Trace Residue Analysis

Trace Residue Analysis

Chemometric Estimations of Sampling, Amount, and Error

David A. Kurtz, EDITOR

The Pennsylvania State University

American Chemical
Society Library
1155 16th St. N. W.
Washington. D. C. 20036

ACS Symposium Series

M. Joan Comstock, *Series Editor*

Advisory Board

FOREWORD

The ACS SYMPOSIUM SERIES was founded in 1974 to provide a
medium for publishing symposia quickly in book form. The
format of the Series parallels that of the continuing ADVANCES
IN CHEMISTRY SERIES except that, in order to save time, the
papers are not typeset but are reproduced as they are submitted
by the authors in camera-ready form. Papers are reviewed under
the supervision of the Editors with the assistance of the Series
Advisory Board and are selected to maintain the integrity of the
symposia; however, verbatim reproductions of previously pub-
lished papers are not accepted. Both reviews and reports of
research are acceptable, because symposia may embrace both
types of presentation.

PREFACE

THE SYMPOSIUM UPON WHICH THIS VOLUME is based was organized originally because of the perpetual need to better formalize both understanding and error in the analytical methods used in quantitative analytical work. In this field, problem areas occur in sampling, recovery, and quantitative measurement. These analyses involve the production of numbers or data that describe quantitatively the system under scrutiny. Those who have been a part of this process know the locations of the various errors and have some idea of the size of the error. They may even run appropriate statistical tests to quantitatively determine the amount of error.

However, society likes to have decisions made in a black and white manner and to know whether something is there or not. This situation suggests that the analytical error should drop to zero. While this result is the goal of all analytical work, it is simply not realistic. Our basic need, then, is to simplify error determinations and explanations and to educate the public both for the reasons and for the interpretations of error. The goal of this volume is to further the use of mathematical and statistical tools—the field of chemometrics—for chemical and, specifically, trace chemical analyses of pesticides and environmental contaminants.

Statistics have been used in chemical analysis in increasing amounts to quantify errors. The focus shifts now to other areas, such as in sampling and in measurement calibrations. Statistical and computer methods can be brought into use to give a quantified amount of error and to clarify complex mixture problems. These areas are a part of chemometrics as we use the term today.

Errors in trace analyses are usually hidden to all except those intimately involved in the sample collection and, later, in the bench analysis. In chromatography, especially, it is too easy to hide behind uncertain work because published research does not concern itself with exactly how the chromatographer makes his quantitative decisions. Today, with the advent of the microprocessor and with the use of "black box" instruments, the chromatographer knows even less about his calibration graph or line, or the error associated with it. In these instruments, a single point and the origin may determine the calibration graph. Similar problems exist in other modern instrumental analysis techniques.

This volume addresses these problems directly. The use of statistics enables error to be determined in calibration measurements at a particular

confidence level. Decisions can be made in sample selection, and the limits of detection can be determined in an orderly manner. The knotty problem of outliers can be approached systematically. The symposium on which this volume is based was formatted, first, to outline appropriate and noncumbersome methods for analytical decision making and, second, to make the methods easily understandable to the ordinary bench chemist so that they will actually be used. I visualize this text, actively being used, next to an analytical instrument. I hope it is clear enough so that when used, the bench chemist will be able to obtain more meaningful results that can be interpreted on short notice.

DAVID A. KURTZ
The Pennsylvania State University
University Park, Pennsylvania

March 25, 1985

Statistics: A Child of Our Time?

FRAUKE TSCHILTSCHKE

Department of Philosophy, Christian Albrechts University of Kiel, Kiel, West Germany

> The use of data to represent scientific information
> has been found not only to be of modern use but also
> to have been a part of our society for centuries.
> Science and statistics thus formed have explained
> our complex world in more and more detail but have
> yet to fully explain truth . . . the goal of the
> philosophers.

"Every year 10% of the American chemists spend 40 hours in
conference rooms and use 19 pounds of paper." Even if this
statement is not a truthful one, it expresses one of the well
established forms of scientific statements, namely a statistical
one. We are quite used to dealing with statistics, the collection
and analysis of data and the drawing of conclusions from this data
(1). In a scientific way, this mode constitutes no problem. On
the other hand compare these two statements: "Get a shot against
the flu because only very few of the inoculated people will get
the flu," versus "Get a shot against the flu, because only 3% of
the inoculated people will get the flu". The second statement
provides more precise information than the first.
    Statistics seems to be a "magic" word of our time!
Statistical use has rapidly increased in our century which has
indicated that a strong belief is now present. In earlier times,
this belief did not exist, but are we not sure that the use of
statistics was not present then? Can it be traced back to the
early civilizations?
    This paper will show that statistics has been with us for a
long time. The techniques have gradually developed from simple
counting or information gathering to explaining complex phenomena
with only limited information. Not to be forgotten will be the
ultimate interactive role of philosophy.

Current address: Klodtstrasse 12, 2408 Tinmendorfer Strand, West Germany

Scientific Thinking

In former times philosophers proposed generalities to describe
life. They failed in some respects because they were out of touch
with reality and did not look at facts. The explanations were
only logically derived and suggested errant directions.
     Later scientific endeavor gained favor because of its
self-correcting nature. Making an observation, collecting
information, analyzing it, and drawing conclusions which express
the reasons for the common behavior in a scientific law is exactly
the way scientific thinking works. If facts found later were
inconsistent, the generalities were modified to include them. In
short the understandable reason for an observed event was
attempted to be found.
     Too often enough information simply was not available for
adequate scientific laws to be fashioned. In this area,
statistics was found to have an ever-increasing role. A
confidence level was devised that allowed limited facts to express
larger generalities. This narrowed the amount of work scientists
had to do to come to good conclusions (and sometimes increased the
work they thought they had to do to reach a desired confidence!).
Conclusions or hypotheses are never absolute but are more and more
certain as the number of facts available increases. Nonetheless,
astounding amounts of good conclusions, again at a given
confidence level, can be drawn from limited facts using
statistics.

Statistics in Former Times

The early role of statistics was essentially only in collecting
facts and assembling them in an orderly way. It seems to have
been a valuable method since we have seen such an increase in its
use. For example, we have found reports about weather, stars,
sun, moon, and change of day and night in all of the old cultures.
Even without a complex language script like ours it was possible
to cut marks in stones and sticks, which allowed counting. In
4241 B.C. Egyptians had a fairly precise calendar; even the leap
year was known (2). Other examples of scientific observations
include registrations of populations, harvests, and tenure.
Around 3700 B.C. Seneferu, a mighty warrior, raided and captured
7,000 men and 200,000 sheep, cattle, and goats (3). Wherever
people lived together and depended on each other, it was necessary
to make plans for the use of their land and available water for
producing food. For example, in Egypt the cultivation depended on
the flood area of the Nile River. They, therefore, had
observation stations along the river to measure the water level.
From this measuring they made the following very precise
predictions: only a 21 foot level meant famine. A 23 foot level
meant imperfectly watered land. However, at 26.5 feet the whole
country had plenty of water (4).

One of the best examples of a precise statistical estimation of counting, analyzing, and drawing conclusions was the exact forecast of the eclipse of the sun in 585 B.C. made by Thales of Millet (5).

New Beliefs in Numbers

Over the years people have switched their beliefs from the explanations of the gods of former times to the explanations of the gods of modern times, the scientists and statisticians. It is easy to see how people have done this. Modern people began to believe in numbers and data because they represented nature so well. The theories of science and statistics were explained, and the ideas of science and statistics became better founded and entrenched in their thinking. However one mistake was made and that was a big one: facts and data were taken as truth and reality.

The change in the beliefs of the people did not happen overnight. As facts and data began to substantiate the theories and methods of science, the beliefs of the people slowly evolved away from the more general explanations of the philosophers. The feeling that the thinking of the philosophers represented truth and reality was lost. In former times the friends of truth – which is the translation of the Greek word for philosopher – tried to find basic explanations from which they could explain all the natural phenomena in the world. However, the study proved to be too complex. In chemistry and physics, for example, there is the belief that the world is built of basic elements, but people kept finding smaller and smaller elements: atoms, neutrons, and now neutrinos and quarks. Our thinking just became shrouded with facts so that the wholeness of the world became lost.

Wholeness Thinking Lost

The truth the philosophers searched for was strongly influenced by the idea of wholeness. Wholeness gives a broad direction in life. However, our people have often felt that scientific thinking has lost its connection to the idea of truth and wholeness. The world is so complex and detailed that people have become only specialists instead of generalists. The latter category includes the philosophers. We switched over to the idea that specialists can be the only ones that bring truth. The different directions, such as math, physics, art, and philosophy, lost their connections to each other and were removed far away from their original study areas and the idea of truth and wholeness. Today the different departments build up such large realms of special knowledge that year-long studies are necessary to find one's way through. For each of these scientific realms people developed their own language which was almost like the event of the building the tower of Babel.

With statistics as an example, I have tried to show how easily a progressive thought, although valuable in explaining nature, loses its context. To be sure, statistics has aided the development of science tremendously especially in recent times. In spite of this, however, explanations of nature that aid life have fallen short of this mark. It is now necessary to find again one common "language" so that we are able to put the results of the different fields together and to bring increased understanding of our world.

Today people wish for security. Their orientation is expressed by the way they believe in science, namely, in something god-like that should be able to rule the world. Our society puts all emphasis on a scientific education and scientific research, and eliminates at the same time all other possible methods which can offer explanations. Therefore, it might be a good start to organize a different sort of conference, where scientists from all disciplines sit together and discuss ways to cooperate with each other. Here philosophers should lead the discussions. Lively discussions will ensure new directions, ideas, and goals, which will again be close to their original thought, the friendship to truth!

Acknowledgments

Literature Cited

1. Snedecor, G. W.; Cochran, W. G. in "Statistical Methods"; 7th Ed.; Iowa State University Press: Ames, IA, 1980.

2. Breasted, J. H. in "A History of the Ancient Egyptians"; Smith, Elder and Co.: London, 1908; Vol. V, pp. 35-36.

3. Budge, E. A. W. in "A Short History of the Egyptian People"; J. M. Dent and Sons, Ltd.: London, 1914; p. 38.

4. Budge, E. A. W. in "The Nile"; Thomas Cook and Son, Ltd.: London, 1902; pp. 76-77.

5. Rousseau, P. in "Man's Conquest of the Stars"; Jarrolds Publishers, Ltd.: London, 1959; p.43.

Sampling for Chemical Analysis of the Environment: Statistical Considerations

B. KRATOCHVIL

Department of Chemistry, University of Alberta, Edmonton, Alberta, Canada T6G 2G2

A statistically valid sampling plan requires
careful design and execution so that
generalizations based on mathematical probability
can be drawn from a small number of test
portions. Guidelines are given for estimation of
the minimum number and size of sample increments
needed to achieve a given level of confidence in
chemical analyses.

Accurate sampling for pesticides and pesticide residues in the
environment presents formidable problems. The population of
interest is likely to be complex. It may consist of such
diverse matrices as air, water, vegetation, soil, sediment,
fish, or wildlife. Furthermore, concentrations of the sought-
for substance may be low and unevenly distributed. The
necessity for a sound sampling program in any study of
pesticide distribution in the environment is generally
recognized. Yet programs are often so designed as to be
statistically unsound, or a valid, well designed plan is
compromised by expediency or carelessness. The effort expended
on evaluation of sampling designs for pesticide monitoring is
usually exceedingly small compared with that expended on the
analytical measurements. In only a few cases have general
considerations for statistically sound environmental sampling
plans been discussed [1,2].

An example of a thorough sampling study is the
investigation of fungicide persistence in soil by a randomized
sampling plan [3]. Other authors have presented general
criteria for sampling matrices such as soils [4], plants and
soils [5], and air [6]; a more general review on sampling for
chemical analysis is available [7]. A useful discussion
containing much practical information has been provided by the
monitoring panel of the Federal Working Group on Pest
Management in the U.S.A. [8]. This group observed that most

recorded data on deleterious substances in the environment have
not come from programs designed according to statistical
principles, and so the reliability of extrapolations from the
results cannot be assessed.

The reliability of any environmental analytical data
depends upon the reliability of sample quality. To generalize
from analytical results on a small portion of material to a
larger population requires careful planning and execution if
bias is to be avoided. This article considers the general
problems involved in sampling heterogeneous bulk populations
such as soil, air, and natural waters; specific details for
particular types of materials are not included. These problems
include the heterogeneity of most environmental materials; the
costs in time, manpower, and effort required for collection of
real samples; and the need to avoid contamination or
decomposition of samples after collection.

A set of definitions of terms frequently used in sampling
is provided because usage sometimes differs among
statisticians, chemists, and others. The definitions have been
selected after consideration of the recommendations of various
standards organizations.

Background

Sources of error in an analysis may be classified as random or
systematic. Systematic errors generally bias a result in one
direction in a relatively reproducible way and are not usually
amenable to statistical treatment. Random errors vary in a
nonreproducible way around the true value and can be treated
statistically by the laws of probability. Therefore in this
discussion we shall deal only with random errors, keeping in
mind that most errors are partly random and partly systematic
and that systematic errors in the analytical operations can be
controlled by proper use of blanks, standards, and reference
samples. Because poor samples are not identifiable by such
checks, sampling uncertainty is often treated separately. For
random errors the overall variance s_o^2 is the sum of the
sampling variance s_s^2 and the variance of the remaining
analytical operations s_a^2: $s_o^2 = s_s^2 + s_a^2$. The value of s_s^2
may be obtained by subtraction of s_a^2 (known if a measurement
process is in statistical control) from s_o^2 (obtained by
analysis of the samples). Alternately, a series of replicate
measurements or samples can be designed to evaluate both
standard deviations. Reduction in the overall uncertainty
requires, therefore, attention to both sampling and analytical
operations. Once the analytical standard deviation s_a is one
third or less of the sampling standard deviation s_s, further
reduction in s_a has little effect on s_o (9).

An example of the importance of sampling is in the
determination of aflatoxins (a class of highly toxic compounds

produced by molds) in peanuts (10,11). Because the
distribution of contaminated kernels is typically patchy and
uneven, and because the tolerable level of contamination is so
low (about 25 ppb), sampling is the major source of analytical
uncertainty even with samples of over 20 kg.

The overall analytical process can be divided into five
steps-- construct a model, design a plan, take samples, perform
analyses, and evaluate results (12). The model defines the
population to be studied, the substances to be measured
(including speciation), the extent of distribution within the
population, and the level of precision required. The sampling
plan specifies the number, size, and location of the sample
increments, the extent of combining of increments
(compositing), and the procedures for reduction of the bulk or
gross sample to a laboratory sample and to test portions
(subsampling). The plan should be written as a detailed
protocol before work begins and revised as warranted by new
information. It should include on-site criteria for collection
of a valid sample, such as whether a substance should be
considered foreign to the population and rejected. A discarded
piece of metal or plastic in a field, for example, might be
considered foreign for a soil analysis and therefore
legitimately rejected. It should also include information on
procedures for protection of the sample from contamination
before and after collection, for preservation, and for labeling
and recording of all appropriate information.

Field sampling operations are often costly in time and
manpower. Those collecting samples should be aware of the
possibility of bias and contamination.

Random and Systematic Sampling

In devising a model for an analytical operation, we identify a
target population to which we want our conclusions to apply.
This will differ from the parent population from which the
samples are actually taken. The difference may be reduced by
random selection of individual portions (increments) for
analysis so that each part of the population has an equal
chance of selection. Genuinely random sampling is difficult
because bias, unconscious or deliberate, is readily
introduced. Untrained individuals often have difficulty in
accepting that an apparently unsystematic sampling pattern must
be followed to be valid.

For simplicity and convenience, sampling at evenly spaced
intervals over a population is often used in place of random
sampling. For example, a field may be divided into uniform
segments, and a sample taken from the center of each segment.
This procedure is generally subject to more bias than random
sampling. Should periodicity in the population be present or
suspected, segments to be sampled should be selected with the

aid of a table of random numbers (13). The sampling site
within each segment should then be selected by further division
into imaginary subsegments, each assigned a number, and the one
to be sampled selected from a table of random numbers.
Sometimes random sampling is difficult to execute, as when a
stream is being monitored with a time-activated automatic
remote sample collection device. Under such conditions a
random start or other superimposed random time element may be
substituted. The efficiency of systematic sampling improves as
the population becomes better understood. Both theoretical and
experimental studies of this point have been made (14).

    When the component of interest is distributed in a
segregated way, special sampling precautions may be needed.
Thus, a pesticide may have been distributed in higher
concentration in one part of the area under study or may have
undergone more rapid degradation in a low wet portion of a
field. To obtain a valid sample of a stratified material, the
procedure recommended (15) is to (i) divide the population into
segments (strata) based on the known or suspected pattern of
segregation, (ii) further divide the major strata into
subsections and select the required number of subsections to be
sampled by use of a table of random numbers, and (iii) collect
samples proportional in number to the relative size of the
major strata. Stratified random sampling is preferable to
unrestricted random sampling, provided the number of major
strata is kept sufficiently small that several increments can
be taken from each.

Composite Samples

When only the average properties of a population, and not the
variability or distribution of the sought-for component, are of
interest, a composite sample may be prepared and analyzed.
Distinction should be made between composite and representative
samples. A representative sample is frequently defined as one
that possesses the average properties of a population; a
composite sample is usually produced by homogenizing in any of
several ways one or more sample increments, and it constitutes
one approach to producing representative samples. Compositing
usually means fewer analyses are required, and sample storage,
recording, and handling are simplified once compositing is
completed. But much useful information may be lost in
preparing a composite sample. Analysis of individual samples
collected by a properly designed and executed sampling plan
permits determination of the between-sample and within-sample
variability as well as the average composition. This
information helps to establish the heterogeneity of the
population, identify anomalous samples, and evaluate
differences within and between laboratories. Thus composite
samples provide limited information and should be employed only
after careful consideration of the disadvantages involved.

Subsampling

If the sample increment is larger than the amount (test portion) needed per measurement, subsampling is necessary. This operation may be simple, as with many liquid or gaseous materials, or complex, as with certain bulk solids. The work required to produce a uniform subsample depends on the heterogeneity of the original material. Subsampling of solids may require several steps of particle size reduction and mixing; much has been written on this topic. Particle size reduction is important when the particles differ appreciably in composition because sampling error may occur even in a well mixed sample if too few particles are taken for analysis. One approach to determining the extent of the reduction needed is to treat the sample as a two-component mixture, with each component containing a different amount of the substance of interest (16,17). This treatment is based on a binomial distribution of the two kinds of particles. Because it has been covered in detail elsewhere, it will not be considered here.

Distributions Found in Nature

For the purpose of sampling for chemical analysis three types of distributions can be considered. These are the Gaussian (also known as the normal, Laplace, or DeMoivre), the Poisson, and the negative binomial. Knowledge of the type of distribution is useful in devising the most efficient sampling design. Gaussian and Poisson distributions are both closely related to the binomial distribution, which applies to the probability of whether or not an event will be observed in a series of independent observations. [The binomial distribution is based on the probability of an event or property being observed $\underline{p}$, or not observed $1-\underline{p}$, in a series of $\underline{n}$ independent observations. The distribution of the number of times the event is observed, $\underline{x}$, in $\underline{n}$ trials is given by

$$\underline{p}(\underline{x}) = \frac{n!}{\underline{x}! \, (\underline{n}-\underline{x})!} \cdot \underline{p}^{\underline{x}} \cdot (1-\underline{p})^{n-\underline{x}}$$

For further information see Reference 18.] The event might be the presence of any particular attribute in a sample, such as the detection of a pesticide. Only two levels of the attribute are possible, present or not present. If many attributes contribute to the result of an observation, the binomial probability distribution approaches a limiting curve whose equation is given by $\underline{y} = (1/\underline{\sigma} \, \sqrt{2\pi}) \, \exp[-(\underline{x}-\underline{\mu})^2/2\underline{\sigma}^2]$. As applied to an analytical measurement of a substance, $\underline{y}$ is the probability of a measurement value $\underline{x}$ being observed, $\underline{\mu}$ is the

true value for the substance, and σ is the standard deviation
in μ. This equation describes the Gaussian distribution. This
distribution is observed for a large fraction of the systems
encountered in chemical analysis; a characteristic is that μ is
greater than σ^2.

The Poisson distribution is closely related to the
binomial, and is likewise derived from consideration of
discrete properties. [The Poisson distribution is given by
$p(x) = e^{-\lambda}\lambda^x/x!$ where $\lambda = Np$ when N is large and p is small.
Thus λ is the expected number of events occurring on any given
observation, $\bar{x} = \lambda$. The Poisson distribution is a limiting
form of the binomial distribution (18).] It applies when the
possible number of values N is large but the probability p of
the attribute of interest being observed is small. One example
is the measurement of radioactive decay, where the probability
of any one of a large number of atoms undergoing decay at a
given time may be small. Another example might be the location
of a weed seedling or a live insect in a field after spraying
with a pesticide. In the field there are a large and
unspecified number of points where a weed plant or insect might
be found, but the probability of finding one at a given point
will be small if the application of pesticide has been
successful. The Poisson distribution is characterized by μ,
the mean or average, being equal to the variance σ_s^2. Thus the
standard deviation s for a set of measurements in a Poisson
distribution is easily obtained by taking the square root of
the average, $s = \sqrt{\bar{x}}$. Each observed event must be independent
for the Poisson distribution to hold.

A third type of probability distribution frequently
encountered in nature is where the occurence of one event at
some location increases the probability of other events being
observed nearby. This leads to clumping or patchiness,
characteristic of many biological systems such as weed or
insect infestations, and mold growth in stored grains.
Although a variety of probability distributions have been
considered for contagious systems, the most successful appears
to be the negative binomial. Here a distinguishing
characteristic is that σ_s^2 is greater than μ.

Major considerations in any sampling plan are the size and
number as well as the location of the sampling increments. The
following sections consider aspects of these points.

Estimation of Minimum Size of Sample Increments

For the determination of a chemical or a pesticide in a field
the sampling increment may be a bulk quantity such as a core of
soil, a volume of air passed through a particulates collector,
or a quantity of vegetation gathered from a single site. A
useful method for relating the amount of sample in an increment
to the sampling uncertainty, developed by Ingamells (19,20)

for mining exploration, can be applied effectively to unsegregated Gaussian distributions. In this approach a sampling constant $\underline{K}_s$, corresponding to the weight of sample required to limit the sampling uncertainty to 1% relative with 68% confidence, is defined by

$$\underline{K}_s = \underline{WR}^2 \qquad (1)$$

where $\underline{W}$ represents the weight of sample taken and $\underline{R}$ is the relative standard deviation in sample composition. For a given population, $\underline{K}_s$ is evaluated by performing a series of analyses on sets of samples of differing size either by calculation or with the aid of a sampling diagram. An example is a study of human liver homogenate prepared by cryogenic grinding at the National Bureau of Standards (21). The effectiveness of the homogenization step was assessed by withdrawing a small portion of tissue, irradiating it, adding it to the remainder of the sample, performing the homogenization operation, and measuring the sodium-24 activity in ten samples each of about 0.1, 1, and 5.5 g. For the first set of ten a value of 13.1 was obtained for the percent relative standard deviation $\underline{R}$, for the second set a value of 5.5%, and for the third 2.53. From Equation 1 values for $\underline{K}_s$ are 17, 30, and 35. Thus the value for $\underline{K}_s$ approaches 35, and this is the best estimate of the sampling constant. From Equation 1, then, we find that the weight of subsample in grams required to hold the sampling uncertainty to 1% relative is 35 g. Equation 1 can be used to estimate the sampling uncertainty for subsamples of other sizes. In the above example, a subsample of 0.5 g would be expected to give a sampling uncertainty of about 8% relative.

Note that preliminary measurements are necessary to establish the degree of heterogeneity of the individual sample increments whenever the properties of the population are unknown. Under such conditions estimation of $\underline{K}_s$ should not be based on a single increment, but on results from several. It is always sound practice whenever possible to perform a preliminary assessment of an unknown population by collecting a few samples and analyzing for the component of interest. These samples can be selected on the basis of experience and judgment. Then on the basis of the preliminary results a refined sampling plan can be designed.

Estimation of Minimum Number of Sample Increments

A second factor to consider in a valid sampling plan is the collection of enough individual sample increments to ensure that heterogeneity on a large scale does not bias the results. Estimation of this number can be made straightforwardly if the component of interest is distributed

throughout the population according to a known statistical
relation.

If the distribution is Gaussian or binomial, the minimum
number of increments can be estimated from

$$\underline{n} = \frac{\underline{t}^2 \underline{s}_s^2}{R^2 \overline{\underline{x}}^2} \times 10^4 \qquad (2)$$

where $\underline{t}$ is the Student's t-table value for the level of
confidence desired, $\underline{s}_s^2$ and $\overline{\underline{x}}$ are estimated from preliminary
measurements on or previous knowledge of the population, and $\underline{R}$
is the percent relative standard deviation acceptable as
sampling uncertainty. Initially, $\underline{t}$ can be set at the value for
95% confidence limits, 1.95, and an initial estimate of $\underline{n}$
calculated. The $\underline{t}$ value for this $\underline{n}$ can then be substituted,
and the system iterated to constant $\underline{n}$.

If the distribution is Poisson, $\underline{s}_s^2 = \overline{\underline{x}}$, and Equation 2
simplifies to

$$\underline{n} = \frac{\underline{t}^2}{R^2 \overline{\underline{x}}} \times 10^4 \qquad (3)$$

For a negative binomial distribution an index of clumping $\underline{k}$
must be incorporated, and Equation 2 becomes

$$n = \frac{\underline{t}^2}{R^2}\left(\frac{1}{\underline{x}} + \frac{1}{\underline{k}}\right)\left(10^4\right) \qquad (4)$$

Both $\underline{k}$ and $\overline{\underline{x}}$ are estimated from preliminary measurements.

Estimation of Number and Size of Increments for a Segregated Population

When the population is segregated, a number of samples should
be taken from each stratum or segment. A guide to the number
of samples to collect under these circumstances has been
developed by Visman (22,23). Through an empirical study,
subsequently put on a theoretical footing by Duncan (24,25),
Visman derived the relation

$$\underline{s}_s^2 = \underline{A}/\underline{w}\ \underline{n} + \underline{B}/\underline{n} \qquad (5)$$

where $\underline{s}_s^2$ is the variance of the average of $\underline{n}$ samples of
individual weight $\underline{w}$, and $\underline{A}$ and $\underline{B}$ are constants for a given
population. The magnitude of $\underline{A}$ depends on the degree of
homogeneity at the local level, and may be calculated from
Ingamell's subsampling constant $\underline{K}_s$ and the average
concentration of sought-for component $\overline{\underline{x}}$ by

$$\underline{A} = 10^{-4}\ \overline{\underline{x}}^2\ \underline{K}_s$$

Once $\underline{K}_s$ and $\overline{\underline{x}}$ have been estimated by $\underline{n}$ preliminary measurements

on a given material, $\underline{B}$ can be estimated by calculating $\underline{s}_s$ for the same preliminary measurements and substitution into Equation 5. The magnitude of $\underline{B}$ depends on the extent of segregation or stratification in the material. Once $\underline{A}$ and $\underline{B}$ are known, and an acceptable level for the standard deviation of sampling decided on, various combinations of $\underline{w}$ and $\underline{n}$ can be chosen to hold $\underline{s}_s$ within the selected value.

   Two other methods of obtaining values for $\underline{A}$ and $\underline{B}$ have been developed. In the first, two sets of samples, one of relatively large and the other of relatively small increments, are collected; the constant $\underline{A}$ is obtained from the measurements on the small samples, and the constant $\underline{B}$ from the large samples. Small samples make the first terms on the right side of Equation 5 larger than the second by emphasizing the effects of local heterogeneity and by making the value of $\underline{w}$ smaller. Large samples have the reverse effect, and when $\underline{w}$ is of such a size that the second term swamps the first, a value for $\underline{B}$ can be calculated.

   If the material being sampled consists of discrete particles such that an average particle mass can be calculated, then still another method is useful. In this procedure the constants $\underline{A}$ and $\underline{B}$ of Equation 5 are obtained from the intraclass correlation coefficient $\underline{r}$ between pairs of small, single-increment samples of equal mass, the increments of each pair being collected near each other and the pairs distributed over the population under study. The value of $\underline{r}$ can be estimated from the relation

$$\underline{r} = \frac{2\Sigma(\underline{x}-\bar{\underline{x}})(\underline{x}'-\bar{\underline{x}})}{\Sigma(\underline{x}-\bar{\underline{x}})^2 + \Sigma(\underline{x}'-\bar{\underline{x}})^2} \tag{6}$$

where the sums are over all pairs $\underline{x}$ and $\underline{x}'$ and $\bar{\underline{x}}$ is the mean of all measurements (26). From this pilot study of 10 to 20 pairs the constants $\underline{A}$ and $\underline{B}$ are obtained by $\underline{A} = \underline{s}^2/(\underline{rm} + 1/\underline{w})$ and $\underline{B} = \underline{rAm}$. Here $\underline{m}$ equals 1/(average particle mass), $\underline{w}$ the mass of the individual sample increments, and $\underline{s}$ the pooled standard deviation for the measurements. An attractive aspect of this approach is that it also allows calculation of a minimum detectable bias (MDB) in the sampling operation for any specified confidence level and number of samples from the relation

$$MDB = \underline{ts}\sqrt{2/\underline{n}}$$

The value for $\underline{t}$ is obtained from a table of student's $\underline{t}$ values (see, for example, Table T-5 in Ref. 13, or Table A-4 in Ref. 26) for the desired confidence level and number $\underline{n}$ of samples taken. The need to estimate the average particle mass limits this method to granular materials.

   An example of a calculation of $\underline{r}$, $\underline{A}$, $\underline{B}$, and MDB is given in the Appendix.

Estimation of Sample Size when Form of Population Distribution is Unknown

In the preceding sections the Gaussian, Poisson, and clumped distributions have been discussed, and methods of calculating the number of samples in each case have been given. When no information is available about a population, however, the question arises as to the best approach to use. If sufficient samples can be collected and analyzed to establish the distribution as one of the three, the problem is solved. If the distribution does not fit one of the above, it should be checked to see whether it can be converted to Gaussian by taking the logarithm of the values. Transformations using functions other than logarithmic may be considered, but are not easily related to most real systems.

For unknown distribution forms where only limited data is available it is possible to draw useful conclusions without knowledge of the distribution. For example, a confidence interval can be established for a set of analytical values by plotting cumulative percent of the number of analyses on the vertical axis against the individual analytical values on the horizontal axis. Then draw lines parallel to this plot at a distance of 100 $d_{1-\alpha}$, the values for $d_{1-\alpha}$ being read from a table for various numbers of samples and confidence intervals (see, for example, Table A-21 in Ref. 18). Tables are also available to determine the number of samples required to be able to state that the population cumulative distribution is within a defined band at a selected confidence level (Ref. 18, Table A-21b). The numbers tend to be large. For example, to be 95% sure of containing the distribution within an interval of ±10% relative 740 samples would be required. Clearly the price required for not knowing the form of the population distribution is more data.

Conclusions

A general theory for sampling a heterogeneous system such as the environment for trace levels of substances such as pesticides is not likely to become available for some time. Although a variety of models have been proposed to describe specific distributions, each requires prior knowledge of the system under study. The best approach appears to be to carry out a set of preliminary sampling and analysis operations based on knowledge of similar systems from past experience. The extent of the preliminary work depends on the time and resources available; the more care and effort expended, the better is the quality of the data ultimately collected. On the basis of this initial information a model and sampling plan can

be developed. It must be borne in mind that the plan may need to be altered as a result of data being collected in the course of the work. Such alteration is valid if statistical principles are carefully adhered to throughout.

Acknowledgments

Appendix
Example of Application of Sampling Theory to Pesticide Analysis

Taylor, Freeman, and Edwards (27) performed a study of the pathways and rate of loss of the pesticide dieldrin from a grass-meadow soil. The large number and variety of samples collected and analyzed allow evaluation of the uncertainty associated with the sampling operation. Briefly, in one part of their investigation a set of three soil cores of differing diameters was collected in a diagonal pattern within each square meter of a 6 m × 6 m square portion of a field (Figure 1). Core depth was 17.7 cm; core diameters were 21, 24, and 44 mm. The cores were each extracted with 1:1 hexane:2-propanol. The extract was washed with water and the residual hexane injected into a gas chromatograph. The precision of the extraction and measurement operations can be estimated to be of the order of a few per cent. Relative to the variability observed in the overall results, these uncertainties can be considered negligible. The results, calculated on an area basis to facilitate comparison, are reproduced in Table I.

The data show a wide range; the values in Columns B and C are relatively high, while those in Column E are relatively low. The authors suggested that these variations may reflect irregularities in the spray application of the pesticide. A second, more local, variability was attributed to incomplete mixing of the soil after the spray application. The result is a large value for the overall standard deviation, 166 mg per square meter.

Given these data, what statements can we make about the number and size of samples that would have to be taken to hold the sampling standard deviation to some prescribed level? Calculation of Ingamells' subsampling constant is not appropriate since segregation is present.

The minimum number of sample increments required can be calculated from either Equation 2 or 5. From Equation 2, assuming that a 50% level of confidence is desired and that an acceptable percent relative standard deviation R is 50, n
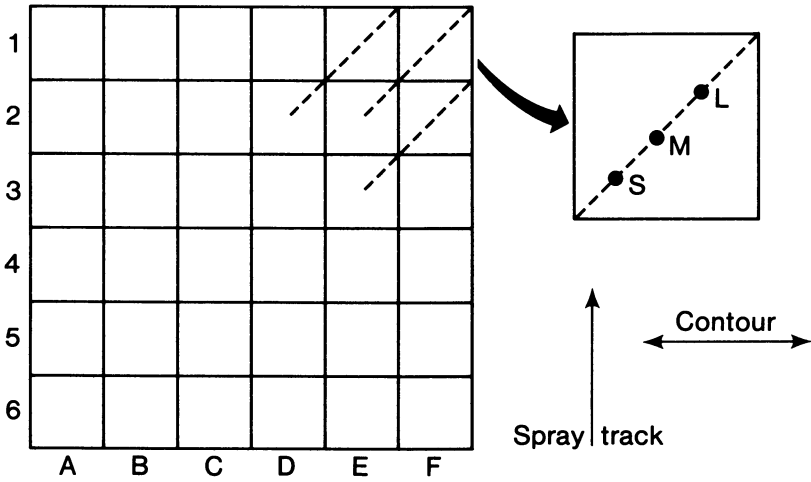
Figure 1. Sampling grid (6×6 m) and location of
individual cores within each grid square. (Large core,
L; medium core, M; small core, S.) Reproduced with
permission from Ref. 26, Copyright 1971, American
Chemical Society.

Table I. Dieldrin Contents (mg/m^2) of Individual Soil Cores
Taken from a Rectangular 6-m^2 Grid of Treated Field Soil.

Row	Core	Column A	B	C	D	E	F	Mean	Std. Dev.
	L	218	353	406	359	50	294		
1	M	51	139	623	127	47	483	250	165
	S	167	315	307	350	34	262		
	L	217	396	215	54	86	89		
2	M	442	456	200	55	39	65	194	141
	S	332	287	244	181	29	106		
	L	263	507	79	22	29	234		
3	M	126	380	110	57	42	222	175	155
	S	145	107	529	117	38	122		
	L	108	152	190	71	48	102		
4	M	143	376	566	79	39	232	174	143
	S	195	251	379	68	39	93		
	L	622	283	125	21	93	276		
5	M	299	385	170	25	59	218	200	155
	S	287	237	175	34	30	335		
	L	115	797	325	16	64	528		
6	M	195	626	212	20	126	160	246	225
	S	153	465	201	28	64	339		
Mean		223	362	281	94	54	231	207	
Std. Dev.		135	172	160	105	25	130		166

Reproduced with permission from Ref. 25, Copyright 1971,
American Chemical Society.

equals 240. This means that if 240 random samples were taken
the true percentage of dieldrin would be expected to fall
within the range ± 0.5 $\bar{x}$ of the average, $\bar{x}$, 50% of the time.
This significant uncertainty for such a large number of samples
reflects the large standard deviation. If only the average
concentration on the field were needed, to save time and effort
it would be worthwhile to combine several sample increments,
mix thoroughly, and analyze subsamples taken from the
composite. This approach would of course sacrifice information
on the dieldrin distribution profile. Note that the relative
standard deviation due to the chemical analysis step, ~5%, is
not a major contributor to the overall uncertainty.
Consideration might be given to a less precise but more rapid
and inexpensive analytical procedure.

