

In pursuit of accuracy: Nomenclature, assumptions, and standards*

L. A. Currie

National Institute of Standards and Technology, Gaithersburg, MD 20899 USA

Abstract Accuracy is the central measure of quality in Analytical Science. Three topics, essential in the pursuit of accuracy, especially in the exposure and reduction of the "big" errors in analytical measurements, are reviewed in this article. The first topic, Analytical Nomenclature, lies at the heart of any effort to realize and communicate the nature and level of accuracy of the Chemical Measurement Process (CMP). The discussion centers about the historic role of IUPAC in this endeavor, together with some outstanding problems in Analytical Nomenclature, particularly as related to detection and identification. The second topic comprises assumptions and standards (materials and data) that are crucial in the search for major errors and in the control of accuracy, both within and between laboratories. Standard (Certified) Reference Materials have long been central to this effort, but more recently they have been joined by Standard Test Data. The latter, data sets having known characteristics, designed to simulate the structure of complex analytical signals, show great promise for the control of quality in the computational phase of the CMP. The final topic relates to Revolutions in Analytical Measurement Science that can make the most profound impact on accuracy: revolutions in sampling, measurement, and computation. Two illustrations are presented: (1) Accelerator Mass Spectrometry, a fundamentally new approach to atomic and isotopic mass spectrometry that makes possible the direct measurement of element and isotope ratios as small as 10^{-16} ; and (2) Exploratory Statistical Graphics, a data visualization tool that permits analytical scientists to directly apply their intuitive "vision" to the assessment of multivariate data quality, and to search for unsuspected relationships in complex chemical datasets.

INTRODUCTION

Accuracy, broadly interpreted, represents the most important goal of Analytical Measurement Science. If our Chemical Measurement Processes are to serve both the practical needs of humankind and excellence in the pursuit of new scientific knowledge, they *must* be endowed with an adequate level of accuracy. Accuracy transcends the conventional measures of quality. Control, and acceptable bounds for imprecision and bias are clearly prerequisites; but scientific conventions (communication) and scientific and technological means for approaching "the truth" must also be considered. These issues are especially appropriate for discussion at the *International Congress on Analytical Sciences 1991* for two reasons: first, accuracy in scientific *communication* is a vital mission of the sponsor, IUPAC; second, the multidisciplinary perspective that is implied in the title of the Congress is mandatory for effecting our approach to accuracy.

Accuracy in communication, among scientists and with the public, begins with nomenclature. Terminology is but the first step; rather, quality nomenclature comprises the fundamental meanings of analytical concepts, together with internationally-accepted terms and symbols. Nomenclature relating to the structure of the CMP, especially its functional and error components, is vital for exposing critical assumptions and capabilities. Although there are many outstanding international nomenclature documents, some problems remain. Diverse nomenclature for analytical detection and diverse practices for the reporting of analytical data, of some consequence for the assessment and regulation of health and safety, are reviewed, together with recommended solutions.

* Contribution of the National Institute of Standards and Technology; not subject to copyright.

COMMUNICATION

Historical perspective

At the 36th IUPAC General Assembly (Hamburg, August 1991) a reprint was distributed that gave a brief history of events in the 19th century chemical world that led to the Karlsruhe Congress of 1860 -- a meeting that foreshadowed the establishment of IUPAC. The central issue, as evidenced by the extracts in Table 1, was scientific communication and nomenclature (ref. 1). Since its first conference in Rome (1920) IUPAC has vigorously pursued the nomenclature problem; the result is the rainbow series listed in Table 2 (ref-2). The effort, of course, has expanded across disciplines. One of the most important multidisciplinary and multinational products, shown at the bottom of Table 2, is the International Vocabulary for Metrology, published in the name of: the International Bureau of Weights and Measures (BIPM); the International Electrotechnical Commission (IEC); the International Organization for Standardization (ISO); the International Organization of Legal Metrology (OIML); the International Federation of Clinical Chemistry (IFCC); and the International Union of Pure and Applied Chemistry (IUPAC) (ref. 3). To these may be added several others of special import to accuracy and the analytical measurement sciences (ref. 4-6).

Despite more than a century of progress, all is not well. As an illustration, Figure 1 shows arsenic results reported to the International Atomic Energy Agency (IAEA) by 16 laboratories participating in an intercomparison exercise for trace elements in environmental/biological materials (ref. 7). (In fairness to the participants, we must point out that other sets of results -- e.g., for Cd -- were quite self-consistent.) Clearly, it is still appropriate to discuss the pursuit of accuracy from the perspective of the 'big' errors in analytical science, especially when difficult, trace level measurements are involved. Apart from the fact that the results in Fig. 1 span nearly 5 orders of magnitude, we observe that *upper* limits (labeled "detection limits") for non-quantitative results are found considerably *below* several of the quantitative results! Two questions present themselves: (1) To what extent is the illogical juxtaposition due to major analytical errors; and (2) to what extent is it due to faulty communication? We shall address the latter question in this section, especially in the contexts of data reporting, and the basic meaning of detection limits.

Detection decisions and limits

Although the detection limit constitutes one of the principal performance characteristics of a Chemical Measurement Process (CMP), and detection decisions are critical in many areas involving human health and safety, there is little uniformity of terminology, symbols, or even common meaning within the scientific community. A glimpse at the extent of the problem is given in Table 3, adapted from a recent overview of historical, societal and technical issues surrounding this topic (ref. 8). The problem is compounded by the fact that a number of the discrepant "official" definitions are employed for regulatory purposes and in communication with the public. Perhaps the most serious issue is the fact that the false negative (β error) is not explicitly recognized in several of the definitions, resulting in a *de facto* ratio of false negative (β) to false positive (α) errors that ranges over a factor of 1000. A particularly insidious effect of varying and unstated definitions for detection limits is found in compilations of analytical method capabilities and low-level databases (ref. 9). The situation is not dissimilar from that preceding the Karlsruhe Congress! Action within IUPAC to address the problem is underway, partly at the request of the CODEX Committee on Methods of Sampling and Analysis, as indicated at the bottom of Table 3.

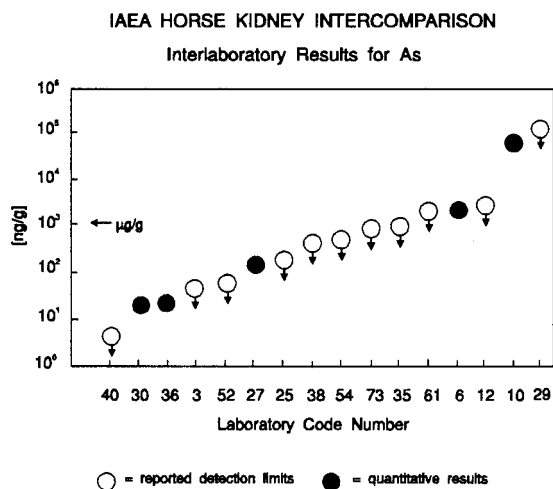


Fig. 1. IAEA Intercomparison results for As in horse kidney (ref. 7). Intermingled "quantitative" and non-quantitative results span approximately 5 orders of magnitude.

Table 1. Karlsruhe 1860 (C. Priesner, 1989)

- "... completely missing common understanding of the notions 'atom', 'molecule', and 'equivalent'. Because a common language of chemistry was missing, it was rather difficult to read and understand chemical publications."
- "One symbol of an element represented different atomic weights and for one substance quite a number of formulas existed. In 1859 August Kekulé listed ... 19 different formulas for acetic acid. Something had to happen ..."
- The Karlsruhe Congress ... the initiative of Kekulé, Wurtz, Weltzien ... a predecessor for IUPAC

Table 3. Detection Limit: Current "official" definitions

Symbol (Name): LLD, IDL, LOD, c_G , L_D , c_1 , MDL
Organization: NRC, EPA, ACS, IUPAC, IAEA, ...
Definitions: $2 \cdot BG$, 3σ , $3\sigma_B$, $3.29\sigma_G$, $2t_{s_0}$, ...
False Positive: $\alpha = 0.05\%$ to 5%
False Negative: $\beta = 5\%$ to 50% [often ignored!]

Committee on Methods of Analysis and Sampling

[CODEX Alimentarius Commission, FAO]

- Request to IUPAC, August 1990, to consider the question of "Limit of Detection" and "Limit of Determination". (Deliberated by the Committee, *since 1982*, for method selection.)

Table 2. Nomenclature Guides

IUPAC

The Green Book: Physical Chemistry
The Blue Book: Organic Chemistry
The Red Book: Inorganic Chemistry
The Orange Book: Analytical Chemistry
The Gold Book: Compendium of Chemical Terminology

VIM: International Vocabulary of Basic and General Terms in Metrology [BIPM, IEC, ISO, OIML, IFCC, IUPAC]

Table 4. Detection -- the concept

Dual Questions:

- Q1 - How little can be detected?
- Q2 - Has something been detected?

Popular Responses:

- **Intuitive** (sound experience, non-quantifiable)
- **Ad Hoc** (rigid formula; dictum, vote, ...)
- **Signal/Noise** (assuming white noise; Q2 only)
- **Avoidance** (small signals not worth considering)
- **Hypothesis Testing** (false positives & negatives)

To develop a quantitative approach to the issue of detection in analytical measurement science, we must pose two basic questions. These are given in Table 4, together with some popular responses. The first question clearly relates to the inherent detection capability of the CMP in question, whereas the second relates to the recipe for making detection decisions. The first gives a measure of method performance, the detection limit; it cannot be answered in the absence of a formalism for answering the second question. Together, the two questions form the basis for the hypothesis testing formulation to detection. This is the best of the quantifiable formulations, for it recognizes both false positive and false negative errors, and it properly distinguishes the decision criterion (and the corresponding *critical level*) from the *detection limit*. The intuitive approach, of course, should not be discounted; sound scientific experience is invaluable for avoiding faulty assumptions which could too easily be incorporated in a simple mathematical expression.

