able to produce results for the six replicates of 22 easily detectable single peaks, less than half of them provided reliable uncertainty estimates. Two-thirds of the participants attacked the problem of doublet resolution, but only 23% were able to provide a result for the most difficult case. (Accuracy assessment for the doublet results was not even attempted by the IAEA because of the unreliability of participants' uncertainty estimates!) Of special import from the point of view of trace analysis, however, was the outcome for the peak detection exercise. The results were astonishing. Of the 22 subliminal peaks, the number correctly detected ranged from 2 to 19. Most participants reported at most one spurious peak, but large numbers of false positives did occur, ranging up to 23! Considering the modeling and computational power available today, it was most interesting that the best peak detection performance was given by the "trained eye" (visual method).

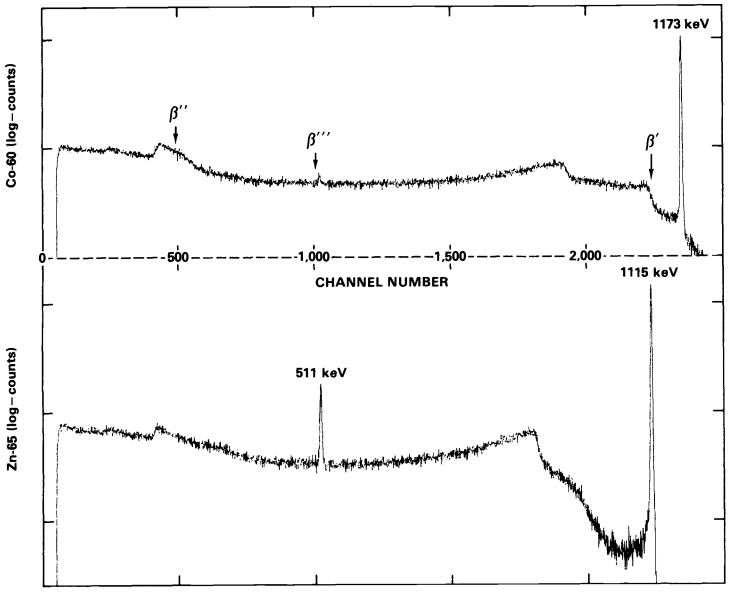



Fig. 14. Ge(Li) spectra of $^{65}Zn$ and the $^{60}Co$ baseline. Arrows indicate three baseline regions and the corresponding shape vectors. The actual baseline in Fig. 15 had shape β', at the $^{65}Zn$ peak location. β'' and β''' were selected to assess the biasing effects of alternative shapes. (Ref. 80).

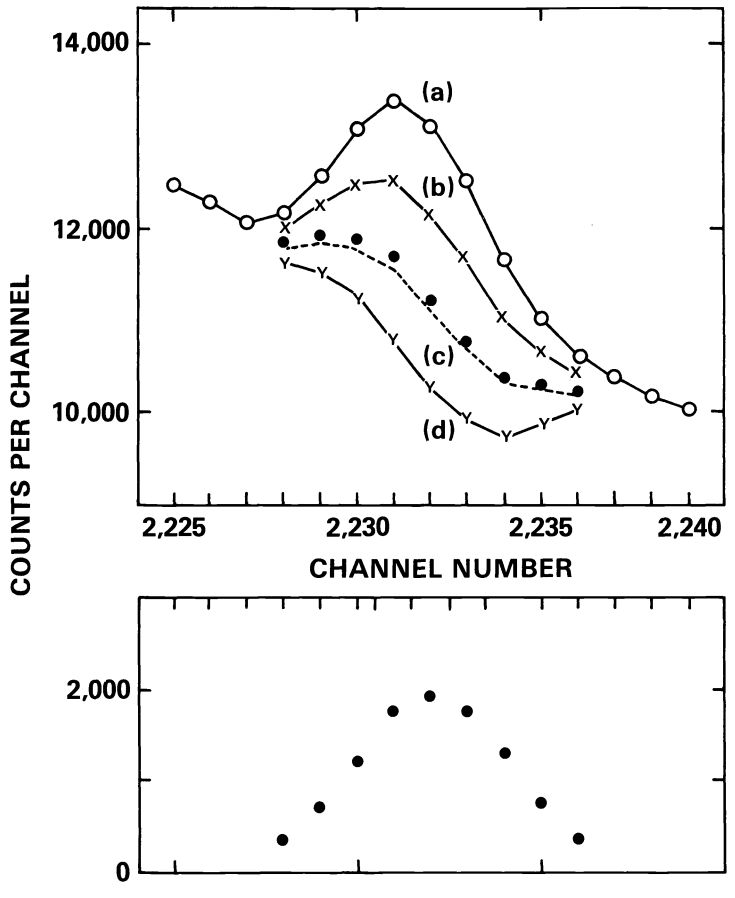



Fig. 15. Composite spectrum: $^{65}Zn$ peak (1115 keV) on $^{60}Co$ Compton edge. Lower portion of the figure gives the observed $^{65}Zn$ peak shape, normalized to $10^4$ counts between channels 2228 and 2236, inclusive. Upper portion shows the actual $^{60}Co$ continuum (solid circles) normalized to $10^5$ counts between the same channels, and the summation spectrum [open circles, curve (a)] without random error. Curves (b), (c), and (d) represent the subtraction of 5000, 10720, and 15000 peak counts from curve (a), respectively. Dashed curve (c) is the Smoothest Consistent Baseline. (Ref. 80).

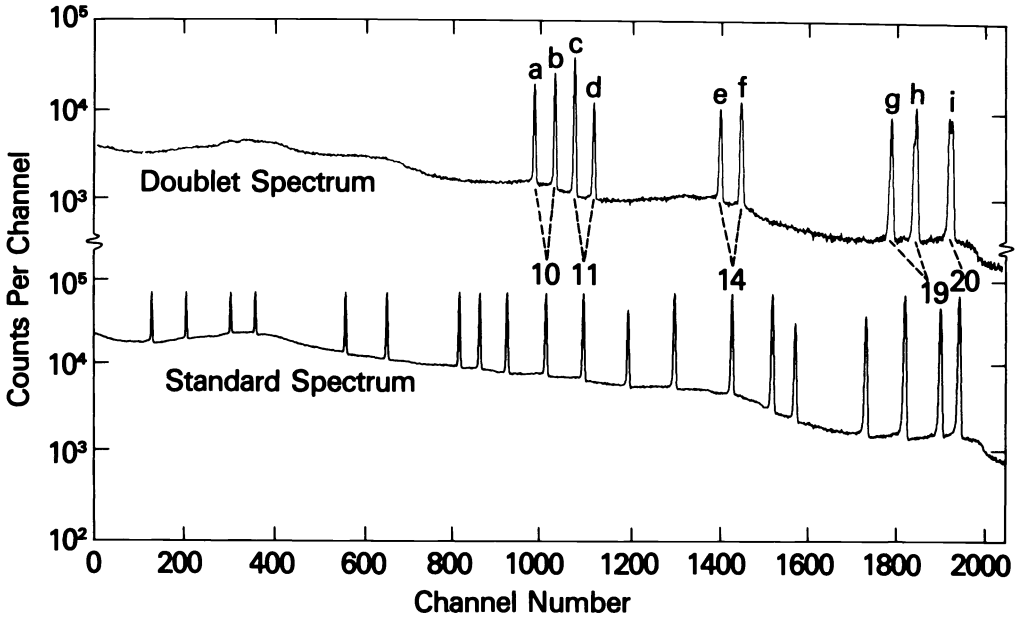



Fig. 16. IAEA test spectra. Singlet peaks from the standard spectrum (lower curve) are used as empirical templates for resolving the nine doublets in the sample spectrum (upper curve). (Refs. 82, 83).

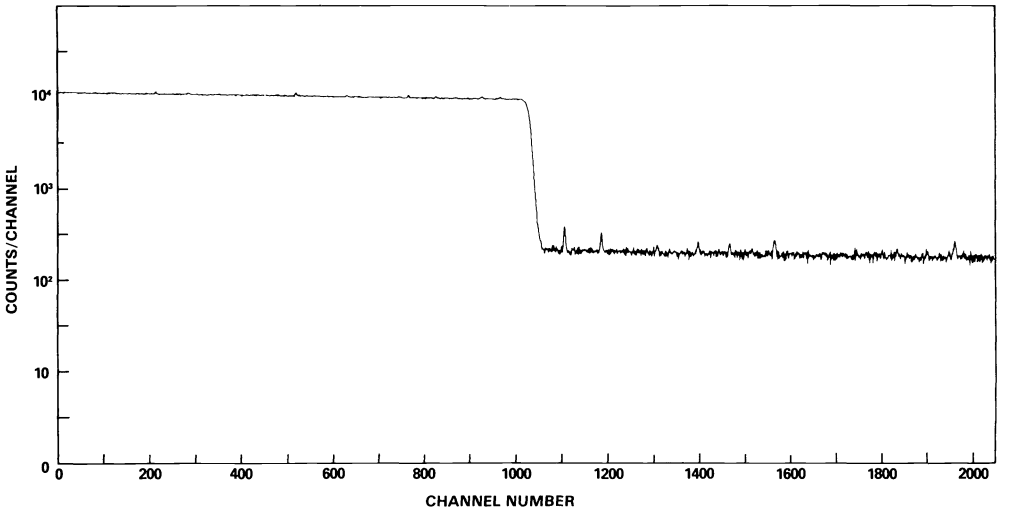



Fig. 17. IAEA test spectrum for peak detection (Ref. 82).