    A second method of estimating the number of sample
increments required is by Visman's approach, applying Equation
5. For estimation of the random sampling component, given by
the first term on the right side of Equation 5, relatively
small samples should be chosen to ensure that the contribution
of the first term is much greater than that of the second.
Repeating the study with samples sufficiently large to reverse
the relative contributions of the two terms allows estimation
of the second term, the segregation component of the overall
variance. Unfortunately the variability is so great in this
system that even the largest samples are not large enough to
allow estimation of the spray-pattern variability, that is,
variability caused by a nonrandom component in the
distribution. Thus adequate values for A and B in Equation 5
cannot be calculated with the data at hand.

    As mentioned earlier, Visman (21) suggested that the
constants A, B, w, and n in Equation 5 may also be evaluated by
selecting adjacent pairs of samples and performing a regression
analysis. Application of this method to the dieldrin in soil
systems is illustrated in Tables II and III, where values of
the constants A and B have been calculated by this procedure,
along with the number of samples needed to obtain the
percentage of dieldrin within 50% confidence limits. In Table
II the large cores from sectors 1A and 1B, 1C and 1D, 1E and
1F, 2A and 2B, and so on were paired. In Table III, large
cores from 1A and 2A, 1B and 2B, 3A and 4A, and so on were
taken. Each set contained 18 pairs. With an assumed soil
density of 2, the weight of the large cores is 537 g, and that
of the small cores 123 g. In a friable and relatively
heterogeneous material such as soil the estimation of average
particle mass is arbitrary. For purposes of illustration an
average particle mass of 0.2 g was assumed. Note that bias
produced by the spraying pattern causes pairs oriented parallel
to the spray direction to yield a lower number of required
samples than pairs oriented perpendicular to the pattern.
Random pairing directions are therefore necessary to avoid this
source of bias.

Table II. Calculation by Visman Equation of Number of Samples Required for Determination of Dieldrin with a Sampling Uncertainty of ±50% Relative Standard Deviation in Test Data of Table I, Using Pairs of Large Cores Taken Perpendicular to Spray-Track Direction.

x_1	x_2
218	353
406	359
50	294
217	396
215	54
86	89
263	507
79	22
29	234
108	152
190	71
48	102
622	283
125	21
93	276
115	797
325	16
64	528

$\bar{x}$ 180.72   253.00

s   151.09   213.82

$\bar{d} = \bar{x}_1 - \bar{x}_2 = -72.28$

$s_{pooled} = (151.09^2 + 213.82^2)^{1/2} = 185.13$

r = 0.0569 (from Equation 6)

$$A = \frac{N\,s^2_{pooled}}{rm+\frac{1}{w}} = \frac{(1)(34274)}{(0.0569)(\frac{1}{0.2})+\frac{1}{537}}$$

$$= 1.20 \times 10^5$$

$$B = rAm = (0.0569)(1.197\times10^5)(\frac{1}{0.2})$$

$$= 3.405\times10^4$$

For a sampling uncertainty at the 95% confidence interval equal to 50 mg/m^2

$$s^2_1 = \left(\frac{50}{1.96}\right)^2 = 651$$

For a sample weight of 1 kg and average particle mass of 0.2 g,

$$n = \frac{\frac{A}{w} + B}{s^2_1} = 52$$

For a sample weight of 500 g, n = 53; for 120 g, n = 54

For determination of minimum detectable bias (MDB) at the 95% confidence level:

$$MDB = t_{.95}s_1\sqrt{\frac{2}{n}} = (1.96)(25.5)\sqrt{\frac{2}{52}}$$

$$= 10 \text{ mg/m}^2$$

Table III. Calculation by Visman Equation of Number of
Samples Required for Determination of Dieldrin as in
Table II, but Using Pairs of Large Cores taken Parallel to
Spray-Track Direction.

x_1	x_2
218	217
353	396
406	215
359	54
50	86
294	89
263	108
507	152
79	190
22	71
29	48
234	102
622	115
283	797
125	325
21	16
93	64
276	528

$\bar{x}$ 235.22   198.50

$\bar{d} = \bar{x}_1 - \bar{x}_2 = 36.72$

$s_{pooled} = 147.15$

$r = 0.2033$

$$A = \frac{(1)(147.15)^2}{(0.2033)(\frac{1}{0.2}) + \frac{1}{537}} = 2.126 \times 10^4$$

$$B = (0.2033)(21,260)(\frac{1}{0.2}) = 2.161 \times 10^4$$

After Table A-I, $s_1^2 = (\frac{50}{1.96})^2 = 651$

Then for a sample of 1 kg,

$$n = \frac{\frac{2.126 \times 10^4}{1000} + (2.161 \times 10^4)}{651} = 33$$

For samples of 500 or 120 g, $\underline{n} = 33$
(same as for 1 kg)

$$MDB = (1.96)(25.51)\sqrt{2/33} = 12 \text{ mg/m}^2$$

The sources of variability in this system were interpreted by Taylor et al. (27) as likely coming from a combination of uneven application (overlap or gaps in coverage during the spraying operation) and variable and incomplete mixing of the applied material into the soil. The sampling was performed about two years after the application, so the variability persists for at least that long. Clearly pesticide concentrations in soils depend in a major way on the uniformity of application and the extent of mixing of the applied material into the soil. The first is a segregation effect, and the second a local homogeneity effect. Accurate estimation of the segregation level require a large number of samples. Reduction of the uncertainty due to nonhomogeneity requires that individual sample increments be much larger. A 500 g sample, corresponding to the largest collected in this study, is not sufficient to provide adequate assessment of the heterogeneity at the local level. A sample of 1 kg (or more if better than ± 50 mg/m^2 uncertainty is required) would be advisable. If a subsample were taken for analysis, the collection of larger sample increments would not be too difficult.

The subsampling step would in itself create an additional source of uncertainty, however. In the core study outlined here the entire increment was treated, so a subsampling error was not present. If a larger sample increment were taken it might be necessary to carry out a subsampling operation. In this event the uncertainty associated with the subsampling step could be evaluated separately by running replicate subsamples from the same increment and calculating the standard deviation. If this value is less than one third that of the sampling step it is not a significant contributor to the overall uncertainty. If it is larger the increments should be homogenized further before subsamples are taken.

It can be concluded that for a system as heterogeneous as this example the number and size of the samples must be large if a significant reduction in the sampling uncertainty is to be achieved. The data also reveal the extent of the variability in spray application. An additional conclusion is that reduction in costs and time in the analytical operations may be possible without sacrificing information, since a less precise procedure would suffice under these circumstances.

Glossary of Terms Used in Sampling

Bulk sampling. Sampling of a material that does not consist of discrete, identifiable, constant units, but rather of arbitrary, irregular units.

Gross sample. (Also called bulk sample, lot sample.) One or more increments of material taken from a larger quantity (lot) of material for assay or record purposes.

Homogeneity. The degree to which a property or substance is randomly distributed throughout a material. Homogeneity depends on the size of the units under consideration. Thus a mixture of two minerals may be inhomogeneous at the molecular or atomic level but homogeneous at the particulate level.

Increment. An individual portion of material collected by a single operation of a sampling device, from parts of a lot separated in time or space. Increments may be either tested individually or combined (composited) and tested as a unit.

Individuals. Conceivable constituent parts of the population.

Laboratory sample. A sample, intended for testing or analysis, prepared from a gross sample or otherwise obtained. The laboratory sample must retain the composition of the gross sample. Often reduction in particle size is necessary in the course of reducing the quantity.

Lot. A quantity of bulk material of similar composition whose properties are under study.

Population. A generic term denoting any finite or infinite collection of individual things, objects, or events in the broadest concept; an aggregate determined by some property that distinguishes things that do and do not belong.

Reduction. The process of preparing one or more subsamples from a sample.

Sample. A portion of a population or lot. It may consist of an individual or groups of individuals.

Segment. A specifically demarked portion of a lot, either actual or hypothetical.

Strata. Segments of a lot that may vary with respect to the property under study.

Subsample. A portion taken from a sample. A laboratory sample may be a subsample of a gross sample; similarly, a test portion may be a subsample of a laboratory sample.

Test portion. (Also called specimen, test specimen, test unit, aliquot.) That quantity of material of proper size for measurement of the property of interest. Test portions may be taken from the gross sample directly, but often preliminary operations such as mixing or further reduction in particle size are necessary.

Literature Cited

1. Eberhardt, L.L.; Gilbert, R.O.; Hillister, H.L.; Thomas, J.M. Environ. Sci. Technol. 1976, 10, 917.
2. Ku, H.H. National Bureau of Standards (U.S.) Spec. Publ. 519 April 1979, p. 1.
3. Polzin, W.J.; Brown, I.F., Jr.; Manthey, J.A.; Probst, G.W. Pesticides Monit. J. 1971, 4, 209.

4. Petersen, R.G.; Calvin, L.D. in "Methods of Soil Analysis"; Black, C.A., Ed.; Am. Soc. of Agronomy: Madison, 1965; p. 4.

5. Temple, P.J.; Willis, R. Methodol. Assess. Air Pollut. Eff. Veg. Handb. Spec. Conf., Am. Pub. Health Assoc.: Pittsburgh 1979, paper No. 13.

6. van Dyk, L.P.; Visweswariah, K. Residue Reviews 1975, 55, 91.

7. Kratochvil, B.; Wallace, D.; Taylor, J.K. Anal. Chem. 1984, 56, 113R.

8. "Guidelines on Sampling and Statistical Methodologies for Ambient Pesticide Monitoring," Federal Working Group on Pest Management, Washington, D.C., Oct. 1974.

9. Youden, J.W. J. Assoc. Off. Anal. Chem. 1967, 50, 1007.

10. Whitaker, T.B.; Dickens, J.W.; Monroe, R.J. J. Am. Oil Chem. Soc. 1974, 51, 214.

11. Whitaker, T.B. Pure Appl. Chem. 1977, 49, 1709.

12. Kratochvil, B.; Taylor, J.K. Anal. Chem. 1981, 53(8), 924A.

13. Natrella, M.G., "Experimental Statistics," National Bureau of Standards Handbook 91, U.S. Government Printing Office: Washington, August 1963, pp. 2-13.

14. Cochran, W.G., "Sampling Techniques"; Wiley: New York, 1977, 3rd ed.; Chap. 8.

15. "ASTM E-300 Standard Recommended Practice for Sampling Industrial Chemicals," American Society for Testing and Materials: Philadelphia, 1973 (reapproved 1979).

16. Benedetti-Pichler, A. in "Physical Methods of Chemical Analysis"; Berl, W.M., Ed.; Academic Press: New York, 1956; Vol. 3, p. 183.

17. Harris, W.E.; Kratochvil, B. Anal. Chem. 1974, 46, 313.

18. Dixon, W.J.; Massey, F.J., Jr. "Introduction to Statistical Analysis"; McGraw-Hill: New York, 1969; 3rd ed., Chap. 13.

19. Ingamells, C.O.; Switzer, P. Talanta 1973, 20, 547.

20. Ingamells, C.O. Talanta 1974, 21, 141; 1976, 23, 263.

21. Harrison, S.H.; Zeisler, R. in "NBS Internal Report 80-2164"; Reimann, C.W.; Velapoldi, R.A.; Hagan, L.B.; Taylor, J.K., Eds.; U.S. National Bureau of Standards: Washington, D.C., 1980, p. 66.

22. Visman, J. Mat. Res. Stand. 1969, 9, November, p. 9.

23. "ASTM D-2234-76 Standard Methods for Collection of a Gross Sample of Coal," American Society for Testing and Materials: Philadelphia, 1976.

24. Duncan, A.J. Mat. Res. Stds. 1971, 11, January, p. 25; 1971, 11, August, p. 32.

25. Duncan, A.J. J. Mat. 1971, 7, 345.

26. Snedecor, G.W.; Cochran, W.G. "Statistical Methods"; Iowa State University Press, Ames, IA, 1967; 6th ed.; p. 295.

27. Taylor, A.W.; Freeman, H.P.; Edwards, W.M. Agric. Food Chem. 1971, 19, 832.

Sampling and Variance in Measurements of Trifluralin Disappearance from a Field Soil

A. W. TAYLOR, J. H. CARO, H. P. FREEMAN, and B. C. TURNER

Beltsville Agricultural Research Center, Agricultural Research Service, U.S. Department of Agriculture, Beltsville, MD 20705

Trifluralin was applied to separate field plots either as a spray of emulsifiable concentrate (EC) or by spreading a granular formulation (GF). Both formulations were incorporated and mixed to the 7.5 cm depth. Soil concentrations were measured by eightfold analysis of five field samples obtained from each plot (40 analyses per plot): elaborate precautions were taken to ensure complete mixing and random subsampling in the laboratory. Coefficients of variation were 10% in the EC data and 38% in the GF. Analysis of variance showed that the difference reflected irregular distribution of the herbicide in the laboratory subsamples due to localized high concentrations at granule sites. In later samplings at about 20 day intervals, this irregularity was reduced as the granules became mixed with the soil. Regression analysis indicated a 50% disappearance time of 81 days on the EC plot and 106 days on the GF, but this difference was not significant at the 90% probability level. Regression of all data from both plots gave a 50% disappearance time of 91 days with a 90% probability that the true value lay between 79 and 108 days. The implications of these observations in experimental design and sampling strategies in field studies and environmental work are discussed.

The work reported in this paper was part of a larger project whose principal objective was to measure the rate of volatilization of the herbicide trifluralin to the atmosphere from a field soil after surface application and incorporation to a depth of 7.5cm by cultivation with a disc harrow. In this experiment it was necessary to measure as accurately as possible the amount of the trifluralin residue in the soil on several days during the growing season and to calculate the rate of disappearance of the

residues. Two treatments were compared, one in which the
herbicide was applied as a spray of water-based emulsifiable
concentrate (EC) and a second in which it was spread as a dry
granular formulation (GF). Both plots were otherwise treated and
managed identically.

Earlier investigations of the problems of measurement and
sampling of pesticide residues in field soils (1,2) have shown
that a detailed understanding of the significance of such data
requires a sampling and analytical scheme designed to measure the
variance of the original data and permit estimates of the degree
of confidence that can be placed on the final results. This paper
discusses the statistical analysis and experimental significance
of the results obtained with the sampling scheme described below,
which was initially designed to obtain the maximum information
available that could be obtained with the resources available for
this part of the work. Details of the work on trifluralin
volatilization will be published elsewhere.

Experimental Design and Procedures

Field Layout and Management. The experiment was done on a level
field of about 5 hectares on the Harford Teaching and Research
Center of Cornell University at Harford, N.Y. The soil was a
Howard gravelly loam with essentially uniform characteristics over
the experimental area. The two formulations were applied to
separate rectangular experimental plots of 2.97 hectares each,
(244 x 122m) located side by side with an untreated 3m strip
between them on the long axes.

The granular treatment was applied on May 2nd, 1974 by
spreading Treflan 5G commercial formulation from a conventional
gravity fed spreader at a nominal rate of 0.7 kg/h of active
ingredient. On May 8th, the emulsion-plot received 0.66 kg/h
trifluralin as a spray of Treflan 4EC emulsifiable concentrate
diluted with water according to commercial practice. (Note:
Mention of a proprietary product is for purposes of identification
only and does not constitute endorsement by the U.S. Department of
Agriculture over products of a similar nature not mentioned).
This application was made between 0700 and 0830 a.m. in cool (air
temp., 10°C) clear and windless weather. Both granular and
emulsion formulations were incorporated to the 7.5cm depth with a
disc cultivator within 15 minutes of application. A second
cultivation to the same depth, but at right angles to the first,
was completed within 2-3 hours. On May 20th both plots were
spring-tooth harrowed for seed-bed preparation for soybean
planting on the same day. The only subsequent soil disturbance
was a row cultivation in late June to control a severe infestation
of quack-grass: control of broadleaf weeds was essentially
complete with the trifluralin.

The first samples were taken from the GF plot on May 2nd, 4
hours after application and from the EC plot on May 8th, 7 hours

after application. Subsequent samples were taken from both plots on May 21st, June 25th, July 30th and September 24th. On the GF plot these correspond to sampling times of 0.16, 18.8, 53, 88 and 144 days, and 0.29, 13.3, 48, 83 and 139 days on the EC plot.

Sampling Procedure. A set of twenty-five sampling points was marked out on each plot on a rectangular 5 x 5 grid. The distance between points was thus 40.5m on the long axis of the grid and 20.2m between rows of points. Each sampling point was permanently identified by number.

On each sampling, five field samples were taken from each plot by combining soil samples from each of five sampling points. The five points combined in each sample were selected each day from a table of random numbers. Since a new selection of points was made on each day the individual field samples numbered I through V in the Tables do not represent soil taken from the same points on the field. Each field sample can be regarded as composed cf soil from a randomly chosen set of sampling points that equally represent the surface of the entire field plot.

The sample taken from each point was obtained by combining 16 separate cores, all within an area of 50 x 50 cm and all taken with the same specially designed corer with an internal diameter of 4.4 cm; all cores were taken to 15.0 cm depth. Each field sample thus included soil taken from 80 cores representing a composite block of soil of $0.122 m^2$ in field area and 15.0 cm deep. On removal from the field this sample was weighed and mixed by tumbling for 5 minutes in a power driven portable concrete mixer. The whole sample, usually weighing between 20 and 25kg, was then divided and sub-divided by four passages through a riffler to obtain a representative field subsample of about 2.5kg weight for laboratory analysis. These subsamples were returned to the laboratory within 12 hours and stored at a temperature below $10^{o}C$ until analyzed.

On each sampling day the precise positions from which the cores were taken was marked with a stake. On the next day the cores were taken from a similar area adjacent to the first so that all the five sample sites at each point were in a line along the long axis of the field. This was also the direction in which the sprayer and spreader moved during application.

Analytical Procedure. After thawing, each field sample was immediately divided into 4 quarters and subsamples of approximately 100g of moist soil taken from each for analysis. A single subsample was oven dried at 105 ^{o}C overnight to measure the water content.

Each analytical sample was shaken for 30 minutes in 250ml of analytical grade benezene containing 80ml of isopropanol. After settling, an aliquot of the extractant was washed with distilled water to remove the isopropanol and the trifluralin concentration measured by direct injection into a gas chromatograph equipped

with an electron capture detector. Testing with samples spiked at
the 0.1ppm level indicated that trifluralin recoveries with this
procedure were consistently 95 to 100%. The limit of detection of
trifluralin in soil was less than 10 ppb. Soil concentrations
were never less than 50 times greater than this.

Four replicates from each field sample were analyzed for
trifluralin content, except for those from the first day when
eight replicates were analyzed in order to obtain more extensive
data for the initial statistical analysis. The results obtained
may therefore be classed as a two-stage hierarchical sampling
structure, with four (or eight) analytical determinations on each
of five field samples each of which can be regarded as providing
independent estimate of the trifluralin content in the entire
field soil. Since the amount of trifluralin in each of these
field samples can be regarded as that in a representative
composite block of soil of $0.122m^2$ area, all the data can be
expressed as milligrams of herbicide per square meter of soil area
in the field.

Results

The results of the eight analyses of the five field samples taken
from both field plots on the day of application are presented in
Table I together with means and standard deviations. The values
of the Grand Mean and its standard deviation were calculated from
the sum and variance of the 40 individual samples. The grand mean
of the EC data (Table I) has a coefficient of variation of 10%.
Comparison with earlier experience (1) suggests that this
approaches the best precision that can be expected in field
experiments of this type. Inspection of the standard deviations
of the individual sample means reveals that (expressed as C.V's)
they range from 2.8% to 4.2%, indicating a high degree of
reproducibility in the laboratory subsampling and analytical
procedures. A detailed analysis of variance (3) gives a value of
14.6 as the component of variance between field samples in
comparison with 2.0 for the component between laboratory
subsamples. This shows that in the EC plot the dominant source of
uncertainty is due to variations between the five field samples
and not due to variations in the laboratory subsamples or in the
actual chemical analyses. A similar conclusion can be drawn by
inspection of the standard deviations presented in Table I.

The samples from the plot receiving the granulated formulation
(Table I) revealed a different situation. The overall variability
was much larger with a coefficient of variation of 39%, with
increased variability both between and within the five samples.
Analysis of variance gave components of 140 between field samples,
and 264 between laboratory subsamples. Since the analytical
procedures were identical with those used in the EC plot samples,
where reproducibility was good, these results clearly indicate a
much greater irregularity of the distribution of the herbicide in

the GF plot samples. This irregularity in the individual analyses themselves probably reflects the number of granules present in the 100-gram subsamples withdrawn for analyses. The granular formulation consisted of irregular grains between 0.6 and 1.0mm diameter which probably survived as individual units through the mixing and riffling procedure in the sample preparation.

Table I. Replicate analyses of trifluralin content of field samples of soil receiving emulsifiable concentrate and granular formulation. (Samples taken on day of application.)

Trifluralin Content (mg/m^2)									
Emulsion					Granular				
Sample					Sample				
I	II	III	IV	V	I	II	III	IV	V
37	39	34	35	46	97	46	22	39	52
37	37	34	35	47	61	36	82	50	57
40	41	35	35	44	58	22	46	33	40
37	39	37	35	46	50	17	53	49	48
37	39	37	39	44	85	75	32	37	45
35	37	34	37	44	64	47	48	40	50
35	41	35	37	44	55	35	43	67	39
37	41	36	37	44	107	20	43	44	60
Mean 36.9	39.3	35.3	36.3	44.9	72.1	37.3	46.1	44.9	48.9
S.D. 1.55	1.67	1.28	1.49	1.25	21.3	19.1	17.5	10.7	7.5

Grand Mean and S.D. (of all samples)
              38.5 ± 3.80   49.9 ± 19.4

Analysis of Variance
Component of variance
Between samples   14.6   140
Within samples   2.0   264

In terms of confidence limits the two Grand Means can be written as 38.5 ± 6.4 mg/m^2 for the EC plot and 49.9 ± 32.7 mg/m^2 for the GF plot at the 90% level. This statement emphasizes the extent to which sampling variability can affect the confidence with which an analytical result is known. Unless the sampling program is designed to measure and identify the source of the variability much effort towards improvement of the quality of the chemical analyses can be wasted (4). The difficulty of improving the sampling procedures to reduce the variability is illustrated by calculation of the number of samples that would have to be analyzed to obtain estimates known to have an uncertainty less than 10% at the 90% confidence level (4). This would require 106 analyses from the EC plot and 2140 from the GF. Both sample sizes

were unrealistic in terms of the resources available for the present work. The only viable course was therefore to perform the maximum number of analyses possible and to subject the results to a full statistical analysis to quantify the uncertainty.

Change in Variability With Time

Table II contains the mean values of the quadruplicate analyses of each of the five samples taken from the EC plot on days 13, 48, 83 and 139, together with standard deviations and the results of the analyses of variance. These may be compared with the data from the GF plot given in Table III.

The results in Table II show that, as in the data of Table I, the differences between the five field samples from the EC plot remained the main source of variation in the data throughout the season: this is strikingly evident on the 139th day, where the variance between samples was 10.8 while the average value for the chemical analyses within samples was only 0.4. Thus, although there was a decline in trifluralin content of the samples with time there was no consistent evidence of redistribution of the herbicide to give a more uniform content of the soil, and the field sampling was always the principal source of uncertainty. It may be noted that since the choice of sampling points in the field was re-randomized each day, a particular number (I through V) does not represent the same set of points: temporal changes in mean values (or standard deviations) associated with each sample number do not therefore have physical meaning.

Table II. Mean values (with standard deviations) of four trifluralin analyses on five field samples of soil receiving emulsifiable concentrate, sampled from 13 to 139 days after application.

	Trifluralin Content (mg/m^2)							
Age(days)	13		48		83		139	
Sample	Mean	SD	Mean	SD	Mean	SD	Mean	SD
I	38	1.4	23	1.7	17	1.3	8.3	0.2
II	29	4.2	21	1.0	24	0.5	13.0	0.8
III	27	0.5	16	1.0	23	2.1	7.7	0.6
IV	29	2.6	30	1.4	17	1.4	10.0	0.7
V	33	1.0	32	0.6	2.0	4.3	16.0	0.6
Mean	31.2		24.4		20.2		11.0	
S.D.	4.4		6.6		3.3		3.5	
Variance								
Between		17		41		9.4		10.8
Within		5.6		1.4		5.3		0.4

The data for the samples from the GF plot presented in Table III show a different picture. Although the variability in these samples is higher than that in the EC plots throughout the season, the situation found in the first sampling, where the dominant contribution to the variance was due to the irregular distribution of the herbicide within the laboratory samples did not persist. On the eighteenth day the analysis of variance and the standard deviation presented in Table III show that the between samples variation was dominant and remained so throughout the season. This reflects the breakdown of the trifluralin granules and some redistribution throughout the soil. How far this represents actual herbicide movement in the soil is uncertain because it may merely reflect a decay of the strength or the integrity of individual granules so that they became more readily mixed throughout the bulk of the field soil sample as this was stirred during the vigorous rolling action in the concrete mixer.

The analyses of variance summarized in Tables II and III show that the irregular distribution of the trifluralin contributed to the higher overall variability in the GF plot throughout the experiment, even though the variations between the means of the field samples of both emulsion and granular plots were not dis-similar (coefficients of variation of 32% and 28% respectively) after about 140 days.

Table III.   Mean values (with standard deviations) of four trifluralin analyses on five field samples of soil receiving granular formulation, sampled from 18 to 144 days application.

Age(days)	Trifluralin Content (mg/m^2)							
	18		53		88		144	
Sample	Mean	SD	Mean	SD	Mean	SD	Mean	SD
I	26	4.8	23	5.0	27	1.7	17	5.2
II	56	7.3	43	6.2	30	3.5	22	1.7
III	56	9.7	45	6.2	16	2.5	19	1.4
IV	26	2.6	19	2.4	29	3.0	11	0.8
V	31	6.6	32	1.9	23	2.5	25	1.0
Mean	39.0		32.4		25.0		18.8	
S.D.	15.7		11.6		5.7		5.3	
Variance								
Between		208		134		29		25
Within		51		22		7.3		6.7

Estimation of Trifluralin Disappearance Rate. The trifluralin disappearance rate can be estimated by regression analysis in two ways, either by using the five individual sample means obtained

each day, or from the values of the grand means of all the data
from each day. In practical terms this represents the use of the
results from all five separate samples obtained each day or a
single determination that would have been obtained if all five
were bulked and mixed before analysis. The latter proceedure is
one that could be adopted to reduce the number and cost of
chemical analyses. Statistically the difference between these
approaches is that in the first the final estimate of the
uncertainty in the regression includes the whole variance (except
that due to analytical variation within samples), while the second
reflects only the deviation of the daily grand means from the
regression line. Comparison of the two calculations reveals the
amount of information lost in the second procedure.

Comparison of the regressions based on a simple linear
equation $P = P_i - c.t$ (where P_i is the initial trifluralin
concentration, t is time in days and c is a constant) with
regression upon a logarithmic curve, $\log P = \log P_i - c'.t$,
showed that the latter equation gave slightly better correlation
coefficients in both cases. The data were therefore analyzed in
terms of the logarithmic equation. It must be emphasized that this
does not imply that the trifluralin disappearance followed
first-order kinetics, but only that the data were best described
by the logarithmic equation. The use of more complex exponential
decay curves was explored but no improvement in correlation
coefficients was found. The use of such curves would introduce
unnecessary complexities into the analysis. It may also be noted
that regression analysis using all the data points instead of
those from sample means gave no meaningful improvement.

For simplicity, both sets of data were normalized by
setting the initial trifluralin concentration (P_i) at 100%.
This normalization was done by performing a regression on the raw
data to determine the best value of the intercept at t = 0, and
then expressing all the analyses as a percentage fraction of this

Table IV. Regression and Correlation Coefficients (with Confidence
          Limits) for decrease in trifluralin content of field
          soil for 140 days after application, based upon the
          equalion $\log P = \log (100\%) - c'.t$.

Emulsion	c'	($+$) confidence limits			r
		95%	90%	80%	
Daily means	0.0037	0.0006	0.0005	0.0004	0.992
Sample means	0.0038	0.0008	0.0007	0.0005	0.909
Granular					
Daily means	0.0028	0.0005	0.0004	0.0003	0.988
Sample means	0.0028	0.0010	0.0009	0.0007	0.745

value. These intercepts were 37 mg/m^2 for the emulsion and 44
mg/m^2 for the granular. This normalization reduces the data to
a form where the two sets can be easily compared.

The results of the regressions of the normalized data are
presented in Table IV. In both cases the values of the decay
constant are not affected when the full set of sample means are
substituted for the daily grand means. The values of the
regression coefficients however decrease considerably, reflecting
the larger scatter of the individual sample results. The values
of the confidence limits to be assigned to the decay constant
representing the limits within which the "true" value can be
expected to lie at the 95, 90 and 80% probabilities are also wider
for the individual sample data. These differences illustrate very
clearly how an artificial increase in the confidence to be placed
upon a result can be obtained by the use of averaged figures that
reject inherent variability in the original data. This is
particularly important in the present case because the most
striking feature of Table IV is the difference between the decay
constants of the two formulations. Presented in terms of
"half-lives" these represent 81 days for the emulsion and 106 days
for the granular formulation. If real, the difference of 25 days
represents an important effect of formulation upon the
environmental behavior of trifluralin. It is essential to enquire
further into the confidence that can be placed in this
observation.

An analysis of covariance ($\underline{3}$) using the daily grand means gave
an "F" value of 5.37 with 1/6 degrees of freedom: this indicated
that the "c" values are significantly different at the 90% but not
the 95% probability level. A similar analysis using the sample
means however gives F = 2.307 at 1/47 degrees of freedom,
indicating significance at 75% but not at 90% probability. Since
the analysis using the sample means takes into account the full
variance present in the original data, and includes a far higher
number of degrees of freedom, it represents a much more critical
test. It must be concluded that at the 90% level there is no
significant difference between the two estimates of the
disappearance rates of the two trifluralin formulations.

As a final result, a single regression using the sample means
for both emulsion and granulated plots gives the overall
regression equation log P = 2.00 - 0.0033 ($\pm$ 0.0005).t with
confidence limits of 90% probability: this represents a best
estimate for the half life of 91 days, with a 90% confidence that
the "true" value is between 79 and 108 days. In physical terms
this represents the decay rate of trifluralin in Harford silt loam
during the growing season in a cool and rather wet summer in
upstate New York: since the decay rate will change radically
during the severe winter conditions typical of this location, this
estimate cannot be extrapolated to obtain an estimate of the
disappearance rate over a much more extended time period.

Discussion

The results presented and discussed in this paper are an example
of how irregularities in herbicide distribution in a field soil
cause sampling variations so great that they severely limit our
confidence in how well we can measure the amount of herbicide
present in the field at any time and its rate of disappearance.
Comparison of the results from the two plots in the experiment
reveals that this irregularity of distribution can come from
different causes or combinations of causes. The most uniform, but
far from perfect, distribution was observed after the herbicide
was applied as a spray of water-based emulsion to the soil surface
followed by a double tillage to incorporate the herbicide into the
soil Here the irregularity probably results from irregularities
in the original spray pattern coupled with incomplete and
irregular mixing into the soil (5): other work with persistent
soil insecticides applied in the same way (1) has revealed similar
irregularities in distribution with large variations in
insecticide contents of individual cores taken 35 to 40cm apart.
     The results also show much greater variability in a plot
receiving a surface application of granules 0.6 to 1mm in diameter
followed by incorporation. Even through there was some
improvement with time as the granules decayed and were more easily
mixed throughout the whole soil during sample preparation, the
variability in the data was much higher owing to the superimposed
irregularity of distribution of the herbicide in the laboratory
subsamples.
     These differences emphasize the advantages in sampling
strategies which permit calculation of confidence limits for the
data obtained and the identification of sources of error. With
this information it is then possible to recognize that there is a
limit to how well the herbicide content of the soil in the field
can be measured in practice and that increased effort may only
lead to a better measurement of the uncertainty of the estimate
rather than the number itself. Attempts to increase the precision
of the analytical procedure may also be irrelevant and the most
important issue will be to design a sampling procedure that will
represent the most effective employment of the available
resources. Following the procedure of Snedecor and Cochran (3) it
may be estimated that the variance in the final estimate of the
trifluralin content of the field soil in the first day could be
reduced by about half in the EC plot and by about four in the GF
plot if, on each sampling, twenty-five separate samples were taken
and each analyzed twice. This would represent a considerably
increased cost, increasing the number of determinations from forty
to fifty and, more importantly, increasing five-fold the number of
sampling operations such as mixing, drying and weighing. Whether
this is justifiable depends upon the objectives of the
experiment. In the present case this objective was to estimate
the disappearance rate of trifluralin applied in a regular

commercial manner to a large experimental field. When the results obtained were compared with data from the remainder of the project it was clear that this objective was achieved within the framework of the overall project. Allocation of increased resources to reduce the uncertainty in the final result could not therefore have been justified. If, however, the objective had been to measure the disappearance rate of the trifluralin with the greatest precision possible, it is clear that an entirely different experimental program, with intensive sampling of smaller and more carefully controlled experimental plots would have been appropriate.

Even in more precise measurements it is however important that the sampling program be designed to give an estimate of the confidence that can be placed upon the result obtained. Not only does the recognition and measurement of the existence of variability and uncertainty in field measurements add to their authenticity and value, but it is an essential element in the competent design and management of an experimental program. This becomes important not only in the choice of a sampling protocol, but in the choice of a sampling schedule. Without measurement of the standard deviation resources may be wasted by too frequent samplings if the interval between them is not long enough to allow a statistically measurable change to have taken place. The choice of sampling schedules will of course require some preliminary estimate of the disappearance rate, and sampling at shorter intervals will be essential for more unstable materials. A multiple range test of the results in the present case reveal that the differences between samplings were significant at the 5% level except between those of the 53rd and 88th days on the EC plot. In other work with much more persistent materials (6,7) much longer intervals may be required to find significant differences.

Literature Cited

1. Taylor, A. W.; Freeman, H.P.; Edwards, W. M. J. Agric. Food Chem. 1971, 19, p. 832-6
2. Caro, J. H.; Taylor, A. W. Proc Int. Conf. on Environmental Sensing and Assessment. 1976. Inst. Electrical and Electronic Eng.: Washington, D.C., pt 1-3, pp. 1-5.
3. Snedecor, G. W.; Cochran, W. G. "Statistical Methods"; 6th edit. Iowa State University Press: Ames, Iowa, 1967.
4. Kratochvil, B. The Role of Chemometrics in Pesticide/ Environmental Residue Analytical Determinations. 1984. ACS Symposium Series. (This volume).
5. Read, K.; Gebhardt, M. R.; Day, C. L. Trans. Am Soc. Agric. Eng. 1968, 11, 155-8.
6. Taylor, A. W.; Barrows, H. L. Proc 2nd IUPAC Cong Pesticide Chem. 1971, IV, 457-75.
7. Freeman, H. P.; Taylor, A. W.; Edwards, W. M. J. Ag. Food Chem. 1975, 6, 1101-5.

4

Processing Outliers in Statistical Data

JOHANN A. MÜHLBAUER

Continental Gummiwerke AG, Postfach 169, D-3000 Hannover 21, West Germany

This paper presents a method to decide the handling of seemingly inconsistent data (outliers). The univariate and multivariate methods recommended are strongly based on statistics and the experience of the author in using them.

What does one do with seemingly inconsistent data? Almost everyone concerned with the analysis of experimental data has been confronted at one time or another with this problem. Figure 1a gives a graphical representation of this subject. There is a set of observations or objects of observation which seem to be "inconsistent" with the main body of the data. Such suspicious observations will be referred to as <u>outliers</u> throughout this paper.

Certainly the results of an investigation can be influenced to a high degree by such outlying observations. How does one handle these observations?

Basic Philosophy

There are four main strategies concerning the processing of outliers. Figures 1b to 1e give a graphical interpretation of these strategies.

<u>Rejection.</u> The first strategy is to remove the suspicious datum from the data. Then, the analysis and the conclusions to be drawn are based only on the remaining values. This certainly is the way to deal with outliers which result from human errors, gross errors of measurement or something similar (Figure 1b).

<u>Incorporation.</u> Incorporation of the suspicious observation in the analysis is in some ways the opposite of our first strategy. This

type of action will sometimes result in a totally different view
of our initial problem (Figure 1c).

Concentration.   Neglecting all the nonsuspicious values and con-
centrating the further analysis on the outlying data is a strategy
which is sometimes very useful in process optimization, quality
assurance or archeology (Figure 1d).

Accommodation. The philosophy of this strategy is to include the
outlying observations in the analysis. Methods are then used to
define the final actions which are only slightly influenced by the
presence of outliers (Figure 1e).   Such statistical methods are
developed under the name of "robust statistics."

Influencing Factors

The choice of the strategy to be used depends on the particular
situation.   The choice of the strategy might also depend on the
ability to answer the question: Are the outliers really inconsis-
tent with the remainder of the data?   Unfortunately, not only the
final action but also the method by which we will define whether
or not an outlier is really inconsistent depends on the situation.