A simple illustration of the links among measurement precision, false positive (α) and false negative (β) errors, and a societal or regulatory problem is given in Fig. 2 (ref. 8). The fictitious relation in the upper part of the figure symbolizes the sociopolitical or socioeconomic generation of a maximum tolerable or regulatory level (L_R), here put in the context of earthquake (precursor) detection. The lower part indicates that the detection limit (L_D) of the method employed must not exceed L_R ; where L_D is defined by the acceptable β -error, the probability density function (pdf) for the observed net signal when the true net signal is L_D , and the critical level (L_C) for making detection decisions. L_C depends in turn on the choice of α and the pdf for the observed net signal when the true net signal is zero. For simple detection -- with normal random measurement error; constant, known variance (σ^2) over the range in question; and α and β each equal to 0.05 -- $L_C = 1.645 \sigma$, and $L_D = 2 L_C$, where σ is the standard deviation of the estimated net signal. (1.645 is the 1-sided critical value of the z-statistic. When σ is not constant, it is important to distinguish between σ_0 and σ_D [ref. 8].) When σ^2 is estimated as s^2 , the factor 1.645 must be replaced by Student's-t for the appropriate number of degrees of freedom. L_D in this case is more complicated, requiring use of the non-central-t, but it is approximately equal to $2t\sigma$. (Since σ is unknown, an interval estimate must be given for L_D in this case.) Further mathematical treatment is beyond the scope of this article; ref. 8 should be consulted for a more comprehensive discussion and review of the literature.

A summary of the most crucial, accuracy-related issues involving detection is given in Table 5. The first is the issue of common nomenclature. The second relates to the very important, difficult problem of communicating in this area with the lay public. The name "Delaney" which appears in Fig. 2 is an allusion to the Delaney Amendment to the U.S. Food, Drug, and Cosmetic Act. It has become symbolic for the mistaken view that zero concentration could, in principle, be detectable (ref. 8). The misuse of L_D and faulty reporting of non-quantitative data comprise the third issue in Table 5. Space does not permit a discussion of the reporting of low-level data, but incomplete or even biased reporting constitutes one of the most serious sources of information loss and distortion, especially in the transition from the laboratory to the public (ref. 8-10). The only guaranteed solution is to report the observed value and its uncertainty, *even when the decision is "not detected."* The last issue is the big one. Mistaken assumptions, and failure to correctly and quantitatively take into account blank variability, and interference or matrix effects, can lead to completely erroneous results, for *both* non-quantitative and quantitative data. A very common oversight, in this regard, is the effect of interference on detection limits.

Extension: discrimination and identification

Accuracy in distinguishing between chemical species or pollutant sources, for example, is no less important than accuracy in detecting them. Fortunately, the very same principle, hypothesis testing, is applicable to both. A glimpse at the approach and application is given in Table 6. For the univariate case, rather than

REGULATORY LEVELS (L_R) AND DETECTION LIMITS (L_D)

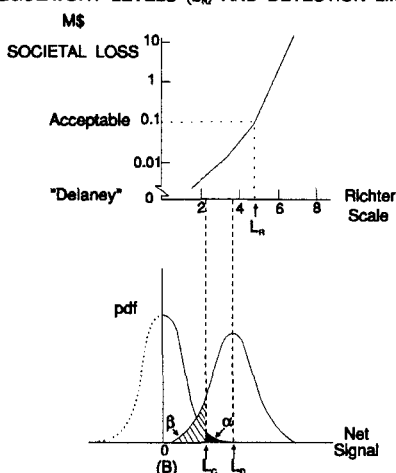


Fig. 2. The link between regulatory limits (L_R) and detection limits (L_D). Upper curve is dictated by societal needs; lower one, by measurement capability.

comparing the observed signal for an unknown with that for the blank as in Fig. 2 (detection decision), one compares the observed magnitude of an identifying variable for an unknown with that for the standard (discrimination decision). Unlike detection where true signals may be continuous and tests are 1-sided, the identifying variable takes on characteristic values for the universe of possible substances, and tests are 2-sided. It is appropriate to speak of "identification" in contrast to "discrimination" when all possibilities (alternative substances) can be tested. Examples of identifying variables are: melting points, spectral wavelengths or energies, chromatographic retention times, and element or isotope ratios. As with detection, given normality and constant variance, Student's- t (2-sided), may be used for discrimination decisions and the non-central- t for establishing performance

Table 5. Most crucial issues

- Adoption of a common, meaningful definition, encompassing *both* false positive and negative errors
- Communication to the Public that the null hypothesis (zero concentration or amount) is unprovable
- Application of L_D as a Performance Characteristic, *not*:
 - as a level for making detection decisions
 - as a quantity to be reported when results are "negative"
- A presumed L_D may be *totally inaccurate*, without attention to:
 - model and assumption validity
 - *variability of the blank*, interference, matrix effects

Table 6. Discrimination and identification limits

- Same principles apply when using an "identifying" variable (element or isotope ratio) rather than a "quantifying" variable (element or isotope amount)
- For multiple identifying variables (multivariate pattern recognition), the simple t -test is generalized by Hotelling's T^2 (vector distance and covariance treated)
- Illustrations from our current work:
 - Authenticity of drug formulations (proof of "sameness")
 - Detection of excursions from an urban air pollution signature (intruding, external air mass)

(discrimination) limits (ref. 8). Multivariate discrimination decisions must take into account covariance among variables; this is accomplished by the multivariate analog of Student's-*t*, namely Hotelling's- T^2 . It is given by

$$T^2 = (\mathbf{x} - \mathbf{x}_0)' \mathbf{S}^{-1} (\mathbf{x} - \mathbf{x}_0)$$

where \mathbf{x} and \mathbf{x}_0 are the unknown and standard chemical (pattern) vectors, respectively, and \mathbf{S} is the estimated covariance matrix. Critical values of T^2 depend on degrees of freedom and the α -error. In the limiting case (multivariate normal, fixed covariance matrix), multivariate discrimination *limits* may then be constructed using also the β -error, and the non-central F distribution. Note that this parametric approach is just one of many for multivariate discrimination; a nonparametric (graphical) approach is presented in part 3 of the article. For an introduction to the vast literature on this topic, see refs. 11 (chemical) and 12 (general).

Some of the most serious errors in chemical analysis have arisen when the observed signals do not come from the presumed substance present. Major gains in accuracy have come from instrumentation and separation advances that have markedly improved identifying-variable resolution, thus reducing or eliminating such unsuspected interference. (See part 3 of this article.) In many cases this is not sufficient, especially when resolution is dictated by intrinsic sample variability. Then multivariate discrimination, or "chemical fingerprinting," becomes interesting. This is the situation for the two problems noted at the bottom of Table 6. A brief introduction to the first problem is given in Table 7, which highlights the importance of the problem of drug authenticity, and the planned approach, using chemical fingerprinting (ref. 13, 14). We conclude this section with a glimpse at the multivariate discrimination of a characteristic urban air mass from samples of intruding or otherwise contaminated air.

Table 8 contains isotopic and chemical data from a study of air pollution sources in the small town of Elverum, Norway. The motivation for the study, discussed in ref. 15, was to apportion quantitatively wintertime aerosol carbon pollution to fossil and woodburning sources. That was accomplished with univariate discrimination, using the isotope ratio $^{14}\text{C}/^{12}\text{C}$ as the discriminating variable. (f_w in the table indicates the percent of woodburning carbon, as deduced from the isotope ratio.) We now wish to ascertain which of the samples are *characteristic* of

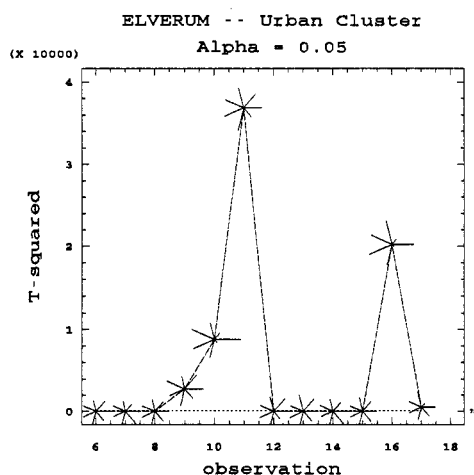


Fig. 3 "Urban plume" T^2 test. The 7 samples on/below the dotted line are consistent with the urban signature; the 5 above, are not. Individual patterns are displayed as 7-pronged glyphs.

Table 7. Drug formulation authenticity

● **The Problem:** Small changes in storage, production, or even trace contaminants in *inactive ingredients*, can produce life-threatening effects.

● **Example:** An arthritis medication (isoxicam) was associated with deaths from Lyell's Syndrome, but only in France. A \$6 million investigation eventually proved that a "minute, unseen contaminant" induced a severe immune system reaction. It was discovered that the contaminant came as an unknown byproduct during synthesis at the plant supplying the product for the French market (ref. 13).

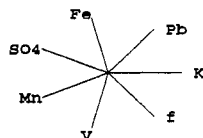
● **Solution:** The U. S. Food and Drug Administration has launched a "scientific fingerprinting" program, using a vast array of analytical and computational techniques, to detect trace and ultra-trace deviations from authentic drug patterns (ref. 14).