In view of the difficulties with peak detection, even given an ideal simulated data set, and because of its overriding importance at the limits of trace analysis, it should be instructive to look more closely at the peak detection process. This is outlined in Fig. 19. At the outset we may note some difficulties *which are absent:* we know that there is (at most) a single peak at each location (no overlap), that the baseline is for the most part linear, and the errors are Poisson and independent. The peak selected in Fig. 19 is one of intermediate detectability – about half of the participants found it. Although numerous detection algorithms may be employed, one of the simplest is illustrated here, a simple digital filter which uses six central channels for gross signal estimation and two symmetric three channel wings for baseline estimation.

**DATA EVALUATION--IAEA**
**$\gamma$-RAY INTERCOMPARISON**

**[Parr, Houtermans, Schaerf - 1979]**

| Peaks | Participants | Observations |
|---|---|---|
| 22 - Singlets (m = 6) | 205/212 | • uncertainties: 41% (none), +17% (inaccurate) |
| 9 - Doublets | 144/212 | • most difficult (1:10, 1 ch.) --49 results |
| 22 - Subliminal | 192/212 | • correctly detected: 2 to 19 peaks<br>• false positives: 0 to 23 peaks<br>• best methods: visual (19), 2nd derivative (18), cross correlation (17) |

Fig. 18. Data evaluation – IAEA $\gamma$-ray intercomparison. Column two indicates the fraction of the participants reporting on the six replicates for 22 single peaks, 9 overlapping peaks, and 22 barely detectable peaks. Column three summarizes the results, showing (a) the percent of participants giving inadequate uncertainty estimates, (b) the number of results for the doublet having a 1:10 peak ratio with a 1 channel separation, and (c) the results of the peak detection exercise (Ref. 82).

In the spectral region shown, this filter yielded a maximum between channels 285 and 286, giving an estimated net signal $\hat{S}$ of 1291 ± 362 counts (one Poisson standard deviation). As $\hat{S}/\sigma_0$ exceeds the critical level of the test statistic, 3.45, we reject the $H_0$-hypothesis and conclude that a real peak is present. ($\sigma_0$ is the standard deviation of $\hat{S}$ for the null case, when S=0.)

Multiple peak detection, as in this IAEA exercise, adds a new facet to detection theory as previously presented. That is, the $H_0$-test statistic ($z_C$ or equivalently $S_C=z_C\sigma_0$) must take into account both the false positive risk $\alpha$ and the number of decisions *overall* being made $n$. Because of the large number of channels being searched, taking into account the energy resolution, $n\tilde{\sim}200$. The effective single-decision risk $\alpha'$ is thus decreased from 0.05 to $1-(1-\alpha)^{1/n}\tilde{\sim}2.6 \times 10^{-4}$. $z_C$ is accordingly increased from 1.64 to 3.45. An analogous increase takes place in the detection limit $S_D$. In this case there are k=22 opportunities to produce false negatives, so $\beta'$ is decreased to $2.3 \times 10^{-3}$. The effects on the critical level and detection limit of these multiple statistical tests is shown graphically in the figure (84).

## THE ANALYTICAL MEASUREMENT PROCESS (AMP)

We now return to the initial topic of this manuscript, the relationship of high quality analytical chemistry to the external environment especially as related to man's technological development. The broader issue thus concerns the societal context in which analytical questions are formed and the manner in which the analytical data are interpreted for decision making. The process is necessarily iterative, with initial questions addressing the overall objective of the sampling and analytical measurement, the appropriate

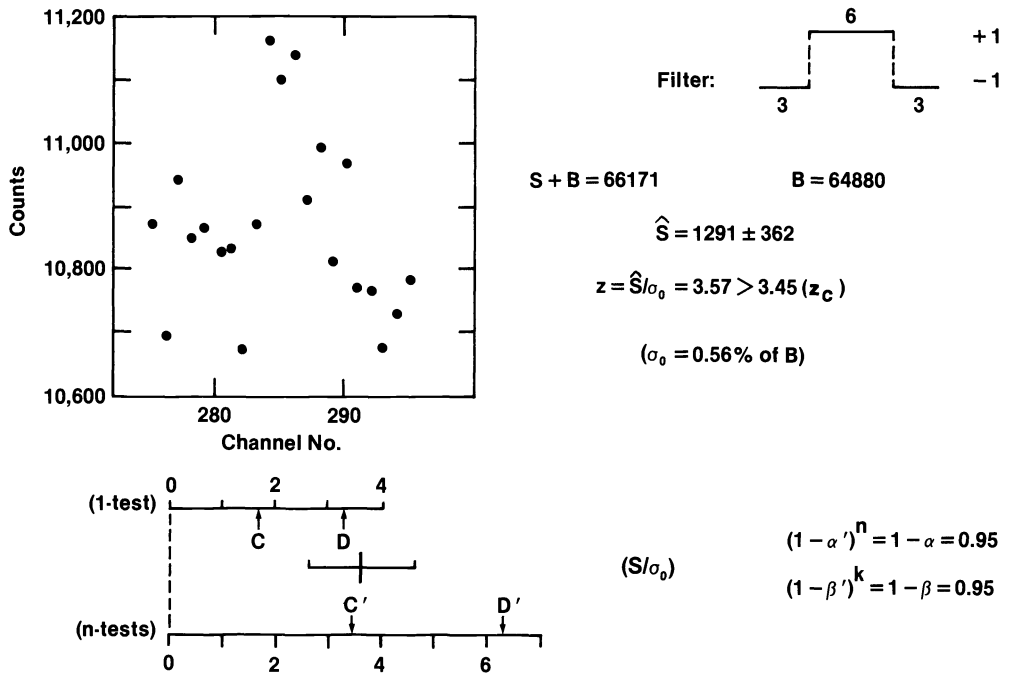



Fig. 19. Multiple-decision hypothesis testing as applied to the IAEA peak detection exercise. (About half of the participants "detected" this peak.) S + B and B represent the number of gross counts and baseline counts given by applying the digital filter to the γ-ray peak. The test statistic (z) exceeds the critical level ($z_C$), yielding the decision "detected". The critical level and the detection limit are both substantially increased because there are approximately 200 (n) detection decisions, and 22 (k) peaks to be detected.

external (environmental) model, and in effect the why, what, and how of the analytical chemical solution.

The first step in designing an appropriate Analytical Measurement Process is a Systems Evaluation (or Technological Assessment) of the overall socio-chemical problem taking into account the complementary expertise of leaders from all relevant disciplines (economists, technologists, sociologists, ...). In this process an initial identification will be made of critical information which may be obtained from scientific measurements and modeling, and the task of designing an adequate AMP may follow. In the best of circumstances this may involve measurement of a single species or property; but much more commonly, the analytical scientist is faced with the need to design a process to deconvolve what nature (or man) has blended together. Very often this requires an astute sampling design (in time and space) together with the use of "chemical fingerprints" to identify or resolve environmental processes or sources (See *Note g*). The question of what and how to measure thus reduces to one of matching the identified needs with CMP's having adequate "informing power" (85). (The frequent alternative of simply applying the CMP at hand, regardless of its performance characteristics or range of chemical species, often yields ambiguous or misleading information.) Following the introduction of a schema to define the AMP we shall examine one specific problem (Receptor Modeling) and consider the role of Archival Samples.

---

*Note g.* The term "environmental" is used here in its most general sense — i.e., to describe all parts of the external (non-CMP) system to which chemical analysis may be applied. It may thus refer to a biological, industrial, geologic, or even cosmic process.

Structure of the AMP

A representation of the Analytical Measurement Process is given in Fig. 20. The input to the AMP is a quantifiable vector $\theta$ for which society needs an estimate. There may be just one pre-specified component for $\theta$, or it may consist of any one of several alternative components (1, 2, ...P) [classification], or it may consist of a superposition of several (time and space varying) components [resolution]. Examples range from a particular disease to be diagnosed to "reading" the historical (and pre-historical) record of natural and anthropogenic pollutant sources as archived in glaciers or sediment.