      There are several different but interdependent factors which
will influence significantly the whole process of handling out-
liers.   One must consider the distinctions

      - between deterministic and statistical (or rather unknown)
        causes of outliers,
      - between univariate or multivariate data sets, i.e., the
        nature of the data,
      - between different specific probability models like the
        normal or the exponential distribution,
      - between different forms of statistical analysis in which
        the outliers have to be encountered, like ANOVA, random
        sampling and so on,
      - between single or multiple outliers, and,
      - most fundamentally, between the different aims and pur-
        poses that one may have in studying outliers.

The Decision Procedure

Figures 2 to 4 describe the recommended procedure for processing
outliers.   These flowcharts could be used also to create a com-
puter program.   The explanation of some of the terms used in these
charts follows:

Automatic Processing of Standard Data.   The main characteristics
of this procedure are that the

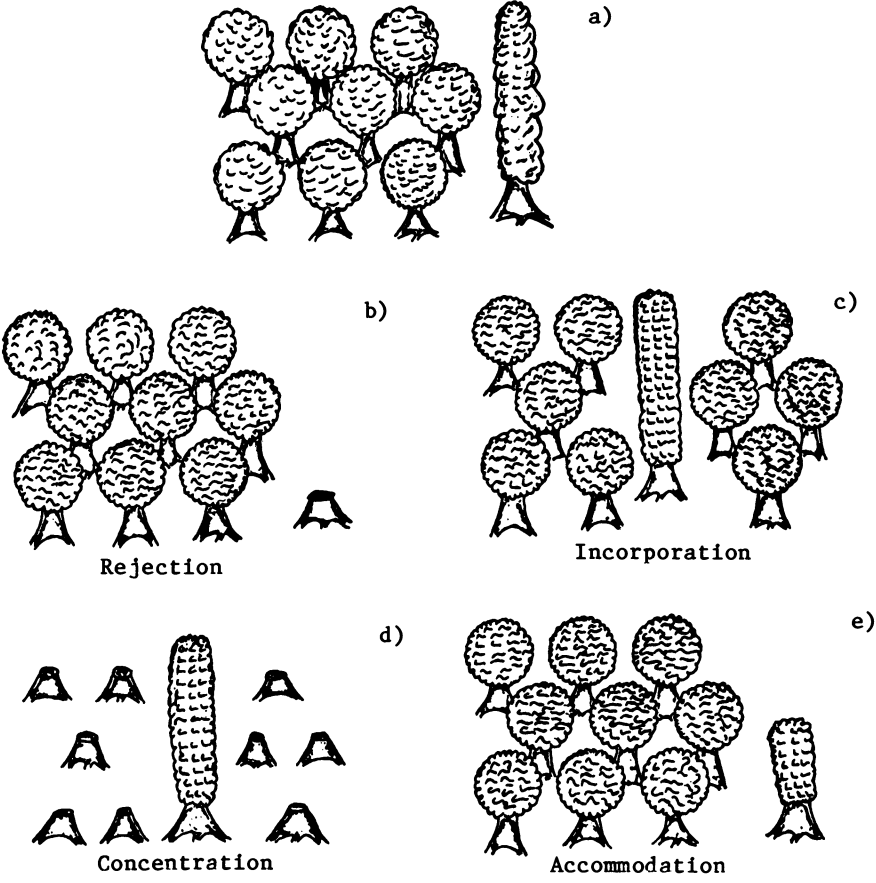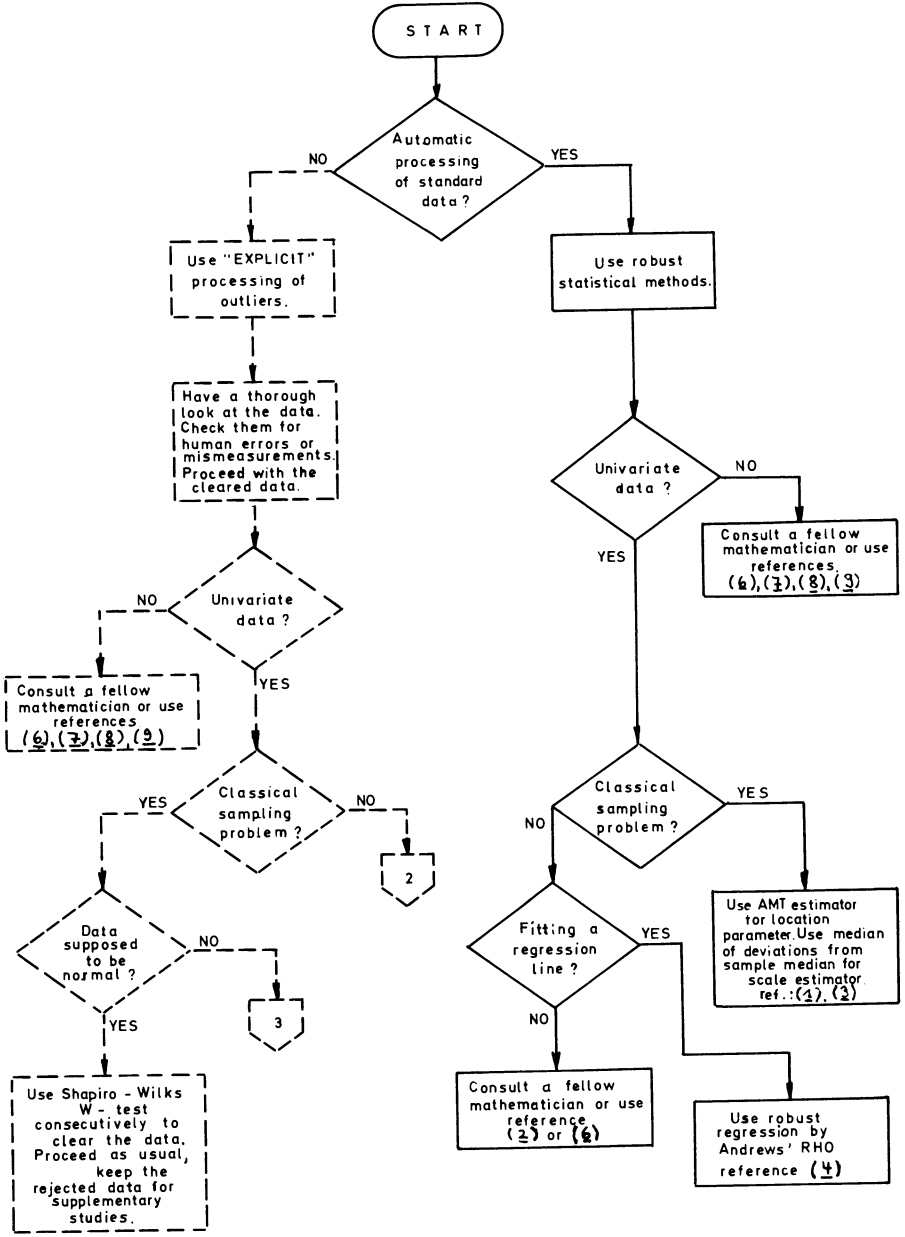   - data is produced and processed routinely without any

a)

b) Rejection

c) Incorporation

d) Concentration

e) Accommodation

Figure 1. Basic concepts
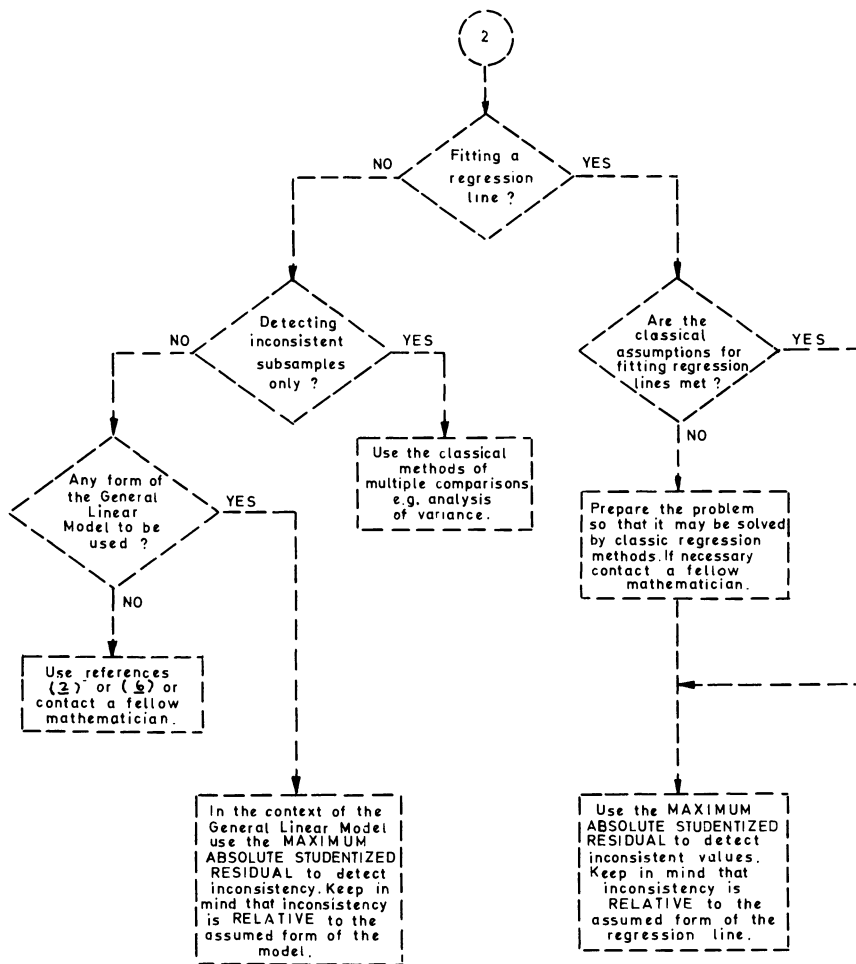
Figure 2. Decision flowchart part 1
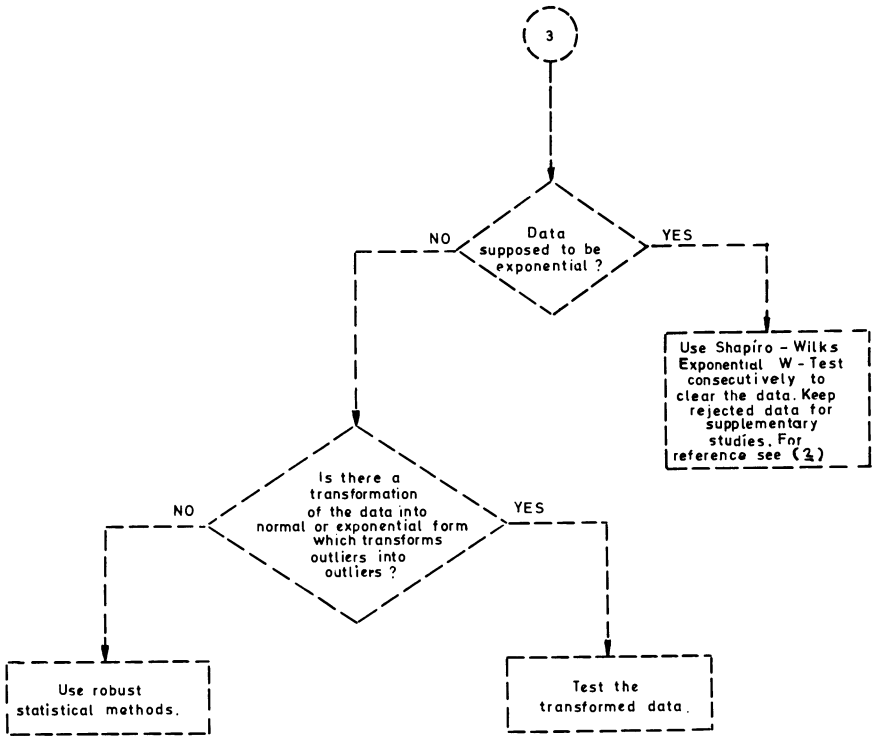
Figure 3. Decision flowchart part 2

```
                                         ( 3 )

                           NO        Data        YES
                                 supposed to be
                                 exponential ?

                                                      Use Shapiro - Wilks
                                                      Exponential  W - Test
                                                       consecutively to
                                                      clear the data. Keep
                                                       rejected data for
                                                        supplementary
                                                         studies. For
                                                      reference see ( 1 )

                              Is there a
                            transformation
                           of the data into
               NO      normal or exponential form    YES
                          which transforms
                           outliers into
                            outliers ?

        Use robust                              Test the
      statistical methods.                   transformed data.
```

Figure 4. Decis.   Flowchart part 3

analysis. The data is not part of any special set or
special environment. Sometimes the data is collected
by automated devices and sometimes by independent
service organizations.
- results of the analysis are given only in a summarized
  manner, such as a mean value, a standard deviation, the
  slope of a regression line, etc.

Univariate Data - Multivariate Data. If one deals only with one
variable under study, e.g., the concentration of a particular
chemical in the water of a river, this is a univariate problem.
It is univariate even when the variable under study depends on
several other variables such as temperature and location of
sampling.

On the contrary, if more then one variable is under study
simultaneously, this would be called a multivariate problem. An
example of a multivariate problem is in determining water quality
using several analyzed variables.

Classical Sampling Problem. If one is only interested in estimat-
ing the location and scatter parameters of a population, this is a
classical sampling problem.

Classical Assumptions for Fitting Regression Lines. The dependent
variable y might be expressed in the following way:

$$y = f(x_1, \ldots, x_n; b_1, \ldots, b_n) + e$$

In this formula, f is a function of the independent variables x_1
to x_n and the unknown parameters b_1 to b_n which is linear in the
parameters. The function

$$f(x_1, x_2; b_1, b_2) = b_1 + b_2 x_2$$

is the classical example, but the function

$$f(x_1, x_2; b_1, b_2) = b_1 \sin(x_1) + b_2 \cos(x_2)$$

is also possible. The error, e, is supposed to be normally
distributed with mean 0 and standard deviation sigma. As a
consequence this means that the various measurements for y are
(stochastically) independent and the associated e's come from an
identical population (they have homoscedasticity or equal variance
over the full range).

Detecting Inconsistent Subsamples Only. A positive response to
this choice results from an analytical problem involving an inter-
laboratory comparison. The main interest is to find those labora-
tories which produce inconsistent results. The results of each

laboratory form a subsample. We can find inconsistent subsamples and, therefore, inconsistent laboratories.

Is There a Transformation of the Data into Normal or Exponential Form? Many data sets are distributed according to probability laws that are not the common normal distribution law. Transformations are possible to convert such data sets to a normal or a nearly normal distribution. It is evident that transforming the data is only appropriate when the original problem, for example, deciding whether two populations are different or not, is not affected by the transformation. Several cases are possible. The following transformation,

$$y = (t+3/8)^{0.5}$$

where t = number of occurrences, will transform Poisson data to normal. This next formation,

$$y = \arcsin [(t+3/8)/(n+3/4)]^{0.5}$$

where n = number of runs, will transform binomial data to nearly normal. Finally,

$$y = \text{arc sinh} [(t+3/8)(n-3/4)]^{0.5}$$

will transform negative binomial data to nearly normal.

Calculation and Processing Procedures; the Processing Flow Chart

There are various methods to process the data which are mentioned in the flow chart. All of them are covered only by citations. You will find the basic references in Table I.

Table I. Mathematical Methods Included in the Flow Chart

Method	Reference Number
AMT-estimator	1
Shapiro Wilks W-test for normal data	2A
Shapiro Wilks W-test for exponential data	2B
Maximum studentized residual	2C
Median of deviations from sample median	3
Andrew's rho for robust regression	4
Classical methods of multiple comparisons	5
Multivariate methods	6-9

<u>Example</u>. For an example of the use of this decision procedure, I will use DATASET D (see Appendix I). The data set is to be used to prepare a calibration graph in chromatographic analysis. It contains a number of excessively high values in the lower levels due to the presence of an overlapping contaminant.

We start at the top of the Decision Flow Chart, part 1, shown in Figure 2.

Decision diamond: Automatic processing of standard data? Since the answer is "NO", the left branch is followed. Instructions are met to have a thorough look at the data. There are several numbers which seem to be inconsistent. However, with no additional data available to this author, I will proceed.

Decision diamond: Univariate data?; "YES"

Decision diamond: Classical sampling problem? As the answer is "NO", I restart at the top of Figure 3.

Decision diamond: Fitting a regression line? "YES"

Decision diamond: Are the classical assumptions for fitting regression lines met? "NO" Clearly the measurements at the different x-levels differ in their variability. This can be shown by using the F-test. Another method is outlined in another chapter of this text (10). In this case weighted least squares will resolve the problem of heteroscedasticity or unequal variance across the graph. I have chosen weights of 1, 1, 0.1, 0.01 and 0.01 for the resolution of this problem.

Table II. Fitting DATASET D Data to the First Order Regression Model, $y = a + bx$

Quantity	Calculated Values		Critical Values	
	t	Max ASR	t	Max ASR
Equation Coefficients (1)				
a = 0.13	0.28		2.10	
b = 26.42	8.66		2.10	
Max ASR (2)		2.29		2.78

(1) Correlation coefficient for the regression fitting is 0.90.
(2) Max ASR occurs at $x = 0.5$, $y = 44.1$.

Decision command: Use the maximum absolute studentized re-
sidual method to detect inconsistent values.

When this method is used, Table II shows the results when the
regression model is the normal first order linear model.   Since
the maximum absolute studentized residual (Max ASR) found, 2.29,
was less than the critical value relative to this model, 2.78, the
conclusion is that there are no inconsistent values.

It is evident that the calculated t-value for the constant
value, a, is less than the critical t-value.   From the statistical
viewpoint this value, then, is negligible.   The data can then be
recalculated according to the first order model without a constant
value.   Table III shows the result of this recalculation.   There
are no changes relating to the conclusions made concerning the
author determination.

Three critical points can be made in this analysis.   The
first one is located at the "thorough look" instruction.   This
examination in reality involves a critical analysis of the experi-
mental protocol and the data produced from it.   For example, it
was quite evident in collecting the standards data from DATASET D
that values were well out of line with previous determinations.
See other DATASETS, especially DATASET E in the Appendix, for
confirmation of this idea.   The second critical point is at the
"Preparation of the problem" instruction.   In this case hetero-
scedasticity must be removed before submitting the data to regres-
sion analysis.   Weighted least squares of several types (11) and
power transformations (10) can be used.   The third critical point

Table III.   Fitting DATASET D Data to the First Order Regression
Model,   y = bx

Quantity	Calculated Values		Critical Values	
	t	Max ASR	t	Max ASR
Equation Coefficients (1)				
b = 27.14	17.2		2.09	
Max ASR (2)		2.36		2.78

(1) Correlation coefficient for the regression fitting is 0.97.
(2) Max ASR occurs at x = 0.5, y = 44.1.

is at the same instruction and is the decision of the regression
model used for the calibration graph. First order, higher order,
and spline (12) methods can all be used for this model. All these
choices will significantly influence the decision concerning the
reality of inconsistant values.

Literature Cited

1. Andrews, D. F.; Bickel, P. J.; Hampel, F. R.; Huber, P. J.;
    Rogers, W. J.; Tukey, J. W. "Robust Estimates of Location:
    Survey and Advances"; Princeton Univ. Press: Princeton, NJ,
    1972; pp. 5, 15, 39-44, others.

2. Barnett, V.; Lewis, T. "Outliers in Statistical Data";
    Wiley:New York, 1978; pp. (A) 89-103, (B) 76-88, (C) 234-265.

3. Huber, P. J. "Robust Statistics"; Wiley: New York, 1981; pp.
    107-109.

4. Lawson, J. S. J. Quality Technology 1982, 14, 19-33.

5. Miller, R. G. "Simultaneous Statistical Inference"; McGraw-
    Hill: New York, 1966; p. 98.

6. Beckman, R. J.; Cook, R. D. Technometrics 1983, 25, 119-163.

7. Campbell, N. A. Applied Statistics 1980, 29, 231-237.

8. Maronna, R. A. The Annals of Statistics 1976, 4, 51-67.

9. Schwager, S. J.; Margolin, B. H. The Annals of Statistics
    1982, 10, 943-954.

10. Kurtz, D. A.; Rosenberger, J. R.; Tamayo, G., Chapter 9 in
    this book.

11. Mitchell, D. G., Chapter 8 in this book.

12. Wegscheider, W., Chapter 10 in this book.

5

The Many Dimensions of Detection in Chemical Analysis

with Special Emphasis on the One-Dimensional Calibration Curve

LLOYD A. CURRIE

Center for Analytical Chemistry, National Bureau of Standards, Washington, DC 20234

Simple detection decisions generally involve the comparison of scalar quantities (gross signal, blank). Conventional chromatography and spectrometry, on the other hand, involve one-dimensional variables (time, mass, wavelength, energy) where signal and baseline traces may be examined to decide whether a peak is present at a given location. Linked techniques, such as GC-MS or two-parameter nuclear spectroscopy, raise the question of detection in two dimensions. Finally, problems wherein a set of samples is characterized by many independent chemical and physical observations raise the issue of multidimensional detection. A unified approach for all such problems is given by the statistical theory of hypothesis testing. Following a brief review of underlying assumptions and techniques for applying the theory to detection decisions and detection limits, primary attention is given to a one-dimensional (reduced from two) problem involving the calibration curve and the pesticide, Fenvalerate. Other topics addressed include information-loss through faulty reporting (at trace levels) and its impact on regulatory issues, and chemometric quality assurance through standard interlaboratory test data sets.

One of the fundamental performance characteristics of any analytical procedure is the Limit of Detection. Just as with the imprecision (standard deviation), with which it is intimately connected, the Detection Limit (L_D) is undefined unless there

exists a fully-specified Chemical Measurement Process (CMP) in a
state of complete control. When these requirements are met, it
is convenient to define L_D in accordance with the statistical
theory of Hypothesis Testing ([1,2]). Although this theory is
well established there continues to be a great diversity of
terminology and formulations which generate needless confusion
in our discipline. Still more serious is interdisciplinary
confusion, when analysts are called upon to provide valid
methods and critical data for regulatory, clinical, or environ-
mental decision-making ([3]). The objectives of this review will
be to summarize the basic concepts of detection in Analytical
Chemistry, with the development following a stepwise increase in
dimensionality. Prime emphasis is given to the assumptions
which must be met, and to illustrations having differing
dimensions. In keeping with the Symposium title and in response
to the invitation of the Symposium organizer, a detailed
exposition is presented for the trace detection of a pesticide
(Fenvalerate) by gas chromatography -- an exercise which
highlights the relationship of the calibration process to the
detection characteristic, and which exposed a surprising (and
unnecessary) limitation to the detection capability.

Treatment of a real, imperfect calibration data set
revealed the full complexity and breadth of the calibration
curve -detection limit problem, ranging from varying statistical
weights to an uncertain model and data containing possible
blunders to an artificially imposed response threshold.
Attempts to simplify an actually complicated situation were
rejected in favor of a full exposition including an Appendix
containing worked-out numerical examples.

SIMPLE HYPOTHESIS TESTING - SCALAR SIGNALS

The basic detection concepts can be presented for the "zero-
dimensional" case where detection decisions and detection limits
are established simply from the characteristics of the chemical
signal (instrument response), without giving detailed attention
to other dimensions such as time, wavelength, analyte concentra-
tion, etc. Actually, higher dimensional situations (multipara-
meter separations or detector responses) reduce to this case
either through sequential classification schemes or via algo-
rithms which operate directly on the multidimensional data.

Our basic task is to distinguish the blank or background
(H_0, null hypothesis), from a signal at the detection limit (H_1,
alternative hypothesis). A straightforward probabilistic
formulation can be given provided that the observed signals
(arising from an underlying "true" signal) are random, independ-
ent and stationary. To completely specify the false positive
(α) and false negative (β) risks, we must know the form of the
distribution and its parameters. For most analytical situations

we assume the distribution to be normal (Gaussian), and the dispersion parameter is simply the imprecision (standard deviation, σ).

As shown in Reference (2), it is sufficient to have an estimate of the blank (B) and its standard deviation (σ_B) plus the variation of σ_y with the signal magnitude (y) to specify a decision criterion or level (L_C) given α, and a detection limit (L_D) given L_C and β. (See especially Figure 2 in Reference (2)). If σ_y is independent of signal magnitude (at and below the detection limit), and if y is normally distributed, one concludes that

$$L_C = z_{1-\alpha}\sigma_0 \qquad\qquad (1a)$$

$$L_D = L_C + z_{1-\beta}\sigma_0 \qquad\qquad (1b)$$

where $z_{1-\alpha}$, $z_{1-\beta}$ are the one-sided normal standard percentiles, and $\sigma_0 = [\sigma_y^2 + \sigma_B^2]^{1/2} = \sigma_B \sqrt{2}$ for "paired" observations. (When the net signal equals zero, $\sigma_y = \sigma_B$.) If 5% risks are acceptable, then $L_C = 1.645\sigma_0$, and $L_D = 2L_C$. A large literature exists giving refinements to the above exposition -- for example, Student's-t with the appropriate number of degrees of freedom should replace $z_{1-\alpha}$ when σ is estimated by replication; and interference will increase the magnitude of σ_0 -- but the basic concepts remain unchanged.

One source of confusion deserves emphasis. That is, L_C is a threshold or critical point (decision level) against which an observed net signal (y-B) should be compared to judge significance. If y-B > L_C, one may conclude that the false positive risk in making the decision "detected" does not exceed α. On the other hand, L_D is a performance characteristic of the CMP, which marks the true (net) signal which will be detected with probability (1-β), given L_C (or α). Although L_D is crucial for specifying the detection capability of the CMP, it is often ignored in favor of the decision level, L_C. (It has been implied that "definitions... concerned with errors of the second type [β] are [too] complex" for common understanding and use (4)!)

If the calibration curve, y = f(x), is exactly known (and monotonic) then the signal detection limit can be transformed directly into the concentration or analyte detection limit, x_D. Alternatively, if a calibration factor is given within certain systematic error bounds, one can specify a corresponding uncertainty range for the concentration detection limit. There is no need to transform L_C in this case, for it is simply a test level for comparison of the observed signal with that arising from the blank. (Note that, even though the magnitude of the concentration detection limit is uncertain, a quite definite value exists which would yield signals exceeding L_C with a probability of exactly 1-β.)

[A note concerning terminology: L_D (Ref. 2) and S_D have been used interchangeably to denote the detection limit for the net signal (y-B); x_D is used here to denote the analyte detection limit (concentration or amount). L_C (or S_C or x_C) denotes the decision level; it is also called the critical point or level, test level, or threshold by various authors. The directly observed gross signal (y) is here referred to as the response.]

THE ANALYTICAL MODEL, ASSUMPTIONS, AND PRACTICAL IMPLICATIONS

Before stepping through the several dimensions, it is worthwhile to examine the general analytical model which applies and, through that, consider the implications of the necessary assumptions in practical applications. To begin, let us express the observed response (y) and its error (e) in terms of the blank (B) and concentrations of all contributing analytes (x_j).

$$y_i = f(B_i, x_j) + e_i \qquad (2a)$$

Here, we represent the one-dimensional situation (y is a single-subscripted vector) so that more than one component-j may be considered. Estimated concentrations $\widehat{x}_j$ are then

$$\widehat{x}_j = \underline{Op}(y_i) = x_j + e_j \qquad (3a)$$

where $\underline{Op}$ represents an operator (or algorithm) applied to the data vector. It should be noted in passing that alternative, valid operators generally may be applied to the same data set. Such Op's differ primarily in efficiency and robustness. A simple example, for isolated chromatographic peaks, would be: O_1 ~ peak area estimator, O_2 ~ peak height estimator, O_3 ~ a digital filter. Obviously, the standard error (and detection limit) depends upon the algorithm selected.

In Equation 3a, we take e_j to be random, with zero mean and known distribution, in order to apply the probabilistic theory of hypothesis testing. Selection of the operator and the nature of e_j are governed by (our perceptions of) the structure of Equation 2. Assumptions concerning $\underline{f}$ and e_i are crucial. In the best of circumstances $\underline{f}$ is linear (in the x_j and B) and e_i is normal, independent and unbiased. Then,

$$y_i = B_i + \sum_{j=1}^{P} A_{ij} x_j + e_i \qquad (2b)$$

(P equals the number of components, and e_i is normal with standard deviation σ_i.) Then, $\underline{Op}$ can be a linear operator (on the y_i), such as that associated with linear least squares estimation (or, for non-interfering peaks, O_1-O_3) -- and the

e_j's have the desired properties (random, unbiased, normal). Note, however, that unlike the e_i's, the e_j's are not generally independent. If the model is correct and σ_i known, the σ_j can be computed by error propagation, or Monte Carlo trials for a complex model. If σ_i is unknown (but constant), and d.f. (degrees of freedom) > 0, then the dispersion about the fitted values $(\hat{y}_i)$ yields an estimate for σ^2. In every case, independent experiments and replicates are vital for external validation of the presumed σ_j's -- e.g., with the aid of the Analysis of Variance.

Deviations from the Ideal Model. In all real situations, the error terms have a structure qualitatively represented by Equation 4,

$$e = \delta + \Delta + h(t) + b \tag{4}$$

containing the desired random error term δ, a systematic error term Δ, a non-stationary or lack-of-control term $h(t)$, and a blunder or mistake (or outlier) term b. If δ is non-normal, but of known distribution, reliable confidence intervals and detection limits may still be set; but the non-random terms in Equation 4 make it impossible to treat detection from a strictly probabilistic viewpoint -- at best one may simply set approximate bounds based on very carefully considered "scientific judgment".

   A complete exposition on the assumptions which must be satisfied to assure the validity of (statistical) hypothesis testing is beyond the scope of our discussion, as are the implications of estimated (vs. known) variances and relative (statistical) weights, but a brief summary, framed in the context of Equations 2-4, is given in Table I. (See Reference (5) for further details.)

   One final point deserves emphasis. The astute reader may already have considered a "higher" level of detection -- i.e., detection of deviations from the ideal assumptions via appropriate test statistics (χ^2, F, ...) or independent measurements. Unfortunately, unless there are severe deviations or large numbers of degrees of freedom, such tests generally lack sufficient power to signal significant (unanticipated) error. For example, a) in (radioactivity) counting experiments a non-Poisson random error component, equal in magnitude (variance) to the Poisson component, will not be detected until there are 46 degrees of freedom (6), and b) it was necessary for a minor component in a mixed γ-ray spectrum to exceed its detection limit by ~50%, before its absence was detected by lack-of-fit (χ^2, model error) (7).

Simple Analyte Detection and Estimation. Using the simplest form for Equation 2b

Table I. Possible Erroneous Assumptions

Nature	Consequence
Lack of control; blunders	Measurement process non-existent
Model error	
$f(B,x) \neq B + \Sigma Ax$	
P ~ wrong	Biased results
A_{ij} ~ wrong	
B_i ~ wrong[a]	
Error distribution (e_i)	
Δ, systematic uncertainty	Non-random uncertainty bounds on results ($\hat{x}_j$)
e_i, non-normal, non-independent	Non-normal e_j
Non-linear operator (Op)	Non-normal e_j
A_{ij} ~ random error[a]	
Faulty reporting[a]	Information loss

[a]Further discussion appears in the text.

$$y_i = B_i + A_i x + e_i \qquad (2c)$$

we can consider the matter of the detection limit for four important, common situations. (See the following section for detailed mathematical development, and the Appendix for worked-out numerical examples.)

Case I - Signal Detection (A ~ ignored). $(y-\hat{B})$ gives an estimate of the net signal (Ax), and the signal detection limit is determined from $\sigma_y^2 + \sigma_B^2$ as discussed in the first section of this article. Although a definite analyte concentration (x_D) corresponds to the signal detection limit, its magnitude is unknown in the absence of information concerning the slope (A) of the calibration curve. Since the operation $(y-\hat{B})$ is linear, the estimated net signal is normally distributed, and percentiles of the normal standard variate (z) are appropriate for calculating the decision and detection limits. It should be noted that signal detection alone is generally of little interest if there may be two or more contributing analytes.

Case II - Analyte Detection (A ~ assumed). Here, the analyte- rather than signal-detection limit is calculated, but the systematic error in A, applied in the estimation of x from Equation 2c imposes systematic error bounds which must be applied to the analyte detection limit. The limit is no longer purely probabilistic in nature (8).

Case III - Analyte Detection ($\hat{A}$ ~ random). One of the most reliable modes of simple (single) analyte detection obtains when y, B and A are each measured (observed) for every sample processed. Such is often the case, for example, in radiocarbon dating where the age of each unknown artifact is estimated from sequential measurements of the sample, the background, and the radiocarbon dating standard. Thus,

$$\hat{x} = \frac{y-\hat{B}}{\hat{A}} = x + e_x \qquad (3b)$$

The beauty of this completely random approach to the analyte detection limit is the direct applicability of the statistical hypothesis testing formalism. Also, long-term trends in calibration slope or backgrounds have little influence. One important assumption is made: that the form of the calibration curve [Equation 2c] is fixed. Also, a subtle change has occurred, the operation is no longer linear, with $\hat{A}$ in the denominator. Thus, the distribution of $\hat{x}$ is only asymptotically normal, as the relative standard deviation of $\hat{A}$ becomes smaller.

Case IV - Analyte Detection (calibration curve determination). If a series of known concentration samples is measured, one can estimate (or verify) both the form and the parameters

(B, A) for the calibration curve. Detection under this circum-
stance is the principal topic of the next major section treating
the detection of the pesticide Fenvalerate by gas chromatography
(GC). Some subtle issues will be noted here, however.

 (a) If the calibration curve is determined once, and then
its estimated parameters and form repeatedly applied to a series
of samples, we are faced with the mixture of random and system-
atic error again, as in Case II above. Errors in $\hat{A}$ and $\hat{B}$ which
were once random become systematic with respect to the series of
samples, and the probabilistic detection limit has imposed upon
it systematic error bounds.

 (b) If the calibration curve is redetermined for each
sample, we have the analogue of Case III, where $\hat{x}$ is indeed
random, but not necessarily quite normal. Deviations from
normality here, and effects of systematic error bounds [above
-(a)] may be quite negligible if the calibration curve is stable
and its measurement very precise compared to subsequent sample
measurements. As noted in the next section, however, a poor
design (concentrations of known samples) can make the analyte
detection limit non-existent for any sample concentration!

 (c) The effects of assumptions concerning B_i Equation 2c
can be profound. In the case of calibration, B is generally
assumed constant, independent of analyte concentration, and
there may often be two independent estimates for B: that
derived from a "blank sample", and that corresponding to the
intercept of the calibration curve. Discrepancies between these
two estimates can sometimes be very large, and that implies
serious questions both about the presumed measurement process
(model) and its detection limit (9). Two illustrations where
assumed constancy of B has been shown seriously in error have
been given by Patterson (10) who stressed the importance of
correcting reagent blanks for chemical yield, and Scales (11)
who noted the differences in liquid scintillation quenching
behavior of background and analyte radiations. Finally, we note
that $\hat{B}$ and $\hat{A}$ derived from the same calibration sample set are
necessarily correlated, in contrast to independent observations
of the slope and the blank [Case III].

Reporting, Regulations, and Practical Implications. Diverse
reporting practices for small or marginally detectable signals,
particularly when related to regulatory issues or matters of
public safety, lead to much needless confusion and loss of
information (3). Specific practical illustrations are just now
being identified in an investigation of radioactivity detection
capabilities in the nuclear industry. For example, we find that
many nuclear laboratories are making detection decisions at the
detection limit, such that $L_C = L_D$, or $\beta = 0.50$ (50% chance of a
false negative). Following such decisions the experimental
results are often quoted as upper limits, "ND", or even zero!
(The mode of reporting will sometimes be adjusted according to

the audience.) For example, it seems universally assumed that
the public cannot comprehend estimated results which carry
negative signs, even though the associated uncertainty greatly
exceeds the absolute value of the negative number.

Consequences of such varied practices are that results from
different organizations cannot be intercompared, and proper
averages (e.g., quarterly) cannot be derived even from the
results of a given organization.

Other practices which tend to underestimate the true
detection limits and add confusion to the uniform evaluation of
results by the public include: varied (or no) treatment of
interference, avoidance of systematic error bound estimation,
and consideration of Poisson counting errors only. A further
problem which has emerged with the prevalence of microprocessors
and proprietary computer software, is the effect of hidden
algorithms and inaccessible source code, so that data evaluation
operations (Op) are not known to the user, and possible source
code deficiences and blunders cannot be readily assessed.

Actually, the situation is not quite so bleak as suggested
above. Many laboratories are addressing the issues of detection
decisions, detection limits and reporting with considerable
care. Also, improvement will come about, as a result of the
current recognition of the above problems by regulatory agencies
such as the Food and Drug Administration and the Nuclear
Regulatory Commission (12). Education, within our discipline
and externally, holds the greatest hope. Clear understanding of
the concepts of measurement uncertainty and detection limits
will discourage the further introduction of views such as (some
interpretations of) the Delaney Amendment to the Food, Drug and
Cosmetic Act, which suggests that chemists should be able to
demonstrate the complete absence of a given analyte (e.g., a
pesticide) in a given sample (13).