Table 8. Multivariable particulate data: Elverum, Norway*

Sample	f_w	C	K	Pb	Fe	SO ₄	Mn	V
6	63	15	121	263	20	730	4	1
7	78	50	424	607	33	1540	6	1
8	64	30	328	497	62	1620	10	2
9	64	9	197	213	27	2710	4.5	8
10	32	5.1	218	54	33	2800	6	6
11	62	6.9	140	160	72	2320	6.5	14
12	54	12	179	262	57	910	4.5	2
13	65	26	246	547	179	1300	8	2
14	83	22	231	325	56	1490	2.5	2
15	67	25	245	419	110	1410	4	2
16	26	11	339	94	150	5340	15.5	17
17	61	10	99	70	59	310	2	1

*Units: f_w (percent woodcarbon); C ($\mu\text{g}/\text{m}^3$); all others (ng/m^3)

the wintertime polluted air and which are different. The problem is thus completely analogous to the use of chemical tracers or chemical patterns to identify an "urban plume" or to identify a "regional signature" (ref. 16). Carbon normalized patterns were used for the multivariate test, the results of which are given in Fig. 3. The figure was constructed using "prior estimates" for x_0 and S (for didactic purposes), and the following glyph code.



The result is that 7 of the 12 samples are consistent with the assumed pattern for the urban plume. The last two samples are especially interesting, in that they were collected at the same, nearby background site. The very different glyph (chemical) pattern of sample 16 indicates a near absence of the urban aerosol species, woodcarbon and lead, and an abundance of elements that might be associated with coal burning (absent from the town). Independent meteorological data showed this sample to be associated with long range transport (ref. 15). This example highlights another very important point, regarding model accuracy. That is, unless the data being modeled, statistically or mechanistically, belong to a homogeneous subgroup, erroneous conclusions can follow (ref. 11, 17). In a sense, sample 16 is a multivariate outlier, which cannot be modeled correctly with multivariate patterns of sources that are indigenous to the town of Elverum.

ASSUMPTIONS AND STANDARDS

Two essential activities in the pursuit and control of analytical accuracy are: (1) attention to the nature and validity of assumptions employed in sampling, measurement, and data evaluation; and (2) judicious use of appropriate standards to expose faulty measurements or faulty assumptions. There are at least 4 classes of standards that are relevant to these endeavors: Standard Reference Data (SRD), Standard (or Certified) Reference Materials (SRM), Standard Test Data (STD), and Standard Field Studies (SFS). SRD and SRM (or CRM) are generally available from national or international standards laboratories; they provide a common basis for critically-evaluated physical and chemical properties, and expertly-characterized physical and chemical materials, respectively. Accuracy, including exhaustive assessment of uncertainties, is a keyword in such evaluation and characterization, because the data and materials serve as the foundation for accuracy in analytical measurement science. The scope of SRD and SRM may be seen, for example, from catalogs available from NIST (ref. 18). STD represent a relatively new means to assure accuracy in analytical data evaluation. These data are designed to simulate actual analytical data (signals), but they are created with fully-known characteristics. The last category SFS differs from the others, in that it is not a product that can be obtained from a standards organization. The impact of SFS on the assessment of modeling accuracy of physicochemical processes occurring in the "field" (complex real systems) can be considerable, however. A recent illustration is the ANATEX study in which the accuracy of a number of atmospheric chemical transport models was evaluated in an experiment spanning three months and 3000 km. The evaluation was accomplished using known releases of several perfluorocarbon tracers plus downwind monitoring over a grid of 77 stations (ref. 19).

Structure of the chemical measurement process (CMP)

The relevance of SRM, STD, and assumptions to analytical accuracy can be conveniently discussed with the aid of a diagram portraying the main components of the CMP (Fig. 4). The input and output of the CMP are, respectively, samples and results. The former have unknown composition, denoted x (scalar for single sample, single component; vector for single sample, multicomponent); the latter consist of estimated (measured) concentrations and uncertainties. Within, the CMP box contains sample preparation and instrumental measurement steps that yield a signal y , and a data evaluation step that operates on the signal. The internal structure may or may not be transparent to the operator. An SRM serves to test the quality of the overall CMP, whereas the STD test the data evaluation step by injecting simulated signals having known characteristics. The latter type of standard is of increasing importance for the assessment of algorithm accuracy, with the continuing trend toward automation and hidden or proprietary software. The simplest possible analytical model is shown below the diagram.

The CMP is generally imbedded in a larger (biological, environmental) system that we wish to characterize or control. Sampling is the critical link between the two, and as such, it often provides the greatest opportunity for imprecision and bias. Unfortunately, SRM use is usually restricted to the post-sampling phase of analysis. Some problems in sampling, such as heterogeneity, an inadequate sampling plan or design, losses, and contamination are well appreciated by chemists. Other problems, however, require

multidisciplinary expertise for their revelation. An eloquent discussion of this point has been given by Iyengar, in the context of biological trace element research (ref. 20). Iyengar's point is that sampling must be both biologically and analytically valid, and that complementary expertises are required to achieve such an end. A classic unidisciplinary pitfall in this regard is the analysis of analytically-convenient, but potentially biologically-misleading, whole blood samples for selected trace elements. For example, based on whole blood, the activity of the selenoenzyme glutathione peroxidase was found to be lower in pregnant rats; yet, when based on hemoglobin, it showed no significant difference (ref. 21). Bias and imprecision result from insufficient or non-representative sampling. The large scatter of trace element dietary data for just one of three countries, shown in Fig. 15 (part 3 of this article), has been attributed to such deviations from the sampling protocol.

Model error, like sampling error, *will* lead to erroneous conclusions. Table 9 captures some of the potential assumption pitfalls in applying the simplest CMP model, by contrasting Nature's model with "our" model

Table 9. Models and assumptions

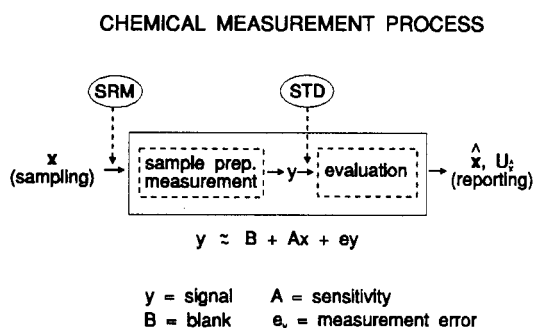


Fig. 4. Schematic diagram of the Chemical Measurement Process.

[y = Signal B = Blank x = Analyte amount]

Nature's Model: $y = f(B, q, A, x) + e$

Our Model: $y \approx B' + \sum A'x + e'$ [sum: 1 to q']

Potential Pitfalls:

- \approx : assumed linear model
- B' : assumed blank
- q' : assumed no. and identity of components
- A' : assumed sensitivity matrix
- e' : assumed error distribution (also: blunders)

for the signal-composition relationship. The latter treats this relationship as linear, with simple additive, (normal) random errors. Errors in B are assumed random and are propagated, and errors in A are assumed negligible. Furthermore, the identities and number of components are assumed known. Excellent methods exist for testing assumptions -- though such methods themselves rest on assumptions; and special statistical/numerical techniques exist for treating non-normal error, random error in the sensitivity matrix, and non-linear models (ref. 11, 22). Unfortunately, "our" model is at best an approximation of Nature's, and some of the special statistical/numerical techniques are misunderstood or misapplied in chemistry, leading to a false sense of security (ref. 23). In the end, every erroneous assumption or model is manifest as systematic error. That error component, of course, cannot be estimated by replication, yet realistic bounds for systematic error are crucial for a meaningful estimate of uncertainty. One route to estimating the systematic component of uncertainty, within the capacity of all expert analytical scientists, is exhaustive scientific analysis of the measurement process. Another is comparison with standards and comparison among methods or laboratories.

Standards and testing of assumptions

Interlaboratory and intermethod comparisons provide an excellent means for checking comparability; if an SRM is used, or if a method or laboratory can be considered "definitive", then accuracy also may be assessed. Important guidance in these areas has been given by international standards organizations (ref. 5, 6). One of the simplest, yet most important techniques for assessing interlaboratory accuracy is the Youden plot (ref. 24). By making a scatterplot for pairs of measurements (x, y) of equivalent samples by multiple laboratories, one can instantly discern systematic and random components of interlaboratory error. This is one of the earliest methods of statistical graphics, which has recently become so popular and so powerful. A useful extension of the Youden 2-sample method has been given by Meglen, for quality control (ref. 25). Quality is monitored by plots of sums and differences, plus the scatterplot, now containing a time marker.

A summary of SRM and STD applications is given in Table 10. One important illustration of SRM use is the detection of bias. A document including the treatment of bias detection decisions, bias detection limits, and bias confidence intervals has been prepared recently by NIST (ref. 26). Bias detection is similar to analyte detection; the same hypothesis testing principles apply. A small difference is that bias can be either positive or negative, so a 2-sided test is appropriate for bias detection decisions. The detection limit for bias ($\alpha, \beta = 0.05$) is $3.60 \sigma/\sqrt{n}$ for n -replicates, where σ is the known standard deviation of the method being tested for bias. The limit increases if the SRM uncertainty is not negligible, or if σ is estimated (as s). Thus, bias detection is not easy unless the bias is considerably larger than the method- σ . Complementary estimation of bias bounds through exhaustive scientific analysis of the measurement process is therefore strongly recommended.

Table 10. SRM, STD Applications

Use of Standard Reference Materials
Assurance of long-term measurement compatibility

- Calibration; internal, external quality assurance

Analytical method development and validation

- Development of reference test methods
- Bias detection; method validation

Use of Standard Test DataDevelopment and validation of data evaluation steps

- Detection of errors in modeling, computation, uncertainty estimation
 - Algorithmic improvement
-

STD have been deemed sufficiently important for computational quality control that the IAEA regularly includes a set of simulated gamma ray spectra in its CRM catalog (ref. 27). The spectra, having known peak and error structure, are designed to test peak detection, resolution, location, and quantification algorithms. When first introduced, as unknowns, these STD produced surprising results in an intercomparison involving some 200 laboratories (ref. 28). For example, the peak detection exercise (Fig. 5 - test spectrum no. 200) had the remarkable conclusion that, despite the application of several sophisticated computer techniques, a "visual" method produced the best result -- 19 true positives and 0 false positives, given 22 real peaks. Computational techniques produced abundant false positives (up to 23), perhaps because of failure to take into account the effect of multiple null tests (ref. 8).