Having identified the environmental vector of interest, we next must consider just what chemical species ($x_1$, $x_2$, ...$x_m$) will remain following the external (environmental) process for the reliable estimation of the components of $\theta$. In an ideal world each $\theta_i$ would correspond to a specific $x_j$, and each $x_j$ would be robust (with respect to environmental perturbations and uncertain assumptions and models) and adequately quantifiable. Occasionally this happens: a specific $\theta_i$ may be the only source of a unique "marker element" $x_k$. More commonly several $x_j$'s are associated with several other $\theta_i$'s, so the task becomes one of deducing adequate chemical vectors or patterns ($x_1$, $x_2$, ...$x_m$) for each of the source vectors $\theta_i$. *It is not an exaggeration to emphasize that these steps constitute the most important and most difficult part of the effort — identifying the critical environmental components, and specifying the chemical (or physical) quantities most likely to lead to a unique and reliable solution to the environmental problem.* It is at this stage, and in the construction of an adequate sampling scheme, that the analytical scientist can make his most crucial contribution, using his

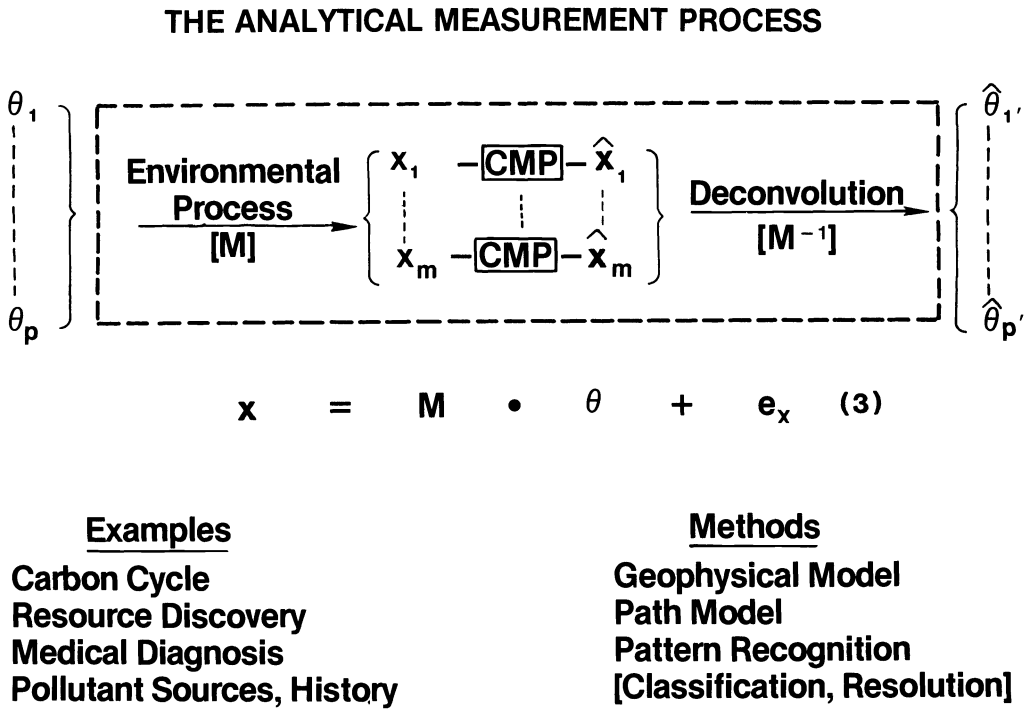



Fig. 20. The Analytical Measurement Process. The external quantities to be estimated (e.g., pollutant source strengths) are indicated as $\theta$. Following external processes (*M*: transport, reaction, ...), chemical (or physical) "fingerprints" $x$ are examined using appropriate *CMP's*, and deconvolution of the estimated values $\hat{x}$ yields estimates for the external quantities of interest $\hat{\theta}$.

knowledge of the chemical and physical characteristics of the external components and of the likely effects of external processes on the resultant chemical vectors.

The selection of an appropriate (set of) CMP's is next, given the sampling strategy and the nature and expected concentrations of the $x$'s. From the perspective of the AMP, these CMP's must be absolutely appropriate and reliable. Technical expertise in this area on the part of the analyst is mandatory and his ability to produce quality data is a foregone conclusion by others in society who would enlist his aid in attacking the external problems. (The difficulty in meeting this mandate, the subject of the preceeding section, unfortunately is not fully appreciated by large segments of society.) After the selection and execution of the CMP-set, one has an estimated x-vector $\hat{x}$, and the classical analytical chemical task is complete.

The final scientific task, however, remains: deconvolution of the estimated chemical concentrations, in order to derive an estimate for an external vector $\hat{\theta}'$. This is the realm of modeling. Success in estimating the initial vector $\theta$ (as opposed to a model-created vector $\theta'$) depends upon the validity of the assumptions underlying the model, and whether the model presents an adequate representation of reality. In this regard, one of the most important contributions of quality analytical data is in the area of model validation (28, 86). Methods of deconvolution range from the use of: (a) complex, hierarchical deterministic and stochastic geophysical models (5,87); to (b) models which presume only environmental paths or linkages and use multivariate analytical data to deduce unobservable or "latent" variables (88); to (c) chemical pattern recognition techniques which treat observed ambient chemical profiles as linear superpositions of a presumed (or inferred) set of external sources (89).

Environmental Receptor Modeling

One current illustration of an AMP in which analytical scientists are deeply involved – both in the definition of the external vector components and their chemical signatures, and in the estimation of the components from multispecies (and multivariate) analytical data – is Receptor Modeling. This expression has been given to the process of fitting chemical patterns as obtained at a (sample) "receptor site" to the set of "fingerprints" characterizing potential sources of environmental pollution (90,91). Complementary techniques involving factor analysis and least squares component fitting are employed to estimate, and occasionally even identify (unknown) contributing sources.

Formally, the matrix equation Equation (3) (See Fig. 20) relating the observed chemical signal $x$ to the source vector $\theta$ is nearly identical to Equation (2), which represented the relation of the analytical signal $y$ (instrument response) to this same chemical vector $x$. *But the difference is enormous.* First, the equations are connected in a hierarchical manner, the solution $\hat{x}$ of the one becoming the input to the other. Second, the model matrix $A$ for equation (2) has a very good chance of being accurately understood, evaluated, and controlled. The model matrix $M$ for Equation (3) suffers in all of these respects, and depends entirely on the luck and perspicacity of the analytical scientist in constructing an adequate representation of the natural process. Of primary importance in this regard, as mentioned above, is the selection of reliable, informative chemical species for analysis. One major distinction between the chemical system (CMP) and the environmental system (AMP) is that we are *not* in control. The assumptions of randomness and stationarity are exceedingly difficult to defend for the environmental system, so we can only hope to circumvent this problem through the selection of robust species and shrewd sampling strategies. Another signal difference is that in the chemical system we frequently have the opportunity to separate (chemically or instrumentally) the various chemical components, yielding interference-free detection limits and decreased dependence on model assumptions. Nature does not provide specificity, however; therefore deconvolution, with all of its limitations and pitfalls is the only route available. Despite these differences, the formal solutions to Equations (2) and (3) necessarily proceed in parallel. One significant difference between the two equations is the error term. In Equation (2), $e_y$ is a random error in the gross signal, and it is generally normally-distributed; $e_x$ in Equation (3) however, refers to the actual chemical concentration variations. Under the best of circumstances, $e_x$ is randomly distributed, but it is likely to be log-normal.

One receptor modeling illustration will be given. Fig. 21 depicts the chemical profile for an ambient sample of fine (<2.5 μm-diameter) atmospheric particles collected in Portland, Oregon during the autumn of 1977. Included in the figure are composition profiles of the four presumed sources θ giving rise to the carbon in the particles. Chemical species $x$ used to characterize each of the patterns are: radiocarbon ($^{14}C/^{12}C$), total-carbon, silicon, nickel, and lead. The objective of the study was to deduce the sources. These measurements formed part of a larger two-year investigation, the Portland Aerosol Characterization Study, whose objective was to assess the relative importance of urban and non-urban (transported) particulate pollution sources to the air quality in downtown Portland (29,92). Carbon is a substance of special interest, because of its effects on health, climate and visibility (93). Of particular importance in this regard also was an assessment of the relative carbon contributions from man's activities and from natural processes. Design of the AMP, including θ and $x$ identification, took these facts into account. It was surmised, for example, (based on an emissions inventory and other "prior" information) that θ should have primarily only the four components shown, and that $^{14}C$, Si, Ni, and Pb might serve as reasonably reliable "unique markers" for each of them. Measurement of total carbon provided one degree of freedom for a consistency check.