PESTICIDE DETECTION - ONE DIMENSIONAL CALIBRATION CURVE

Detection, and the GC calibration process for the pesticide
Fenvalerate constitutes a major focus of this paper. Though
relatively simple compared to possible higher dimensional cases,
it is representative of a large and basic class of problems
(calibration in one-dimensional chromatography and spectros-
copy), and it exhibits already some of the difficulties with
assumption validity.

To gain some perspective on the problem, it is well to
realize that assessment of the detection limit is subject to all
of the assumptions and restrictions of the estimation process.
That is, the functional and error structure of the calibration
curve must be known or assumed, and the respective parameters
and their uncertainties must be estimated. Although large

concentration behavior is less relevant, the estimation and
variability of the blank (and interference) is crucial, as is
the design of the calibration process.

In the text which follows we shall examine in numerical
detail the decision levels and detection limits for the Fenval-
erate calibration data set ("set-B") provided by D. Kurtz ($\underline{17}$).
In order to calculate said detection limits it was necessary to
assign and fit models both to the variance as a function of
concentration and the response ($\underline{i.e.}$, calibration curve) as a
function of concentration. No simple model (2, 3 parameter) was
found that was consistent with the empirical calibration curve
and the replication error, so several alternative simple
functions were used to illustrate the approach for calibration
curve detection limits. A more appropriate treatment would
require a new design including real blanks and Fenvalerate
standards spanning the region from zero to a few times the
detection limit. Detailed calculations are given in the
Appendix and summarized in Table V.

<u>Decision and Detection -- Linear Calibration Curves</u>. Before
examining the actual Fenvalerate GC data, let us consider the
basic linear calibration relations. (What follows was inspired
in part by Hubaux and Vos ($\underline{14}$), to which the reader might refer
for supplemental detail.) If we represent a straight-line
calibration as

$$y = B + Ax + e \tag{2c}$$

where, A represents the slope of the line and e, a random normal
error (y, B, x - as defined previously) then the estimated
concentration is

$$\hat{x} = (y - \hat{B})/\hat{A} = \hat{S}/\hat{A} \tag{3b}$$

where $\hat{S}$ represents the estimated net signal. In principle, and
ideally, the variance of $\hat{x}$ ($V_x = \sigma_x^2$) could be obtained by full
replication as in Case III in the preceding section. Alterna-
tively, it may be estimated following classic error-propagation
formulas ($\underline{15}$) as

$$V_x \approx \frac{1}{A^2} [V_y(x) + V_B + x^2 V_A + 2x V_{AB}] = \frac{V(y-\hat{y})}{A^2} \tag{5}$$

By setting x to zero in Equation 5, x_C (the decision level) may
be computed -- $\underline{i.e.}$,

$$x_C = z_{1-\alpha}\sqrt{V_0} \quad (V_x \text{ at } x = 0) \tag{6a}$$

The detection limit x_D must be determined by iteration:

$$x_D = x_C + z_{1-\beta}\sqrt{V_D} \quad (V_x \text{ at } x = x_D) \qquad (6b)$$

but such iteration takes place quickly since V_x changes relatively slowly below the detection limit, and $x_D \approx 2x_C$ when $\alpha = \beta$.

If the standard deviation of the observed response σ_y is constant and if $\alpha = \beta$, then an analytic solution can be given as follows. From Equations 5 and 6,

$$(x_D - x_C)^2 = \frac{z^2}{A^2} (V_y + V_B + x_D^2 V_A + 2x_D V_{BA})$$

By substituting $x_C = z\sigma_0/A = z(V_y + V_B)^{1/2}/A$ and rearranging, we obtain

$$x_D = 2x_C(K/I) \qquad (7)$$

where: $K = 1 + \rho_{BA}(\sigma_B/\sigma_0)(z\phi_A)$; and $I = 1 -(z\phi_A)^2$

[In the expression for K, V_{BA} has been replaced by its equivalent $\rho_{BA}\sigma_B\sigma_A$, and ϕ_A represents the relative standard deviation (RSD) of A -- i.e., $\phi_A = \sigma_A/A$.]

Important consequences follow immediately from the form of Equation 7. First, if independent estimates $\hat{B}$, $\hat{A}$ are obtained, the correlation coefficient ρ_{BA} is necessarily zero; so K = 1. Second, if $\phi_A \ll 1$ (precise $\hat{A}$ or A-known), both K and I = 1; so $x_D = 2 x_C$. Third, since ρ_{BA} is negative and $|\rho_{BA}|<1$, and $\sigma_B \lesssim \sigma_0$, the ratio K/I may be written as $[1-\epsilon(z\phi_A)]/[1-(z\phi_A)(z\phi_A)]$ where $0 \leq \epsilon < 1$. Thus, there can be no analyte detection limit ($x_D \to \infty$) if

$$\phi_A > 1/z = 1/1.645 = 0.608$$

Also $x_D = 2x_C$ if $\epsilon = |\rho_{BA}|\sigma_B/\sigma_0 \to z\phi_A$, and in the limit, $\phi_A \to 0$. If $\epsilon < z\phi_A$, $x_D > 2x_C$ and the converse. The minimum in the ratio x_D/x_C occurs when the design $\{x_i\}$ is such that $\bar{x} = x_D$ (see Table II and Equation 10).

More generally, the calibration curve can be represented as a matrix equation

$$y = M\theta + e \qquad (8)$$

whose weighted least-squares solution is

$$\hat{\theta} = (M^T W M)^{-1} M^T W y \qquad (9a)$$

$$V_\theta = (M^T W M)^{-1} \approx (M^T M)^{-1} V_y \qquad (9b)$$

The approximation for the variance of $\hat{\theta}$ in (9b) holding when V_y is approximately constant. A summary of the application of Equation 9 to the linear calibration curve derived from known analyte concentrations $\{x\} = (x_1, x_2 \ldots x_n)$ and corresponding statistical weights (inverse variances) is given in Table II.

Table II.   Decision and Detection (weighted least squares)

$$(y = M\theta = B + Ax)$$

$$M^T = \begin{pmatrix} 1 & 1 & \ldots & 1 \\ x_1 & x_2 & \ldots & x_n \end{pmatrix} \qquad\qquad \theta^T = (B\ A)$$

$$M^T W M = \begin{pmatrix} \Sigma w & \Sigma wx \\ \Sigma wx & \Sigma wx^2 \end{pmatrix},$$

where $w_i = 1/V_y(x_i)$
$[k/V_y$, if k-replicates$]$

Then,

$$V_x \approx \frac{1}{A^2} \left[V_y(x) + \frac{1}{\Sigma w} + \frac{(x - \bar{x}_w)^2}{\Sigma w(x - \bar{x}_w)^2} \right]$$

Decision: $x = 0$ Detection: $x = x_D$

Expressions for the variances and covariance of $\hat{B}$ and $\hat{A}$ follow from the inverse matrix $(M^T W M)^{-1}$. See the discussion of "case-f" from Table V in the Appendix, for explicit formulas.

(Note that $\bar{x}_w$ represents the weighted mean of the $\{x\}$.) Given the defining expressions for decision and detection limits together with the calibration design $\{x\}$, the equation for V_x in Table II immediately yields the desired quantities for the linear calibration curve. For equal weights (V_y = const.) and taking roots, the expression simplifies to

$$\sigma_x \approx \frac{\sigma_y}{A} \left[1 + \frac{1}{n} + \frac{(x - \bar{x})^2}{\Sigma(x-\bar{x})^2} \right]^{1/2} \qquad (10)$$

Non-linear curves may be treated using Equation 9 directly, using the techniques of non-linear least squares, when appropriate. (Note that a non-linear calibration curve does not necessarily imply non-linear least squares. The latter is necessary only if the problem is non-linear in the estimated parameters (16). For example, $y = a+bx+cx^2$ and $y = a+bx^c$ are both non-linear functions, but only the latter is non-linear in the parameters.)

Fenvalerate Data. Calibration data for the GC measurement of Fenvalerate were furnished by D. Kurtz (17). Average responses for five replicates at each of five standard concentrations are given in Table III. It should be noted that the stated responses are not raw observations, but rather on-line computer generated peak area estimates (cm^2). (Had we started with the raw data [chromatograms], the problem would actually have been two-dimensional, including as variables retention time and concentration.) The stated uncertainties in the peak areas are based on a linear fit ($\sigma = a+bx$) of the replication standard deviations to concentration; and the "local slopes" [first differences] in the last column of Table III are presented

Table III. Fenvalerate (GC) Data - Set B

(averages of 5 replicates)

Response ($\bar{y}$, cm^2)[a]	Amount (x, ng)	$\Delta\bar{y}/\Delta x$
1.18 ± 0.02$_4$ [0.023]	0.05	23.6
7.08 ± 0.06$_8$ [0.365]	0.25	29.5
29.68 ± 0.23 [0.18]	1.00	30.1
209.0 ± 1.1$_1$ [1.87]	5.00	44.8
920.6 ± 4.4$_0$ [4.32]	20.00	47.4

[a]Uncertainties represent standard errors, based on the fitted equation $\sigma(y) = (0.028 + 0.49\ x)/\sqrt{5}$. Quantities in brackets are the observed standard errors.

simply to indicate the extent of non-linearity in the calibration curve. (This is not so easy to grasp from a plot, because of the very wide dynamic range.)

In order to calculate V_x, and therefore the detection limit, it is necessary first to estimate V_y as a function of concentration and then to use this information to estimate the parameters of the calibration curve using weighted least squares (WLS) fitting. Rigorous application of WLS requires knowledge of relative weights, but the technique is already considered adequate when n ≥ 5 (18).

In Table IV we present the results of fitting alternative models to the pattern of weights and the calibration curve. Before using the results in Tables III and IV to calculate detection limits,

Table IV.   Alternative Calibration Models

Model	$\hat{B}$	$\hat{A}$	$\chi/\sqrt{df}$
(1) $y = B + Ax$[a]	0.042	22.8	--
(2) $y = B + Ax$	-1.04 ± 0.02	38.81 ± 0.12	32.7
(3) $y = B + Aq$[b]	0.042 ± 0.024	32.46 ± 0.10	9.64

[a] This model is taken to be exact -- it uses $\hat{B}$ from model-3 together with the initial point, $(x,\bar{y}) = (0.05, 1.18)$, to derive A.

[b] We take q to be exactly $x^{1.12}$ to account for the non-linearity in the curve; the two parameters (B, A) are then estimated by linear least squares, using weights as indicated in Table III.

a number of observations should be made:

(a) The observed SE's (Table III) are generally monotonic (certainly not constant) with increasing concentration and consistent with the linear model, with the exception of the value at x = 0.25 ng.

(b) The initial observation (at x = 0.05) has a response already > forty times the zero-point standard deviation

(1.18/0.028); thus, it is clearly way in excess of the detection limit. A very large extrapolation is therefore necessary to estimate both the background (B) and standard deviation (σ_y) in the region of the detection limit. This is the basis of introducing model-1, for illustrative purposes (Table IV).

(c) Model-2 (Table IV) obviously is inadequate; the significantly negative intercept and poor fit rule it out (over the entire data range). Not shown are simple polynomial fits, which are also inadequate.

(d) Model-3 is better. The intercept is consistent with zero (to be expected from the technique of calculating net GC peak area). The fit, however, implies an additional (non-replication) error source. Again, for illustrative purposes, the function q(x) has been taken exact in order to avoid the distributional perturbations of non-linear least squares (not justified in view of the foregoing limitations of the data).

Before turning to the question of detection, it is illuminating to examine a plot of the data, and the residuals from the fit of model-3. These are shown in Figure 1. The principal observations which derive from the residual plot are that the assumed shape of the curve and variation of statistical weight with concentration are generally acceptable. The magnitude of the residuals and dispersion for certain replicates and concentrations are not. That is, there is additional scatter about the fitted curve, unaccounted for by the replication error; and certain replicates, especially (● and □) in the 0.25 ng and 5 ng samples are more widely separated than the others. Queries which followed these observations led to suggestions that some untoward dilution errors may have been involved in preparing two of the standards, and random errors in "x" (concentrations of standards) may not be negligible. Thus, a detailed evaluation of the calibration process would require scrutiny (or restandardization) of standard solutions for possible blunders (outliers), and the difficult task of fitting the calibration data taking into account errors in both variables (19).

Fenvalerate Detection Limits. To the extent that detection limits require knowledge of the calibration curve and random error (for x) as a function of concentration, all of the foregoing discussion is relevant -- both for detection and estimation. However, curve shape and errors where x $\gg$ x_D, are relatively unimportant at the detection limit, in contrast to direct observations of the initial slope and the blank and its variability. (It will be seen that the initial observation in the current data set exceeded the ultimate detection limit by more than an order of magnitude!)

To give some perspective to the above remarks a set of alternative decision and detection limits are given in Table V, derived from appropriate information in the preceding three tables. First, we observe that there are two broad classes of
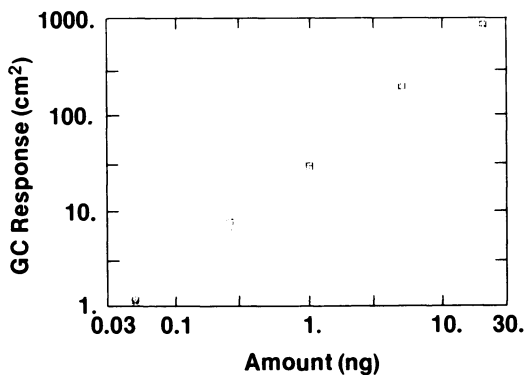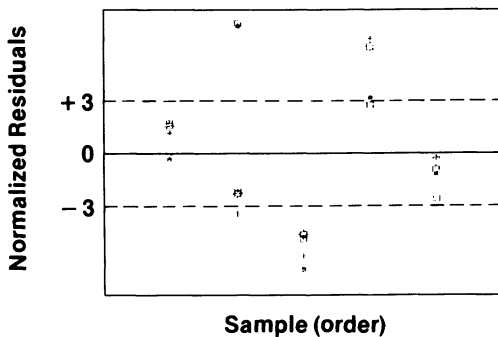
Fenvalerate--Set B

Figure 1. Top portion shows a plot of the observed Fenvalerate
response vs. the mass (ng). Lower plot gives ordered, normalized
residuals from the fit of model-3 to the data (Table IV) using
the weights given in Table III. (Symbols indicate the five
replicates, and the plotted residuals are normalized by the
standard deviations for these individual replicates. The
"goodness of fit" residuals of the model to the means of the
replicates are larger by $\sqrt{5}$, because they are normalized by the
standard errors at each concentration.)

Table V. Fenvalerate Detection Limits

Condition	Amount (pg)	
	x_C	x_D
No Threshold ($y_C = S_C + B \approx 0.09$ cm^2)		
(a) $\sigma_y = 0.028$	2.02	4.04
(b) $\sigma_y = 0.028 + 0.49x$ [x in ng]	2.02	4.19
(c) $\sigma_y = 0.028$ [B, A - estim.]	2.30	4.60
Threshold ($y_C = 1.00$ cm^2)		
(d) $\sigma_y = 0.028$	42.0	44.0
(e) $\sigma_y = 0.028 + 0.49x$	42.0	45.7
(f) $y = B + Aq$ [$\sigma_y = f(x)$]	43.0	46.4

See text and the Appendix for detailed comments on the
individual conditions (a)-(f), and calculations of results.
(Note that all except (f) use the model, $y = B + Ax$. For
Case-f, q is defined as $x^{1.12}$.)

limits: (1) the "natural" (variability-based) lower limit
having no imposed artificial response threshold; (2) an order-
of-magnitude higher set of limits, resulting from a minimum
response of unity (y_C = 1.00) imposed by the data evaluation
"black box" (hardware-software). The effect of such a high,
artificial discriminator is to make α (false positive error)
utterly negligible, but at the cost of considerable sensitivity
(i.e., detection capability).

Added effects of the y-threshold, if there are uncertain-
ties in B and A, are to produce (generally small) uncertainties
in α and x_D. The latter is given by error propagation:

$$\phi^2_{x_D} = \phi^2_A + 2\rho_{BA}\phi_A\phi_B(B/S_D) + \phi^2_B (B/S_D)^2 \qquad (11)$$

where the ϕ's are relative standard deviations, and S_D is the
net signal (y_D-B) at the detection limit.   (See the Appendix,
Case-f, for the application of Equation 11.)

Within each of the two classes in Table V, the first two
sets of limits ((a), (b), (d), (e)) use the constant and
variable weights, respectively, and assume B and A are exactly
known (model-1 in Table IV). The remaining limits involve
estimated parameters, based on the design {x} and the equations
of Table III. Method (c) utilizes the parameters of Model-1
and constant weight; method (f) uses Model-3 and variable
y-errors (weight).

Principal conclusions to be drawn from this exercise,
displayed graphically in Figure 2, are that:
  ⊙ The "black-box" threshold imposes a large and unnecessary
    increase in detection limit.
  ⊙ In the region of the detection limit, for this data set,
    the alternative weighting scheme or model selected has
    little effect.
  ⊙ The additional, non-replication, scatter about the fitted
    calibration curve -- perhaps due to random error in the
    x-variable -- does show a substantial effect. (See last
    paragraph, Appendix.)
  ⊙ Optimal assessment of the minimum detection limit would
    require a design {x} well below the current standard con-
    centrations and including the blank.

The scope of this article does not permit the consideration of
physical vs. empirical models for the calibration curve, nor the
effect of new designs on the detection limit, but these are
extremely important issues in calibration. For example, it can
be shown that with an inadequate design the detection limit (for
$\alpha = \beta = 0.05$) may not even exist! ($x_D \rightarrow \infty$.)
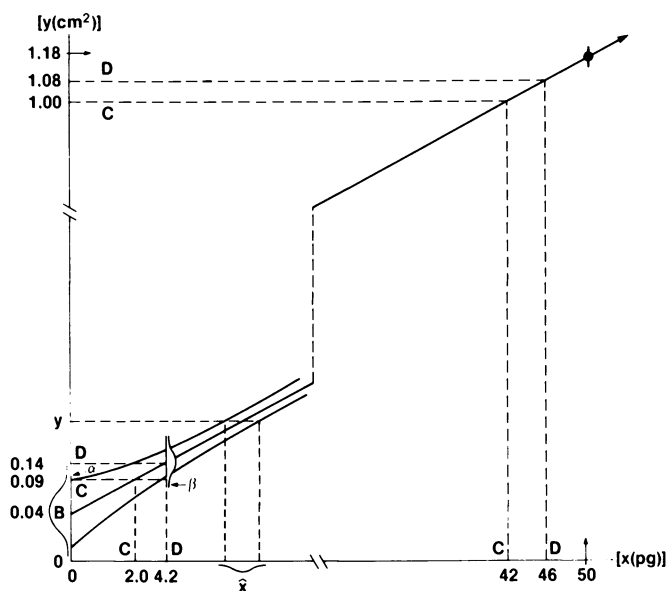
Figure 2. Calibration curve for Fenvalerate in the Regions of the Detection Limits. Numerical values of decision (C) and detection limits (D) are shown for the "No Threshold", case (b) and "Threshold", case (e) from Table V. For the former, α- and β-errors are indicated qualitatively. The first (lowest concentration) data point is shown at $(x,y) = (50, 1.18)$. (Though topologically correct, the scales have been deliberately distorted to encompass both cases, and especially near the origin to dramatize the effect of some designs on the uncertainty of the intercept and the ratio x_D/x_C.)

HIGHER DIMENSIONS: EXPLORATION AND VALIDATION

Space remains for only a brief glance at detection in higher
dimensions. The basic concept of hypothesis testing and the
central significance of measurement errors and certain model as-
sumptions, however, can be carried over directly from the lower
dimensional discussions. In the following text we first examine
the nature of dimensionality (and its reduction to a scalar for
detection decisions), and then address the critical issue of
detection limit validation in complex measurement situations.

Physicochemical Analysis vs. Chemometric Resolution. Once we
pass beyond the one-dimensional calibration of a pure substance,
we enter the realm of mixtures, and the analyzing dimensions of
chromatography, spectrometry, relaxation times, morphology,
chemical "fingerprinting", etc. When one has sufficient
resolving power, whether by means of a simple dimension of
extreme resolution or a linked ("hyphenated") series of inde-
pendent dimensions yielding the product of their individual
resolving powers, then the problem reduces to the zero-
dimensional case. That is, one simply measures the signal in
the appropriate hypercube in multidimensional space which marks
the location of the species of interest. An outstanding example
of such multispectral sorting is the new technique of Accele-
rator Mass Spectrometry (20), which has led to a revolution in
measurements for radiocarbon dating, isotope geophysics, nuclear
geology, etc. Here, for example, ^{14}C atoms and clusters are
initially mass analyzed as high energy (~2 MeV) negative ions,
after which all molecular fragments are destroyed and most
electrons removed; then additional acceleration and mass
analysis occurs with $^{14}C+3$ ions, and final discrimination takes
place with 8 MeV $^{14}C+3$ ions on the basis of ionization density
(dE/dx) and energy (E) or range. The resolving power is so
enormous that one can isolate the signal of one ^{14}C atom from
the associated 10^{12} -10^{14} ^{12}C atoms. A subtle dimension in this
spectroscopy is time, in that the overwhelming background of ^{14}N
is eliminated by the decay of N^- during the initial acceleration
phase. Final quantitative estimation comes from integrating
counts in the appropriate region of the dE/dx, E - plane.

More commonly, we are faced with the need for mathematical
resolution of components, using their different patterns (or
spectra) in the various dimensions. That is, literally,
mathematical analysis must supplement the chemical or physical
analysis. In this case, we very often initially lack sufficient
model information for a rigorous analysis, and a number of
methods have evolved to "explore the data", such as principal
components and "self-modeling" analysis (21), cross correlation
(22). Fourier and discrete (Hadamard, . . .) transforms (23),
digital filtering (24), rank annihilation (25), factor analysis
(26), and data matrix ratioing (27).

Under the best of circumstances we can express the multi-dimensional signal (y) as a linear function of the unknown concentrations (x_k), such as decaying nuclear or optical spectra:

$$y_{ij} = \Sigma\ y_{ijk} = \Sigma\ A_{ijk}\ x_k + e_{ij} \qquad (12)$$

where

$$A_{ijk} = U_k(\lambda_i)\ e^{-t_j/\tau_k} \qquad (13)$$

(The first factor $U_k(\lambda_i)$ is the spectrum of species-k vs wavelength (λ_i); the second is the decay curve vs time (t_j) with mean life τ_k.) If the e_{ij} are normally distributed with known (relative) variances, and we know the spectra and lifetimes for all components, then weighted, linear least squares will provide estimates for x_k and σ_{x_k} (28). Since each $\hat{x}$ is a linear sum of the normally distributed observations, it too is normal, and it is (almost) straightforward to compute the decision level (x_C) and detection limit (x_D) for species-k. In principle, the quantities would require the evaluation of σ_x as x increases from zero to its detection limit. If the signal is relatively weak ($x \lesssim x_D$) and σ_x is not too rapidly varying, however, we can approximate x_C as 1.64 σ_x (x = 0) and $x_D \approx 2x_C$ (assuming σ is known). Thus, weak estimates may be tested approximately, by comparing the relative standard error to ~60% (1/1.64). (Note that prior knowledge of σ or A_{ijk} is not required if one has a suitable multivariate data set. Larger uncertainties and increased detection limits follow, however.)

Many multidimensional problems yield a different decomposition for y_{ijk}. That is, one acquires a data matrix (y_{ij}) in which (presumably) the same components are present in varying proportions (x_{kj}), and where each sample-j is measured using the same method (or dimension)-i. Then,

$$y_{ij} = \Sigma\ A_{ik}\ x_{kj} + e_{ij} \qquad (14)$$

when the number of samples is substantially larger than the number of components, multivariate techniques such as "self-modeling" can be used to estimate both the A's (e.g., spectra) and the x's. Application of such methods carry their own special requirements (assumptions), and obviously the uncertainties (for x) and detection limits are strictly dependent on the particular technique (algorithm) adopted. Regardless of the technique, however, the decision level and detection limit are defined by the acceptable α- and β-errors and derived from expressions (Equation 3a) which relate the uncertainty of x to its magnitude for the particular mix of components (chemical matrix) present.

A brief comment on dimensionality is in order at this point. As used here, the number of dimensions is taken equal to the number of subscripts on the data matrix. Thus, an optical or mass or nuclear spectrum is one-dimensional, but if different samples or sampling times are involved it is considered two-dimensional, as in GC-MS. In this context, we treat the vector representation of a spectrum or a multielement analysis as single dimension, though it is frequently viewed as "a point in hyperspace."

Multidimensional Data Intercomparisons. Estimation of reliable uncertainty intervals becomes quite complex for non-linear operations and for some of the more sophisticated multidimensional models. For this reason, "chemometric" validation, using common, carefully-constructed test data sets, is of increasing importance. Data evaluation intercomparison exercises are thus analogous to Standard Reference Material (SRM) laboratory intercomparisons, except that the final, data evaluation step of the chemical measurement process is being tested.

Two extensive tests of this sort have taken place during recent years -- the first based on simulation γ-ray spectra generated by the International Atomic Energy Agency (IAEA; Reference (29)), and the second based on multisample simulated air particulate compositional data generated by NBS (30). In both cases the data were constructed from realistic mixtures of known components incorporating known random errors. Although the organizers knew the "truth", the participants were provided only with information concerning possible components, calibration data and random error structure (distributions, parameters).

Outcomes of both exercises were revealing. In each case wide variations in bias, precision and detection capabilities were exhibited among the participants -- in some cases even when they employed the same data evaluation technique! (Certain [all?] techniques are known to require the input of "scientific judgment.") By way of illustration, one portion of the IAEA intercomparison data, the "detection limit" spectrum, is given in Figure 3. Of the 212 participants, 90% returned results for this spectrum, with the best performance (19/22 peaks correctly detected) coming from the "visual" method! The number of correct identifications ranged from 2 to 19, and false positives ranged from 0 to 23. Closely following the visual method in performance were those using the second derivative and cross correlation. Because of the non-linear characteristics of some of the methods applied, the distribution of $\hat{x}$ (especially when x = 0) may differ from normal. This, together with the fact that the exercise involved multiple detection decisions, adds a certain complication to the calculation of detection limits (5,31).
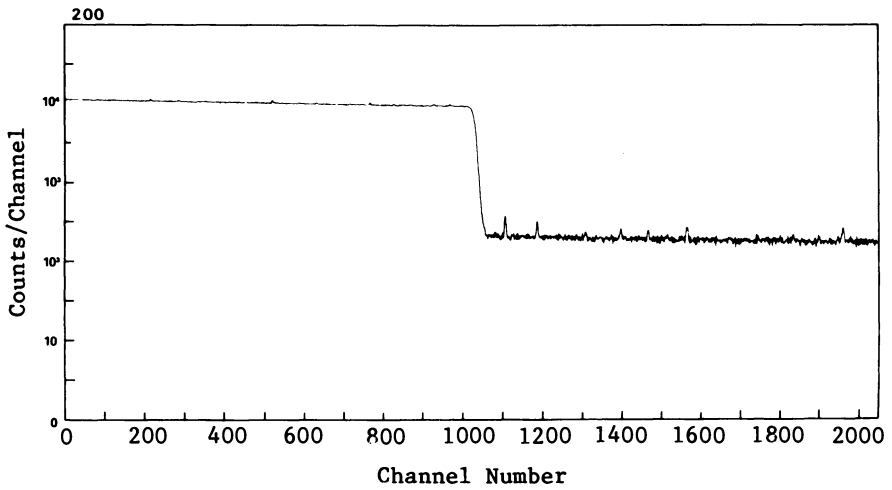
Figure 3   IAEA Test Spectrum for Peak Detection (Ref. 29).

FINAL REMARKS

Decision levels and detection limits are relatively easy to
define and evaluate for "simple" (zero dimensional) measure-
ments. The transition to higher dimensions and multiple
components introduces a number of complications and added
assumptions related to the number and identity of components,
shapes and parameters of calibration functions and spectra, and
distributional consequences of non-linear estimation.
        Major attention has been given to a one-dimensional problem
-- evaluation of detection limits for the measurement of the
pesticide, Fenvalerate, by gas chromatography -- using an actual
set of calibration curve data. Beyond developing the equations
and reviewing the treatment for this case, we encountered some
rather special problems: (a) slight non-linearity in the curve;
(b) extra variability (beyond replication variance) about the
curve, presumably due to errors in the x-dimension (concentra-
tions of standard samples); and (c) quite a high instrumental
(software) threshold, preventing the automatic acquisition of
data for the blank or data in the region of the ultimate
detection limit. This last matter is the most serious. It is
perhaps representative of many problems which accompany the
benefits of the microprocessor revolution -- i.e., "black boxes"
with hidden algorithms all too frequently limit our knowledge or
control of the data handling phase of our measurements. In this
case, the threshold-imposed detection limit exceeded the natural
detection limit by about a factor of ten, and it forced us to
extrapolate the data over this wide interval to estimate the
blank (intercept), its standard deviation, and the natural
detection limit.
        The primary factor which governs the detection limit, the
uncertainty of the blank, can differ markedly depending upon
whether it is estimated from a model (intercept, baseline), or
from a direct observation, or from "blind" interlaboratory
comparison. Inappropriate blank evaluation, together with
inadequate reporting of data when at or below the detection
limit are seen as two of the most critical sources of informa-
tion loss or information distortion, with serious potential
effects in environmental and regulatory areas (3).
        Chemometric quality assurance via laboratory and method
intercomparisons of standardized test data sets, finally, is
becoming recognized as essential for establishing the validity
of detection decisions and estimated detection limits, espe-
cially when treating multidimensional data with sophisticated
algorithms including several chemical components.

APPENDIX -- CALIBRATION CURVE DETECTION LIMITS
(Calculation of Table V Results)*

<u>Case - a -- Intercept (B) and Slope (A), Known; σ_y = Constant.</u>

$$B = 0.042 \text{ cm}^2 \qquad A = 22.8 \text{ cm}^2/\text{ng} \qquad \sigma_y = 0.028 \text{ cm}^2$$

From Eq (5),
$$\sigma_x = \sigma_y/A = 1.2_{28}\times10^{-3} \text{ ng}$$

DECISION (CRITICAL) LEVEL:

$$x_C = z_{1-\alpha}\sigma_o = 1.645(0.00123) = 0.00202 \text{ ng } [2.02 \text{ pg}]$$

DETECTION LIMIT:
$$x_D = x_C + z_{1-\beta}\sigma_{x_D}$$

Since $\alpha = \beta = 0.05$, and σ_y = const; From Eq (7)

$$x_D = 2x_C \text{ (K/I)} \qquad\qquad (7)$$

Since $\phi_A = 0$ [A-known], K and I = 1, and

$$x_D = 2x_C = 4.04 \text{ pg}$$

<u>Case-b -- Intercept and Slope Known; σ_y = f(x)</u>

$$B = 0.042 \text{ cm}^2 \qquad A = 22.8 \text{ cm}^2/\text{ng} \qquad \sigma_y = 0.028 + 0.49x$$
$$\sigma_x = \frac{1}{A} (0.028 + 0.49x) \qquad\qquad (A1)$$

DECISION LEVEL: (unchanged)

$$x_C = 1.645(0.028)/22.8 = 0.00202 \text{ ng } [2.02 \text{ pg}]$$

DETECTION LIMIT: (must iterate; but rapid convergence)

$$x_D = x_C + z_{1-\beta}\sigma_{x_D} = 0.00202 + \left(\frac{1.645}{22.8}\right)(0.028 + 0.49x_D) \to 0.00419 \text{ ng } [4.19 \text{ pg}]$$

<u>Case-c -- σ_y, constant; Intercept and Slope Estimated (unweighted</u>
<u>L-S)</u> Here, for illustrative purposes, the model is taken to be
(rigorously):

$$y = B + Ax, \text{ with } B, A \text{ and } \sigma_y \text{ as in Case-a}$$

---

*The calculations given below all assume that σ_y or σ_y = f(x) is
known. An approximate treatment for σ-estimated would involve
the substitution of ts for zσ, where t is Students-t.

The design (x_1 through x_5) is taken to be the same as for the Fenvalerate calibration data, but without replication.

[Note: This is a very poor design for this purpose, since no observations are made in the region of the detection limit. Also, it should be recognized that the above model gives a poor representation for the full range of the calibration curve. It may be quite adequate at and below the detection limit, but new observations would be necessary to test this conjecture.]

From Eq (10),

$$\sigma_x = \frac{\sigma_y}{A}\left[1 + \frac{1}{n} + \frac{(x-\bar{x})^2}{S_{xx}}\right]^{1/2} \quad ; \quad S_{xx} \equiv \sum(x-\bar{x})^2 \quad (10)$$

for $\{x\} = (0.05, 0.25, 1., 5., 20.)$:

$$n = 5 \qquad \bar{x} = 5.26 \text{ ng} \qquad S_{xx} = 287.7 \text{ ng}^2$$

DECISION LEVEL: [set $x \to 0$ in Eq. (10)]

$$x_C = \left(\frac{1.645(0.028)}{22.8}\right)\left(1 + \frac{1}{5} + \frac{(-5.26)^2}{287.7}\right)^{1/2} = 0.00230 \text{ ng}$$

$$[2.30 \text{ pg}]$$

DETECTION LIMIT:

$$x_D = x_C + \frac{z_{1-\beta}}{A}\sigma_{x_D} = x_C + \frac{z_{1-\beta}\sigma_y}{A}\left(1 + \frac{1}{n} + \frac{(x_D-\bar{x})^2}{S_{xx}}\right)^{1/2} \quad (A2)$$

The solution may be obtained by iteration of the above equation, or again by use of Eq (7) -- since $\alpha = \beta$ and $\sigma_y = $ const.

$$x_D = 2x_C \ (K/I) \approx 2x_C = 4.60 \text{ pg} \qquad (7)$$

($K, I \approx 1$ because ϕ_A is quite small, ca 7×10^{-5}.)

Case-c' -- Non-negligible Errors in B and A

In order to illustrate the treatment of intercept and slope uncertainties, let us re-examine case-c with σ_y amplified by a factor of 100; i.e. -- $\sigma_y = 2.80 \text{ cm}^2$. Obviously (Eq. 10), σ_B and σ_O and x_C (and σ_A, ϕ_A) would be scaled by the same factor.

$$\sigma_B = 2.80\left(\frac{1}{5} + \frac{(-5.26)^2}{287.7}\right)^{1/2} = 2.80\sqrt{0.296} = 1.52 \text{ cm}^2$$

$$\sigma_{s_O} = A\sigma_O = (\sigma_y^2 + \sigma_B^2)^{1/2} = 3.19 \text{ cm}^2$$

DECISION LEVEL:

$$x_C = \left(\frac{z_{1-\alpha}}{A}\right) \sigma_{S_O} \approx \left(\frac{1.645}{22.8}\right) (3.19) = 0.230 \text{ ng } [230 \text{ pg}]$$

DETECTION LIMIT:

This could be re-computed (iteratively) from Eq (A2), or directly from Eq (7), using $K = 1 + \rho_{BA} (\sigma_B/\sigma_O)(z\phi_A)$ and $I = 1 - (z\phi_A)^2$.

From the design $\{x_i\}$, we may calculate ρ_{BA} and ϕ_A:

$$\rho_{BA} = -\bar{x}/x_{rms} = -0.570$$

$$\phi_A = \frac{\sigma_A}{A} \approx \frac{\sigma_y}{A} \sqrt{1/S_{xx}} = \frac{2.80}{22.8} (1/287.7)^{1/2} = 0.00724$$

[Thus: $\hat{B} \pm \sigma_B = 0.04 \pm 1.52 \text{ cm}^2$, and $\hat{A} \pm \sigma_A = 22.8 \pm 0.2 \text{ cm}^2/\text{ng}$]

Therefore, $z\phi_A = (1.645) (0.00724) = 0.0119$, and

$$I = 1 - (0.0119)^2 = 0.99986$$

$$K = 1 + (-0.57) \left(\frac{1.52}{3.19}\right) (0.0119) = 0.9968$$

So, $K/I = 0.9969$, and

$$x_D = 2x_C (K/I) = (1.9938) x_C = 458.6 \text{ pg}$$

Notes

[i] A slightly more rigorous expression for σ_O may be obtained by keeping an additional term in the Taylor expansion for the variance of the ratio,

$$\hat{x} = (y-\hat{B})/\hat{A}. \quad \text{That is,}$$

$$V_O = \sigma_O^2 \approx \frac{1}{A^2} (V_{yo} + V_B) J, \text{ where } J = 1 + 3\phi_A^2 \quad (3b)$$

In the present case of $\phi_A \simeq 0.00724$, so the inclusion of J would increase σ_O and therefore x_C by less than 1 part in 10^4. The existence of such a factor is important in principle, however, for it signifies the contribution of e_A to the variance of the estimated net analyte concentration (or amount) even when that concentration is zero.