One important implication of the outcome is the value of expert "vision" for the exploration of complex data, suitably presented. (See part 3 of this article.)

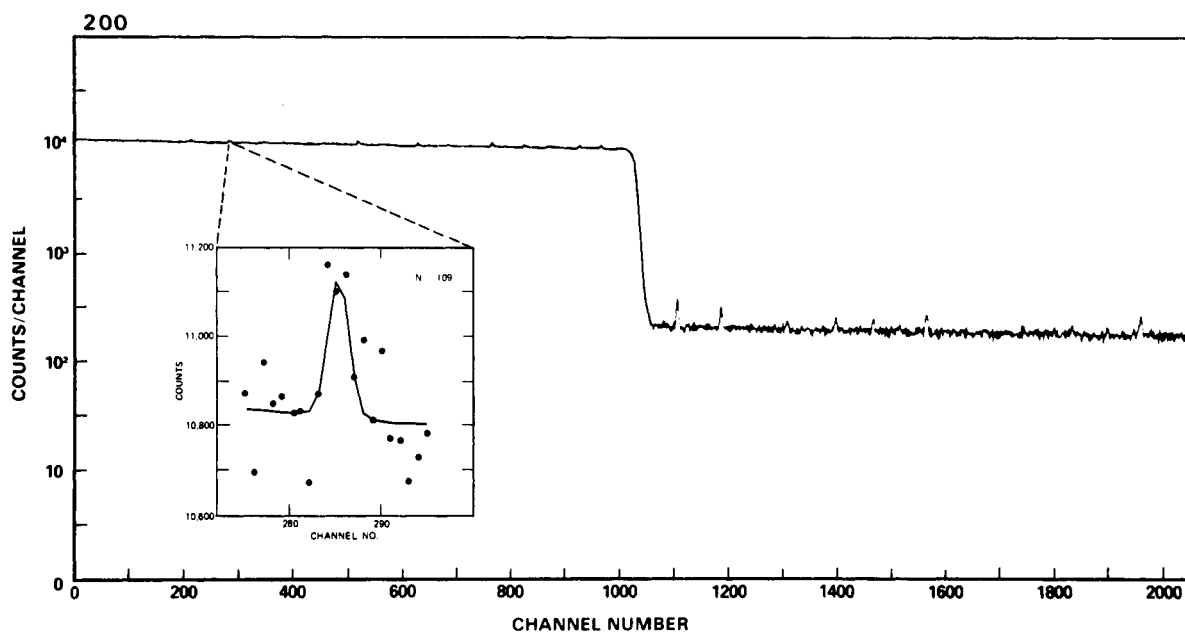


Fig. 5. IAEA γ -ray STD for testing detection algorithms. Inset (ref. 29) shows a peak detected by half the participants. Best result ("visual") found 19 (of 22) peaks, with no false positives.

Creation of realistic STD for measurement science has continued. Multivariate analytical data prepared to assess the accuracy of aerosol receptor modeling methods also led to surprises, in an intercomparison exercise (ref. 30). Fig. 6 shows, for example, absolute normalized deviations from the "truth" for 3 similar (chemical mass balance) methods. Had the deviations been consistent with random normal error, roughly half of the results from each group would have fallen above and below the horizontal median centerline. A conclusion was that only 1 of the 3 laboratories gave realistic (accurate) uncertainty estimates, the others giving serious under- or over-estimates (ref. 29). Currently, at NIST, there is a major project to develop test data for software performance evaluation of automated coordinate measuring machines (ref. 31), and a new project to develop STD suitable for spectroscopic or chromatographic methods and instrumentation incorporating significant computational steps. The need for this aspect of analytical quality control is likely to increase with the increase in multivariate "black boxes" possibly incorporating questionable assumptions.

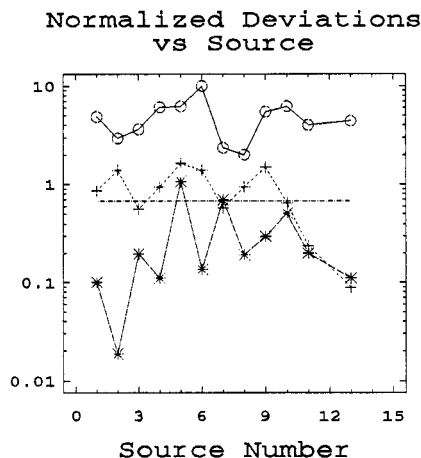


Fig. 6. Receptor modeling STD – absolute, normalized deviations from the "truth". Only 1 of the 3 Labs (+) was consistent with the expected median (horizontal centerline).

(R)EVOLUTIONS IN MEASUREMENT SCIENCE

Major changes affecting accuracy

The conduct of Analytical Science has been profoundly affected in recent decades by fundamental developments in at least three areas: sampling, measurement technology, and computation. Progress in each of these areas has had broad impacts on our discipline, including such practical matters as the cost and speed of doing business. The revolutionary impact that is pertinent to this text, however, is the impact on accuracy. This comes about largely because innovations in the three areas have made it possible to sample the previously unsampleable, measure the previously unmeasurable, and compute the previously uncomputable. As a result, *direct* information may now be obtained for problems that previously depended (in part) on assumptions for their solutions.

Since it is impossible to treat adequately the accuracy-related developments in the above areas in the space available, I shall limit my remarks to some brief comments on sampling, plus slightly more extensive discussion and illustrations from one of the advances in each of the other two areas. One of the most interesting and important developments in sampling in the past two or three decades has been the remarkable progress in capabilities for sampling astronomical bodies, including the planet earth and its moon. Accuracy of our knowledge of lunar composition naturally took a quantum leap, once actual samples were retrieved. Prior to that our knowledge was limited to indirect inference, derived from meteorites, radiation, and theory. The accuracy of arguments set forth by Harold Urey, based heavily on thermodynamic and geochemical principles, was remarkable. Lunar students would do well to study his work of the 1950s (ref. 32). The lunar sampling program was initiated with the landing of Apollo 11 in 1969; we shall return to this briefly in connection with an important "visual metaphor." The ability to retrieve samples, to perform *in situ* measurements or to remotely sense the composition of the stratosphere, deep sea sediment, and polar ice cores, for example, has led to enormous improvements in the accuracy of our knowledge of these environmental "compartments" -- an area that constitutes one of the major new foci of IUPAC (ref. 33). Such compositional information is crucial for understanding the chemical history of the earth and its atmosphere, monitoring current changes, and forecasting. Perhaps the most dramatic recent example is the discovery of the "ozone hole"; prior views on the mechanisms and rate of destruction of stratospheric ozone were shown to be quite incomplete (ref. 34).

Measurement advances – accelerator mass spectrometry (AMS)

No less dramatic than the advances in sampling, are those involving new concepts in measurement, new technology, and automation. To illustrate the impact on accuracy, I shall discuss just one such development, the invention of Accelerator Mass Spectrometry (AMS). AMS was originally conceived and applied to the analysis of mass-3 nuclei by Luis Alvarez in 1939 (ref. 35). It was re-discovered in 1977, more or less simultaneously by one group working with a cyclotron (ref. 36) and one working with a tandem accelerator (ref. 37). The invention of AMS constituted a revolution in mass spectrometry because it made possible the *direct* measurement of atoms (ions) of long-lived radionuclides which previously could only be analyzed by counting their decay products.

The "AMS Advantage" is illustrated, for ^{14}C , in Table 11. (To date the largest application of AMS has been the measurement of radiocarbon, with applications to archaeology, geochronology, biomedicine, oceanography, environmental studies, and geology.) The "advantage," a factor of ca. 3×10^4 for ^{14}C , can be interpreted as a relative increase in signal (counts) for a given sample size, or a relative decrease in requisite sample size for a given counting precision. The latter has had major impact, for it has opened the field of radiocarbon research to previously unmeasurable sub-milligram carbon samples and/or chemical fractions. The impact of AMS on longer-lived nuclides is even greater. As shown in Table 11, the "advantage" is directly proportional to the half-life (or, mean life τ which equals half-life/ $\ln 2$). The consequence for ^{10}Be ($\tau = 2.3 \times 10^6$ yr) and ^{36}Cl ($\tau = 4.3 \times 10^5$ yr) has been to create two new fields of geochronology, with extensive ramifications in environmental and geochemistry, geology, and geophysics (ref. 38).