The problem represented in Fig. 21 is an example of Chemical Mass Balance, and source strength estimates $\hat{\theta}$ result simply from inverting (or least-squares fitting) the matrix equation 3. Details of the solutions are given in Fig. 22. The conclusion is that vegetative carbon $C_v$ accounts for about one-fourth of the total particle mass and 80% of the carbonaceous mass.

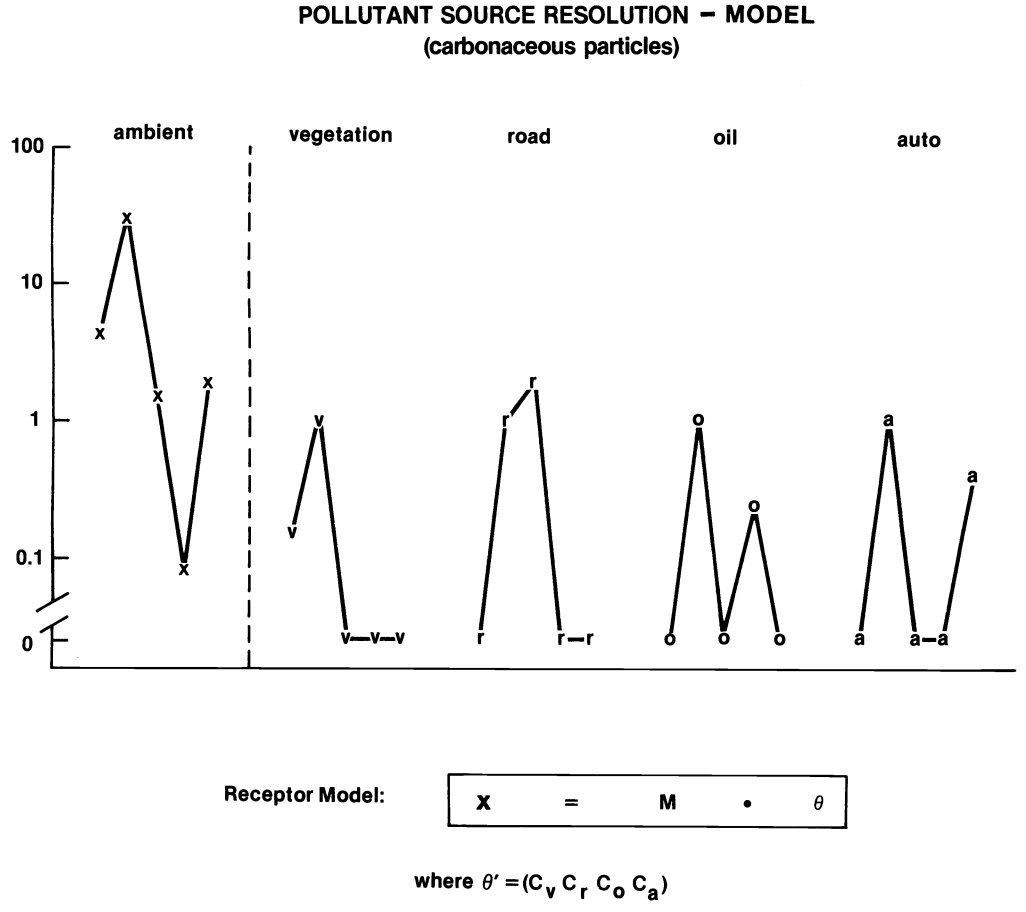



Fig. 21. Pollutant source resolution-model (carbonaceous particles). The ambient air particulate sample was taken to represent a linear combination of the four carbonaceous pollutant sources. The five measured "chemical" features were $^{14}C$, total-C, Si, Ni, and Pb-for each of the samples. (Ref. 94).

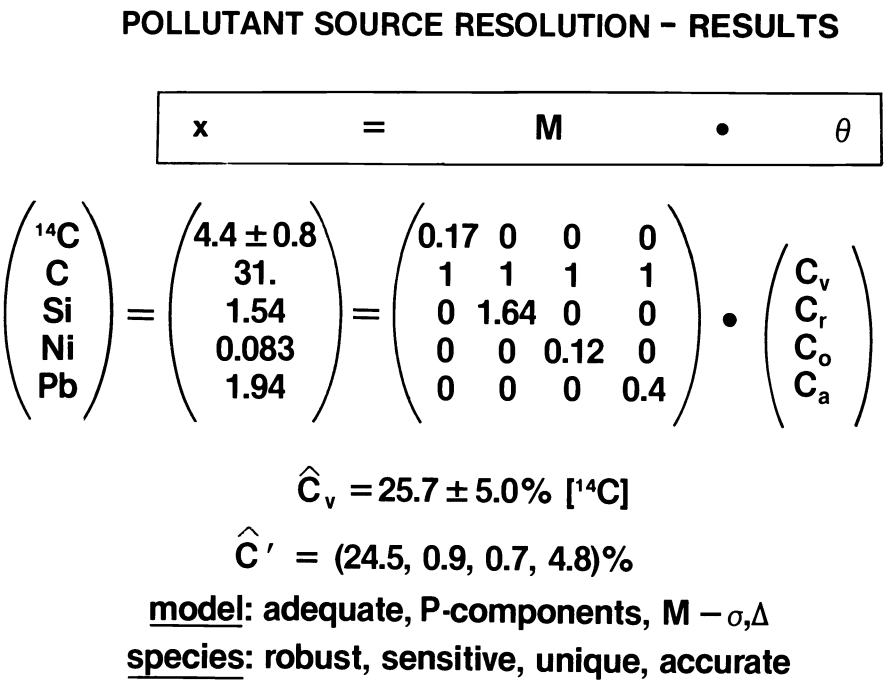



Fig. 22. Pollutant source resolution-evaluation. Results are given for the estimated percent of mass in the ambient particles due to carbon from each source. (Ref. 94).

Carbon from motor vehicles (∿5% of the total particle mass) is the only remaining significant source of the carbon in the sample. The importance of this conclusion, for Portland, is due to the fact that major sources for potential vegetative carbon pollution are associated with man's burning of wood and grass in connection with grass seed ("field burning") and timber ("slash burning") industries and with the use of wood as a residential fuel. At the time of sampling (October 18, 1977) major slash burning was underway in a neighboring forest, and the receptor model results were therefore decisive in indicating the importance of this source to the fine particulate pollution in the city. By multiplying the $\theta_1$ result ($C_v$) by the fine particle concentration, this contribution was found to be quite significant, 13 $\mu g/m^3$-C (or 21 $\mu g/m^3$-fine particles) from vegetative burning (94).

The independent confirmation of $\theta_1$ (i.e., $C_v$), through the measurement of radiocarbon, had major importance. That is, since the $^{14}C$ concentration in vegetation is equivalent to that in all living matter and the concentration in fossil fuel is nil, this isotope gives a unique measure of the biogenic component. The second estimate for $C_v$ (24.5%) is indirect, and therefore subject to a number of possible model (and parameter) uncertainties not affecting the radiocarbon estimate. For this reason, and because of the increasing concern with man's contributions to carbonaceous gases and particles in the atmosphere, environmental radiocarbon measurements are assuming considerable importance (29).

The apparent simplicity of this Receptor Modeling example is deceptive. The availability and consistency of both independent direct and indirect receptor modeling estimates is, in fact, quite rare. Difficulties which characterize current efforts in this developing field are largely contained in this example, as presented in Fig. 22. First, the model *must be* complete (all components represented) and matrix elements $M_{ij}$ must be correct if unbiased results are to be obtained. Ideally the $M_{ij}$ should have negligible (systematic and randon) error and the component signatures should be sufficiently dissimilar that the matrix does not approach singularity. "Unique" tracers, such as $^{14}C$, Si, Ni, and Pb in this example, are extremely helpful in this respect.

Completeness of the model was not so difficult here, because relatively few major sources of carbon were active at the time of sampling.

The identification of all significant components for the total particulate mass is very much more difficult, and it probably constitutes the major limitation to accurate receptor modeling at the present time (90,91). The difficulty of providing accurate parameter values is illustrated by each of the unique tracer matrix elements in Fig. 22. $M_{11}$ ($^{14}C$) depends on the age span of the wood being burned, primarily as a result of the increase in biospheric radiocarbon during the last few decades of nuclear testing (94); $M_{32}$ (Si) actually depends both on particle size and the origin (rural, urban) of the road dust component; $M_{43}$ (Ni) is also particle-size dependent, and the uncertainty in its value is further compounded by the fact that the oil source may exhibit considerable variations in composition depending upon its origin (95); lastly, $M_{54}$ (Pb) may vary by as much as a factor of two depending upon the assumed mix (catalyst, non-catalyst, diesel) and speed of traffic (96).