[ii] Strictly speaking, x_C and x_D as calculated above must be viewed as approximations (though extremely good ones) since A is

not exactly known (denominator of Eq. 10). A useful viewpoint is to consider x_C as exactly $z_{1-\alpha}'$ $\sigma_0/\tilde{A}$, where $\tilde{A}$ is a single, selected outcome $\hat{A}$.

$$z_{1-\alpha}'/\tilde{A} = z_{1-\alpha}/A_T = z_{0.95}/A_T \quad [A_T = \text{"true" slope}]$$

It follows that $\sigma_z/z \approx \sigma_A/A$ for small σ_A/A. For the present case this (ϕ_A) is equal to 0.72%, so we find a (1σ) confidence interval for $z_{1-\alpha}$ of 1.645 ± 0.011$_9$. It follows that

$$\alpha \approx 0.0500 \pm 0.0012$$

Exactly the same argument follows for $z_{1-\beta}'$ and β'; i.e.,

$$\beta: \quad 4.88\% \text{ to } 5.12\%$$

[A confidence interval could be given for x_D, for $\beta = 5.0\%$ and fixed x_C (α-uncertain); but a common interval treatment or scaling for both α and β seems more desirable.]

[iii] This example is very close to the treatment by Hubaux and Vos (H & V, Ref 14). The principal differences are that we take a fixed value for x_C and treat α as uncertain, whereas H & V fix S_C and $x_C = S_C/\hat{A}$ varies with the various estimates $\hat{A}$. Their upper limits (or "detection limits") similarly vary with $\hat{A}$, whereas again we consider β as uncertain but x_D fixed. When $\hat{A} = A$, the results are identical.

Case-d -- Response Threshold (y_{th} = 1.00); σ_y = const. B, A-known. (This case is trivial to calculate, given the considerations for Case-e, so it will not be treated explicitly.)

Case-e -- Response Threshold (y_{th} = 1.00); σ_y = f(x); B, A-known

$$B = 0.042 \text{ cm}^2 \qquad A = 22.8 \text{ cm}^2/\text{ng} \qquad \sigma_y = 0.028 + 0.49x$$

DECISION LEVEL:

$$y_C = 1.00 \text{ cm}^2 \qquad x_C' = (y_C-B)/A = 0.0420 \text{ ng } [42.0 \text{ pg}]$$

Decisions are now made on the basis of the directly observed response (y in comparison to $y_C = y_{th}$). Though a corresponding concentration or "amount" level x_C' may be calculated, it does not carry any direct operational meaning.

With so high a threshold ($y_C \gg \sigma_y$), the α-risk is extremely small:

$$z_{1-\alpha} = \frac{y_C - B}{\sigma_y\text{o}} = \frac{1-0.042}{0.028} = 34.2 \quad [\alpha \ll 1]$$

DETECTION LIMIT:

As with the α-risk, the β-risk is related to the distribution of observed response (y) given x. [α: distribution of $y(x=0)$; β: distribution of $y(x=x_D)$.] The minimum detectable response (y_D), given the probability (β) of failing to exceed y_C, is:

$$y_D = y_C + z_{1-\beta}\,{}^0y_D$$

that is,

$$y_D = 1.00 + 1.645\,(0.028 + 0.49\,x_D)$$

Since $x_D = (y_D - B)/A$, the solution may be found by iteration for y_D by substituting in the above equation, or the entire equation may be transformed:

$$x_D = \frac{y_D - B}{A} = \frac{1}{A}\,(y_C - B + z_{1-\beta}\,{}^0y[x_D])$$

$$x_D = \frac{1}{22.8}\,(1 - 0.042 + 1.645[0.028 + 0.49\,x_D]) \rightarrow 0.0457\,\text{ng} \quad [45.7\ \text{pg}]$$

Case-f -- $\underline{y_{th} = 1.00;\ \sigma_y = f(x)}$; weighted least-squares (WLS) Solution for the Model:

$$y = B + Aq, \text{ where } q = x^{1.12}$$

$$[\hat{B}\colon 0.042 \pm 0.024;\ \hat{A}\colon 32.46 \pm 0.10]$$

This is perhaps the "best solution <u>for the given data set</u>, and it is certainly the most interesting. It is not offered as a rigorous solution, however, for the lack of fit ($\chi^2/\text{df} = [9.64]^2$) implies additional sources of error, which may be due to: additional scatter about the calibration curve (σ_y - "between" component), residual error in the analytic model for the calibration function, or errors in the "standard" x-values. (We believe the last source of error to be the most likely for this data set.) For these reasons, and because we wish to avoid complications introduced by non-linear least squares fitting, we take the model $y = B + Ax^{1.12}$, and the relation $\sigma_y = 0.028 + 0.49x$ to be exact; and then apply linear WLS for the estimation of B and A and their standard errors.

STANDARD ERRORS, CORRELATION, AND WEIGHTS:

Before considering x_C' and x_D, it will be fruitful to examine the foregoing quantities. This can be beneficial in practice because a rapid calculation of the detection limit can

often be made (e.g., using Eq's (3b) and 5)) if the results of WLS are available. Thus, in Table IV we find already the quantities σ_B, σ_A which are needed for calculating the detection limit. (Not shown, but generally also available is the correlation coefficient ρ_{BA}, which here equals -0.292.) In case these quantities were not directly available, they could be calculated from the general expressions given in Table II, as follows:

Thus, for this model (model-3 of Table IV):

$$w_i = 5/\sigma_{y_i}^2 = 5/(0.028 + 0.49 \, x_i)^2 \text{ for 5-replicates}$$

(The same w's follow directly from the calculated response standard errors given in Table III.)

$$\sigma_B = \left[\frac{1}{\Sigma w} + \frac{(\bar{x}_w)^2}{\Sigma w(x-\bar{x}_w)^2}\right]^{1/2} = \left[\frac{1}{1972} + \frac{(0.0892)^2}{94.18}\right]^{1/2} = 0.0243$$

$$\sigma_A = [1/\Sigma w(x-\bar{x}_w)^2]^{1/2} = 1/\sqrt{94.18} = 0.103$$

$$\rho_{BA} = -\bar{x}_w/\sqrt{\overline{x_w^2}} = -0.292$$

DECISION LEVEL:

As with Case-e, the response decision level is pre-fixed:

$$y_C = 1.00 \text{ and } z_{1-\alpha} \approx \frac{y-\hat{B}}{\sigma_y 0} = \frac{1-0.042}{0.028} = 34.2$$

Again, $\alpha \ll 1$, but now its value is uncertain, because of the uncertainty in B. Specifically, the standard deviation of $z_{1-\alpha}$ is $\sigma_B/\sigma_y 0 = 0.024/0.028$.

Thus, we estimate

$$\hat{z}_{1-\alpha} = 34.2 \pm 0.9$$

Finally, $\hat{q}_C' = (y_C-\hat{B})/\hat{A}$; and $\hat{x}_C' = (\hat{q}_C')^{1/1.12} = 0.043$ ng
$$[43.0 \text{ pg}]$$

DETECTION LIMIT:

$$y_D = y_C + z_{1-\beta}\sigma_y D = y_C + z(0.028 + 0.49x_D)$$

Transforming: $q_D = (y_D-B)/A = x_D^{1.12}$, we have:

$$x_D \approx \left[\frac{y_C-\hat{B}+z_{1-\beta}\sigma_y D}{\hat{A}}\right]^{1/1.12} = \left[\frac{1.000-0.042+1.645(0.028+0.49x_D)}{32.46}\right]^{1/1.12}$$

This equation may be solved by iteration to yield,

$$\hat{x}_D = 0.0464 \text{ ng} \quad [46.4 \text{ pg}]$$

Unlike the preceding case where B and A were known, we now have only an estimated x-detection limit, even though y_D is (almost) known. ("Almost", because of the x-dependence of σ_y. Though this is a second order effect when there is a high threshold, it would be preferable in this case to estimate σ_y as a function of y rather than x. Then y_D would be quite independent of $\hat{B}$ and $\hat{A}$.)

DETECTION LIMIT CONFIDENCE INTERVAL:

A confidence interval for x_D may be calculated from y_D, using Eq (11), upon the substitution of $\hat{q}$ for $\hat{x}$. That is

$$\phi_q^2 = \phi_A^2 + 2\rho_{BA}\phi_A\psi_B + \psi_B^2 \quad (11')$$

where

$$\psi_B = \phi_B\left(\frac{B}{y_D-B}\right) = \phi_B\left(\frac{B}{S_D}\right) \text{and} \quad \phi_{x_D} = \phi_{q_D}/1.12$$

Using the previous results, we find

$$\rho_{BA} = -0.292$$

$$S_D = y_D-B = y_C-B+z_{1-\beta}\sigma_{y_D} \approx 1-.042+1.645(.028+0.49[0.0464])$$

$$\psi_B = \phi_B\left(\frac{B}{S_D}\right) = \left(\frac{\sigma_B}{B}\right)\left(\frac{B}{S_D}\right) = \sigma_B/S_D \approx 0.0243/1.041 = 0.0233$$

$$\phi_A = \sigma_A/A \approx 0.103/32.46 = 0.00317$$

Thus, $\phi_{q_D} = 0.0226$; so $\phi_{x_D} = 0.0202$

The ("1σ") confidence interval for the detection limit is thus,

$$\hat{x}_D: 46.4 \pm 0.94 \text{ pg}$$

A symmetric and normal confidence interval is a good approximation, since the uncertainty is dominated by σ_B (numerator of Eq 3b).

Finally, <u>if</u> the poor fit simply reflected proportionately extra y-variance, then:

$$w_i \rightarrow w_i/(9.64)^2 \text{ and } \sigma_{\bar{y}} \rightarrow \sigma_y \cdot (9.64)/\sqrt{5}$$

and ϕ_A, ϕ_B would be increased by a factor of 9.64

The resulting estimate for x_D would be

$$x_D: 59.1 \pm 11.5 \text{ pg}$$

This result, however, should not be taken too seriously, because the poor fit may <u>not</u> be simply related to extra y-variance.

Literature Cited

1. Kaiser, H. Two papers on the Limit of Detection of a Complete Analytical Procedure. London:Hilger, 1968.
2. Currie, L. A. Anal. Chem. 1968, 40, 586.
3. Rogers, L. B. Subcommittee dealing with the scientific aspects of Regulatory Measurements, American Chemical Society, 1982.
4. Ingle, J. D., Jr. J. Chem. Educ. 1974, 51, 100-5.
5. Currie, L. A. Pure & Appl. Chem. 1982, 54, 715-754.
6. Currie, L. A. Nucl. Instr. Meth. 1972, 100, 387.
7. Currie, L. A. in "Modern Trends in Activation Analysis"; DeVoe, J. R.; LaFleur, P. D., Eds.; Nat. Bur. Stand (U.S.) Spec. Publ. 312, 1968; p. 1215.
8. Currie, L. A. in "Treatise on Analytical Chemistry"; Elving, P.; Kolthoff, I. M., Eds.; J. Wiley & Son:New York, 1978; Vol 1, Chap. 4.
9. Ingle, J. D., Jr.; Wilson, R. L. Anal. Chem. 1976, 48, 1641.
10. Patterson, C. C.; Settle, D. M. 7th Materials Res. Symposium, Nat. Bur. Stand (U.S.) Spec. Publ. 422, 1976; p. 321.
11. Scales, B. Anal. Biochem. 1963, 5, 489.
12. Horwitz, W. (FDA); Meinke, W. W. (NRC), personal communications, 1982. See also Reference (3).
13. Horwitz, W.; Kamps, L. R.; Boyer, K. W. J. Assoc. Off. Anal. Chem. 1980, 63, 1344.
14. Hubaux, A.; Vos, G. Anal. Chem. 1970, 42, 849.
15. Ku, H. H. J. Res. Natl. Bur. Stand. 1966, 70c, 263.
16. Brownlee, K. A. in "Statistical Theory and Methodology in Science and Engineering"; J. Wiley & Son:New York, 1960.
17. Kurtz, D. A., personal communication, 1982, 1983.
18. Jacquez, J. A.; Mather, F. J. and Crawford, C. R. in "Linear Regression with Non-Constant, Unknown Error Variances"; Biometrics 1968, 24, 607.
19. Golub, G. H.; van Loan, C. F. J. Numer. Anal. 1980, 17, 883.
20. Purser, K.; Russo, C.; Gove, H.; Elmore, R.; Ferraro, R.; Beukens, K.; Chang, L.; Kilius, L.; Lee, H.; Litherland, A. Chapt. 3 in Symposium on Nuclear and Chemical Dating Techniques, Currie, L. A., Ed.; American Chemical Society: Symposium Series No. 176, Washington, D.C., 1982.

21. Lawton, W. H.; Sylvestre, E. A.; Maggio, M. S. Techno-metrics 1972, 14, 3.
22. Horlick, G. Anal. Chem. 1973, 45, 319.
23. Marshall, A. G., Ed.; "Fourier, Hadamard, and Hilbert Transforms in Chemistry", Plenum Press:New York, 1982.
24. Savitzky, A.; Golay, M.J.E. Anal. Chem. 1964, 36, 1627.
25. Ho, C.-N.; Christian, G. D.; Davidson, E. R. Anal. Chem. 1981, 53, 92.
26. Malinowski, E. R.; Howery, D. G. in "Factor Analysis in Chemistry"; J. Wiley & Son:New York, 1980.
27. Fogarty, M. P.; Warner, I. M. Anal. Chem. 1981, 53, 259.
28. Nicholson, W. L.; Schlosser, J. E.; Brauer, F. P. Nucl. Instr. Meth. 1963, 25, 45.
29. Parr, R. M.; Houtermans, H.; Schaerf, K. Computers in Activation Analysis and Gamma-ray Spectroscopy Ed., Conf. -780421 1979, p. 544.
30. Currie, L. A.; Gerlach, R. W.; Lewis, C. W. Atm. Environment 1984, 18, 1517.
31. Liggett, W. ASTM Conf. on Quality Assurance for Environmental Measurements, STP (in press) 1984, Boulder, CO.

Introduction to the Theory of Correlation Chromatography

RAYMOND ANNINO

The Foxboro Company, Foxboro, MA 02035

The technique of correlation chromatography is described through text and figures in a step-by-step manner. The description explains how a Pseudo Random Binary Sequence (PRBS) can control multiple and overlapped input injections into a chromatograph and again be used to sort out the detector data to give a correlogram. Correlograms mimic chromatograms but represent chromatographic data at a much higher sensitivity. The method is used to immensely increase the signal-to-noise ratio of a chromatograph. Problems in sampling and non-linearity are also discussed.

One of the difficulties in writing about Correlation Chromatography (CC) is in providing the reader with enough conceptual understanding of this elegant procedure to appreciate both its potential and the aggravating problems which still hamper its use. Many of us require a physical model to aid us in understanding the elements of a problem. A mathematical formula is of little help unless we can associate it with some physical picture. For this reason, we will attempt in this paper a largely pictorial but still rigorous presentation of Correlation Chromatography (CC).

Background

Before proceeding to the main subject of CC, it is necessary for the reader to gain some understanding for the basic process of "correlation." Correlation is a mathematical procedure for measuring the similarity of two different signals or the spectral characteristics of one signal.

Consider the two time-varying signals shown in Figure 1. Is there any similarity between the two signals? To answer this question we might cut out the signal 1b so that after we place it underneath the other we can shift it a little along the x axis (which is a time axis), to visually compare it with the other. If
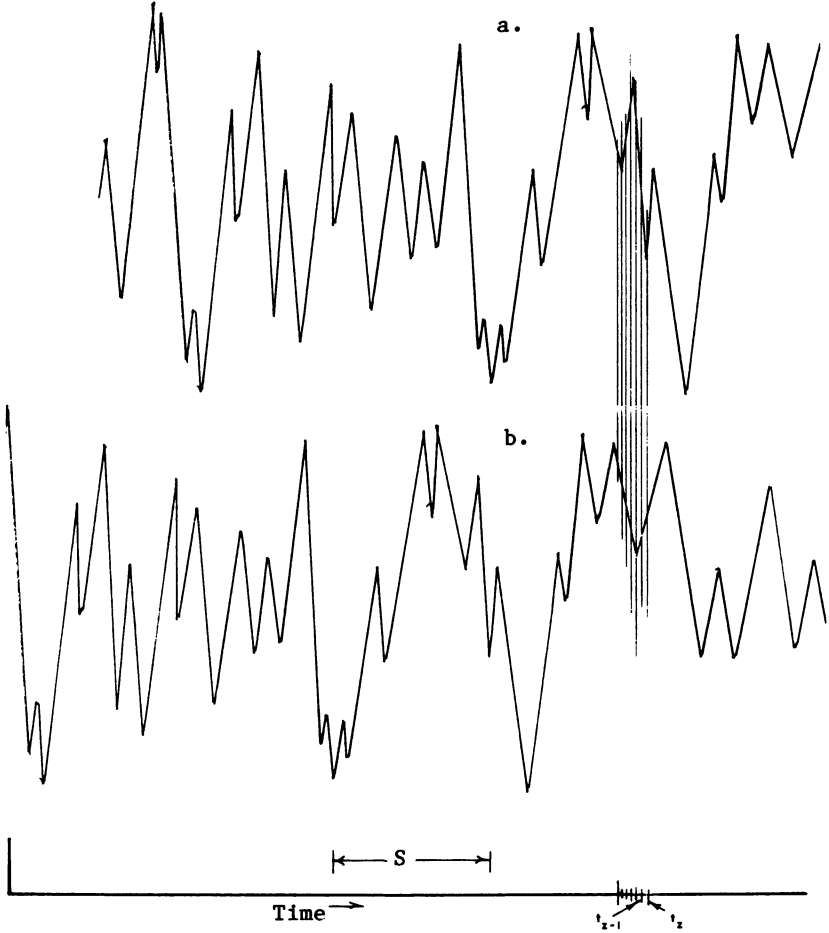
Figure 1. Signals to be compared. Signal a. and b. are the same. Signal b. has been delayed S units in time.

we were to do this, we would find that when we had shifted it S
units, the two signals would appear to be identical. Our conclu-
sion is that the only difference in the two signals is that one of
them is shifted in time with respect to the other one.

How can we document this comparison procedure mathematically?
Suppose we again examine Figure 1. Let us digitize the two
signals into the units we have shown along the x - axis. Now at
each of these units multipy the intensity value of one signal with
the corresponding intensity value of the other. Do this for each
digitized unit, sum all of these individual products and divide by
the number of units, T, in the sum to obtain an average value for
all these products. To express this operation mathematically we
write:

$$R_{xy}(0) = 1/T \sum_{t=0}^{t=T} x(t)y(t)$$

The number obtained in this way is called a correlation
coefficient, R_{xy}, and its magnitude furnishes us with a measure of
the correlation between the two signals. Small values of the
correlation coefficient indicate little or no correlation while
large values are obtained when the two signals match. What we
would like to do now is shift the signal in the same manner that
we did when we cut it out, and repeat the above multiplication,
summing, and averaging procedure to calculate another correlation
coefficient. In other words, each intensity value of signal 1a at
a certain time value, t_z, will be multiplied with the intensity
value in 1b found at $t_z - 1$. We will proceed with our multiplica-
tion, summing, and averaging in this manner for all values of t
within the interval we have selected for examination, and thus
produce a new value for the correlation coefficient, R_{xy}, at this
time shift of one unit. Thus we have:

$$R_{xy}(1) = 1/T \sum_{t=0}^{t=T} x(t-1)y(t)$$

We continue this shifting and calculation process until
we have shifted the desired interval. Our process is one of
moving past events in signal 1b into the present for compari-
son with 1a. If we call the time shift tau, we can write a
general expression for the calculation of the correlation
coefficient at any value of tau.

$$R_{xy}(\tau) = 1/T \sum_{t=0}^{t=T} x(t-\tau)y(t)$$

The limit of this expression gives us the familiar definition for the cross-correlation function with the limits of integration redefined for a distribution about zero.

$$R_{xy}(\tau) = \lim_{T \to \infty} (1/T) \int_{-T/2}^{T/2} x(t-\tau)y(t)dt$$

Following the calculation of the correlation coefficient at a number of values of tau, we can plot the correlation coefficients against tau. Such a plot might look like the one shown in Figure 2. Notice that the value of the correlation coefficient is small except at the time shift where everything in both signals lines up.

Suppose the signals in Figure 1 were random in nature. We know from our elementary statistics that if we had a set of totally random numbers centered around some value, say zero, that the average value of the sum of these numbers would be zero. Thus, for a random signal centered about zero our above operation for calculating the correlation coefficient produces a correlation coefficient close to zero (the value about which the signal is centered) except when the signals are lined up. Since these are random signals, this will occur at only one value of tau, the time delay between the two signals. Our correlation plot would then look like the one in Figure 3. Notice that this correlation plot is noise-free as compared to the one shown in Figure 2 where some correlation exists at a number of time shifts. This noise-free correlogram is a characteristic of the cross- or autocorrelation of a random signal.

Correlation Chromatography

<u>Single Peak CC</u>. Let us now turn our attention to mating this procedure with chromatography in an effort to increase the detector sensitivity.

Suppose we were to set up a chromatograph in such a way that either sample or carrier gas will be flowing through the column. This set up is shown schematically in Figure 4. Which gas is being injected will be determined by the position of the two position valve, V. Arbitrarily, we will label one position of
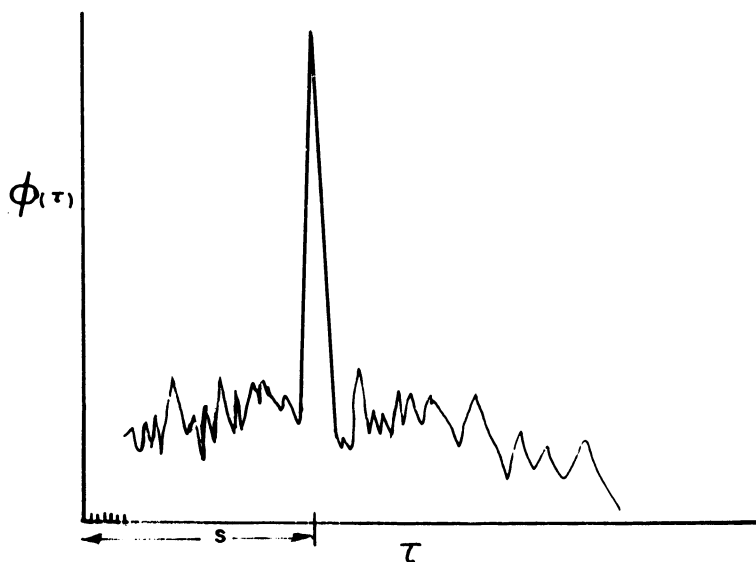
Figure 2. Plot of correlation coefficients vs. tau.
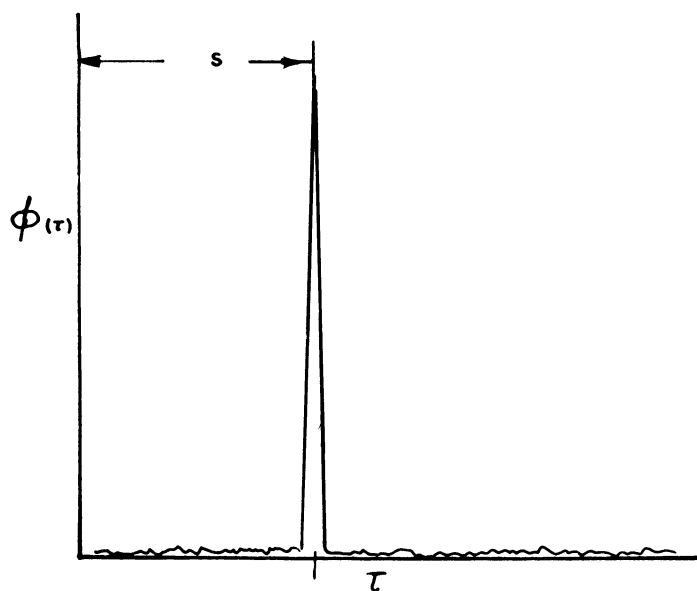
Figure 3. Plot of correlation coefficients vs. tau for random signals S units out of phase.
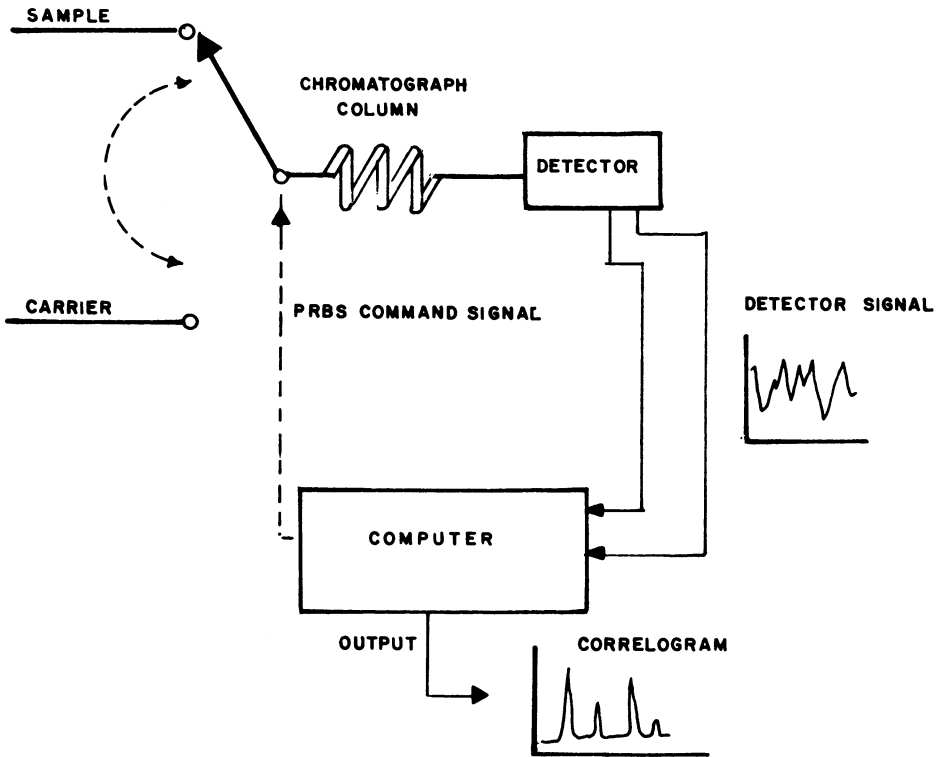
Figure 4. Schematic of a correlation chromatograph.

this valve as 1, (sample gas being injected) and the other position as -1, (pure carrier gas being injected). Assume the single component in the sample reservoir elutes with a time of 3 units and a signal height of 3D units.

Also, in this example, it is assumed that the column will not distort the signal -- only delay it in time. Thus, if we were to switch the valve according to the PRBS code shown in Figure 5a, we would expect, some time later, to see a detector signal such as that also shown in Figure 5a. If we compare this output signal to the input code, we see that there is not a one-to-one correspondence at a given time. Obviously the detector signal is not going to be exactly synchronized with the valve position code because of the column delay.

Let us now cross-correlate the detector signal with the valve position code in the manner just discussed for the signals of Figure 1. We have illustrated this procedure in Figure 5b. An array of detector signals has been memorized and the array of valve positions associated in time with detector signals is shown above it. If you multiply each detector signal with the valve position shown above it and sum each of these products, you will obtain the correlation coefficient shown at the right. The valve position code is then shifted one code unit (given by the value of tau) and the process repeated. The plot of these correlation coefficients vs. tau gives us the correlogram shown in Figure 5c.

We have just performed correlation chromatography.

Notice that this procedure converts the time varying detector signal to a single pulse even though we are feeding sample to the column more or less continuously. This pulse is, in theory, identical to a conventional single pulse chromatogram. Actually, the pulse is modified into a Gausian-like signal by the column just as the column modifies a single pulse chromatogram. However, there is an important difference. Because it is constructed from a number of single pulses and represents the average of these single pulse chromatograms, it, in theory, has a much larger signal-to-noise ratio than its single pulse analog. Again using our elementary statistics we predict that the signal strength has been increased by the number of sample injections made during the period of the code, while the noise has only increased by the square root of the number of injections. This then is the basis for using this procedure for trace analysis where one is usually working at the limit of the signal-to-noise ratio of the system.

One could consider the correlogram to be a snapshot of past events. We must memorize the detector signals and valve positions over a period of time and then calculate a correlogram with this data. To obtain a more up-to-date snapshot we must return to the array of data and calculate another correlogram based on the newest information which has been placed in the array.

As mentioned previously, the correlation of random signals yields clean baselines. The valve position code shown in Figure 5a has been chosen with this property in mind. It is from a set

a.
CHROMATOGRAPHIC DETECTOR SIGNAL

TIME ⟶

3
2 D
1

LENGTH OF CODE UNIT

1   SAMPLE INJECT
-1 CARRIER INJECT

PRBS COMMAND SIGNAL TO VALVE

CALCULATION OF CORRELATION COEFFICIENTS

b.

VALVE POSITION DETECTOR SIGNAL		$\phi(\tau)$
$\tau = 0$	− + − − + + + 3 3 3 0 3 0 0	0
$\tau = 1$	+ − + − − + + 3 3 3 0 3 0 0	0
$\tau = 2$	+ + − + − − + 3 3 3 0 3 0 0	0
$\tau = 3$	+ + + − + − − 3 3 3 0 3 0 0	12
$\tau = 4$	− + + + − + − 3 3 3 0 3 0 0	0
$\tau = 5$	− − + + + − + 3 3 3 0 3 0 0	0
$\tau = 6$	+ − − + + + − 3 3 3 0 3 0 0	0

c.

12

$\phi(\tau)$

1   2   3   4   5   6

τ

Figure 5. Construction of correlograms. a. Detector
signal generated by the sample injected over a period of
time according to the PRBS shown below. b. Calculation of
correlation coefficients. c. Correlogram.

of codes which are called pseudo random binary sequences (PRBS), and can be generated on demand from published algorithms (1,2,3). These codes have the interesting property that within the interval defined by the sequence they behave like a totally random sequence of binary events. This is a convenient code to use. It depicts binary events which is exactly our situation with a sample valve; it is either injecting sample or it is not. The shortest time interval for injection will establish the primary unit of the code and the selected code will determine the sequence of these units. A number of codes can be generated which will allow us the flexibility to select the desired resolution within the sequence.

Correlation of the code with itself (autocorrelation) yields only one correlation point in the time domain defined by the sequence and the unit code interval (see Figure 5c) and an otherwise clean baseline. Since the detector in our chromatogram just follows what the sample valve is doing, it also should be a pseudo random sequence and the cross-correlation of input and output is really an autocorrelation and thus yields the single pulse correlogram with an otherwise clean baseline.

Suppose the detector signal contains random noise which has been contributed from other parts of the system. This random noise has a different source from our artificially created, randomly varying detector signal; therefore it is not correlatable with ours. It is not in phase with our signal. Our procedure for calculating the correlogram averages this noise to zero. This will only be true for truly random noise and provided that you have selected a large enough period for the calculation, i.e. a statistically valid sample. In practice, this means that you must select a PRBS which is long enough so that the data array contains enough samples to average out the random noise.

Multiple Peaks. Finally, we must answer the question of how this method yields chromatograms containing many peaks and whether the correlogram intensity is related to component concentration?

Notice in our previous examples our detector signal has been given the arbitrary units of 3D. The type (voltage, current) and intensity of this output will depend on the particular detector (TC, FID, EC, etc.) and the concentration of the species in the sample. The input code is just that. It is not a signal. The +1 and -1 values are used to depict a position of the injection valve.

If the sample contains more than one species with different retention times, the effect of the column will be to transmit the same version of the switching code but at different delay times, according to the retention times of each of the components. This can be visualized then as a number of delayed versions of the sampling pattern transmitted through the column simultaneously -- all originating simultaneously from the same injection code.

We will select a PRBS of 15 for this next example for increased resolution of the two solutes whose separation we are

going to demonstrate. Consider, for example, a sample composed of
two components.   Component A has a retention time of 3 and, if
injected over one of the time units of our code (a single pulse),
would yield a detector signal of 2. Similarly, component B has a
retention time of 7 and a single pulse detector intensity of 7.

        If we were to use two detectors, one specific for only A and
one specific for only B, the signals would appear, shifted in
time, as shown in Figure 6a. The position of the injection valve
at the time these detector values were recorded is also shown.
Similarly, if we were to use on non-specific detector, the sum of
the signals due to A and B would yield the detector signal labeled
in Figure 6a as Combined Signal.

        Now let us take a time period of detector signals large
enough to encompass the length of the pseudo random binary
sequence injection code which produced it, and cross-correlate it
with this injection code of −1 and 1.

        We have shown in Figure 6a an array of the combined signal
values for one sequence of our code and underneath, in Figure 6b,
the corresponding positions of the sample valve at the moment each
detector value was recorded. Also, we have calculated a value for
the correlation coefficient at each value of tau as we shifted the
injection code from the past into the present. Finally, we have
plotted the value of the correlation coefficients vs. tau in
Figure 6c.

        Notice the simple cross-correlation operation has deconvo-
luted the two chromatograms from the rather nasty detector signal
of overlapped chromatographic peaks. Also, if you examine the
PRBS injection code of 15 which we used, you will see that there
are a total of eight +1 positions and seven −1 positions. Thus we
have injected sample eight times before repeating our code. The
summation procedure at time shifts of 3 and 7 should reflect this,
and they do. A value of 16 is obtained at tau equals 3. This is
eight times the single pulse response of 2 obtained for this
compound at this concentration. Similarly, at tau equal to 7,
compound B has a correlation coefficient of 56 (eight times
seven). Notice that we have set this example up so that the
average value of the detector signal baseline is zero and a simple
summation accomplishes the averaging.

Noise Addition. To illustrate the reduction of out-of-phase
random noise which is possible with this procedure, let us put
some random noise on the single pulse chromatogram of the sample
used for the above example. We have plotted the single pulse
chromatogram with a noise band (minimum to maximum noise signal)
of 2 in Figure 7b. A similar noise band added to the signal from
our correlation detector shown in Figure 6a produces the signal
shown in Figure 7a. The noise signal is also reproduced here, and
you can see it is really a shifted PRBS code of 15. This was done
only to keep the calculations within the paper and pencil domain
as we have throughout this paper. The results will be the same
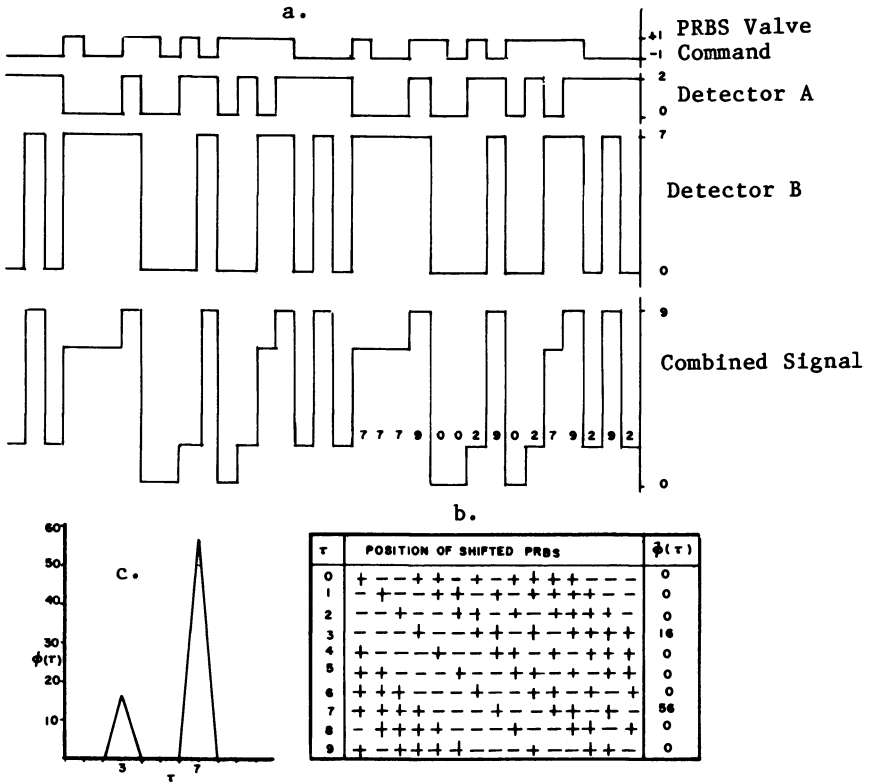for truly random noise.

Figure 6. Construction of a correlogram for a two component sample.

a.