Before discussing some specific impacts of AMS on accuracy in analytical science, it will be useful to glance at the technology. The bottom line of Table 11 shows that the utility of the method requires it to go far beyond the isotope ratio capability of conventional mass spectrometry, which is limited to ratios of perhaps 10^{-10} due to molecular ion interference. (Note that biospheric ["living"] radiocarbon has a $^{14}\text{C}/^{12}\text{C}$ ratio of about 1.2×10^{-12} .) The enhanced capability of AMS depends on unique means for markedly reducing both molecular ion and isobaric ion interference --

using techniques borrowed from nuclear physics research. This is illustrated in Figures 7 and 8. The first figure is a photograph, taken by the author, of the 12 MV model MP tandem Van de Graaff accelerator at the Nuclear Structure Research Laboratory at the University of Rochester, where the first tandem ^{14}C experiments were performed (ref. 37). The photo shows the main acceleration chamber, where low energy ions enter at the right and high energy ions exit at the left of the figure. The second figure is a schematic plan of the tandem accelerator at ETH in Zürich (ref. 39). Carbon samples, in the form of elemental or graphitic carbon, are placed in the Cs^+ sputter ion source (upper left of the diagram), where negative atomic and molecular ions are produced. The negative ions are accelerated to 40 keV and magnetically mass analyzed before entering the main acceleration chamber. Ions further accelerated by a potential of 4.5 MV impinge on a gas stripper where molecular ions are destroyed by the "coulomb explosion." The 4.5 MV potential then acts again on the resulting *positive, atomic* ions, resulting in 18 MeV C^{+3} at the high energy end of the machine. Electrostatic and magnetic mass analysis of the high

TABLE 11. The AMS Advantage

<ul style="list-style-type: none"> ● Accelerator Mass Spectrometry counts atoms: Counts = Atoms x Efficiency = $N \times (0.01)$ ● Radioactivity measurement counts decays: Counts = Counting Time x Atoms / Mean life = $t \times N / \tau$ ● AMS Advantage: $[(0.01) \tau] / [t]$ For C-14 ($t = 1$ day, $\tau = 8266$ years), this equals 30,000 ● <i>But</i>, one must measure isotope ratios (14/12) down to 10^{-16}
--

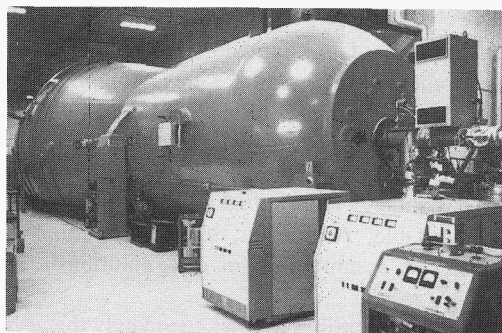


Fig. 7. Photograph of the University of Rochester tandem accelerator.

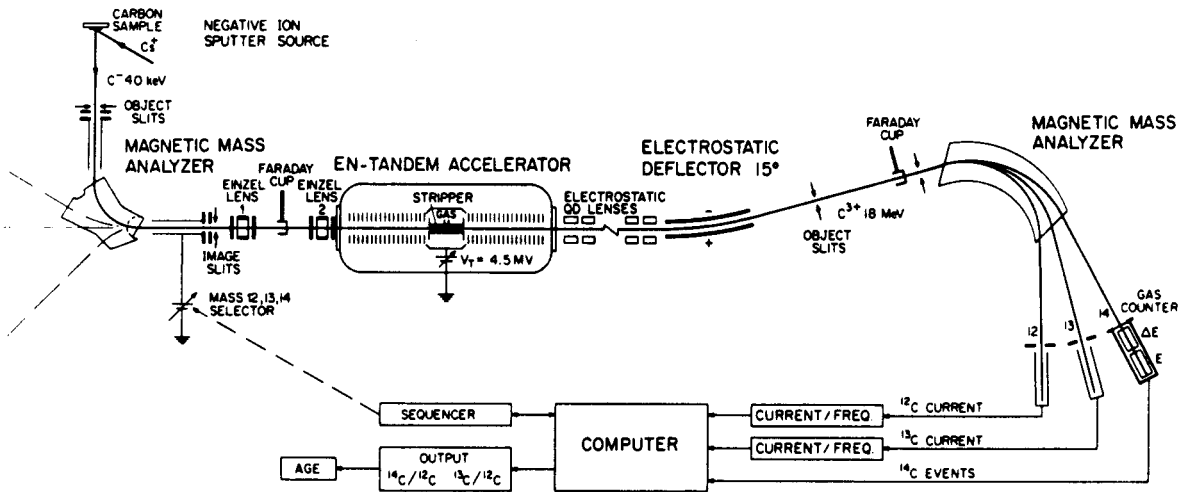


Fig. 8. Schematic diagram of the AMS dating facility at ETH (reprinted from ref. 39 with permission of RADIOCARBON, New Haven CT).

energy ions is followed by isobar resolution by two parameter (ionization density, energy) spectrometry. The key to both molecular ion and isobar resolution is the megavolt accelerating potential. An added benefit for ^{14}C accelerated as a negative ion, is the instability of N^- , which essentially eliminates ^{14}N isobaric interference.

AMS applications. Major advances in a number of disciplines have occurred as a result of experiments that were previously impossible or impracticable. New ground has been broken both for radionuclide analysis and for *ultra-trace element (isotope) analysis*, for both benefit from the freedom from molecular ions. Some classes of applications, drawn from the continuing series of International Conferences on Radiocarbon (ref. 40) and Accelerator Mass Spectrometry (ref. 41), are given in Table 12.

Regarding our central topic, accuracy, the $^{10}\text{Be}/^{36}\text{Cl}$ "failure" demonstrated that assumptions about the atmospheric transport of these nuclides were invalid (ref. 42). Subsequent research suggests an interesting link between ^{36}Cl production and transport in the stratosphere and heterogeneous chemical processes, such as those associated with the "ozone hole" (ref. 34, 43). Inaccuracy, related to faulty assumptions or unsuspected contamination, has been enormously reduced for ^{14}C studies. For example, the ability to date extremely small (submilligram) samples of carbon has made possible the accurate dating of ancient buried

Table 12. Some AMS applications (*many more*)

Successes

C-14 [$t_{1/2}$: 5730 yr] – Absolute, unique tracer for environmental carbon; Validation of elemental & molecular tracers

Be-10 [$t_{1/2}$: 1.5×10^6 yr] – Cosmic ray history of meteorites and the earth's atmosphere

Cl-36 [$t_{1/2}$: 3.0×10^5 yr] – Ground water dating (nuclear waste isolation)

Ultra-trace elements – trace impurity microanalysis of high purity metals and electronic materials

"Failure"

Be-10/Cl-36 Geochemical Clock (polar ice cores) (Failure turned into stratospheric chemistry success)

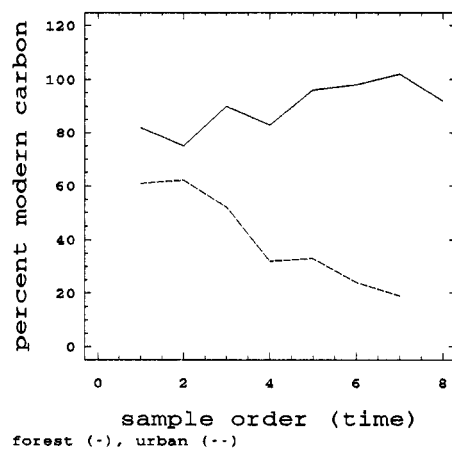


Fig. 9. Biospheric contribution to aerosol carbon for rural (solid) and urban (dashed) regions, based on ^{14}C measurements (ref. 45).

bones through AMS analysis of individual amino acids isolated from large amounts of contaminating (non-contemporaneous) carbon (ref. 44). Similarly, ^{14}C "dating" of atmospheric aerosol particles yields assumption-free apportionment of pollutant carbon. Unlike *assumed* source/transport models, or non-absolute *indirect* tracers (e.g., Pb and K), ^{14}C gives a direct and absolute apportionment of fossil/biospheric carbon. The impact is illustrated in Fig. 9, where ^{14}C measurements generated two "surprises": (1) in the forested region (Shenandoah Valley) high levels of pollutant sulfate led to the (faulty) assumption that the aerosol carbon would be fossil; (2) in the urban region (Houston), the assumed emissions inventory contained *only* fossil carbon. AMS ^{14}C measurements showed that both assumptions were wrong. In the first instance, gas (SO_2) vs aerosol (C) transport differences are likely the culprit; in the second, long range transport appears to have introduced biospheric aerosol carbon from agricultural burning (ref. 45).

Data exploration and visualization

The revolution in computational capacity has benefited analytical accuracy in at least three broad areas: simulation-modeling of complex physicochemical processes (ref. 46); computer-intensive estimation of precision and bias (ref. 47); and the graphical representation of multivariate analytical data. Only the last, popularly known as "statistical graphics," will be discussed here.

The significance of "visual" data exploration was demonstrated already in the previous discussion of Standard Test Data, where the expert visual search for subliminal gamma ray peaks proved more accurate than the most

sophisticated computer algorithms employed. More recent advances in the visual representation (ref. 48) and visual exploration (ref. 49) of complex data have greatly enhanced our ability to look and think about possible meanings, without the constraint of a rigid, pre-conceived, possibly erroneous model. In other words, visual data exploration is "robust" and admits the possibility of discovering unsuspected relationships and unsuspected blunders. A set of guidelines, to be illustrated later, is given in Table 13. The 8 rules in this Table grew out of a set of 7 presented in ref. 50, where a more detailed discussion may be found. They might be further expanded by adding a 9th rule: that statistical tests will generally be no more powerful than visual tests of data, appropriately displayed. Regarding the first three rules: Rule 1, offered by James Filliben, creator of the graphical data analysis package DATAPLOT (ref. 49), is at the heart of exploratory statistical graphics. Rule 2, from John Tukey of Exploratory Data Analysis fame (ref. 51), symbolizes the initial step in graphical residual analysis. Rule 3 captures the basic meaning of exploration, for any field; it is the key to discovering unsuspected structure as well as possible mistakes or blunders. The reference to Nathaniel Southgate Shaler, an eminent 19th century Harvard geologist, is appropriate because of his method of opening the minds of young students. He would present a student with a rock, as in Fig. 10, and require him to submit the results of all possible (visual) observations on the rock. After "all possible" observations were made, Prof. Shaler would direct the student to repeat his efforts, again and again, in the effort to discover still more -- a superb educational technique.