A second difficulty relates to the measurements, themselves (the $x$ vector). This is primarily a question of the quality of the CMP, but it nevertheless is an outstanding problem for receptor modeling as currently practiced. Because of the small sample sizes, trace concentrations, and multiple opportunities for contamination and other sampling artifacts, the production of reliable measurements having adequate discriminating power is still a major limitation for the method. In fact, despite the importance of carbonaceous species to particulate pollution, suitable measurements are relatively rare (97). It should be emphasized that these types of difficulties and uncertainties are typical for receptor modeling currently. Much larger data sets and models having up to perhaps 10 components are common, but in every case model uncertainties impose such severe limitations that the results must be considered, at best, suggestive or semi-quantitative. (Agreement to within a factor of two between observed and fitted results is considered "good".) New directions are underway, however, for improving: (a) the quality of the models using techniques such as "target transformation" factor analysis (98) and "wind trajectory analysis" (90), and (b) the quality of the process through the preselection of robust and unique tracers such as $^{14}C$.

This brief illustration of Receptor Modeling, taken as a case study of the present state-of-the-art of at least one facet of the Analytical Measurement Process supports the following conclusions.

- "Chemical fingerprints" have considerable potential for the quantitative identification of sources of environmental contaminants. They promise greater reliability than earlier approaches based on emission inventories and dispersion modeling (92).

- The formal similarity between Equation (2) [CMP], and Equation (3) [AMP] implies the conditions for reliable deconvolution of the environmental source contributions.

- Difficulties which were recognized in the CMP case – reliable knowledge concerning the number of components, their matrix elements ($A_{ij}$), the blank and contamination, the error distributions and control of the system – are enormously amplified for the environmental system.

- The best chance for quality results for such a complex and open system lies with the skill and knowledge of the analytical scientist in designing powerful sampling strategies, in the selection of robust and "informing" (85) species for analysis, and in designing the CMP to match the problem (AMP) rather than simply utilizing a readily available multielement method in the hope that its performance characteristics (species, precision) will yield a unique solution.

- The societal stakes are high. The issue of the burning of wood and other solid residential fuels, for example, is the subject of an international, interdisciplinary conference this year (99). It is representative of a large class of sociochemical problems in which the energy/environment balance is the key issue, and for which the quality of our future life is dependent, at least in part, on the quality of our trace analytical data and the design and validation of the Analytical Measurement Process.

Archival Samples — The Environmental Specimen Bank

One of the most powerful forms of insurance to compensate for our present limitations in analytical methods and for our (necessary) ignorance concerning sociochemical issues of the future is the archival sample. The archival sample is one which comprises representative materials (environmental, biological, geological, ...), collected at a particular point in time (and space), which may be stored reliably for analysis in the future. Such substantial "records" include both natural and anthropogenic collections. Major knowledge concerning the history of the earth and its climate, for example, is evolving through the analysis of rocks, ocean sediment, and ice cores. This last natural archive, as represented by the polar ice caps, is one of the most exciting from the viewpoint of trace analysis. Concentrations at and below $10^{-9}$ g/g are the rule; the matrix is pure and conditions of storage are nearly ideal; and contamination-free information is stored for up to $10^5$ years on temperature variations (through oxygen isotopic analysis (100)), volcanic activity and the accumulation of terrestrial and extraterrestrial dust (through "glaciochemical dating" (24)), solar activity (through cosmic-ray produced nuclides (101)), and ancient atmospheric compositions (through trapped air and carbon dioxide (16-18)). Yet, it is only in recent years that sampling and measurement techniques have made it possible for us to tap this vast resource which captured the "recent" ($10^5$ years) history of the earth. The classic "anthropogenic" archives, the physical records of ancient man, form the basis of modern archaeology. It is interesting that the archaeological record, also, is under the scrutiny of modern trace analysis (under the label "archaeometry") for the purpose of generating inferences concerning ancient communities, trade routes, and nutrition (102).

During the last century there have been several dramatic human influences on the environment, most of which have been a direct consequence of technological development. Given our current understanding of the consequences of some of the resulting environmental insults, and the urgent need to project the influence of our current activities on the future, it would be quite desirable to have available large and well documented environmental and biological samples covering this era. Some exist. To a limited extent tree rings or museum artifacts (of human hair, for example) can provide useful information. Vintage wines have provided valuable data on tritium and radiocarbon (19, 103). But because such archives are limited both in scope and reliability, and because of unanticipated needs and analytical methods of the future, major "sample banking" activities have been initiated on an international scale. One of the most important of these was started in a single laboratory, about 20 years ago: the collection of air samples for assessing trends in atmospheric carbon dioxide (104). A more recent effort is the Environmental Specimen Bank (15).

The Environmental Specimen Bank represents the first major international effort designed to provide a long-term biological-ecological sample archive for the purpose of preserving the (material) record of man's impact on the biosphere. The U.S. program began in 1975, and during recent years there has been a significant cooperative effort with the Federal Republic of Germany. By its very nature, this program is a superb illustration of the complete Analytical Measurement Process. It is focussed from the very beginning on a major societal issue (man's chemical influence on the present and future biological environment) of such scope that it transcends national interests as well as local boundaries in space and time. The international and interdisciplinary aspects of the problem are such that the analytical scientist must work closely already at the system design phase with many others having complementary expertise. And the scientific rigors of the program equal or surpass any that we have considered above. Severe matrix effects and the enormous range of substances and concentrations make the CMP extremely challenging. Also, chemical and biological sample integrity must be guaranteed at the time of collection as well as for an indefinite period of future storage, including trace and ultra-trace inorganic and organic species which may not as yet even have been defined! Highlights of the U.S. program, from the joint perspectives of the AMP, the CMP and quality, will be presented below.

The goals of the current National Bureau of Standards' current five-year Pilot Program are summarized in Table 3 (105). The skeletal structure of the Analytical Measurement Process, ranging from problem definition to feedback, is shown in Fig. 23 (106). Upon examination of the various phases of the overall process, two aspects recur: the breadth (chemical,

TABLE 3. Major Tasks of the Pilot Specimen Bank Program (5 year program)

1. Collection of four types of environmental samples/sampling protocol development.
   - Human accumulator – liver
   - Marine accumulator – mussels
   - Food accumulator – grain/milk
   - Air accumulator – lichen/moss/air filters
2. Real-time analysis of a representative fraction of the samples.
3. Continued research on analytical methodology and sample preparation.
4. Evaluation of long-term storage under different conditions.
5. Retrospective analysis of stored samples.
6. Data management and dissemination.

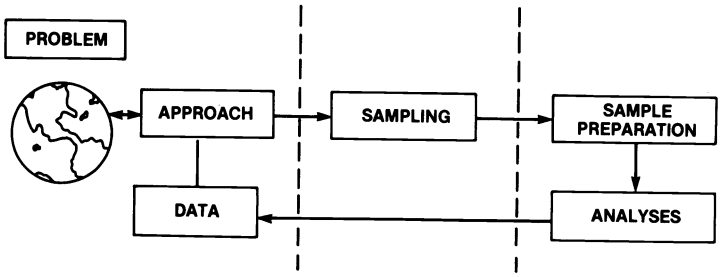



Fig. 23. Solving an environmental problem – interfaces in the Analytical Measurement Process. (Ref. 106).

interdisciplinary) of the work, and the quality assurance component. In the *sampling* phase, for example, it was first necessary to develop a carefully constructed sampling protocol, in cooperation with those performing autopsies, in order to obtain contamination-free human liver samples. Within the bounds of practicality and following a period of education and coworking, autopsy procedures were altered to achieve this end. Different disciplinary views (and the need for very careful communication) were early evident from the perceived meanings of such a term as "clean" – i.e., interpreted as "sterile" vs. "non-contaminated" chemically. Some of the measures established to prevent contamination during sampling include the use of talc-free gloves (a change from conventional autopsy procedures); use of water of extreme purity for rinsing of samples and apparatus; handling, transfer, and storage using minimally-contaminating material (polytetrafluoroethylene "FEP Teflon") (*Note h*); and use of a high purity titanium-bladed knife for all cutting operations. The selection of the titanium knife was an interesting exercise in contamination prevention. Pure titanium was chosen so that, if cutting contamination should occur, it would be limited to one element – that one (Ti) being currently believed to be neither essential nor toxic to man. Had stainless steel been used on the other hand, one can consider the effect of a 10 μg chip becoming mixed with a liver sample. Since steel contains about 10% nickel, and the concentration of nickel in liver ranges from about 0.01 to 0.2 μg/g, the minute chip from the knife could lead to a blank exceeding the signal by a factor of 100 (105).