PRBS Code

Combined Signal

7 7 7 9 0 0 2 9 0 2 7 9 2 9 2

Noise

Signal
with Added Noise

7 7 9 11 0 2 2 11 2 4 4 9 2 9 4

SINGLE PULSE CHROMATOGRAM
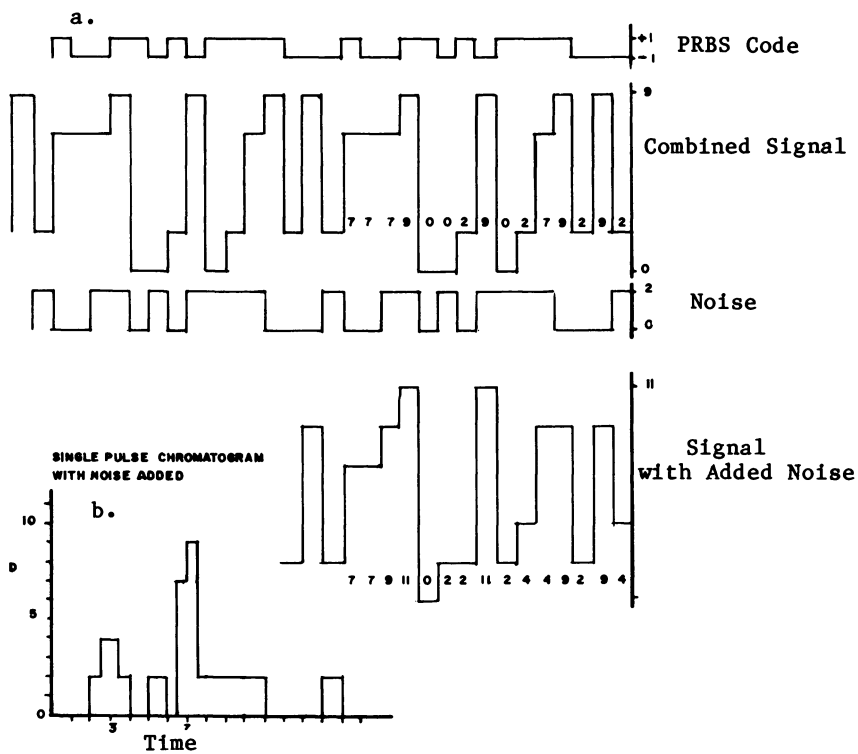WITH NOISE ADDED

b.

Time

Figure 7. Addition of noise to the detector signal of a
single pulse chromatographic run and a run using a PRBS
injection sequence.

We have kept the phases the same so you can cross-correlate
the shift table shown in Figure 6b with the noisy combined detec-
tor signal shown in Figure 7a to obtain the corresponding correla-
tion coefficients.

Exactly the same correlation coefficients are obtained for
the signal in Figure 7a as for the noise free detector signal
shown in Figure 6a. The correlograms are the same. The noise has
vanished! Clearly, the correlogram of Figure 6c represents a much
better data set from which to calculate the concentration of the
two components than the chromatogram shown in Figure 7b. You can
understand the excitement of early researchers in the field when
they realized the potential of this technique. Large increases in
the signal-to-noise ratio were possible simply by encoding the
sample injections properly. Also, it might be possible to turn
the batch-operated chromatograph into one giving more or less con-
tinuous answers. The calculations required to decode the results
were trivial and thus would not require a large computing power.

Problems

Now that we have discussed the theoretical basis for correlation
chromatography, let us examine some areas which may cause problems
or at least, in practice, limit its application.

<u>Sampling</u>.   Notice that the correlation procedure produces a tri-
angle whose base width is two basic time units wide. In order not
to degrade the peak shape produced by a single pulse width, our
triangle should be of the same width. Thus, in trace analysis we
are limited to a unit code width which is one half the sample
volume allowed for a single pulse chromatogram of the same sample.
This may have implications on the sampling switch design if very
fast switching is necessary to maintain resolution of closely
spaced peaks. In this regard, we concluded in earlier work (<u>4</u>)
that ensemble averaging a number of single pulse chromatograms
might be a more effective way of increasing signal-to-noise ratios
in fast chromatography of closely spaced peaks than by using CC.

In selecting parameters for correlation work, two factors
need to be considered. First, the length of the pseudo random
sequence is determined; it must be longer than the longest eluting
peak. Second, the maximum unit pulse width is calculated to main-
tain a given chromatographic integrity. Both of these factors are
then used as criteria to select the pseudo random code that one
must use in a particular analysis.

The sample valve must be made so that it can be activated
automatically at the command of the switching code. Also, in
terms of hardware design, CC demands a reliable sample switch
which must switch many more times over a given analysis cycle than
it would have to if run in the single pulse chromatography mode.
Since the sampling valve is probably the most unreliable component

of an automated chromatograph, you can understand our interest in
no moving part fluidic valves (4,5). Other researchers have
recognized the sample valve problem and offered their own contri-
butions (6,7).

Non-linearity. One of the most aggravating problems in CC which
was recognized early in the research is associated with the
so-called non-stationary nature of the system. In all of the
examples shown above we have assumed linearity of system response.
This means that peak shapes do not change with concentration (all
the moments of the peak remain the same); only the magnitude of
the response changes and this does so linearly with the concentra-
tion. Only if this property is maintained will our multiplica-
tion, summing and averaging work out to cancel everything except
at the correct time shift. Otherwise, the operation produces
significant values for the correlation coefficients at other
points, and therefore, a noisy baseline. We have termed this
noise "correlation noise" to identify its source. It is a noise
component which will not be reduced by cross correlation. It is
always present to some extent in CC and can many times be confused
with real peaks.
    Also, the concentration of the analyzed species must remain
the same during the time of the analysis or, again, we will
produce correlation noise.

Further Discussion. For a detailed explanation of the problems of
CC, backed by both experimental data and the results of computer
modeling, you are referred to our first papers on the subject
(8,9). On the basis of the results, we proposed at that time
guidelines for the effective use of this technique which I believe
are still valid. We have repeated some of these ideas in a
largely tutorial paper that we published some time later (3) and
in a review of signal enhancement techniques (10).
    All of the above problems have been demonstrated in the
literature. You can verify them yourself by taking the examples
we have worked, moving the position of some of the peaks in the
standing waves of each of the components and then adding them
together to produce the final detector signal. Cross-correlation
of this detector signal with the sampling code will not produce a
nice clean correlogram. Similarly, you can demonstrate that the
same thing happens if the intensity of the standing waves changes
during the run, as would be the case if the concentration of the
sample species were changing.
    Burke and his students (11) have published a proposal for
solving the non-linearity problem associated with CC and the
consequent correlation noise. They used a constant frequency
multiple injection signal; while this occurred, this frequency was
modulated. Before each injection, a random number was generated
to determine the magnitude and sign of the deviation from the
carrier frequency for the next injection time. Thus, the next

injection would come a little earlier or a little later than the time dictated by the base frequency governing the multiple injection. By this means, the sample concentration in the column was kept at an almost constant value and, thus, within the linear range of the isotherm. This was an extremely clever proposal, and it is not clear why it did not work better than it did, since correlation noise was still present. It may be that this was due to the hardware implementation of the technique rather than a conceptual problem.

Recognizing the problems associated with CC, allows one to select the application for which it is best suited. Trace analysis using CC is a natural since it is under these very dilute solution conditions that chromatography is apt to be a linear phenonema. However, as we have pointed out a number of times, it only makes sense if the trace material is dissolved in the solvent that we are using as carrier fluid. In gas chromatography, this means ambient air analysis with nitrogen as a carrier gas (12) and in liquid chromatography it probably means water pollution applications with water as a carrier fluid (13), although there may be some other applications in organic solvents which fortuitously can be used as a carrier fluid or which will not change their chromatographic solvent properties when diluted by one-half (remember, in essence, we inject our sample for 50% of the time and pure carrier fluid for the rest of the time).

The model we have presented for CC appears to be a simplified representation, but it is quite rigorous in terms of the correlation process. It can be further refined to yield the common chromatographic peak shapes by convoluting each unit pulse with a Gausian-like function before summing them to give the detector signal. The output of this simple model then looks like that shown in Figure 8. However, the above operation merely makes the appearance of the correlogram more chromatographic-like; it does not make it a model which will mimic the results obtained in the laboratory. The computer model we have used to demonstrate the laboratory experimental results in CC was one we had previously developed to explain the anomolies in finite difference chromatography (14) which we modified to take the encoded sample input of a correlation chromatograph.

For the reader who is interested in pursuing the mathematical details of CC, I recommend the papers of Smit and his co-workers (15). These researchers published early in the area of CC (16) and continue to contribute regularly to the field (17). Phillips has also been active in the field of CC, which he considers to be a subset of what he calls, multiplex chromatography (18). Some workers have used on-line correlation chromatography to study the thermal decomposition of polymers and compared the results against those using conventional injection procedures (19), while others have applied it to the study of gas-solid adsorption (20).

In addition, for those of you who may wish to do some further reading on the general subject of correlation in analytical
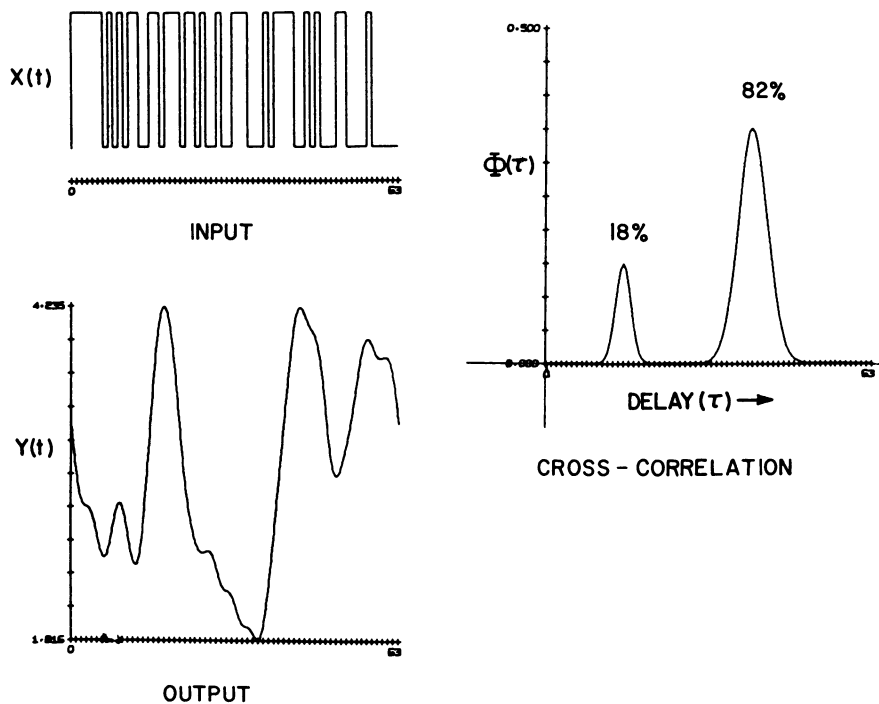
Figure 8. Plots of computer generated representation of the input code, x(t), detector output, y(t), and final correlogram output, $\phi(\tau)$, for a sample containing 18 and 82% concentration of two components. Reproduced with permission from Ref. 9, copyright 1973, "American Chemical Society."

chemistry, the chapter by Horlick and Hieftje (21) is an excellent overview.

Literature Cited

1. Davies, W. D. T. Control 1966, 10, 302, 364, 431.
2. Scholefield, P. H. R. Electronic Tech. 1960, 389.
3. Annino, R.; Grushka, E. J. Chromatog. Sci. 1976, 14, 265.
4. Annino, R.; Leone, J. J. Chromatog. Sci. 1982, 20, 19.
5. Annino, R.; Gonnord, M.-F.; and Guiochon, G. Anal. Chem. 1979, 51, 379.
6. Laeven, J. M.; Smit, H. C.; Kraak, J. C. Anal. Chim. Acta 1983, 150, 253.
7. Valentin, J. R.; Carle, G. C.; Phillips, J. B. J. High Resolut. Chromatog. Chromatog. Commun. 1982, 5, 269.
8. Annino, R.; Bullock, L. E. in "Gas Chromatography 1972"; Perry, S. G.; Adlard, E. R., Eds., Applied Sci. Publisher: London, 1973; pp. 171-186.
9. Annino, R.; Bullock, L. E. Anal. Chem. 1973, 45, 1221.
10. Annino, R. in "Advances in Chromatography"; Giddings, J. C.; Grushka, E.; Cazes, J.; Brown, P. R., Eds.; Marcel Dekker Inc.: New York, 1977; Vol. 15, pp. 33-67.
11. Villalanti, D. C.; Burke, M. F.; Phillips, J. B. Anal. Chem. 1979, 51, 2222.
12. Moss, G. C.; Kipping, P. J.; Godfrey, K. R. in "Gas Chromatography 1972"; Perry, S. G.; Adlard, E. R., Eds.; Applied Sci. Publisher: London, 1973; pp. 187-197.
13. Lub, T. T.; Smit, H. C.; Poppe, H. J. Chromatog. 1978, 149, 721.
14. Annino, R.; Franko, J.; Keller, H. Anal. Chem. 1971, 43, 107.
15. Lub, T. T.; Smit, H. C. Anal. Chim. Acta. 1979, 112, 341.
16. Smit, H. C. Chromatographia 1970, 3, 515.
17. Smit, H. C.; Duursma, R. P. J.; Steigstra, H. Anal. Chim. Acta. 1981, 133, 283.
18. Phillips, J. B. Anal. Chem. 1980, 52, 468A-478A.
19. Kal'yurand, M. R.; Kullik, E. J. Chromatog. 1979, 186, 145.
20. Phillips, J. B.; Burke, M. F. J. Chromatog. Sci. 1976, 14, 495.
21. Horlick, G.; Hieftje, G. M. in "Contemporary Topics in Analytical and Clinical Chemistry"; Plenum Press: New York, 1978; Vol. 3, p. 153.

Developments in Correlation Chromatography
Application in Trace Analysis

H. C. SMIT

Laboratory for Analytical Chemistry, University of Amsterdam, Nieuwe Achtergracht 166, 1018 WV Amsterdam, The Netherlands

The use of the chemometric technique of correlation chromatography has been demonstrated in liquid chromatography. The sensitivity compared to normal liquid chromatography has been extended up to 100 fold. An analysis time period of 2 hours was required. Estimated time for conventional enhancement to a similar signal-to-noise ratio is 50 days. Correlograms very similar in shape to chromatograms of 50 fold higher in component concentration are shown. The advantages of correlation chromatography in ultra trace analysis are unmistakable despite the relatively large amount of sample needed for high enhancement of signal.

The instrumental analytical techniques, developed in the last three or four decades, are almost all based on the limited signal and data processing capabilities of relatively simple analog instruments, and utilize a limited or simple theoretical basis for calculations. Apart from the rather advanced application of statistics, only a modest use of mathematical techniques in analytical chemistry has been used in these traditional analyses.

The computer, with its enormous power in data processing and its possibilities in automation and control, has added a new dimension both to the instrumental analytical method and the application of mathematics and statistics in analytical chemistry. The introduction of the computer was one of the main factors initiating a new analytical subdiscipline, chemometrics, which has a strong mathematical character.

In most applications chemometric methods are applied to analytical data in an off-line mode; that is, data has already been obtained by conventional techniques and is then applied to a particular chemometric method. Examples of this use are in cluster analysis and in pattern recognition. They are applied to spectroscopic, chromatographic, and other analytical data.

A more active kind of chemometrics aims at the integration of statistical and mathematical techniques with the analytical procedure. The conventional analytical process is modified or a completely new process is developed in studying reactions, transport processes, adsorption, absorption, etc. The ultimate aim is to obtain more and better information in an optimum way.

Correlation Chromatography (CC) can be considered a typical example of an active or on-line chemometric technique. Impossible without computers, it shows promising results in (ultra) trace analysis. This paper will describe two directions that utilize correlation techniques: a semi-continuous kind of chromatography (1) and an extension of the limit of detection in trace analysis (2). Correlation Chromatography will be shown to be a powerful method for application in (ultra) trace analysis.

Principals of Correlation Chromatography

Classically, the chromatogram is the response of a chromatographic system that utilizes an impulse or single injection of a sample. Correlation chromatography, on the other hand, utilizes semi-continuous injection of sample over a period of time. An example at this point is needed to describe this comparison more clearly. In normal chromatography an injection is made at a given instant. The introduction of the sample is a single, discreet action. The output of such an injection that has been carried through the separation column is a chromatographic recorder trace which consists of a series of (hopefully) well defined "peaks", where each peak represents (hopefully) a single compound. Hence, a single force or "impulse" of injection produces a response of a single peak (for each compound) which could be called an "impulse response". It is essentially single impulse chromatography.

In contrast, correlation chromatography is multiple impulse chromatography. It utilizes sample that is discreetly added many times in a random way . Since any number and length of injections can be made in liquid chromatography before the compound of interest may elute from the column, the response (the total signal) of such chromatography is a massive group of fused peaks that looks like a lot of noise often with a greatly raised baseline. See Figure 1. To the naked eye it is impossible to visualize separated peaks. However, to the computer which knows the injection function, the output response has a lot of sense. It can resolve a peak from noise and produce a "correlogram" which is very similar to a normal chromatogram. The longer the system is run, the larger the sought peaks will be. In trace analysis the result is the detection of trace compounds otherwise not attainable by impulse techniques. The cost of such work is the larger amount of sample needed (in mL instead of microliter quantities) and a longer analysis time than from single impulse chromatography.

Of course, the impulse response of a system can be determined by measuring the response on an impulse-shaped input signal, but an alternative way is to determine the cross-correlation function of a suitable stochastic (random) input signal and the resulting output. Omitting the mathematical proof, which is given in ($\underline{2}$), we will describe this process by following the reasoning given in ($\underline{3}$).

Some basic definitions are necessary. The definition of a cross-correlation function (CCF) of two non-zero average power signals, $x(t)$ and $y(t)$, is:

$$R_{xy}(\underline{\tau}) = \lim_{T \to \infty} 1/T \int_{-T/2}^{T/2} x(t-\underline{\tau})y(t)dt \qquad (1)$$

The cross-correlation function $R_{xy}(\tau)$ for two signals, $x(t)$ and $y(t)$, describes the general dependence (correlation) of the amplitude of one signal to the other as a function of the time displacement $\underline{\tau}$. For example, $x(t)$ can be the input signal and $y(t)$ the resulting output signal of a system. The correlation between the input and the output signal is determined by the properties of the system. If in this system there exists only a pure delay, τ_d, without affecting the signal (mathematically $y(t) = x(t - \tau_d)$), then of course the maximum correlation is found at $\tau = \tau_d$.

The definition of $R_{xy}(\tau)$ is given for non-zero average power signals, i.e., signals theoretically not limited in time, like noise and periodic signals. Peaks and other time limited signals are zero average power signals. $R_{xy}(\tau)$ is the average product of signal $y(t)$ and a time τ delayed version of signal $x(t)$.

The autocorrelation function (ACF) of a non-zero average power signal $x(t)$ is defined by

$$R_{xx}(\underline{\tau}) = \lim_{T \to \infty} 1/T \int_{-T/2}^{T/2} x(t-\underline{\tau}) \, x(t)dt \qquad (2)$$

$R_{xx}(\tau)$ is the average product of $x(t)$ and a time $\underline{\tau}$ delayed version of $x(t)$. The autocorrelation function is a basic function in the characterization of a stochastic signal. Considering the ACF as a function of τ of a relatively fast fluctuating (stationary) stochastic signal with an average value equal to zero, one can observe a much faster decrease of the ACF compared with an identical but slowly fluctuating stochastic signal.

It is impossible to predict the amplitude of a <u>stochastic</u>
signal at a certain time in the future in contrast to a
<u>deterministic</u> signal like a sine wave. Only a statistical
<u>description,</u> for instance by distribution functions and
autocorrelation functions, can be given. Most kinds of noise have
a stochastic character.

An ACF is always an even function, symmetrical with respect
to τ=0. The fast decreasing ACF of a very fast fluctuating
stochastic signal can be considered as an impulse.

In correlation chromatography a special kind of a stochastic
input signal is used, a binary noise. In binary noise only two
amplitude levels can occur, high or low (see Figure 1). Neverthe-
less, it is a stochastic signal because it is unpredictable which
of the two levels will be present at a certain time in the future.
On the average each of the two levels has a probability of 0.5.
Generally, the binary noise is generated artificially by a
generator controlled by an internal clock. The clock period $\underline{\Delta}$t
determines the minimum time that one of the two states will exist.
During the determination of the ACF $R_{xx}(\tau)$ of a binary noise
with amplitude levels of +1 and -1, if the time shift τ is greater
than the clock period $\underline{\Delta}$t, then the <u>average</u> product of x(t) and
x(t - τ), being $R_{xx}(\tau)$, will be zero; the probability of each
of the products +1 and -1 is 0.5. However, if τ <Δt, then the ACF
as a function of τ is linearly decreasing from 1 (τ=0) to 0 (τ=$\underline{\Delta}$t
or τ=-$\underline{\Delta}$t). Hence, the ACF is a triangle with a base width of two
times the clock period.

A special kind of random noise, pseudo random noise, has the
special property of not being really random. After a certain time
interval, a sequence, the same pattern is repeated. The most
suitable random input function used in CC is the Pseudo Random
Binary Sequence (PRBS). The PRBS is a logical function, that has
the combined properties of a true binary random signal and those
of a reproducible deterministic signal. The PRBS generator is
controlled by an internal clock; a PRBS is considered with a
sequence length N and a clock period Δt. It is very important to
note that the <u>estimation</u> of the ACF, if computed over an integral
number of <u>sequences,</u> is exactly equal to the ACF determined over
an infinite time.

Figure 1 shows an example of a short PRBS (N = 15 clock
periods) and the procedure to determine a point of the ACF. The
product function (shaded area) is integrated.

Figure 2 shows a basic diagram of a correlation
chromatograph. The input of the column is connected in turn with
a sample stream and an eluent stream, controlled by a suitable
PRBS input pattern. Cross-correlation of the PRBS and the
resulting detector output gives the impulse response of the
system, in this case the chromatogram. The following reasoning
makes this plausible:

Suppose we have an ideal chromatographic column, no
dispersion, and a linear partition isotherm. The result is only a
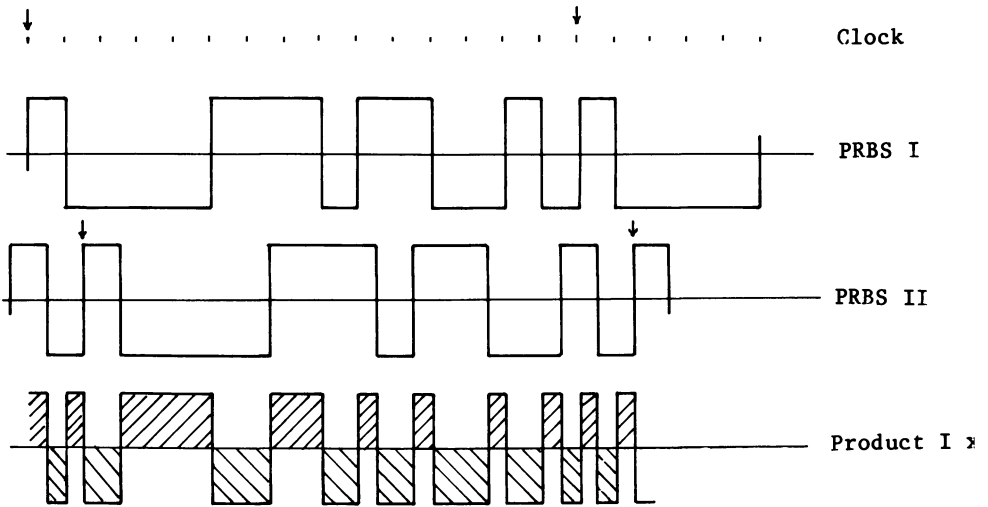
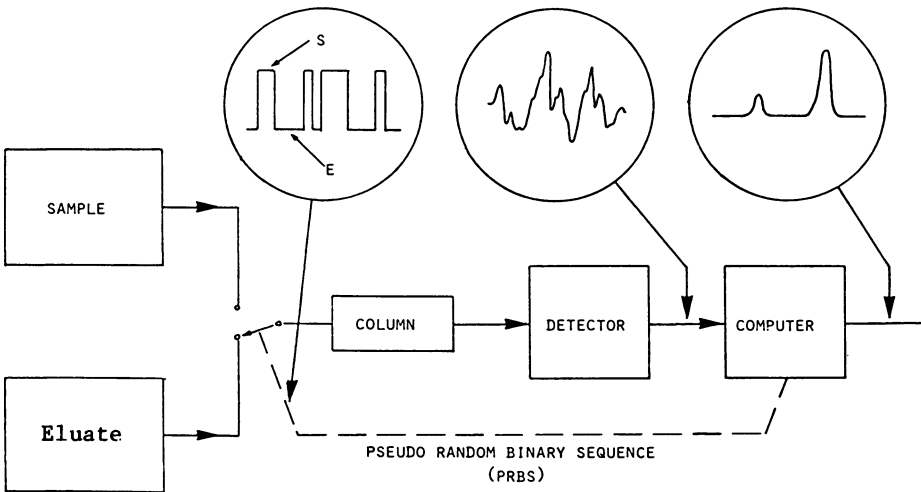Figure 1.   Example of a PRBS. Determination of one point of
R(τ).

Figure 2.   Basic diagram of a correlation chromatograph.

delay for every input pattern. A conventionally pulse-shaped
injection (one component) causes a delayed pulse at the detector;
the delay time is t_R. Of course the result of a PRBS input
(one component) is a PRBS output, t_R seconds delayed. In that
case the cross-correlation function (CCF) of input and output is
identical to the ACF of the PRBS, however, t_R seconds shifted
in time. The ACF in this case is again a triangle, however not at
$\tau=0$ but at $\tau=t_R$. If the clock period is small compared with
the delay time, this triangle can be considered as an impulse.
Hence, the correlogram is identical with the chromatogram. In
both cases the amplitude is proportional to the input
concentration of the component.

A sample with $\underline{n}$ components, each with a certain concentration
and its own retention time, results in $\underline{n}$ summed PRBS functions at
the detector output, each with an amplitude dependent on the con-
centration. Cross-correlation gives again a correlogram similar
to the chromatogram. Considering a normal chromatogram to be made
up from $\underline{n}$ points, each representing a "component" with a certain
amplitude and retention time, leads by the same reasoning to the
same conclusion, a correlogram is identical to the chromatogram.

The PRBS is to be preferred to other random inputs with
approximately impulse-shaped autocorrelation functions for the
following reasons:
1) It is a binary noise with only two levels (+1 and -1 or 1 and
0, respectively). The levels can be used to control simple on/off
valves.
2) The function can be easily generated and reproduced.
3) Its special properties offer the possibility of reducing the
so-called correlation noise, caused by a limited correlation time.
CC is essentially statistical by nature. The system noise
(detector noise) is not correlated with the input PRBS; the noise
in the correlogram, resulting from the detector noise, is
converging to zero with increasing correlation time.

Set Up of a Correlation Chromatograph

A correlation chromatograph requires only a modification of the
injection system of a conventional chromatograph. Figure 3 shows
a design for a correlation high pressure liquid chromatography
system. It is suitable for high pressure up to 500 bar and for
use with corrosive samples. An extensive description is given in
($\underline{4}$). This chromatograph is intended for use in ultra trace
analysis and research in CC. Apart from the injection system,
essential extensions of a chromatographic set-up to allow CC that
are required are:
a) The pattern generator which is necessary for generating either
a single pulse (normal chromatography) or a PRBS adapted to the
sample (CC). It is used for stimulation of the column via
switching of valves and for calculation of the CCF from the
detector output and the pattern.

b) The data sampler, which is used to sample the filtered electric detector signal and to convert it to a digital value.
c) An arithmetic unit to calculate the CCF.
d) A display for the results.

Some extensions are not essential for CC, but greatly improve its capabilities. Interfaces to a data storage device and to a hard copy unit are valuable. Some facility for data processing afterwards (baseline correction and peak area determination) is desirable.

A microprocessor is ideally suited for CC. In our laboratory a microprocessor based instrument, a correlator, has been developed which meets all the mentioned requirements for CC. Details can be found in (5).

Results in Trace Analysis

The first experiments directed to trace analysis were carried out in correlation gas chromatography (2). However, in the recent years much attention was paid to correlation HPLC, because the detection is generally more a problem than in GC and because injection is inherently easier. Results with a first experimental set-up and an off-line computer calculation of the CCF were very promising.

Table I. Listing of Solutes Present in the Chromatogram (Fig. 4).

Peak No.	Solute	Capacity Ratio	Concentration (ppm)
1	impurities, THF	-	-
2	2,3-dichlorophenol	5.52	10.5
3	2,6-DCP	6.34	12.6
4	3,4-DCP	7.07	10.9
5	2,5-DCP	7.78	10.0
6	2,3,4-trichlorophenol	8.53	10.1
7	2,3,6-TCP	9.70	11.8
8	3,5-DCP	10.52	9.9
9	3,4,5-TCP	11.94	10.4
10	2,4,6-TCP	12.76	10.0
11	2,3,4,5-tetrachlorophenol	15.52	10.1
12	2,3,5,6-TCP	17.21	10.3
13	pentachlorophenol	24.96	10.4

A lowering of the detection limit by a factor of 100 was
achieved in the analysis of phenol and dimethylphenol (6). The
efficiency of CC compared with well known signal enhancement
methods was demonstrated in these experiments. The total
correlation time, required to achieve a factor of 100 lowering of
the detection limit, was about two hours. Signal enhancement
(summing of a number of chromatograms) would have required about
1200 hours or 50 days (!) to achieve the same result. In the
signal enhancement technique the signal increases proportionally
to the number of summed chromatograms. However, the noise
increases only by the square root of the number of summations.

The microprocessor-based correlator has been used in a
correlation "separation" of a more complex mixture (7). Figure
4 shows a plot of a conventional chromatogram representing the
HPLC separation of twelve chlorinated phenols. The concentration
of each component is about 10 ppm or 10 mg/L (see Table I).

Figure 5 represents a correlogram of this analysis obtained
with the correlator and a modified HPLC system. The concentration
of each component is only 0.2 ppm, an enhancement of 50. In
Figure 6 the response trace leading to the correlogram of Figure 5
is shown.

CC is essentially a differential method; negative peaks
caused by components in the eluent and positive peaks caused by
components in the sample can both be present. Apart from minor
differences in resolution, caused by a slight modification in
separation conditions (column temperature), two extra peaks appear
in the correlogram compared with the chromatogram. The small peak
at 138 seconds in the correlogram is presumably an isomer. The
large peak at 35 seconds was probably a late-eluting peak from a
previous injection or a peak that was folded due to incorrect
correlogram duration time. The periodicity of the PRBS caused an
overlap in the correlogram if the length of the chromatogram
exceeds the length to the chosen PRBS.

Figure 7 shows an enlarged part of a complex correlogram,
displayed after correlation by the correlator. The area of one of
the peaks is shown.

The analytical performance of CC is demonstrated by extending
the calibration graph of measured phenol with data found by CC.
See Figure 8. HPLC equipment using reversed phase separation with
fluorometric detection was used. Injection for correlation HPLC
was accomplished with a newly developed device (4). The total
range of phenol measured was five decades of concentration: 0.01 -
100 μg/L. The two higher concentrations (10 - 100 μg/L) were
determined solely by conventional methods.   The two lower
concentrations (0.01 - 0.1 μg/L) were determined by correlation
HPLC with 16 and 3 sequences correlation time, respectively.
Measurements at the 1 μg/L level were carried out both by
conventional and by correlation HPLC (1 sequence). The bars
indicated on the calibration graph represent the peak area +3
σ_I (arbitrary units), where σ_I is the standard deviation
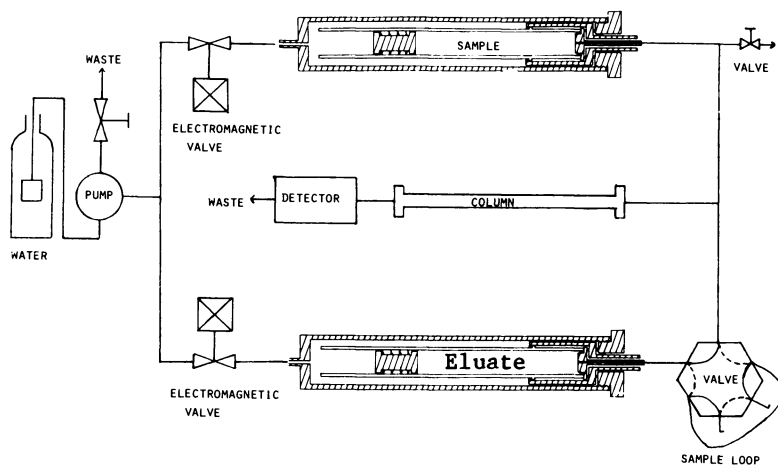
Figure 3. Set-up of a correlation HPLC system. The constant water flow is controlled by a PRBS pattern which directs the flow to either the sample or the eluent reservoir driving the appropriate plunger forward. A 6-way rotary valve is placed at the outlet of the eluent reservoir to allow single injection experiments.
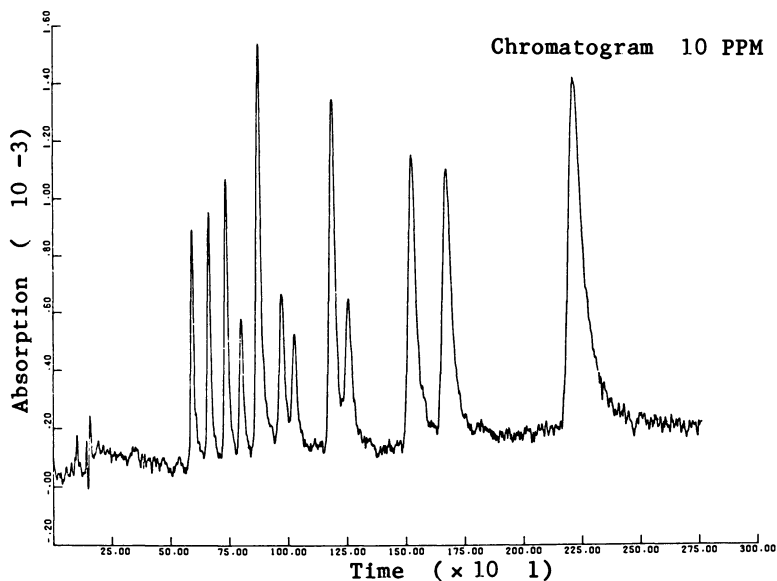

Figure 4. Separation of twelve different chlorinated phenols by conventional HPLC. The solutes are listed in Table I.
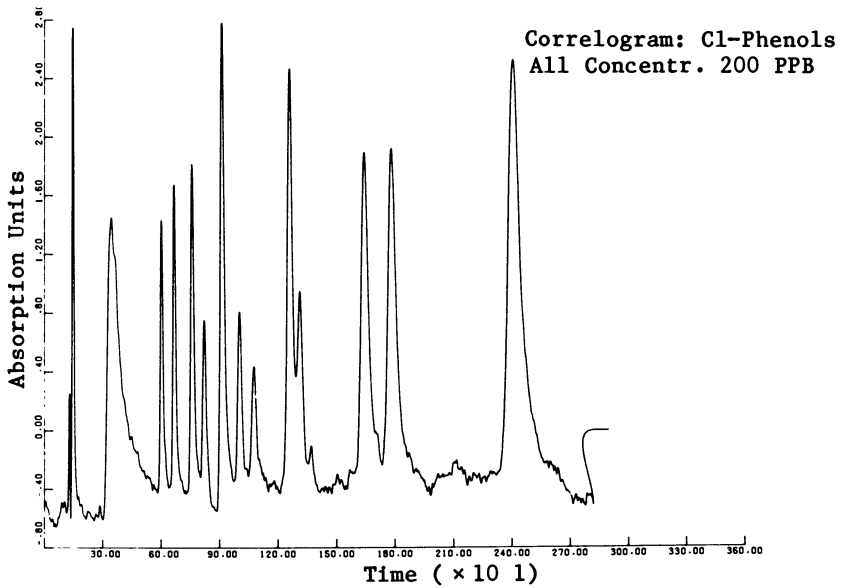
Correlogram: Cl-Phenols
All Concentr. 200 PPB

Figure 5.   Correlogram corresponding to Figure 4 with
slightly different separation conditions.   The concentration
of each component is 0.2 ppm.
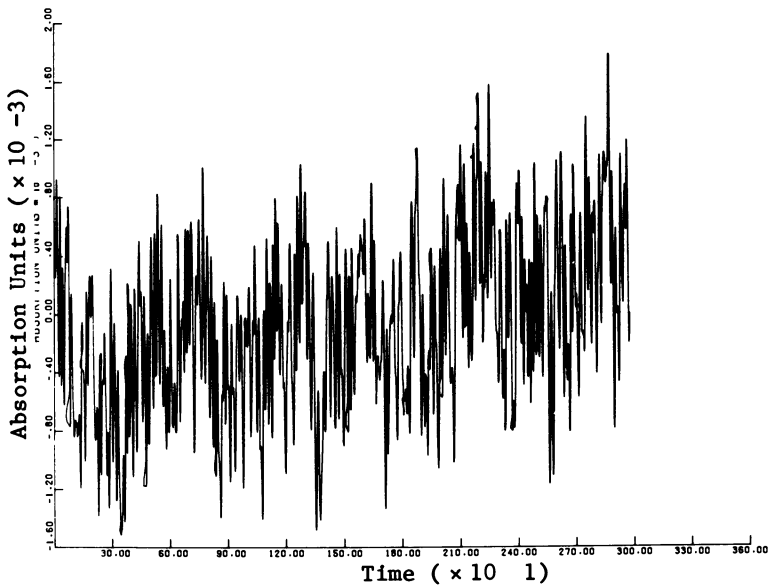
Figure 6.   Detector output during correlation leading to the
correlogram of Figure 5.