The rock pictured in Fig. 10 has a dual role in this section. It is, in fact, a lunar sample retrieved in the Apollo 11 mission, and sent to Brookhaven National Laboratory for exploratory chemical analysis of trapped noble gases. Thus, it is archetypical of the revolution in sampling, noted at the beginning of this section. The photo is offered also as a visual metaphor, to draw an analogy between the detailed exploration of a physical object and the detailed exploration of a data "object"; in both cases, the goal is discovery and understanding of the underlying structure. To complete the analogy, a data object that will be the subject of the remainder of this section is portrayed in Fig 11. In the case of the lunar rock, cursory visual examination reveals a heterogeneous structure, which suggests the value of deeper exploration. Also, the Brookhaven plan, to "look at" trapped noble gases, exemplifies the use of appropriate chemical

Table 13. The 8 rules of exploratory data analysis

-
1. Plot the data [*visualization*] (Filliben)
 2. Remove what you know (Tukey)
 3. Examine what remains from every possible perspective (Filliben, Shaler)
 4. Outliers are generally present: ANOB (Currie)
 5. $df = \text{observations} - \text{parameters} < 0$, *always*;
- Scientific Intuition is essential (Currie)
 6. Perform subset analysis (Filliben);
- Be alert to data heterogeneity (Currie)
 7. Univariate - multivariate exploration (Filliben)
- Understand low-dimensional structure first
- But be aware of new, multivariate perceptions
 8. Use Real Variables (RV) and Principal Components (PC)
-RV: Outlier resistant; interpretable
-PC: Efficient display of multivariable space
-

ANOB = Analysis of Blunders

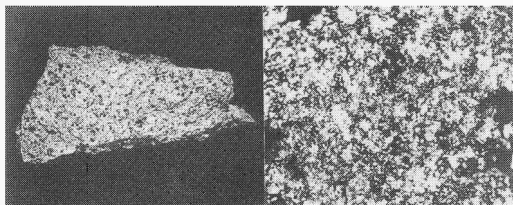


Fig. 10. Photograph of Apollo 11 lunar sample #10057 (Brookhaven National Laboratory, Oct. 1969, courtesy of R. Davis, Jr.)

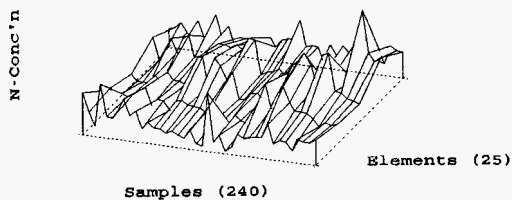


Fig. 11. 3D section of 5D compositional data from the IAEA Daily Diet Study. The plot shows only a portion of the total (6300) 5D data. Concentrations (z) are median-normalized.

instrumentation to extend the reach of normal human vision. Both issues carry over to exploration of the multidimensional data object. That is, we ask at the outset: (1) whether the data object contains non-random structure, and whether it is heterogeneous; and (2) what data visualization methods may be employed to extend our view beyond the simple 3 dimensional plot in Fig. 11. Still another analogy exists, that is central to our topic; namely *quality* — the quality of the lunar sample, and the quality of the multidimensional data. The two issues, data structure and data quality (accuracy), are the foci of the case study that follows.

The three dimensional (3D) "object" in Fig. 11 represents a 3D section (or slice) of a 5D structure, the remaining dimensions being laboratories and analytical methods. (For simplicity of presentation, only 40 of the 240 samples and 5 of the 25 elements, with median-normalized concentrations ranging from 0.084 to 4.36, have been used to construct the figure.) Sectioning (and projecting) are key exploratory tools for a data object, just as they are for a physical object. This will become apparent as we illustrate visual approaches (Table 13) to assess the structure and quality of this relatively complicated, multivariate chemical dataset.

Case study: The IAEA Coordinated Research Programme on Daily Dietary Intakes. The data represented in Fig. 11 derive from an international project, initiated by the International Atomic Energy Agency (IAEA) in 1985 (ref. 52), involving the analysis of minor and trace elements plus fiber, phytate and energy in daily diets of 20 population groups in 11 countries. Up to six analytical methods have been employed; and data quality has been monitored by reference samples, plus tri-laboratory measurements. Our illustration of computer graphics for accuracy-related data exploration is drawn from a somewhat broader study of the data (ref. 53). It represents a concrete example of the exploratory approach using *some* of the *compositional* data. It does not treat *dietary intake*, per se; and it must not be construed as *representative*, as the database was incomplete.

In viewing the compositional data, we shall consider 3 characteristics: structure, heterogeneity, and blunders. Possible differences between and among groups will be of interest, as well as the dimensionality of structure and blunders. Fig. 11, which presents "a plot of the data" [rule 1], represents a beginning, but instant visual insight for so complex a dataset is not so facile as in the case of the gamma ray spectrum (Fig. 5). Examination of selected *univariate* plots, keyed to countries and subgroups is a good next step [rules 6, 7]. Fig. 12 shows global frequency histograms for 3 characteristic elements: Zn (essential), Hg (toxic), and Se (essential, but toxic at higher concentrations). Median concentrations are 21 mg/kg, 24 μ g/kg, and 116 μ g/kg, respectively. It is immediately apparent that the data are not homogeneous; the distributions are approximately

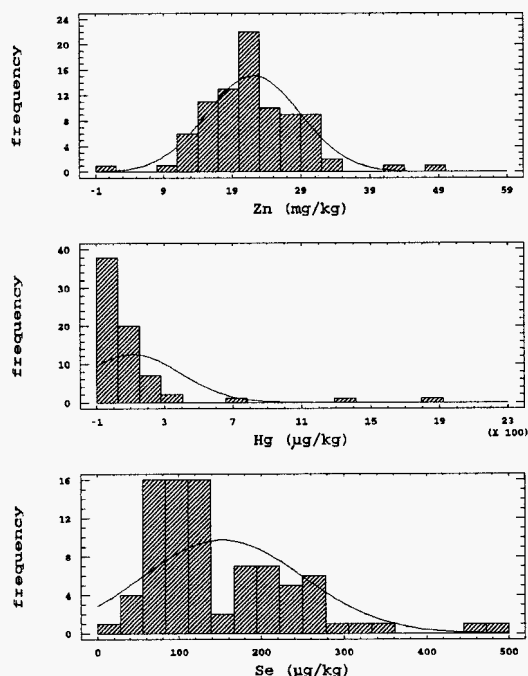


Fig. 12. Global frequency histograms for Zn, Hg and Se (70–86 samples and 7–9 countries, each).

normal, lognormal, and bimodal, with some suggestion of outliers. The next step in a *scientific* investigation would be to enquire as to the reasons for the different distribution types. Although a normal distribution might be expected for a bio-regulated (essential) element, and lognormal, for a non-essential environmental contaminant, these questions deserve expert multidisciplinary investigation (ref. 54).

Greater insight can be gained by examining subgroup distributions; conveniently accomplished with multiple box plots (Fig. 13). Here, a generally dichotomous character for *Ni* concentrations is observed, with higher and more disperse concentrations obtaining for Iran and Turkey, which have vegetarian diets. There is little evidence, however, of intra-regional differences. This figure contains also evidence of serious outliers, in population groups 10, 11, and 15. In the multivariate graphical exploration that follows, we shall gain some clues as to whether such outliers represent unusual samples, laboratory mistakes/contamination, data entry errors, etc.

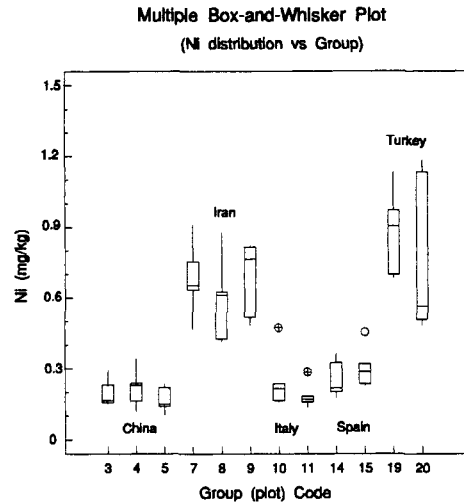


Fig. 13. *Ni* distributions. The boxes indicate the interquartile ranges and medians.

Unlike *Ni*, some elements do exhibit intra-regional heterogeneity. For example, phytate is considerably higher in the daily diet of a wheat eating, rural population group in China than either a rice eating, rural group or a group from Beijing. Similarly, differences are manifest in the concentrations of many of the elements in two Italian population groups -- one from a mountainous area with a traditional diet (IT-2), and one from an area having a high seafood diet (IT-3). This is illustrated in the last univariate plot, Fig. 14. We make two observations: (1) that *As* concentrations for group IT-3 are considerably higher than those for group IT-2; and (2) samples *b* and *j* differ greatly from other members of their respective groups.

Moving on to 2 dimensions [rules 6, 7], we next examine a bivariate *Zn*-*Se* plot for food samples from Iran, Spain and the U.S. (Fig. 15). Three types of discrepancy occur in this case: (1) Sample *A*, which deviates greatly from the Iranian cluster, was confirmed by two laboratories -- hence, it is a "sample outlier", rather than a laboratory or data entry error. The fact that the outlier is bivariate is important; it means that "simple" reagent contamination or data entry (apart from mislabeling) is not the cause.

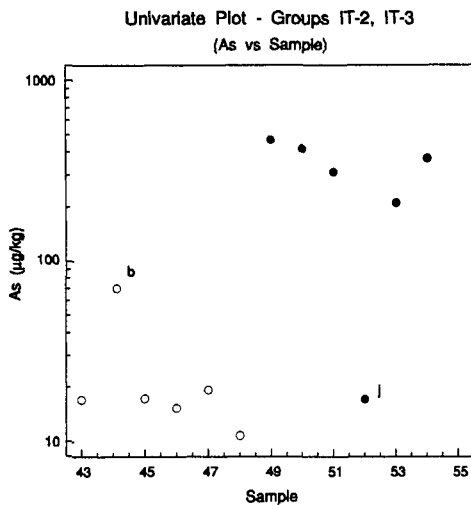


Fig. 14. Univariate plot of *As* for daily diet samples from Italy -- population groups IT-2 (o), IT-3 (●).