Viral contamination, and the resultant hazard to future workers, is a special problem for this program. Screening for hepatitis type-B antigen is performed by the simultaneous collection of a blood sample. Negative results

*Note h.* Mention of commercial product names is for purpose of identification only; it does not imply endorsement by the National Bureau of Standards.

are not necessarily conclusive, however. Because of this, and because liquid nitrogen storage (vide infra) preserves viruses, a histological section also is obtained, and the histological slides are preserved as part of the specimen bank.

Sample *storage* demands are extreme. Following cryogenic homogenization (brittle fracture), sample aliquots are submitted for immediate ("real time") inorganic and organic chemical analysis and four different modes of storage. The storage container material ["PFA Teflon"] was selected only after extensive tests of 12 different plastics for impurities, using isotope dilution mass spectrometry and neutron activation analysis. Samples are stored as follows: (a) freeze-dried at room temperature, and frozen at (b) -25 °C, (c) -80 °C, and (d) in liquid nitrogen vapor at -120 °C to -190 °C. Because liquid nitrogen contains traces of oil and dust, packaging in two sealed FEP-Teflon bags is provided to protect the frozen sample during long-term storage. (See Fig. 24, (105,107)) Evaluation of concentration changes with time under alternative modes of storage is a critical part of the program. This constitutes another example of hypothesis testing – i.e., the null hypothesis (no change) must be tested at a given level of significance taking into account analytical (measurement error) and heterogeneity random errors. For this purpose a diagram has been constructed relating the probability of detecting a given percentage change in concentration to the number of specimens analyzed, the subsample heterogeneity, and the analytical error. If after a three year interval 30 liver samples are analyzed, for example, and if heterogeneity and analytical error (standard deviation) are each 5%, then we find that there is a 98% chance of detecting a relative change in analyte concentration of 6%.

*Analysis* of the samples follows the usual NBS (Center for Analytical Chemistry) quality assurance policy, as established for the certification of Standard Reference Materials. That is, wherever possible each analyte is determined in duplicate by at least two independent analytical techniques, using SRM's for procedure validation. Typical results and analytical methods for inorganic trace elements in human liver are shown in Table 4. An impressive example of one of the goals of the Pilot program is the result for platinum (26). Here, the need to develop improved analytical methodology was so stimulated by the environmental specimens, that an exquisite radiochemical method (including homogeneous precipitation) was developed to decrease our ignorance about the ultratrace element composition of human liver. At $5 \times 10^{-11}$ g/g, platinum has the lowest concentration yet determined in such samples. In stark contrast to such high quality analysis, laboratory intercomparisons of trace levels of hydrocarbons in sediments (108) and in mussel tissue (109) have shown variations by two orders of magnitude, and by factors of three to four, respectively. Intercomparisons, as one very informative measure of quality assurance, form an intrinsic part of both German and American programs.

Finally, one interesting environmental implication has evolved from the recent measurements at NBS. Concentration ratios for lead (∿ µg/g) and thallium (∿ ng/g) have been compared for human liver, bovine liver (NBS-SRM #1577) and the earth's crust. These ratios (Pb/Tl) were found to be approximately $10^3$, 40, and 20, respectively (107), suggesting that biological lead contamination is quite significant, especially for man (as opposed to cattle) with his more direct and lengthier exposure. Exposure evidence and pollutant source identification are stated objectives of the Environmental Specimen Bank, and thus this anthropogenic archive will make possible inferences which will very closely complement those drawn from non-biological environmental samples as discussed above under Receptor Modeling.

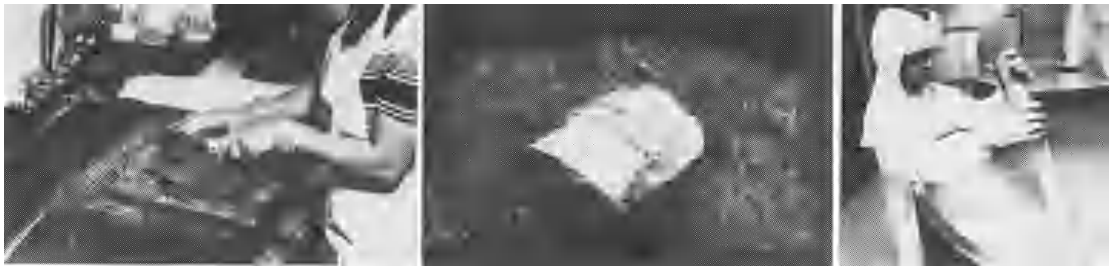


Fig. 24. Sampling, packaging, and storing a human liver sample. (Ref. 105, 107).

TABLE 4. Analysis of Human Liver for Trace Elements (Ref. 105)

| Element | Concentration (μg/g) | Analytical method[a,b] | Priority[c] |
|---|---|---|---|
| Be | ND | | 1 |
| F | 0.06 - 1.4 | | 1 |
| Al | 1.6 - 2.6 | AAS, INAA | 2 |
| V | <0.007 - 0.09 | | 1 |
| Cr | 0.005 - 0.27 | INAA | 1 |
| Mn | 0.5 - 1.9 | AAS, INAA | 1 |
| Fe | 70 - 210 | INAA | B |
| Co | 0.017 - 0.16 | INAA | 1 |
| Ni | 0.009 - 0.32 | AAS | 1 |
| Cu | 3.2 - 14.7 | AAS, ASV/LSV, INAA, IDMS | 2 |
| Zn | 31 - 80 | ASV/LSV, INAA | 1 |
| As | 0.006 - 0.46 | AAS, RNAA | 1 |
| Se | 0.097 - 0.68 | AAS, INAA, RNAA | 1 |
| Rb | 7 - 12 | INAA | B |
| Sr | 0.01 | | 2 |
| Mo | 0.4 - 1.6 | RNAA, INAA | 1 |
| Pd | ND | | 1 |
| Ag | 0.006 - 0.07 | INAA | 2 |
| Cd | 0.5 - 4.9 | AAS, ASV/LSV, INAA, PGAA | 1 |
| Sn | 0.08 - 0.65 | | 1 |
| Sb | 0.01 | INAA, RNAA | 2 |
| Ba | 0.01 | | 2 |
| Pt | 0.00005[d] | RNAA | 2 |
| Hg | 0.005 - 0.25 | INAA | 1 |
| Tl | 0.001 - 0.009 | IDMS | 1 |
| Pb | 0.8 - 2.3 | ASV/LSV, IDMS | 1 |

[a] Analytical method in use at NBS in the pilot specimen bank program.

[b] AAS = atomic absorption spectroscopy;
ASV/LSV = anodic stripping voltammetry/linear sweep voltammetry;
INAA = instrumental neutron activation analysis;
IDMS = isotope dilution mass spectrometry;
RNAA = radiochemical neutron activation analysis;
PGAA = prompt gamma activation analysis.

[c] 1 = First priority elements from "International Workshop on Monitoring Environmental Materials and Specimen Banking" held in Berlin, October 1978.
2 = Additional elements of environmental concern.
B = Biological elements. ND = No data available.

[d] Personal communication, R. Zeisler.

## CONCLUSION

High quality trace analysis is intrinsically linked to the quality of mankind's technological development, and at the same time to the quality of life as influenced by anthropogenic environmental contamination. The nature of these technological and environmental issues is such that their solution demands an interdisciplinary approach, with the analytical scientist intimately involved in every phase, beginning with problem definition and concluding with result evaluation. The overall *Analytical Measurement Process* symbolizes a systems representation of the problem, where, in societal context, the analyst must address the fundamental questions of analytical chemistry: why, what, and how to analyze, and how to quantitatively evaluate the results and their implications. Ill-defined environmental models and imperfect chemical fingerprints make this task exceedingly difficult, but astute selection of species to measure (robust, informing) and Chemical Measurement Processes to employ promise a good chance of producing reliable inferences. Receptor Modeling and Environmental Specimen Banking are relatively new but vital analytical activities which both depend on high quality trace analysis for their execution and promise some hope of understanding the environmental system and man's chemical influence.

The *Chemical Measurement Process* is extremely simple compared to the Analytical Measurement Process (of which it is a central component). Yet it provides numerous opportunities to stumble. Society expects that the CMP will be performed without error; in fact, many do not appreciate the fact that all scientific results possess uncertainty. The analyst therefore has a double responsibility to establish and execute CMP's of the highest quality, and to evaluate and report results with meaningful uncertainty bounds. The most likely route to success with the CMP is to pay attention to each of its critical links: sampling, sample preparation, measurement, data evaluation, and reporting. The most important attributes for high accuracy trace analysis include procedure validation through intercomparisons and the use of Certified Reference Materials, specificity, control of the blank, and careful attention to the error- and evaluation model-structure. Hypothesis testing is central in this effort both with respect to the detection of trace components and for the detection of bad data or erroneous models.