Area =   11.44141 mV·s

Sigma-I =   1.46756 mV·s

                                     595.20

                                           .41669
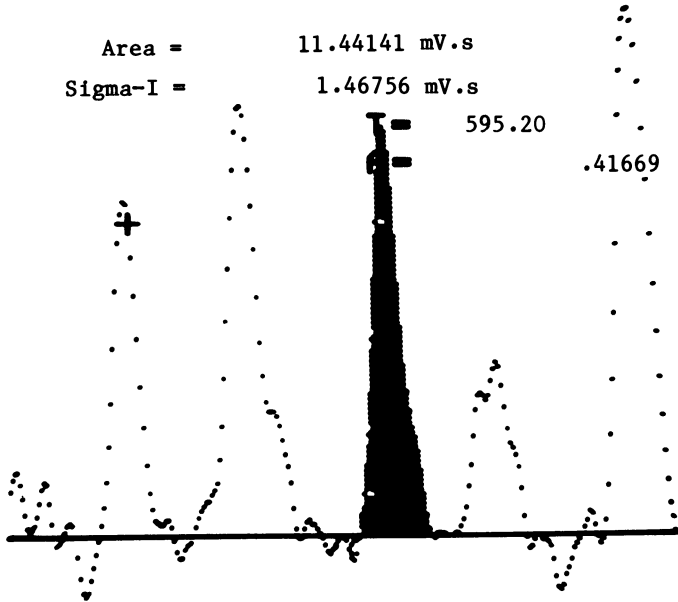
Figure 7.   Enlarged part of a correlogram displayed on a
TV-screen.
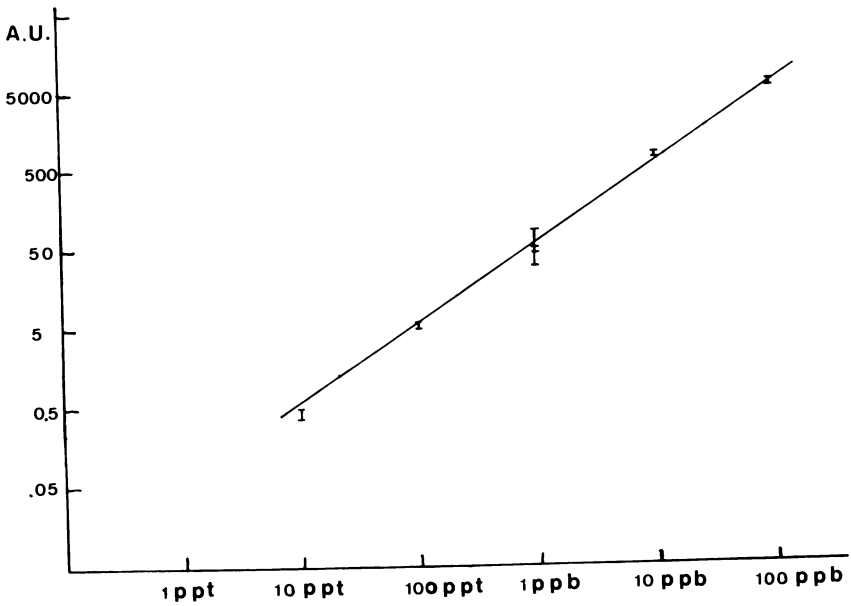
Figure 8.   Calibration graph of phenol with fluorometric
detection.

of the integrated signal plus noise (8). The inner bars at the
1 g/L level represent the correlation results and the outer bars
the single injection results. The detection limits of the single
injection experiments and of the correlation procedure with 10
ng/L concentration, both defined as 3 σ_I, were estimated to be
about 500 ng/L and 3 ng/L (or 3 ppt, parts per trillion),
respectively (Figure 9).
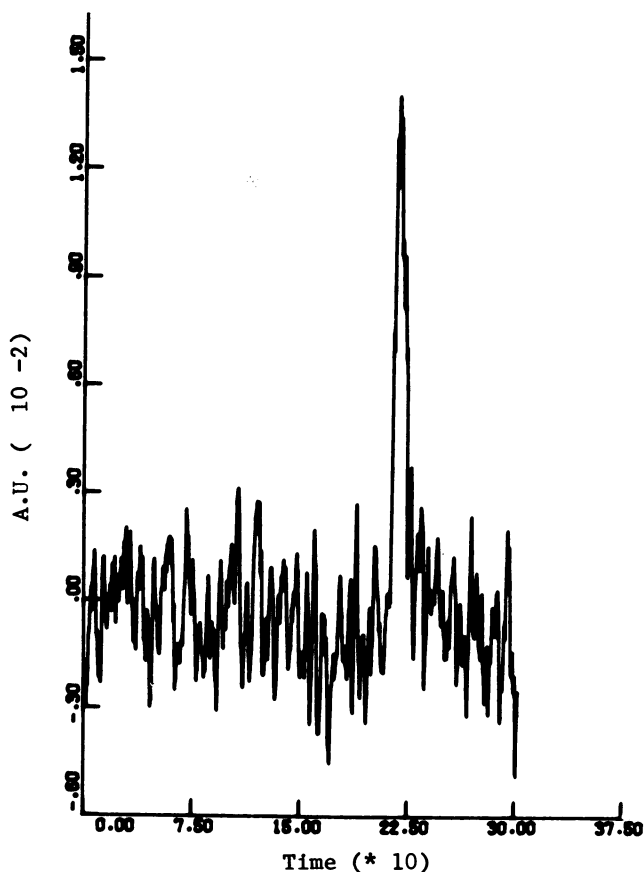
Figure 9.   Correlogram of a 10 ng/L phenol sample.
Detection limit for 80 min correlation time is approximately
3 ppt (3.10 ng/L).

Discussion and Conclusions

The advantages of CC in ultra trace analysis are shown to be unmistakable. The quantitative reliability of the method was demonstrated by the extension of a calibration graph for phenol to two decades of concentration more when compared with conventional chromatography. A considerable improvement of the signal-to-noise ratio can be achieved in a relatively short time. The method offers excellent prospects for ultra trace analysis in cases where preconcentration of the solute fails.

More research has to be done. One area is the influence of non-linearity which causes an extra baseline noise in the correlogram. Injection systems for other kinds of chromatography, in particular GC, have to be developed. The ability to detect very small differences between the solute concentrations in the background and the sample has to be investigated.

Literature Cited

1. Izawa, K.; Furuta, K.; Fujiwara, F.; Suyama, N. Ind. Chim. Belge 1967, 32 , 223.
2. Smit, H. C. Chromatographia 1970, 3 , 515.
3. Smit, H. C. Anal. Chim. Acta 1980, 122 , 201.
4. Laeven, J. M.; Smit, H. C.; Kraak, J. C. Anal. Chim. Acta 1983, 150 ,253.
5. Smit, H. C.; Duursma, R. P. J.; Steigstra, H. Anal. Chim. Acta 1981, 133 , 283.
6. Lub, T. T.; Smit, H. C.; Poppe, H. J. Chromatog. 1978, 149 , 721.
7. Smit, H. C.; Lub, T. T.; Vloon, W. J. Anal. Chim. Acta 1980, 122 , 267.
8. Duursma, R. P. J.; Smit, H. C. Anal. Chim. Acta 1981, 133 , 67.

Calibration-Curve-Based Analysis

Use of Multiple-Curve and Weighted Least-Squares Procedures with Confidence Band Statistics

DOUGLAS G. MITCHELL

Center for Laboratories and Research, New York State Department of Health, Albany, NY 12201

Two procedures for improving precision in calibration curve-based-analysis are described. A multiple curve procedure is used to compensate for poor mathematical models. A weighted least squares procedure is used to compensate for non-constant variance. Confidence band statistics are used to choose between alternative calibration strategies and to measure precision and dynamic range.

This paper describes the use of statistical techniques to improve precision in routine chemical analysis at a modest extra cost and to measure the precision of such analyses.
Note the key words in this aim:

Precision, not accuracy. Accuracy is mainly a chemical problem, whereas precision is a chemical, instrumental and statistical problem.

Routine chemical analysis. This implies analysis of many samples, and use of calibration curves is an economic necessity. In general, the two-standard method, with standards bracketing each sample analyzed, is economical for the analysis of up to about 10 samples. Conventional least squares curve of best fit procedures are economical for analysis of 10 to 500 samples. The procedures described here are cost effective for the analysis of 500 samples or more.

Cost. There is always a trade-off of cost versus data quality. Data quality can be improved by further method development, more extensive calibration, replicate analysis or better statistics.

Measurement of precision. Measurement of data quality is valuable for both the analyst and the data user. Least-squares curve-of-best-fit statistical programs usually provide some information on precision (correlation coefficient, standard error of estimate). However, these are not sufficiently quantitative and often overstate the quality parameters of the data.

To provide optimum data in routine chemical analysis, the analytical method must meet four criteria. The first two are

fundamental and determine the upper limit of method performance.
The last two concern calibration.
     1. The method must be precise. (Without precision we
cannot get accuracy, unless we carry out many replicate
analyses.)
     2. The method must be accurate, or at least free from
unpredictable bias.
     3. The calibration process must not excessively degrade
precision. (Use of a calibration curve will usually result in
less precise analyses, compared with bracketing each sample
measurement with standard measurements.)
     4. The calibration curve must be stable, i.e. free from
drift.
     This paper is concerned with the effects of the
calibration process on data quality.

Risks to Data Quality in the Calibration Process

Over a limited dynamic range, say 1 to 10, use of least squares
procedures has little adverse effect on data quality. With
calibration over wide dynamic ranges and with non linear curves,
precision is lost because:
     The mathematical model may not closely fit the data. For
example, Figure 1 shows calibration data for the determination of
iron in water by atomic absorption spectrometry (AAS). At low
concentrations the curve is first-order, at high concentrations
it is approximately second-order. Neither model adequately fits
the whole range. Figure 2 shows the effects of blindly fitting
inappropriate mathematical models to such data. In this case, a
manually plotted curve would be better than either a first- or
second-order model.
     Calibration curves yield the best precision at the mean
concentration of the standards. For example, a curve based on
standard with concentrations of 1, 4 and 10 yields best precision
at 5 (assuming constant variance). To achieve maximum precision
the standards should be selected so that their mean concentration
is equal to the most important sample concentration, such as an
action level. The curve will yield increasingly poor precision
with increasing distance from this mean.
     The least-squares curve-of-best-fit procedure implicitly
assumes the same variance (standard deviation) at all concentra-
tions. This assumption is rarely correct. Figure 3a shows
hypothetical replicate standard analysis data with constant
variance. This pattern is almost never seen in routine chemical
analyses. Figure 3b shows a much more realistic pattern in which
the variance increases with concentration.

## Proposed Solutions to Calibration Problems

There are several approaches to minimizing the loss of precision
inherent in calibration-curve-based analysis. We have chosen a
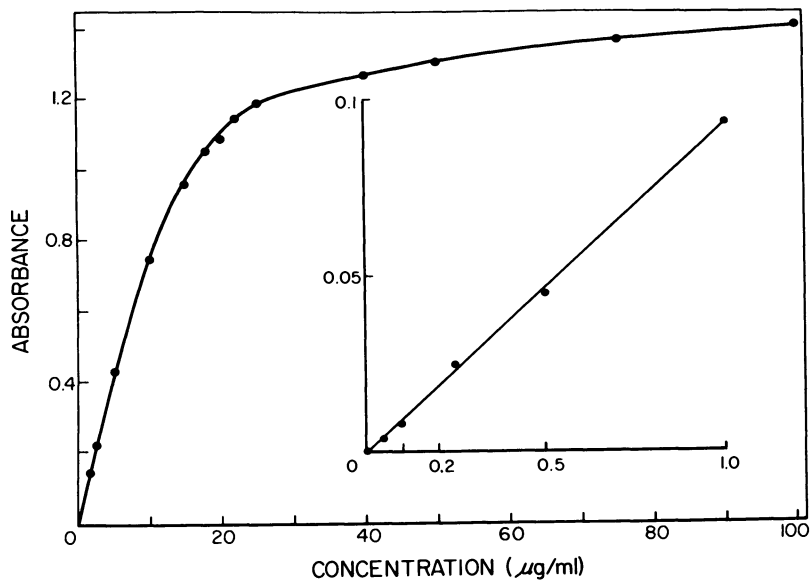
Fig. 1.
Calibration data for determination of iron in water by AAS
(Reprinted with permission from D. G. Mitchell and J. S. Garden,
Talanta, 1982, <u>29</u>, 921–929.)
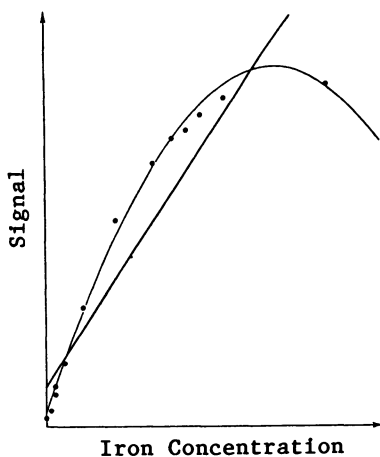
Fig. 2.
Inappropriate use of first–and second–order least
squares–curves–of– best–fit procedures with part of the data
shown in Figure 1.
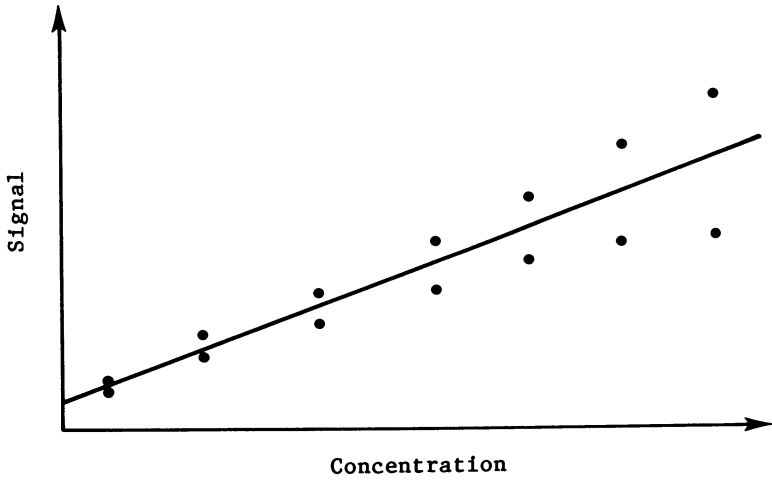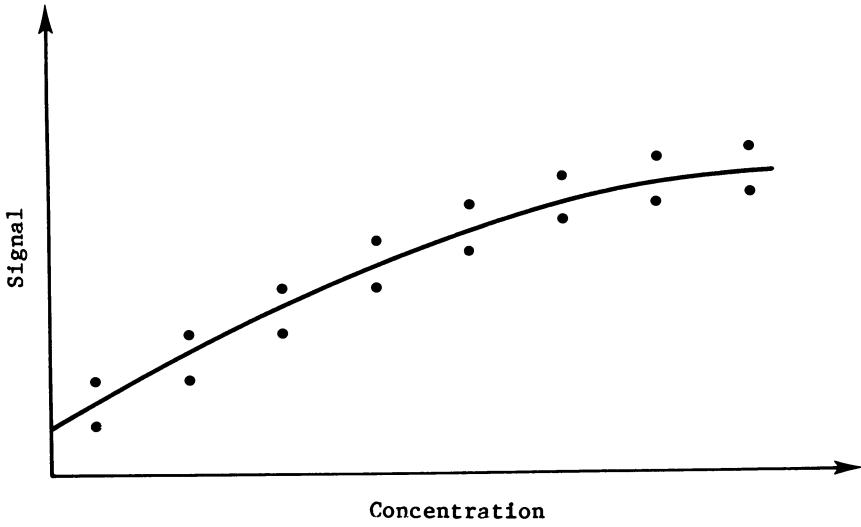
Fig. 3.
Hypothetical calibration data showing replicate standard analyses
with (a) constant and (b) nonconstant variance.

hybrid calibration-curve, two-standard method for calibration
over wide dynamic ranges and a weighted least-squares procedure
to compensate for nonconstant variance. Commonly used measures
of precision such as correlation coefficients, standard errors of
estimate and relative standard deviations (RSD) cannot reliably
evaluate alternative calibration strategies. This is beacause
they do not measure precision in terms of predicted sample con-
centration. For example, a correlation coefficient (r) of 0.98
is better than a value of 0.97 providing the same standards are
used. If different standards are used, the high r value curve
will not necessarily yield best precision for sample analysis.
Before proceeding further, it is necessary to choose appropriate
statistical techniques.

Confidence Band Statistics. The confidence-band statistical
approach is described in texts by Natrella (1) and Miller
2) and in three papers from our laboratory (3-5). A computer
program, REGRES, (See Appendix) was used to carry out all the
computations described in this paper.
    The confidence band approach is illustrated in Figure 4.
An appropriate mathematical model is chosen. This is usually a
first order (signal = b_0 + b_1 x concentration) or second order
(signal = b_0 + b_1 x concentration + b_2 x concentration^2)
linear equation. A calibration curve is then calculated using a
least-squares curve-of-best fit procedure (6). Next a
confidence band is calculated around the curve using the regres-
sion band equation from Table I. This band encloses the curve
with a, say, 90% level of probability. A confidence band is then
calculated around the signal using the appropriate signal band
equation (Table I). The two bands are combined as shown in
Figure 4 to yield a confidence band around the predicted concen-
tration. The resulting band around the predicted concentration
gives a conservative estimate of the precision of the analysis,
including the effects of error in both sample and curve.
    There is a debate among statisticians concerning the best
procedures for this application, and our approach may be too con-
servative. To some extent a conservative bias is probably an
advantage, because it could approximately compensate for (uncal-
culable) errors due to minor inaccuracies inherent in many
methods. Even if it is too conservative, the bands will be self-
consistent and should provide accurate estimates of relative
precision.
Improved mathematical models. First or second order linear
equations adequately fit much calibration data. If neither model
is appropriate, the following semi-empirical multiple curve
procedure may be used.
    Standards covering the proposed dynamic range are
analyzed, and the resulting calibration data entered into program
REGRES. Each sample is analyzed, and REGRES chooses the combina-
tion of contiguous standards enclosing the sample which yields
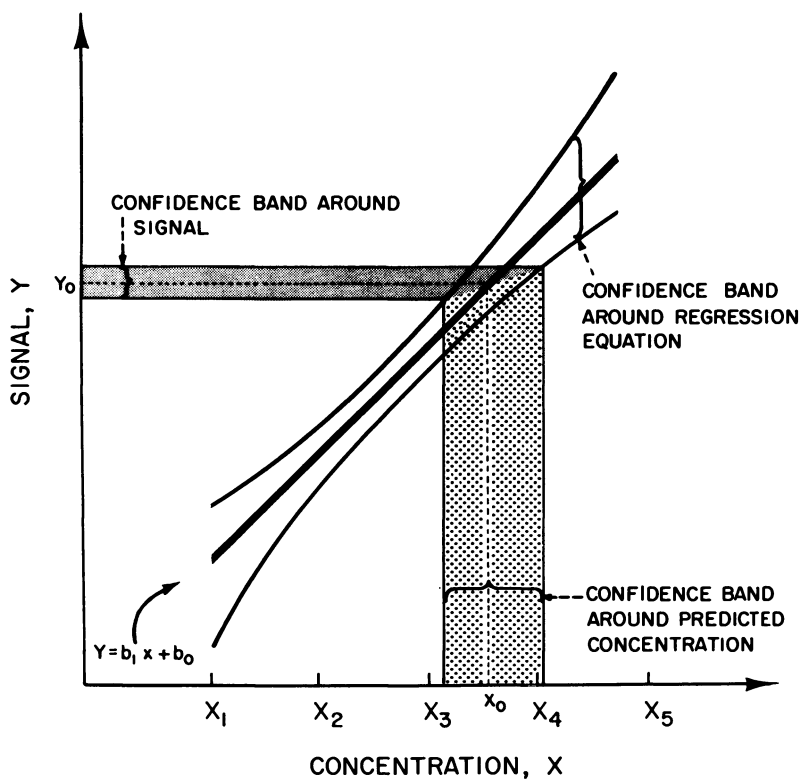
Fig. 4.
Calibration curve with confidence bands around the curve, sample
signal, and predicted concentration. (Reproduced with permission
from D. G. Mitchell, W. N. Mills, J. S. Garden, and M. Zdeb,
Anal. Chem, 1977, 49, 1655-1660, copyright 1977, American
Chemical Society)

Table I. Algebraic Equations for First-Order Regression Calculations

Parameter	Variable-variance data	Constant-variance data
Model	$Y = b_0 + b_1 X + \epsilon; \epsilon \sim N(0, \sigma^2/\omega)$	$Y = b_0 + b_1 X + \epsilon; \epsilon \sim N(0, \sigma^2)$
Mean, $\bar{X}$	$\Sigma\omega X/\Sigma\omega$	$\Sigma X/n$
Intercept, b_0	$(\Sigma\omega Y \cdot \Sigma\omega X^2 - \Sigma\omega X \cdot \Sigma\omega XY)/[\Sigma\omega \cdot \Sigma\omega X^2 - (\Sigma\omega X)^2]$	$(\Sigma Y \cdot \Sigma X^2 - \Sigma X \cdot \Sigma XY)/[n \cdot \Sigma X^2 - (\Sigma X)^2]$
Slope, b_1	$(\Sigma\omega \cdot \Sigma\omega XY - \Sigma\omega X \cdot \Sigma\omega Y)/[\Sigma\omega \cdot \Sigma\omega X^2 - (\Sigma\omega X)^2]$	$(n \cdot \Sigma XY - \Sigma X \cdot \Sigma Y)/[n \cdot \Sigma X^2 - (\Sigma X)^2]$
Signal band, ΔY	$\pm Z \left[\dfrac{n-2}{\chi^2_{\alpha/2}} \cdot \dfrac{m}{n-2} \right]^{1/2} S_0$	$\pm Z \left[\dfrac{n-2}{\chi^2_{\alpha/2}} \cdot \dfrac{m}{n-2} \right]^{1/2} S$
Regression band	$(2F)^{1/2} s \left[\dfrac{1}{\Sigma\omega} + \dfrac{(X_0 - \bar{X})^2}{\Sigma\omega(X_i - \bar{X})^2} \right]^{1/2}$	$(2F)^{1/2} s \left[\dfrac{1}{n} + \dfrac{(X_0 - \bar{X})^2}{\Sigma(X_i - \bar{X})^2} \right]^{1/2}$

Weights: $\omega_i = 1/s_i^2$; s_i = standard deviation at concentrations i. n = number of standard measurements m = number of replicate sample measurement. Standard error of estimate: $s^2 = SS/(\underline{n - 2})$. Sum of squares of residual: $\Sigma\omega Y^2 - b_0 \Sigma\omega Y - b_1 \Sigma\omega XY$, $\Sigma Y - b_0 \Sigma Y - b_1 \Sigma XY$. Band around predicted concentrations: substitute Y values (mean signal $\pm \Delta Y$) in $Y = b_0 + b_1 X +$ band around regression. Solve for X_0.
$\alpha = 0.05$; $z_{1-\alpha/2} = 1.96$.
(Reprinted with permission from Ref. 4.)

the narrowest band around the predicted concentration. Figure 5
shows typical curves selected by this procedure. At low concen-
trations a first-order equation based on, say, the three lowest
standards is chosen. At high concentrations a second-order
equation yields the narrowest band. Note that this procedure
often does not use all available data, an omission which seems
intuitively incorrect. The procedure will improve precision when
the benefits from better mathematical modeling exceed the losses
from not using some data. In general, the multiple-curve proce-
dure produces maximum benefits at the low-concentration end of a
long, nonlinear curve. For example, in typical data for the
determination of fenvalerate by gas chromatography (Table II), use
of the multiple-curve procedure improved the precision of the
analysis by a factor of two at the 1 mcg level, and a factor of
three at the 5 mcg level.

Correction for nonconstant variance. To correct for noncon-
stant variance, it is necessary to weight standard measurements
according to their local variance, S^2. For each standard con-
centration the variance is determined by repetitive analysis at
that level, and a weighting factor, $\underline{w} = 1/s^2$, is calculated.
These factors are used in the equations given in Table I. The
computation requires only that the variance $\underline{ratios}$ be
accurately known. The absolute precision of the method may
change from day to day without affecting the validity of either
the least-squares curve-of-best fit procedure or the confidence
band calculations. (It is not practical to regularly monitor
local variances, and errors may develop in variance ratios.
However, the error due to incorrect ratios will almost always be
much less than the error due to assuming constant variance. Even
guessed values of, say, S^2 α concentration are likely to yield
more precise data.)

An unweighted least squares procedure is often adversely
affected by high concentration standards, with high (absolute)
variances. These may cause large errors in the slope of first
order equations. The line is 'rotated', causing large
$\underline{relative}$ errors at low concentrations. The weighting proce-
procedure deemphasizes these points, thus reducing this effect.
Figure 6 shows data for the determination of lead in blood by
Delves cup AAS. The first-order curve is known to pass through
zero. The weighted least-squares line is close, with an   4
intercept of 1.5, but the unweighted line has been ''rotated''
by a single low value (not an outlier) at 65 µg/dL, giving an
incorrect intercept of 3.3. A sample yielding a signal of 5.3
has a calculated lead concentration of 10 mcg Pb/dl using the
weighted least squares line and 6 mcg/dl using the unweighted
line - a 40% error. Similarly, figure 7 shows part of the
weighted and unweighted least-squares curves for symposium
Dataset B. Standards over the range 0.05 to 20 ng fenvalerate
were analyzed, and the figure shows a range of only 0-1. The
variance at each amount level was known, so both weighted and
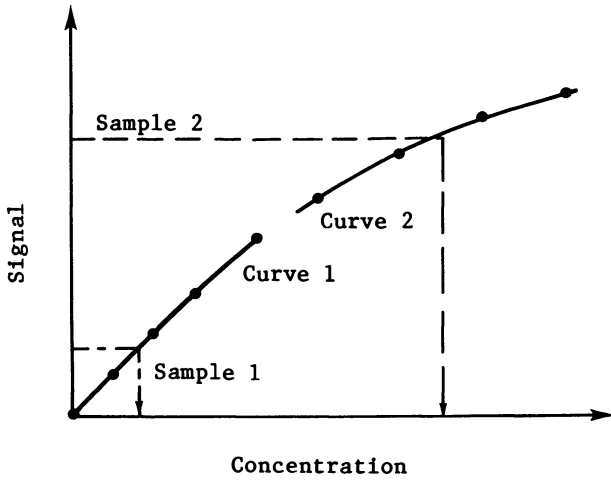
Fig. 5.
Use of multiple-curve procedure. Subsets of calibration data,
each comprising several standards bracketing the samples, are
used to calculate predicted concentrations for high-and
low-concentration samples.
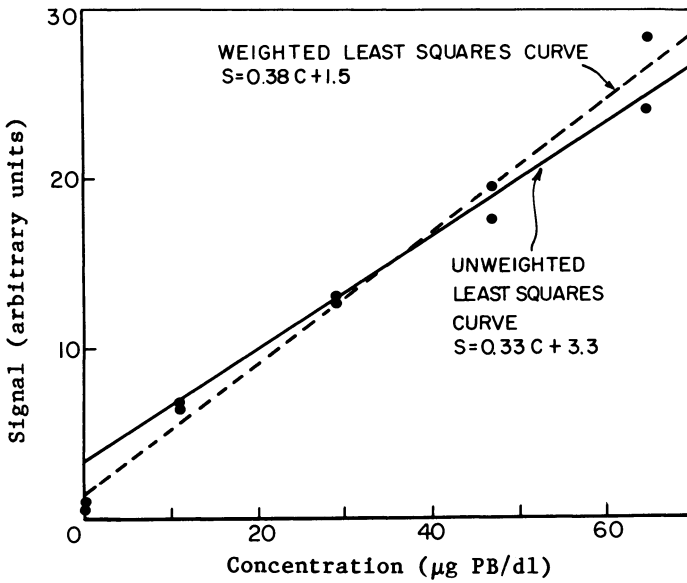
Fig. 6.
Determination of lead in blood by Delves-cup AAS. Both weighted
and unweighted curves of best fit are shown.

Table II. Use of multiple curve procedure to improve precision
           of fenvalerate analysis by gas chromatography

Fenvalerate Amount (mcg)	Single-Curve Calibration RCB (%)	Multiple-Wave Calibration Calibration Range		
		Low	High	RCB%
0.05	40	0.05	1.00	24
0.25	33	0.05	1.00	20
1.00	33	0.05	5.00	17
5.00	26	1.00	20.00	7
20.00	20	1.00	20.00	7

Note: $\alpha = 0.05$; $Z_{1-\alpha/2} = 1.96$
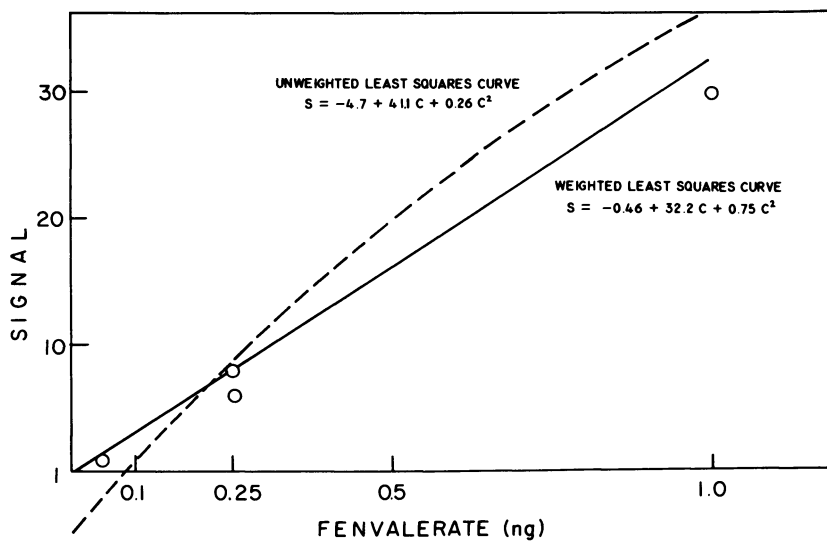
Fig. 7.
Determination of fenvalerate by gas chromatography with DATASET B
showing weighted and unweighted second-order curves.

unweighted curves could be calculated. These data are very
precise, with five replicate measurements at 20 ng fenvalerate
having a range of ± 2.6%. Use of the unweighted procedure
caused significant errors only at amount levels below 1 ng
fenvalerate.

Improved Measurement of Precision

Calibration curve quality. Calibration curve quality is
usually evaluated by statistical parameters, such as the correla-
tion coefficient and standard error of estimate, and by empirical
indexes, such as the length of the linear range. Using confi-
dence band statistics, curve quality can be better described in
terms of confidence band widths at several key concentrations.
Other semi-quantitative indexes become redundant. Alternatively,
the effects of curve quality can be incorporated into statements
of sample analysis data quality.
Sample analysis data quality. Precision of sample analysis
is almost always measured by determining the RSD at two or more
concentrations without using a calibration curve. Such data do
not include the effects of the calibration process on precision.
Much better information is given by the relative confidence
bandwidth (RCB) defined as:

$$RCB(\%) = \frac{(Upper\ band - lower\ band)\ x\ 100}{2\ x\ Predicted\ Concentration}$$

For example, Figure 8 shows both RSD and RCB data for determina-
tion of chloride and lead in water. In Figure 8a, the least-
squares curve of best fit closely fits the lead standard data,
and the calibration process has little adverse effect on
precision. RSD's and RCB's are almost equal. On the other hand,
chloride standard data in Figure 8b does not closely fit the
mathematical model, and the RSD data overstates the precision of
the analysis by a factor of about two.
Minimum reportable concentration. The lower concentration
limit for a method is usually measured by determining the
detection limit. This is basically an instrument signal to noise
ratio, and it does not include calibration effects. At low
concentrations the calibration process often has a major adverse
effect on precision. Detection limits are useful for comparing
the inherent sensitivity of methods, but they are not realistic
indexes of measurable concentrations in routine analysis.
       We suggest using a new parameter, the minimum reportable
concentration, defined as the concentration whose confidence band
just includes zero (5). This parameter is obtained by
reducing the value of signal Yo, figure 4, until the band around
predicted concentration, Xo, just touches zero. For example, for
the determination of iron in water by AAS, (data given in Table
III) the detection limit, defined as the concentration at which the

a

b
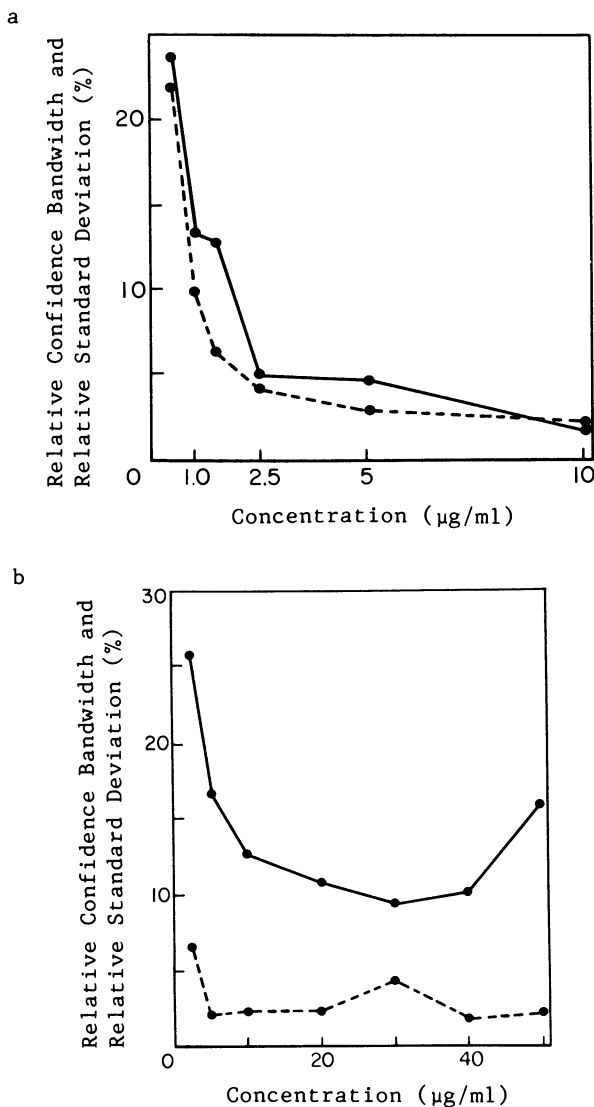
Fig. 8.
Comparison of RCB (—) and RSD (---) for determination of (a)
chloride and (b) lead in water.   (Reprinted with permission from
D. G. Mitchell and J. S. Garden, Talanta 1982, *29*, 921-929,
copyright 1982, Pergamon Press Ltd.)

Table III. Determination of Maximum Concentration of Iron in Water
          by AAS

Standard (µg/ml)	Absorbance at nm	Single-Curve Calibration: RCB (%)[a]	Multiple-Curve Calibration		
			Calibration range (µg/ml)		
			Low	High	RCB (%)
0.05	0.004	40	0.05	1.5	33
0.10	0.008	26	0.05	1.5	19
0.25	0.022	13	0.05	1.5	6
0.50	0.045	9	0.1	1.5	5
1.00	0.093	8	0.1	1.5	4
1.5	0.142	7	0.1	20.0	6
2.5	0.222	7	0.1	20.0	6
5.0	0.430	5	0.1	20.0	4
10.0	0.750	6	0.1	20.0	5
15.0	0.961	11	0.1	20.0	8
18.0	1.054	60	15.0	40.0	11
20.0	1.086	60	15.0	40.0	12
22.0	1.145	60	15.0	40.0	16
25.0	1.191	60	15.0	40.0	25
40.0	1.268	60	25.0	40.0	38
50.0	1.300	60	0.05	100.0	60
75.0	1.360	60	40.0	100.0	60
100.0	1.405	60	40.0	100.0	60

Note: $\alpha = 0.05$; $Z_{1-\alpha/2} = 1.96$

[a] A weighted least-squares technique was used: calibration range
0.05–18 µg/ml.