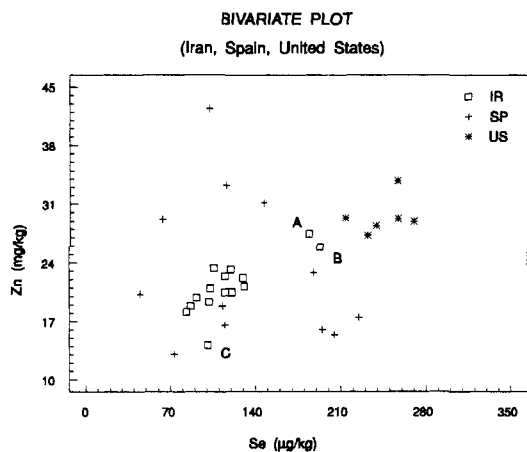


Fig. 15. Bivariate *Zn*, *Se* plot for daily diet samples from Iran, Spain, US.

(2) Sample *B* gave discrepant results between two laboratories; the diagnosis must be *at least a laboratory error*, possibly also a sample error -- reanalysis is in order. (3) The scatter (imprecision) exhibited by the Spanish samples (+) is surprising in view of the Iranian and U.S. data; there is some question as to the quality of the sampling protocol. It should be emphasized that this exercise is *exploratory*; except for interlaboratory discrepancies, unusual sample results are not necessarily wrong. Rather, they merit further investigation.

Enlarging our view to 3 and 4 dimensions [rules 6–8] is still more informative. Fig. 16 shows *Zn* results from 3 laboratories for a common set of samples. This figure represents the extension of the "Youden Plot" concept, for the assessment of interlaboratory quality, to the multilaboratory – multisample (multivariate) regime (ref. 24, 50). The data are plotted as a conventional 3D real variable (RV) plot in the lower part of the figure, and as a principal component (PC) plot in the upper part. Both plots show immediately that one result (*A*) is outlying; the PC plot shows instantly the corresponding laboratory (*R*). This figure makes a good bridge between RV space and PC space. PC projections provide the most efficient lower dimensional view of multivariable space; in this case we see the best projection of the 3–laboratory data on a plane (ref. 11).

PC plots are especially valuable beyond 3 dimensions, as in Fig. 17. Here, data for the 4 toxic elements for Italian groups IT–2 ("traditional diet") and IT–3 ("high seafood") are projected on the PC1, PC2 plane. This type of plot allows us to quickly spot outliers, data heterogeneity (clusters), and linear relationships using all samples and all (4) variables. We see, for example: (1) An extremum; sample-*i* lies isolated at the tip of the *Pb* vector, implying a serious outlier. In fact, the *Pb* concentration in this sample exceeds that of all others by a factor of 100; a result confirmed by all 3 laboratories. (2) A mismatch; sample-*j* falls within the "wrong" group, as it did in the univariate plot (Fig. 14). (3) A discrete separation between the two population groups, IT–3 concentrations exceeding those of IT–2 for *As*, *Hg*, and *Pb*. (4) A more or less linear pattern at least for IT–2, which seems related to the *Cd* and *Pb* variables. (5) Correlation between *Cd* and *Pb*, suggesting that we examine the corresponding 2 dimensional RV plot [rule 8]. That plot is given in Fig. 18.

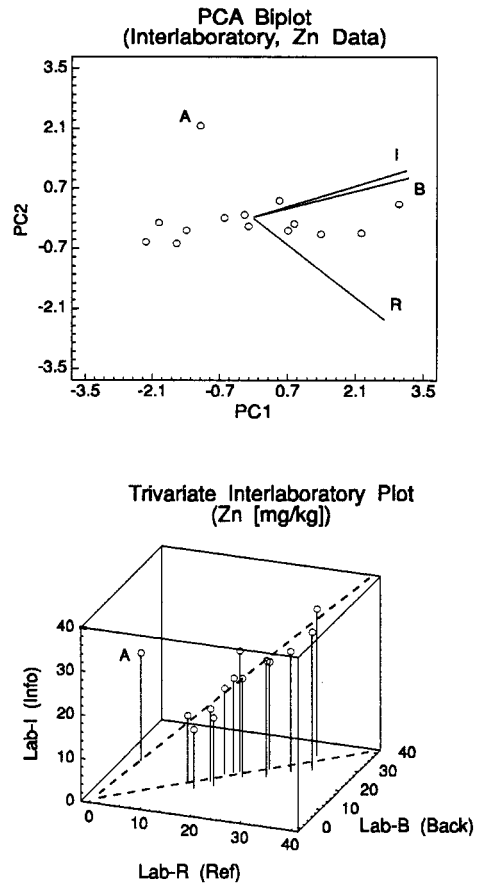


Fig. 16. PC (upper) and RV (lower) plots of *Zn* interlaboratory data. Labs *I*, *B* and all samples except *A* align with the 1st principal component.

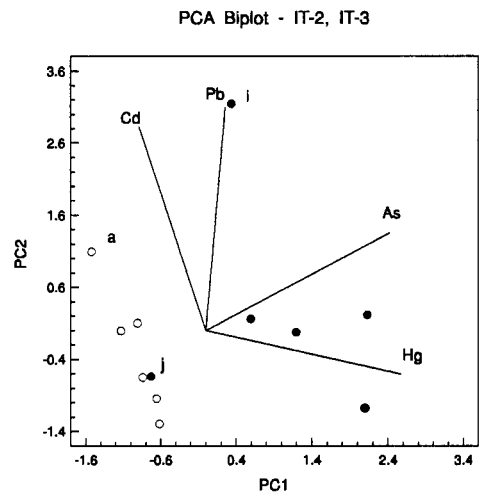


Fig. 17. Toxic element PC plot for Italian groups IT–2 (o), IT–3(●). Circles represent samples; vectors represent elements.

This last figure is very interesting indeed. With the exception of sample-*j*, which is *precisely* in the wrong population group, the two groups display unique linear relationships, such as would be manifest in 2-source (environmental) mixing processes. The scatter about the lines is consistent with measurement error, and far smaller than the overall dispersion. Thus, this small example has demonstrated how graphical multivariate exploration can lead to explicit answers to our initial questions about data structure, data heterogeneity, and data quality. Also, it illustrates that eventually we run up against rule 5, when a large dataset (here, 6300 observations) with rich and interesting heterogeneity is decomposed into homogeneous subgroups.

Some final comments: (1) Further exploration showed sample-*j* to be in the wrong group for *all* variables -- it must have been mislabeled. (2) Selecting the right RV plot (Fig. 18) was not a matter of chance; the insight came from the PC plot (Fig. 17). Similarly, an 8 dimensional PC plot of essential elements for groups IT-2, IT-3 led to a linear relation in 3 dimensions. To find that directly would have required examining as many as $C(8,3) = 56$ three-dimensional plots! (3) Multivariate data exploration is not an end, in itself; rather, it is a beginning. It has been gratifying that after communication of some puzzling patterns in the data to participants, re-examination of some methodology (analytical and data handling) led to important corrections.

CONCLUSION

The pursuit of accuracy in analytical science is a timeless endeavor, and necessarily a multidisciplinary one. Expertise in exposing and controlling the really important errors requires a breadth and depth of knowledge that transcends any one field. The process begins with a common language that we call *Nomenclature*. International scientific organizations, such as IUPAC, have provided a solid foundation in this regard; but work remains to achieve common, and scientifically valid approaches to detection and identification limits, and low-level data reporting -- so important for trace analysis and communication with the public. The largest errors arise from erroneous assumptions and models. For these, the best defense involves a judicious combination of multidisciplinary competence and exemplary reference materials and test data. The most dramatic improvements in analytical accuracy derive from major advances in measurement and computational science that help to replace our dependence on assumptions and models by *direct* observation (analytical and visual).

Acknowledgement Important discussions with G. V. Iyengar and W. Wegscheider are gratefully acknowledged. Thanks are due also to D. B. Klindinst for major assistance with the figures.

REFERENCES

1. C. Priesner, *Chemistry International* 11, 216-238 (1989). See also C. Priesner, *Chemistry and Congress: The Karlsruhe Congress in 1860 and the Development of Chemistry* (distributed at the IUPAC 36th General Assembly, Hamburg, 1991).
2. IUPAC Nomenclature Guides:
 - I. Mills, T. Cvitaš, K. Homann, N. Kallay and K. Kuchitsu, Eds., *Quantities, Units and Symbols in Physical Chemistry* (The Green Book), Blackwell, Oxford (1988).
 - J. Rigaudy and S.P. Klesney, Eds., *Nomenclature of Organic Chemistry* (The Blue Book), Pergamon Press, Oxford (1979).
 - G.J. Leigh, Ed., *Nomenclature of Inorganic Chemistry* (The Red Book), Blackwell, Oxford (1989).
 - H. Freiser and G.H. Nancollas, Eds., *Compendium of Analytical Nomenclature: Definitive Rules 1987* (The Orange Book), Blackwell, Oxford (1987).
 - V. Gold, L. Loening, A.D. McNaught and P. Sehmi, Eds., *Compendium of Chemical Terminology: IUPAC Recommendations* (The Gold Book), Blackwell, Oxford (1987).