The future of trace analysis is exciting. Improvements in specificity and detection have extended the concentration range from about $10^{-6}$ down to less than $10^{-16}$. It is unlikely that a similar extension of the "trace" regime will again take place, for we have already achieved the level where natural concentrations and their fluctuations are limiting, as is the statistical error associated with the counting of individual particles. Control of interference (specificity) and the analytical blank in chemical analysis (CMP) imply the realization of extremely low, interference-free detection limits. This situation will not obtain in the external (environmental, industrial, biological, ...) system, however, so we may expect the basic limitations in extracting reliable chemical information from that system to be our ability to detect small changes (differential trace analysis), the external blank and its variation (environmental noise), and the natural convolution of the various chemical sources. This being the case, the most important way in which the analytical scientist can assist in understanding the environmental system is by contributing his unique expertise to the design of the Analytical Measurement Process – especially regarding sampling, feature (species) selection, CMP selection, and system modeling.

A final comment: the control of quality and system understanding can be enormously enhanced by reference samples. There exists a certain parallelism in this respect in the use of Certified Reference Materials for the Chemical Measurement Process, and the use of archival reference samples (such as the Environmental Specimen Bank samples) for the Analytical Measurement Process.

## ACKNOWLEDGMENT

Grateful appreciation is extended to many colleagues for stimulating discussions and questions, and contributions to several of the subjects covered. Special thanks for insights and data concerning SRM's go to R. Alvarez and G. Uriano; concerning the blank, to H. M. Kingston; concerning Receptor Modeling and $^{14}C$ measurements, to J. Cooper, G. Klouda, and R. Gerlach; and concerning the Environmental Specimen Bank, to R. Zeisler and S. H. Harrison. Critical reading of the manuscript by R. Velapoldi is gratefully acknowledged.

M. Sankar Das is specially thanked for creating the specific opportunity and the stimulus which led me to address the topic. Permission from the appropriate publishers for adapted or reproduced material in figures 2, 4-6, 12-18, 21, 22 is also acknowledged.

REFERENCES

1. L. A. Currie, Anal. Letts. 13, 1 (1980).
2. P. Handler, Pangs of Science, 1978 Washington, D.C. Meeting of the Amer. Assoc. Adv. Science [Reprinted in Chem. Eng. News, 29 (Apr. 17, 1978)].
3. H. Brooks, Potentials and Limitations of Societal Response to Long-Term Environmental Threats, Global Chemical Cycles and Their Alterations by Man, W. Stumm, ed., Dahlem Konferenzen, Berlin (1977).
4. R. A. Howard, J. E. Matheson, and D. W. North, Science 176, 1191 (1972).
5. International Council of Scientific Unions/World Meteorological Organization, The Physical Basis of Climate and Climate Modelling, Global Atmospheric Research Programme, Publ. Ser. No. 16 (1975).
6. U.S. Senate, Bill S.517, A Bill to Amend the Clean Air Act, (Proposes periodic review of the status of the theory of ozone depletion) (February 1981).
7. R. P. Turco, R. C. Whitten, I. G. Poppoff and L. A. Capone, Nature 276, 805 (1978).
8. W. S. Broecker, T. Takahashi, H. J. Simpson and T. H. Peng, Science 206, 409 (1979).
9. J. P. Cali, J. Mandel, L. Moore and D. S. Young, A Referee Method for the Determination of Calcium in Serum, NBS Spec. Publ. 260-36, U.S. Government Printing Office, Washington, D.C. (1972).
10. New York Times, Medical Labs May Not Be All That Accurate, p. E-9, (March 28, 1976).
11. Energy and Environmental Standards, No. 45, Hearings Before the Subcommittee on Energy the Committee on Science & Astronautics, U.S. House of Representatives, Ninety-Third Congress (1974).
12. J. McNesby, Testimony on Assessment - Need for Standards, at Hearing on The Effects of Chronic Exposure to Low-Level Pollutants in the Environment, before the Subcommittee on the Environment and the Atmosphere, House Committee on Science & Technology, U.S. Congress (1975).
13. Hydrogen – A Workshop on Societal Aspects of Hydrogen Energy Systems, L. A. Currie and R. H. Kropschot, eds., Committee on Science and Technology, U.S. House of Representatives (1979).
14. J. Williams, ed., Workshop on Carbon Dioxide, Climate and Society, IIASA, Pergamon Press (1978).
15. Monitoring Environmental Materials and Specimen Banking, N. P. Luepke, ed., Martinus Nijoff Publishers, The Hague, (1979).
16. H. Oeschger, The Contribution of Radioactive and Chemical Dating to the Understanding of the Environmental System, Symposium on Nuclear and Chemical Dating Techniques, L. A. Currie, ed., American Chemical Society, Washington, D.C. (1981), in press.
17. S. Matsuo and Y. Miyake, J. Geophys. Res. 71, 5235 (1966).
18. R. F. Weiss, P. Bucher, H. Oeschger and H. Craig, Earth Planet. Sci. Lett. 16, 178 (1972).
19. S. Kaufmann and W. F. Libby, Physical Review 93, 1337 (1954).
20. E. Marshall, Science 211, 463 (1981).
21. G. S. Hurst, S. D. Kramer and B. Lehmann, Resonance Ionization Spectroscopy for Low-Level Counting, Symposium on Nuclear and Chemical Dating Techniques, L. A. Currie, ed., American Chemical Society, Washington, D.C. (1981), in press.
22. W. Horwitz, L. R. Kamps and K. W. Boyer, J. Assoc. Off. Anal. Chem. 63, 1344 (1980).
23. R. Muller and P. Tans, personal communication (1981).
24. M. M. Herron, Glaciochemical Dating Techniques, Symposium on Nuclear and Chemical Dating Techniques, L. A. Currie, ed., American Chemical Society, Washington, D.C. (1981), in press.
25. G. V. Iyengar, W. E. Kollmer and H. J. M. Bowers, The Elemental Composition of Human Tissue and Body Fluids, Verlag-Chemie, Weinheim, New York (1978).
26. R. Zeisler and R. R. Greenberg, J. Radioanal. Chem. (to be published).
27. R. G. Heath, Interlaboratory Method Validation Study for Dioxins, Environmental Protection Agency (Jan. 5, 1979). [Quoted in Ref. 73].
28. S. N. Davis and H. W. Bentley, Dating Ground Water, Symposium on Nuclear and Chemical Dating Techniques, L. A. Currie, ed., American Chemical Society, Washington, D.C. (1981), in press.