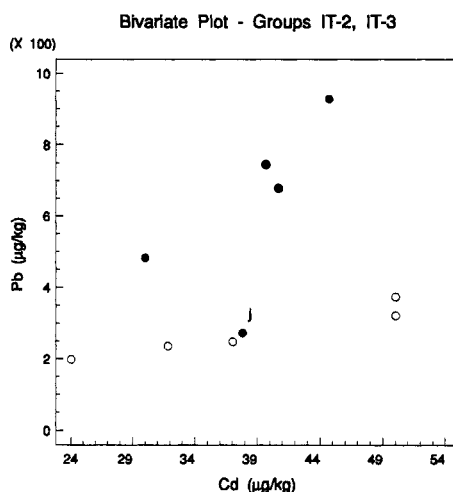


Fig. 18. RV-plot of *Pb* vs *Cd*, for population groups IT-2 [○] and IT-3 [●].

3. International Vocabulary of Basic and General Terms in Metrology (BIPM, IEC, ISO, OIML, IFCC, IUPAC), International Organization for Standardization, Geneva (1984); draft revision (1990).
4. Statistics – Vocabulary and Symbols, ISO-3534, Geneva (1977).
5. Terms and Definitions Used in Connection with Reference Materials, ISO Guide 30; Uses of Certified Reference Materials, ISO Guide 33, Geneva (1989).
6. IUPAC Interdivisional Working Party for Harmonization of Quality Assurance Schemes for Analytical Laboratories: W. Horwitz, Pure Appl. Chem. **60**, 856–864 (1988); W.D. Pocklington, Pure Appl. Chem. **62**, 149–162 (1990). See also ISO-9000 Series of Quality Assurance Standards.
7. R.M. Parr, Intercomparison of Cadmium and Other Elements in IAEA Horse Kidney (H-8), Progress Report No. 2, IAEA, Vienna (1985).
8. L.A. Currie, Detection: Overview of Historical, Societal and Technical Issues, Ch. 1 in L.A. Currie, Ed., Detection in Analytical Chemistry: Importance, Theory, and Practice, ACS Sympos. Ser. 361, Amer. Chem. Soc., Washington, (1988).
9. M.W. Brossman, Rapporteur, Reporting Low-Level Data for Computerized Data Bases, Ch. 17 in L.A. Currie, Ed., Detection in Analytical Chemistry: Importance, Theory, and Practice, ACS Sympos. Ser. 361, Amer. Chem. Soc., Washington, (1988).
10. L.A. Currie and G. Svehla, Recommendations for the Presentation of Results of Chemical Analysis, to be submitted to Chem. Intl. (in review, 1991).
11. D.L. Massart, B.G. M. Van de Ginste, S.N. Deming, Y. Michotte and L. Kaufman. Chemometrics: a textbook, Elsevier, Amsterdam (1988).
12. B. Flury and H. Riedwyl, Multivariate Statistics, Chapman and Hall, London (1988).
13. M. Waldholz, Wall Street Journal, 51 (19 Feb 1987).
14. T. Layloff, Pharmaceutical Technology **15**, No. 9, 146–148 (1991).
15. T. Ramdahl, J. Schjoldager, L.A. Currie, J.E. Hanssen, M. Möller, G.A. Klouda and I. Alfheim, Sci. Total Environment **36**, 81–90 (1984).
16. K.A. Rahn and D.H. Lowenthal, Science **228**, 275–284 (1985).
17. G.E. Gordon, Envir. Sci. Technol. **22**, 1132–1142 (1988).
18. National Institute of Standards and Technology, Standard Reference Data Products Catalog; Standard Reference Materials Catalog; NIST, Gaithersburg (1991).
19. T.L. Clark and R. D. Cohn, Across North America Tracer Experiment (ANATEX), U.S. EPA Report-600/3-90/051 (1990).
20. G.V. Iyengar, Sci. Total Environ. **100**, 1–15 (1991); G.V. Iyengar, Elemental Analysis of Biological Systems Vol. I, CRC Press, Boca Raton (1989).
21. D. Behne, J. Clin. Chem. Clin. Biochem. **19**, 115 (1981).
22. L.J. Gleser, Chemometrics and Intelligent Laboratory Systems **10**, 45–57 (1991).
23. L.A. Currie, Chemometrics and Intelligent Laboratory Systems **10**, 59–67 (1991).
24. W.J. Youden, Statistical Techniques for Collaborative Tests, Assoc. Offic. Analyt. Chem., Washington (1969).
25. R.R. Meglen, Ch. 18 in J.J. Breen and P.E. Robinson, Eds., Environmental Applications of Chemometrics, ACS Sympos Ser. 292, Amer. Chem. Soc., Washington (1985).
26. Quality Assurance Task Force, Center for Analyt. Chem., National Instit. Stands. Technol., Use of NIST Standard Reference Materials for Decisions on Performance of Analytical Chemical Methods and Laboratories, Gaithersburg, MD 20899 (1991).
27. International Atomic Energy Agency, Analytical Quality Control Services: Intercomparison Runs Reference Materials, IAEA, Vienna (1991).
28. R.M. Parr, H. Houtermans and K. Schaerf in Computers in Activation Analysis and Gamma-Ray Spectroscopy, CONF-780421, 544–562, U.S. Dept. of Energy, Germantown (1979).
29. L.A. Currie, J. Res. Natl. Bur. Stand. **90**, 409–419 (1985).
30. L.A. Currie, R.W. Gerlach, C.W. Lewis, W.D. Balfour, J.A. Cooper, S.L. Dattner, R.T. DeCesar, G.E. Gordon, S.L. Heisler, P.K. Hopke, J.J. Shah, G.D. Thurston and H.J. Williamson, Atmospheric Environment **18**, 1517 (1984).
31. M.E. Algeo (NIST), member of Amer. Soc. Mechan. Engin., B89.1.12 Software Performance Evaluation Task Force, personal communication (1991).
32. H.C. Urey, The Planets: Their Origin and Development, Yale Univ. Press, New Haven (1952).
33. V.A. Koptug, Coordinator, Chemistry and the Environment, IUPAC, Oxford (1990).
34. G.P. Brasseur, Rapporteur in F.S. Rowland and I.S.A. Isaksen, Eds., The Changing Atmosphere, 235–256, Wiley, Chichester (1988).

35. L.W. Alvarez and R. Cornog, Phys. Rev. **56**, 379 & 613 (1939).
36. R. Muller, Science **196**, 489 (1977).
37. K.H. Purser, R.B. Liebert, A.E. Litherland, R.P. Beukens, H.E. Gove, C.L. Bennett, M.H. Clover and W.E. Sondheim, Rev. Phys. Appl. **12**, 1487 (1977).
38. J. Klein and R. Middleton (^{10}Be), Nucl. Instrum. Meth. Phys. Res. **B5** 129–133 (1984); P.W. Kubik, P. Sharma, R.T.D. Teng, S. Tullai–Fitzpatrick, S. Datar, U. Fehn and H. Gove (^{36}Cl), Nucl. Instrum. Meth. Phys. Res. **B52**, 238–242 (1990).
39. G. Bonani, H–J. Hofmann, E. Morenzoni, M. Nesi, M. Suter and W. Wölfi, Radiocarbon **28**, 246–255 (1986).
40. A. Long, R.S. Kra and D. Srdoč, Eds., 13th International Radiocarbon Conf.: Radiocarbon **31**, No. 3, 229–1082 (1989).
41. F. Yiou and G. Raisbeck, Eds., 5th International Conf. on Accelerator Mass Spectrometry: Nucl. Instrum. Meth. Phys. Res. **B52**, Nos. 3/4, 211–630 (1990).
42. D. Elmore, N. Conard, P. Kubik, H. Gove, M. Wahlen, J. Beer and M. Suter, Nucl. Instrum. Meth. Phys. Res. **B29**, 207–210 (1987).
43. M. Wahlen, B. Deck, H. Weyer, P. Kubik, P. Sharma and H. Gove, ^{36}Cl in the Stratosphere (Abstract), Radiocarbon **33**, 257 (1991).
44. T.W. Stafford, Jr., P.E. Hare, L.A. Currie, A.J.T. Jull and D. Donahue, Quaternary Research **34**, 111–120 (1990).
45. L.A. Currie, G.A. Klouda, R.E. Continetti, I.R. Kaplan, W.W. Wong, T.G. Dzubay and R.K. Stevens, Radiocarbon **25**, 603–614 (1983); L.A. Currie and G.A. Klouda, NBS Report of Analysis – Shenandoah Samples (19 July 1982).
46. Science **253** (13 Sept. 1991) -- "computer science" issue; Committee on Earth and Environmental Sciences, Our Changing Planet: The FY 1992 U.S. Global Change Research Program, U.S. Office of Science and Technol. Policy, Washington (1991).
47. B. Efron and R. Tibshirani, Science **253**, 390–395 (1991).
48. E. Tufte, The Visual Display of Quantitative Information, Graphics Press, Cheshire (1983).
49. J.J. Filliben, DATAPLOT: Introduction and Overview, NBS Special Publ. **667**, National Bureau of Standards, Gaithersburg (1984).
50. L.A. Currie, Ch. 6 in K.S. Subramanian, G.V. Iyengar, K. Okamoto, Eds., Biological Trace Element Research: Multidisciplinary Perspectives, ACS Sympos. Ser. 445, Amer. Chem. Soc., Washington (1991).
51. J.W. Tukey, Exploratory Data Analysis, Addison–Wesley, Reading, MA (1977).
52. R.M. Parr, et al., Dietary intakes of trace elements and related nutrients in eleven countries: preliminary results from an IAEA co–ordinated research programme, Proc. Sympos. Trace Elements in Man and Animals – TEMA–7, Dubrovnik (May 1990).
53. L.A. Currie, et al., Exploration of intra– and inter–regional multivariable patterns in the IAEA daily diet data, accepted for publication in Chemometrics and Intelligent Laboratory Systems, 1991.
54. K.S. Subramanian, G.V. Iyengar, K. Okamoto, Eds., Biological Trace Element Research: Multidisciplinary Perspectives, ACS Sympos. Ser. 445, Amer. Chem. Soc., Washington (1991).