29. L. A. Currie, G. A. Klouda and J. A. Cooper, Radiocarbon 22, 349 (1980).
30. W. T. Donaldson, Envir. Sci. Technol. 11, 348 (1977).
31. L. L. Lamparski and T. J. Nestrick, Anal. Chem. 52, 2045 (1980).
32. S. A. Borman, Anal. Chem. 53, 37A (1981).
33. J. N. Bahcall and R. Davis, Jr., Science 191, 264 (1976).
34. L. A. Currie and J. R. DeVoe, Systematic Error in Chemical Analysis, Chapter 3, Validation of the Measurement Process, J. R. DeVoe, ed., American Chemical Society, Washington, D.C., 114-139 (1977).
35. W. J. Youden, Anal. Chem. 32(13), 23A (1960).
36. R. Fukai, S. Ballestra and C. N. Murray, Radioact. Sea 35 (1973).
37. L. A. Currie, Sources of Error and the Approach to Accuracy in Analytical Chemistry, Chapter 4, Vol. 1, Treatise on Analytical Chemistry, P. Elving and I. M. Kolthoff, eds., J. Wiley & Son, New York (1978).
38. A. L. Wilson, Talanta 17, 21 (1970); 17, 31 (1970); 20, 725 (1973); and 21, 1109 (1974).
39. R. W. Fennel and T. S. West, Pure Appl. Chem. 18, 437 (1969).
40. C. Eisenhart, Science 160, 1201 (1968).
41. R. J. Howarth, Analyst 98, 777 (1973).
42. W. J. Youden, Phys. Today 14, 32 (1961).
43. L. A. Currie and G. A. Klouda, unpublished data (1981).
44. H. A. Laitinen and W. E. Harris, Chemical Analysis, McGraw-Hill Book Company, Second Edition (1975).
45. H. H. Ku, J. Res. Natl. Bur. Stand. 70c(4), 263 (1966).
46. G. Gordon, R. Greenberg and K. Eberhardt, personal communication (1981).
47. K. R. Nair, Sankhya 4, 551 (1940).
48. W. Feller, An Introduction to Probability Theory and Its Applications, Vol. 1, 2nd edition, Wiley, New York (1957).
49. C. Eisenhart, J. Res. NBS, 67c(2), 161 (1963).
50. L. A. Currie, Detection and Quantitation in X-ray Fluorescence Spectrometry, Chapter 25, X-ray Fluorescence Analysis of Environmental Samples, T. Dzubay, ed., Ann Arbor Science Publishers, Inc., p. 289-305 (1977).
51. T. R. Tyler, L. R. Chapin, R. F. Alvaro and J. L. Cox, J. Radioanalytical Chem. 12, 475-482 (1972).
52. H. F. Hamil and D. E. Camann, Collaborative Study of Method for the Determination of Sulfur Dioxide Emissions from Stationary Sources (Fossil-Fuel Fired Steam Generators), EPA-650/4-74-024 (1973).
53. A. C. Stern, H. C. Wohlers, R. W. Boubel and W. R. Lowry, Fundamentals of Air Pollution, Academic Press, New York (1973).
54. G. A. Cowan, D. G. Ott, A. Turkevich, L. Machta, G. J. Ferber and N. R. Daly, Science 191, 1048 (1976).
55. E. C. Kuehner, R. Alvarez, P. J. Paulsen and T. J. Murphy, Anal. Chem. 44, 2050 (1972).
56. K. F. J. Heinrich, Adv. X-ray Anal. 11, 40 (1968).
57. J. P. Cali, T. W. Mears, R. E. Michaelis, W. P. Reed, R. W. Seward, C. L. Stanley, H. T. Yolken and H. Ku, The Role of Standard Reference Materials in Measurement Systems, NBS Monogr. 148, U.S. Government Printing Office, Washington, D.C. (1975).
58. W. G. Cochran, Sampling Techniques, John Wiley & Sons, Inc., New York (1953).
59. B. Kratochvil, personal communication (1981).
60. W. E. Harris and B. Kratochvil, Anal. Chem. 46, 313 (1974).
61. J. K. Taylor and B. Kratochvil, Anal. Chem. (in press, 1981).
62. G. A. Uriano and S. D. Rasberry, Dimensions 65, National Bureau of Standards (Jan./Feb. 1981).
63. W. W. Meinke, Anal. Chem. 43(6), 28A (1971).
64. A. Cohen, H. S. Hertz, J. Mandel, R. C. Paule, R. Schaffer, L. T. Sniegoski, T. Sun, M. J. Welch and V. E. White, Clin. Chem. 26(7), 854-860 (1980).
65. W. E. May, S. P. Wasik and D. H. Freeman, Anal. Chem. 50, 1 (1978).
66. Standard Reference Material 1580 – Shale Oil, Announcement (Spring 1980).
67. H. T. Yolken, The Role of Standard Reference Materials in Environmental Monitoring, Marine Pollution Monitoring (Petroleum), NBS Spec. Publ. 409, Proceedings of a Symposium held at National Bureau of Standards, Gaithersburg, MD (May 13-17, 1974).
68. K. Purser, C. Russo, H. Gove, D. Elmore, R. Ferraro, R. Beukens, K. Chang, L. Kilius, H. Lee and A. Litherland, The Application of Ultrasensitive Mass Spectrometry to Nuclear Dating, Symposium on Nuclear and Chemical Dating Techniques, L. A. Currie, ed., American Chemical Society, Washington, D.C. (1981), in press.
69. B. Scales, Anal. Biochem. 5, 489-496 (1963).
70. C. C. Patterson and D. M. Settle, 7th Materials Res. Symposium, NBS Spec. Publ. 422, U.S. Government Printing Office, Washington, D.C., 321 (1976).

71. T. Hirschfeld, Anal. Chem. 48, 16A (1976).
72. T. J. Murphy, The Role of the Analytical Blank in Accurate Trace Analysis, NBS Spec. Publ. 422, Vol. II, U.S. Government Printing Office, Washington, D.C., 509 (1976).
73. W. Horwitz, Analytical Measurements: How Do You Know Your Results Are Right?, Special Conference: The Pesticide Chemist and Modern Toxicology (June 1980) (See also reference 27).
74. L. P. Powell, personal communication (1981).
75. H. M. Kingston, personal communication (1981).
76. J. W. Tukey, Exploratory Data Analysis, Addison-Wesley, Reading, MA (1977).
77. V. V. Nalimov, The Application of Mathematical Statistics to Chemical Analysis, Addison-Wesley, Reading, MA (1963).
78. L. A. Currie, Anal. Chem. 40, 586 (1968).
79. J. Op de Beeck, Atomic Energy Review 13, 743 (1975).
80. L. A. Currie, J. of Radioanalytical Chemistry 39, 223-237 (1977).
81. J. Rice, An Approach to Peak Area Estimation, Draft (1981).
82. R. M. Parr, H. Houtermans and K. Schaerf, The IAEA Intercomparison of Methods for Processing Ge(Li) Gamma-ray Spectra, Computers in Activation Analysis and Gamma-ray Spectroscopy, CONF-780421, p. 544 (1979).
83. G. L. Ritter and L. A. Currie, Resolution of Spectral Peaks: Use of Empirical Peak Shape, Computers in Activation Analysis and Gamma-ray Spectroscopy, CONF-780421, p. 39 (1979).
84. L. A. Currie and G. L. Ritter, unpublished work.
85. H. Kaiser, Anal. Chem. 42(2), 24A (1970).
86. Geophysics Study Committee, Energy and Climate, National Academy of Sciences, Washington, D.C. (1977).
87. J. A. Logan, M. J. Prather, S. C. Wofsy and M. B. McElroy, Phil. Trans. Roy. Soc. Lond. B290, 187 (1978).
88. R. W. Gerlach, B. R. Kowalski and H. O. A. Wold, Anal. Chim. Acta 112, 417 (1979).
89. B. R. Kowalski, Chemometrics: Theory and Application, ACS Symposium Series 52, American Chemical Society, Washington, D.C. (1977).
90. G. E. Gordon, Envir. Sci. Tech. 14, 792 (1980).
91. J. A. Cooper and J. G. Watson, J. Air Pollution Control Assoc. 30, 10 (1980).
92. J. A. Cooper, J. G. Watson and J. J. Huntzicker, Summary of the Portland Aerosol Characterization Study: Air Pollution Control Assoc. Mtg, 72nd, paper 79-29.4 (1979).
93. T. J. Kneip and P. J. Lioy, eds., Ann. N.Y. Acad. Sci. 338 (1980).
94. J. A. Cooper, L. A. Currie and G. A. Klouda, Envir. Sci. and Tech. (1981) (in press).
95. G. E. Gordon in reference 93, p. 99 (1980).
96. W. R. Pierson, Proc. Conf. Carbonaceous Particles in the Atmosphere, T. Novakov, ed., LBL-9037, p. 221 (1979).
97. G. T. Wolff and R. L. Klimisch, ed., Particulate Carbon: Atmospheric Life Cycle, General Motor Research Laboratories, (1981), in press.
98. D. J. Alpert and P. K. Hopke, Atmos. Envir. 15, 675 (1981).
99. J. A. Cooper, organizer, 1981 International Conference on Residential Solid Fuels, Portland, OR (June 1981).
100. S. J. Johnsen, W. Dansgaard, H. B. Clausen and C. C. Langway, Nature 235, 429-434 (1972).
101. G. M. Raisbeck and Y. Françoise, Radiocarbon 22(2), 245-249 (1980).
102. G. F. Carter, ed., Archeological Chemistry II, Adv. in Chem. Series No. 171, American Chemical Society (1978).
103. J. S. Lopes, R. E. Pinto, M. E. Almendra and J. A. Machado, Variation of $^{14}C$ Activity in Portuguese Wines from 1940 to 1974, International Conference on Low-radioactivity Measurements and Applications, The High Tatras, Czechoslovakia, p. 265-268 (1975).
104. C. D. Keeling, C. A. Ekdahl, P. R. Guenther, L. S. Waterman and J. F. S. Chin, Tellus 25(5), 174 (1973).
105. S. H. Harrison, R. Zeisler and S. A. Wise, eds., Pilot Program for the National Environmental Specimen Bank – Phase I, Report EPA-600/S1-81-025, National Bureau of Standards (1981).
106. G. M. Goldstein, personal communication (1981).
107. R. Zeisler, personal communication (1981).
108. L. R. Hilpert, W. E. May, S. A. Wise, S. N. Chesler and H. S. Hertz, Anal. Chem. 50, 458 (1978).
109. S. A. Wise, S. N. Chesler, F. R. Guenther, H. S. Hertz, L. R. Hilpert, W. E. May and R. M. Parris, Anal. Chem. 52, 1828 (1980).