Source: Reproduced with permission from Ref. 5.

RSD is 50%, is 0.015 µg/ml. The minimum reportable concentration
is a factor of 2 higher when the method is calibrated over a
narrow, low concentration range (0.05 to 0.1 µg/ml). It is a
factor of 20 higher when the method is (inappropriately)
calibrated over a dynamic range of 100 (0.05 to 5 µg/ml).
<u>Maximum reportable concentration</u>. The upper limit of
measurement for a method is usually defined as the concentration
at which the curve shows a certain deviation from linearity.
This is a valid empirical criterion, since sensitivity and hence
precision decreases as the curve flattens. However, linearity
does not directly measure the performance parameter of interest:
precision. In practice an analyst would accept curves at high
concentrations, providing the precision is still adequate and
providing the method does not have accuracy problems at high
concentrations e.g., because of light scattering in absorption
methods.

     Confidence bands are direct precision data, and the
maximum reportable concentration can be defined as the maximum
concentration at which the method yields adequate precision
(<u>5</u>) (excluding measurements near the minimum reportable
concentration, where poor precision is unavoidable). Table III
shows RCB for the determination of iron in water by AAS. The
analyst may consider a RCB of say, 15% to be adequate. The
maximum reportable concentration would be 15 µg/ml from a single,
weighted least-squares curve, and 20 µg/ml by the multiple-curve
method. Samples containing > 20 µg/ml should be diluted to 1-10
µg/ml and analyzed using standards containing 0.05 - 15 µg/mL.
(Note that it is always better to include a standard above the
maximum desired concentration. The precision of this standard
measurement will be poor, but poor data at this level are better
than none.)
<u>Implications For Method Development</u>. The effects of the
calibration process on precision suggest the need for an
additional step in the development of an analytical method. A
suggested flow chart is shown in Figure 9. The analyst should
first develop a method of adequate accuracy and precision without
using calibration curves. The calibration step is then added,
and the precision is rechecked. If precision has been
excessively degraded, the analyst can choose among alternative
calibration strategies, such as use of more standard measurements
and use of the multiple-curve procedure.

Conclusion

I have described a reasonably complete set of mathematical
techniques for improving the precision of calibration-curve-based
analyses and measuring their precision. Each technique may not
be the optimum solution to each problem, but the overall
philosophy should be correct. We should develop statistical
techniques to measure precision which are self-consistent and
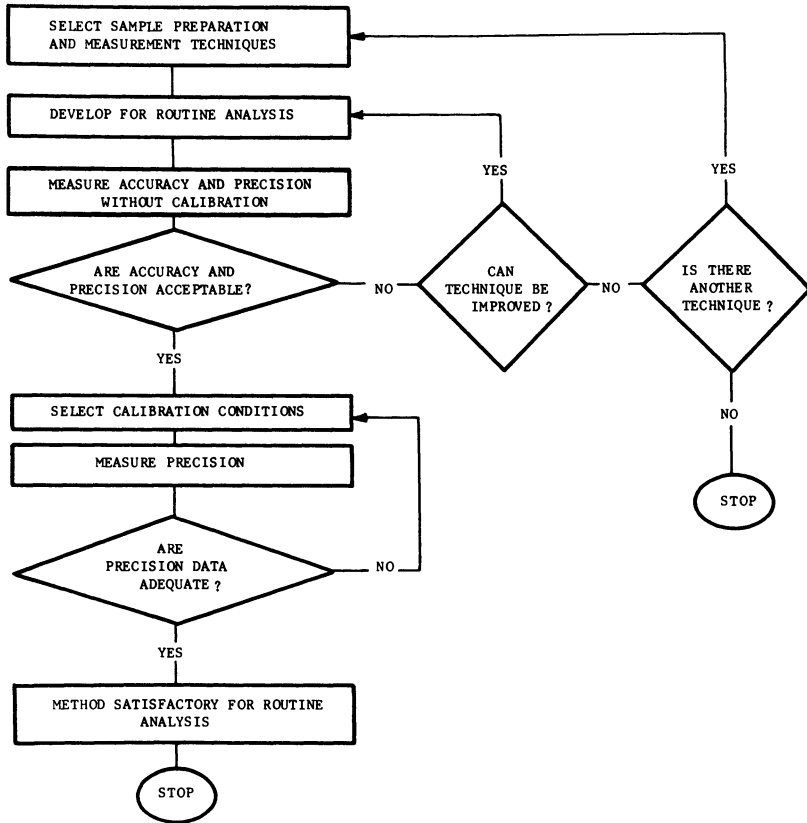
Fig. 9.
Method development procedure for calibration-curve-based
analysis. (Reproduced with permission from D. G. Mitchell and J.
S. Garden, Talanta, 1982, 29, 921-929, copyright 1982,
Pergamon Press Ltd.)

which account for all the factors affecting precision. We should use them to choose optimum calibration strategies and to measure the precision of the resulting data.

The computer program REGRES was written by John S. Gorden, New York State Department of Health. It can be obtained by sending a check for $30.00, made out to Health Research Inc., and a 9-track magnetic tape to John S. Gorden, New York State Department of Health, CSMDP, Concourse Level, Empire State Plaza, Albany, New York 12237.

Literature Cited

1.   Natrella, M.G., ''Experimental Statistics'', National Bureau of Standards Handbook 91 (1963)
2.   Miller, R.G., ''Simultaneous Statistical Inference'', McGraw-Hill, New York (1966)
3.   Mitchell, D. G., Mills, W., N. Garden, J.G., and Zdeb, M., Anal. Chem., 1977, 49, 1655-1660.
4.   Garden, J. S., Mitchell, D. G., and Mills, W. M., Anal. Chem., 1980, 52, 2310-2315.
5.   Mitchell, D. G., and Garden, J. S., Talanta, 1982, 29, 921-929.
6.   Draper, N.R., and Smith, H., ''Applied Regression Analysis'', J. Wiley and Sons, New York, (1966)

The Linear Calibration Graph and Its Confidence Bands from Regression on Transformed Data

DAVID A. KURTZ[1], JAMES L. ROSENBERGER[2], and GWEN J. TAMAYO[2]

[1] Pesticide Research Laboratory, Department of Entomology, The Pennsylvania State University, University Park, PA 16802
[2] Department of Statistics, The Pennsylvania State University, University Park, PA 16802

Linear calibration graphs were constructed from chromatographic response values by means of least squares statistical regression techniques to calculate amount estimates. The amount interval estimates reflect both the uncertainties of measuring the response values and the uncertainty of the calibration graph. The following steps were followed: transformation of response variables to constant variance across the graph using a family of power transformations approach, transformation of the amount variable with similar transformations towards linearity, calculation of the regression coefficients by sums of squares, and solving the regression equation for unknowns. The total range of the amount interval estimation was found by construction of the response confidence interval and the confidence band around the calibration graph. Estimated amounts and amount intervals were calculated from chromatographic analysis of pesticide standards data at 95% probabilities with an overall $\alpha=0.05$. Data were presented that show large errors at the limit of detection using non-transformed or improperly transformed data.

The calibration problem in chromatography and spectroscopy has been resolved over the years with varying success by a wide variety of methods. Calibration graphs have been drawn by hand, by instruments, and by commonly used statistical methods. Each method can be quite accurate when properly used. However, only a few papers, for example (1,2,15,16,26), show the sophisticated use of a chemometric method that contains high precision: regression with total assessment of error.

Difficulties in both the chemical and statistical aspects of the problem have been found to be enormous in utilizing such a

calibration method. Chromatographic detection is accomplished by
flame ionization as well as by many species specific detectors,
such as the electron capture and the flame photometric detectors.
The flame detector is non-specific to the chemical species found,
is mass sensitive since the total sample is burned, and has a
linear range of some 7 powers of ten (3). The electron capture
detector is concentration dependent. It has been linearized
through geometric design and electronic configuration to a range
approaching 3 powers of ten (3). The flame photometric
detector, similar in action with the flame ionization detector,
has been found to be linear to 4 powers of ten in the phosphorus
mode (3). Other known effects on the calibration have included
contamination of the detector and day-to-day variability.
Fulfilling the statistical protocol also requires careful study.
Aspects in this area include model fitting, preparing constant
variance data across the graph, diagnostic tests of closeness of
fit and constant variance achievement, and the construction of
confidence limits.

There are a number of ways to model calibration data by
regression. Most researchers have attempted to describe data with
a linear function. Others (4,5) have chosen a higher order or a
polynomial method. One report (6) compared the error in the
interpolation using linear segments over a curved region verses
using a curvilinear regression. Still others (7,8) chose
empirical or spline functions. Mixed model descriptions have also
been used (4,7).

Ordinary least squares regression requires constant variance
across the range of data. This has typically not been satisfied
with chromatographic data (4,9,10). Some have adjusted data to
constant variance by a weighted least squares method (4). The
other general adjustment method has been by transformation of
data. The log-log transformation is commonly used (9,10). One
author compares the robustness of nonweighted, weighted linear,
and maximum likelihood estimation methods (11). Another has
constructed calibration graphs and confidence limits under the
condition of nonuniform variance (12). On the other hand a
completely different approach for the processing of chromatogra-
phic data has been suggested (13) which involves the use of a
mean slope method.

The need for reporting accuracy and error in the form of
confidence limits when reporting analytical results has already
been well outlined (14). The confidence interval requires
information about the number and distribution of calibration
measurements, the location of the sample and the number of sample
replicates. Agterdenbos has simplified the calculation by
assuming homogeneity of variances and assuming that the variance
of the sample and the calibration standards are the same (15).
Schwartz has calculated the approximate confidence limits of
linear graphs without elaborate digital computation (16).

We will describe an accurate statistical method that includes
a full assessment of error in the overall calibration process,
that is, (1) the confidence interval around the graph, (2) an
error band around unknown responses, and finally (3) the estimated
amount intervals. To properly use the method, data will be ad-
justed by using general data transformations to achieve constant
variance and linearity. It utilizes a six-step process to
calculate amounts or concentration values of unknown samples and
their estimated intervals from chromatographic response values
using calibration graphs that are constructed by regression.

Laboratory Methods and Equipment

Preparation of Standard Solutions. The standards used in the
preparation of the solutions for this work were obtained from the
Health Effects Laboratory, U. S. Environmental Protection Agency,
Research Triangle Park, NC. Dilutions were obtained from
concentrated solutions using wiretrol measuring capillaries
(Drummmond Scientific Co., Bromall, PA). The fenvalerate and
chlorothalonil data sets were prepared with a statistically
equivalent format: Each of the standards at each concentration
level had the same number of dilution steps and should therefore
contain the same variance of dilution. In this case the initial
dilution was used to prepare two working standards, 1 and 2. Each
was diluted once to 1.1 and 2.1. Each of these was used to
prepare three standards for chromatographic injection, 1.11, 1.12,
1.13 and 2.11, 2.12, and 2.13.

The pesticides included in this study were fenvalerate,
chlordecone (kepone), chlorothalonil, and chlorpyrifos. Fenvaler-
ate is a synthetic pyrethroid insecticide used, for example, for
mites on chickens. Its chemical name is cyano(3-phenoxyphenyl)-
methyl 4-chloro-alpha-(1-methylethyl)benzeneacetate. Chlordecone
is an insecticide, no longer used, and has a chemical name
decachloro-octahydro-1,3,4-metheno-2H-cyclobuta(cd)=pentalen-2-one.
Chlorothalonil is fungicide used on tomatoes whose chemical name
is 2,4,5,6-tetrachloroisophthalonitrile. Chlorpyrifos is an
insecticide with a chemical name O,O-diethyl O-(3,5,6-trichloro-
2-pyridyl)phosphorothioate. Chlorpyrifos is the U. S. Food and
Drug Administration chromatographic reference standard since
numerous specific detectors (electron capture, flame photometric
in both sulfur and phosphorus modes, alkali flame, nitrogen
phosphorus, and Hall detectors) are sensitive to it.

Each of the Datasets A-F were also of fenvalerate and were
obtained from an extensive study of fenvalerate residues in
chickens and eggs. They show how much variability in data quality
can be obtained in practice. Table VII describes the number of
calibration levels, replicates at each level, and ranges in ng of
amounts injected into the gas chromatograph. Dataset A is an
"ideal" set, a set that looked ideal at the time it was recorded.
Dataset B is a set of data taken over two days under constant

detector sensitivity. Dataset C is a set of data taken over two
days under changing detector sensitivity. Dataset D has values
where an artifact compound was present in the same peak as
fenvalerate which altered the areas of smaller peaks. Dataset E
has the points containing the artifact removed from the set.
Dataset F has a limited range and was still found to be non-linear
when log-log transformation was performed.

     Gas chromatographic data was obtained on a Tracor Model 220
gas chromatograph equipped with a Varian Model 8000 autosampler.
The analysis column was a 1.7 m "U" column, 4 mm id, filled with
3% SP-2250 packing (Supelco, Inc., Bellefonte, PA) held at
200° C. The injection temperature was 250° and the
nitrogen carrier gas flow rate was 60 mL/min. The detector
temperatures were 350° for electron capture and 190° for
flame photometric. Detector signals were processed by a Varian
Vista 401 which gave retention times and peak areas.

     The symposium Appendix contains all the raw data sets
analyzed in this paper.

General Analytical Plan.   A six step process is described to
calculate the amount or concentration values of unknown samples
using chromatographic response values and calibration graphs that
were constructed by regression. The steps are:

1. Instrumental response values of the standards were transformed
   according to Tukey's simple family of power transformations
   (17) and described later by Box-Cox (18) to a point where a
   statistical test of constant variance was accepted. In this
   work the state of constant variance was tested by the Hartley
   test (19). Response variances were calculated at each amount
   level. The H statistic is then found by dividing the maximum
   variance by the minimum variance, each taken from any level.

2. The amount data corresponding to the response values in 1 above
   were transformed by the same general family of power
   transformations until linearity was obtained. The F-test
   statistic that relates lack of fit and pure error was used as
   the criterion for linearity.

3. The transformed response values were regressed on the trans-
   formed amount values using the simple linear regression model
   and ordinary least squares estimation. The standard deviation
   of the response values (about the regression line) was calcu-
   lated, and plots were formed of the transformed response values
   and of the residuals versus transformed amounts.

4. The Working-Hotelling confidence band (20) was then
   constructed around the estimated regression line.

5. Unknown amounts and error limits of those amounts were
   predicted by the Lieberman, Miller and Hamilton method (21).

6. The predicted transformed amount and amount values and their
   interval limits were transformed back to their original units.

Though the six-step procedure is complicated, it is easily
implemented on a computer. Figure 1 illustrates this process.

A

(Variance Range of Transformed Response vs. Amount Level — low to high; rows a:, b:, c:)

B

(Transformed Response vs. Transformed Amount; curves a:, b:, c:)

C

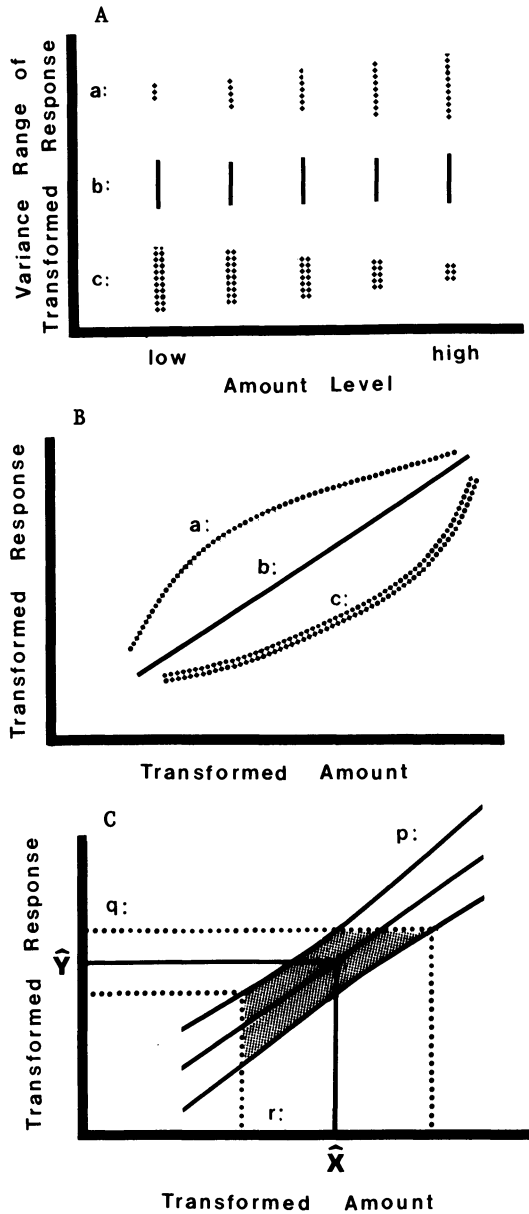(Transformed Response vs. Transformed Amount; p:, q:, r:, $\hat{Y}$, $\hat{X}$)

Figure 1. Plots showing the Calibration Process. A. Response transformation to constant variance: Examples showing a. too little, b. appropriate, and c. too much transformation power. B. Amount Transformation in conforming to a (linear) model. C. Construction of p. confidence bands about the regressed line, q. response error bounds and intersection of these to determine r. the estimated amount interval.

Statistical Calibration

In the calibration problem two related quantities, X and Y, are investigated where Y, the response variable, is relatively easy to measure while X, the amount or concentration variable, is relatively difficult to measure in terms of cost or effort. Furthermore, the measurement error for X is small compared with that of Y. The experimenter observes a calibration set of N pairs of values (x_i, y_i), i=1,...,N, of the quantities X and Y, x_i being the known standard amount or concentration values and y_i the chromatographic response from the known standard. The calibration graph is determined from this set of calibration samples using regression techniques. Additional values of the dependent variable Y, say y_j*, j=1,...,M, where M is arbitrary, are also observed whose corresponding X values, x_j* are the unknown quantities of interest. The statistical literature on the calibration problem considers the estimation of these unknown values, x_j*, from the observed y_j*, and the equally important aspect of calculating the upper and lower bounds for x_j*.

The technique for obtaining interval estimates for X*, discussed in this section, is presented in the paper by Lieberman, Miller, and Hamilton (21) and based on the Bonferroni inequality (22) described below. Other methods are found in the references (23,24).

The simple linear regression model will be assumed throughout this section. That is,

$$y_i = \beta_0 + \beta_1 x_i + e_i \qquad\qquad (1)$$

where i=1,...,N, e_i are independent errors with constant variance, and (x_i, y_i) are observations from the standard calibration sample.

Since data from chromatography standards typically do not satisfy the assumptions of constant variance nor linearity, a procedure described above for fitting a family of transformations on the y_i and x_i will be used. We assume for the rest of this section that the above model is satisfied for the transformed data.

Bonferroni Interval Estimates.

Interval estimates for the unknown X*, referred to as unlimited simultaneous discrimination intervals (21), are based on the estimated regression line of y_i on x_i, and the confidence interval (on the Y-axis) about the response y_j* for an unknown. The resulting interval estimates have the property that for at least $100(1-\alpha)$% of the different calibration sets, at least 100P% of the amount intervals estimated from that calibration will contain true unknown amounts x_j*. The $(1-\alpha)$ confidence refers to the uncertainty inherent

in the estimation of the calibration line,

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

and the response variance, σ^2, whereas the probability P refers to the sampling distribution of the unknown samples y_j*.

The Bonferroni interval estimate of X*, given Y*, is found in three moves. First, the Working-Hotelling confidence band for the regression line

$$E(Y) = \beta_0 + \beta_1 X$$

is obtained (25). Next, the confidence interval (on the Y-axis) for the true value of Y*, say U_{Y*}, is determined. Lastly, the Bonferroni inequality is invoked to combine the two proceeding confidence statements, each made with the confidence $(1-\alpha/2)$, to yield an interval estimate for X* with confidence at least $(1-\alpha)$. The confidence band on the regression line and the confidence interval on U_{Y*} are intersected and the Bonferroni interval estimate of X* is found by projecting the intersection onto the x-axis. Figure 1c illustrates the procedure. If U_{Y*} is in the interval on the Y-axis and if the hyperbolic confidence band contains the line

$$E(Y) = \beta_0 + \beta_1 X$$

then the shaded region must contain the point $(X*, \beta_0 + \beta_1 X*)$ and hence, X* must lie in the interval on the X-axis, the interval estimate for X*.

With confidence $(1-\alpha/2)$ the Working-Hotelling confidence band contains the true line

$$E(Y) = \beta_0 + \beta_1 X$$

and with $(1-\alpha/2)$ confidence the true mean, U_{Y*} of Y*, is contained in the interval about Y*. The Bonferroni inequality guarantees a confidence coefficient of at least $(1-\alpha)$ that both statements are true. The steps in the mathematical construction of these confidence intervals are given below and in Lieberman, Miller and Hamilton (21) and Hunter (26).

Move 1. First, the Working-Hotelling confidence band for the true regression line

$$E(Y) = \beta_0 + \beta_1 X$$

is obtained with confidence coefficient $(1-\alpha/2)$. For any X, the bounds are

$$\hat{Y} - W\, s(\hat{Y}) \; \leq \; \beta_0 + \beta_1 X \; \leq \; \hat{Y} + W\, s(\hat{Y}) \qquad (2)$$

where $\quad W^2 = 2\,F\,(1 - \alpha/2;\; 2,\; N-2)$

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X$$

$$s(\hat{Y}) = \sqrt{MSE\left[\frac{1}{N} + \frac{(X - \overline{X})^2}{\Sigma(x_i - \overline{X})^2}\right]}$$

and $\quad F(P;\; n_1, n_2)$ is the P percentile of the F distribution with n_1 and n_2 degrees of freedom, and MSE is the error variance about the regression line.

Move 2. Next, the confidence interval for the true mean of Y* is constructed such that, for a given Y*, the 100P% confidence interval for U_{Y*} is

$$Y* - Z(P)\sigma \; \leq \; U_{Y*} \; \leq \; Y* + Z(P)\sigma \qquad (3)$$

where $Z(P)$ is the appropriate percentile of the standard normal distribution. That is, each interval given by Equation 3 has probability P of containing the true mean of Y*. However, since σ is not known, an upper bound for σ can be found and inserted into Equation 3. Then an interval with probability at least P of containing the true mean of Y*, U_{Y*}, is obtained.
$(N-2)MSE/\sigma^2$ has a chi square distribution with $(N-2)$ degrees of freedom. To find the upper bound for σ, we use

$$Pr\{\, a \leq (N-2)MSE/\sigma^2\,\} = 1 - \alpha/2$$

where the value $\underline{a}$ is the lower $\alpha/2$ percentile point of the chi square distribution with $(N-2)$ degrees of freedom. Thus,

$$Pr\{\, \sigma^2 \leq (N-2)MSE/a\,\} = 1 - \alpha/2$$

Thus the unknown standard deviation σ can be bounded above by

$$\sigma \leq \sqrt{\frac{(N-2)\,MSE}{\chi^2(\alpha/2;\, N-2)}} \qquad (4)$$

with probability $(1 - \alpha/2)$. By the Bonferroni inequality, both Equations 2 and 4 are correct with probability of at least $(1 - \alpha)$.

Move 3. The Working Hotelling confidence band about the regression line

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X$$

is intersected with the interval on the Y-axis about $y_i{}^*$ and the intersection is projected onto the X-axis. Figure 1c illustrates how Steps 1 and 2 are combined to yield the interval estimate for X*.

If the estimated regression line has positive slope, then the upper bound for X*, denoted B+, is the solution to the equation

$$\hat{\beta}_0 + \hat{\beta}_1 X^* - W\, s(\hat{Y}) = Y^* + Z(P)\sqrt{\frac{(N-2)\ MSE}{\chi^2(\alpha/2, N-2)}}$$

and the lower bound for X*, denoted by B-, is the solution to the equation

$$\hat{\beta}_0 + \hat{\beta}_1 X^* + W\, s(\hat{Y}) = Y^* - Z(P)\sqrt{\frac{(N-2)\ MSE}{\chi^2(\alpha/2, N-2)}}$$

Since $s(\hat{Y})$ is a function of X*, the roots of these equations are given by

$$B- = \overline{X} - \frac{\hat{\beta}_1(Q - Y^* + \overline{Y}) + \sqrt{R\left[(Q - Y^* + \overline{Y})^2 + \dfrac{\sum\limits_{i}^{N}(X_i - \overline{X})^2}{N}(\hat{\beta}_1^2 - R)\right]}}{\hat{\beta}_1^2 - R}$$

$$B+ = \overline{X} + \frac{\hat{\beta}_1(Y^* - \overline{Y} + Q) + \sqrt{R\left[(Y^* - \overline{Y} + Q)^2 + \dfrac{\sum\limits_{i}^{N}(X_i - \overline{X})^2}{N}(\hat{\beta}_1^2 - R)\right]}}{\hat{\beta}_1^2 - R}$$

where

$$R = 2\, F(1 - \alpha/2; 2, N-2)\ MSE\ /\ \sum_{i}^{N}(x_i - \overline{X})^2$$

$$Q = Z(P)\sqrt{\frac{(N-2)\ MSE}{\chi^2(\alpha/2; N-2)}}$$

<u>Discussion of Statistical Methodology.</u>   Previous work using
similar statistical methods to these are given by Garden,
Mitchell, and Mills (<u>4</u>). They use weighted least squares to
resolve the non-constant variance of the response signal for
different concentrations, whereas we transform the response to
achieve constant variance.

A point which may need emphasis, stated clearly in Hunter
(<u>26</u>), is the precise interpretation of the confidence band about
the predicted amount. This is important since without a clearly
understood meaning, the interval will not be useful for assessing
the precision of the predicted amounts or concentrations nor for
comparing the results from various laboratories. Another reason
the user of these methods must understand the interpretation is
because increased precision can be achieved in at least two ways -
by additional replication of the standards, which reduces the
width of the confidence band about the regression line, and by
performing multiple determinations on the unknowns, which reduces
the width of the interval about the mean instrument response of
the unknown. The interval for U_{y*} is then given by

$$\overline{Y*} \; \pm \; Z(P) \sqrt{\frac{MSE}{m} \; \frac{(n-p)}{\chi^2(\alpha/2,n-2)}}$$

where $\overline{Y*}$ is the average of the m determinations of the unknown.

The frequency interpretation of the interval estimates on the
unknown amounts is given by the following (<u>27</u>): With at least
$1-\alpha$ confidence, based on the sampling characteristics of the
observations on the standards, at least P proportion of the
interval estimates made from a particular calibration will contain
the true amounts. The Bonferroni inequality insures the $1-\alpha$
confidence since the confidence interval about the regression line
and the upper bound on σ are each performed using a $1-\alpha/2$
confidence coefficient. Hence, the frequency interpretation
states that at least $(1-\alpha)$ proportion of the standard
calibrations are such that at least P proportion of the intervals
produced by the method cover the true unknown amounts. For the
remaining α proportion of standard calibrations the proportion of
intervals which cover the true unknown values may be less than P.

Results and Discussion

<u>Response Transformation. Step 1.</u>   We found that the calibration
graph response data obtained from gas chromatography seldom have
constant variance along the length of the graph. The data in
Tables I-III clearly show that the larger the response value the
larger the variance of the response at that level. Fenvalerate in
Table I, chlordecone (kepone) in Table II and chlorothalonil in
Table III have the information for untreated data (at a

Table I. Comparison of Transformed Response Variances for Several
Powers of Transformation. Fenvalerate Data on the
Electron Capture Detector.

Amount,ng	Typical Response	Transformation Power			
		1.0	0.5	0.15	Log
0.05	1	0.0042	0.00 1	0.000 10	0.000 85
0.25	6	0.031	0.00 2	0.0000 4	0.000 19
1.0	24	0.090	0.00 1	0.0000 1	0.0000 3
5.0	125	6.8	0.0 16	0.0000 5	0.000 10
25.	750	173.	0.0 71	0.0000 6	0.0000 7
100.	3000	3400.	0. 338	0.000 11	0.0000 8
H Statistic (a) =		>>	301	10.3	24.5

(a) Critical H (0.95, 6 levels, 6 replications) = 18.7

Table II. Transformed Response Variances Converging to a Constant
Level. Kepone Data on Electron Capture.

Amount Level, ng	Transformation Power				
	1.00	0.50	0.30	0.25	0.10
0.01	0.002	0.000 7	0.000 27	0.000 19	0.0000 34
0.05	0.028	0.00 26	0.000 63	0.000 39	0.0000 47
0.20	0.25	0.00 53	0.000 71	0.000 38	0.0000 29
0.38	0.57	0.00 58	0.000 58	0.000 29	0.0000 18
1.0	0.84	0.00 34	0.000 24	0.000 11	0.00000 5
H Statistic (a)	400.	8.5	2.9	3.6	9.0

(a) Critical H (0.95, 5 levels, 3 replications) = 202

Table III. Transformed Response Variances nearly not Converging
        to a Constant Level. Chlorothalonil Data on Electron Capture.

	Transformation Power				
Amount Level, ng	1.00	0.25	0.20	0.15	0.10
0.02	0.01	0.000 2	0.000 1	0.0000 6	0.0000 2
0.10	2.9	0.00 35	0.00 17	0.000 75	0.000 26
0.40	2.6	0.000 4	0.000 2	0.0000 6	0.0000 2
1.0	24.	0.000 3	0.000 1	0.0000 4	0.0000 1
4.0	260.	0.000 5	0.000 2	0.0000 5	0.0000 1
20.	4000.	0.00 24	0.000 7	0.000 19	0.0000 4
H Statistic (a)	1500.	18.	15.	19.	26.

(a) Critical H (0.95, 6 levels, 5 replications) = 29.5

transformed power of 1) shown in the first column. Unpublished
data in the author's laboratory for liquid chromatographic and
visible spectroscopic analyses have shown the same result.
     This situation shows two problems: The application of
ordinary least squares estimation, which requires constant
variance, is not appropriate with untreated data. Then, the large
variance of the largest numbers in such data excessively controls
the direction or slope of the graph.
     The solution to the problem of non-constant variance (or
heteroscedasticity) rests in several suggestions. The simplest is
to limit the range of the graph (1). The range, however, would
be so small that it would be ineffective to use it practically.
One acceptable suggestion is to perform the regression with
weighted response values (4) where the weights are the inverse
of the variance at each level of the calibration. However, since
the true weights are not known, the determination of the weight
values at each level of the calibration is not reliable.
     Another solution to the problem of non-constant variance is
to transform the response data. A common way of transforming data
has been by taking the logarithms of both the response and amount
variables (8-10). However, for all the data we looked at, the
log transformation has been too strong. See Tables I and V.
Table V shows that in some data sets the log transformation is
acceptable, but this is usually as a result of less perfect data
and not an inherent quality.

The authors, Box and Cox, describe (18) a family of power transformations. This family includes the following calculations: The response is raised to the 2, 1, 0.5, 0.25, 0.1 power, the log of the response is taken, and the negative reciprocal of the response is raised to the 0.5 power. We chose to simplify the series by starting with a power of 0.50 proceeding downward to 0.10 in units of 0.05. We also included the logarithm of the response for comparison purposes which has a strength corresponding to the limit as the power goes to zero.

Constant Variances.   Response values from the electron capture chromatographic analysis of the insecticide fenvalerate, were transformed by the process described above. The six response values at each of six different amount levels were transformed by a series of powers, and the variances calculated at each level (Table I). For a transformation power of 0.5 the value of the variances increased from 0.001 to 0.338 as the response increased. When the logarithm of the response was used, the value of the variances decreased from 0.00085 to 0.00008 as the response increased. Raising the responses to the 0.15 power gave calculated variances that remained roughly constant across the range of amounts.

There are a large number of statistical procedures published to test for constancy of variance (28-30). The Hartley Test was chosen for this task because of its simplicity (19). Selected critical values for the H statistic at $\alpha = 0.05$ are shown in Table IV for reference purposes.

Table IV.   Percentiles of the H Statistic Distribution.
H(1- α ; r,n) where P {H <= H (1- α ;r,n)} = 0.95
α = 0.05

	r (a)				
n (a)	2	3	4	5	6
3	39.0	87.5	142	202	266
4	15.4	27.8	39.2	50.7	62.0
5	9.60	15.5	20.6	25.2	29.5
6	7.15	10.8	13.7	16.3	18.7
7	5.82	8.38	10.4	12.1	13.7

(a) r = no. of concentration levels, n = no. of samples/level

From the authors' experience not all real data sets can be
transformed to constant variance using power transformations.
Instrumentation imperfections in our laboratory resulted in data
that had variable variances despite our attempts at
transformation. The transformed chlorothalonil data set, as shown
in Table III illustrates a set where the transformations attempted
nearly failed to give constant variance across the response range;
in this case the Hartley criterion was barely satisfied. The
replications at the 0.1 and 20. ng levels had excessively high
variance over the other levels. An example where constant
variance was easily achieved utilized data of the insecticide
chlordecone (kepone) also on the electron capture detector. Table
II shows that using a transformation power of 0.3 resulted in
nearly constant variance.

<u>Transformation Power of Selected Data Sets.</u>   Hartley statistic
values are shown in Tables I-III for fenvalerate, chlordecone, and
chlorothalonil. In each case a power transformation was found of
sufficient size at a 95% probability which satisfied the H
criterion. For fenvalerate the power of 0.15 was satisfactory for
constant variance. For chlordecone the whole range of powers from
0.50 to 0.10 satisfied the critical H value (listed in order of
increasing transformation power). Despite apparent non-constancy
of data for chlorothalonil shown in Table III, the critical H was
satisfied for the range in power transformation from 0.25 to 0.10.
        The response transformation powers for all data sets studied
for this work are given in Table V. The acceptable range, judged
from the Hartley test, for each individual data set is listed.
The optimal power is also listed even though some sets did not
have any acceptable power of transformation.
        For fenvalerate the minimum H statistic was found to be at a
transformation power of 0.15. The power range of 0.15 to 0.10
gave an H statistic less than or equal to the critical H value.
The various Datasets A-F were all of fenvalerate. We chose to
transform the response values of these sets to the same power as
required for the fenvalerate data set since we wished to use these
data sets as examples of "unknown" data sets or as examples of
poor quality standard sets. If the compound has an inherent
analysis quality relating to the response variance, the quality of
the poor sets is reflected in differences in the error bands. The
acceptable ranges for the Datasets A-F, as shown in Table V, did
include the 0.15 power eventually used in all cases where an
acceptable H value was found.
        The regression for chlorothalonil was different from that of
fenvalerate in its optimal power probably due to its shorter a-
mount range (Table VII), 1000 to 1 where the latter was 2000 to 1.
        Chlorpyrifos had an amount range of only 500 to 1. It was
different from others cited by its flame photometric (sulfur mode)
detection. It is interesting, however, that the power transfor-
mations for both chlorothalonil and chlorpyrifos were so similar.

Table V. Optimal and Acceptable Range of Response Transformation
Power (a) Satisfying the Hartley Test for Data Sets
Determined in Various Detectors.

Data Set	Calibration Set (b) r,n	Mode (c)	Response Transformation Acceptable Range Descending Power	Optimal
Fenvalerate	6,6	EC	0.15 - 0.1	0.15
Dataset A	5,2	EC	0.3 - 0.1	0.3
Dataset B	5,5	EC	none	0.3
Dataset C	5,4	EC	0.3 - log	0.2
Dataset D	5,4	EC	none	0.7
Dataset E	4,4 (d)	EC	none	0.6
Dataset F	4,2	EC	0.3 - log	0.2
Chlorothalonil	6,5	EC	0.3 - 0.1	0.2
Chlorpyrifos	5,6	FP	0.3 - 0.1	0.2

(a) transformation power performed in steps of 0.05
(b) r = number of concentration levels, n = samples/level
(c) EC = electron capture, FP = flame photometric
(d) unequal number of standards, n, at each level

Amount Transformation. Step 2.   The amount transformation was
performed in a way similar to that of response by use of a power
series but for a different reason. In this case linearity was
desired in order to use a simple linear regression model. This
transformation therefore required a test for satisfactory
conformity. One can use a variety of criteria including the
correlation coefficient or visual examination of the plot of
residuals verses amount. We chose the F test for lack of fit,
$F^* = MSLF/MSPE$, based on the ratio mean square for lack of fit
(MSLF) over the mean square for pure error (MSPE) (31). F^*
follows the F distribution with (r-2) and (N-r) degrees of
freedom. A value of $F^* < F(1-\alpha ; r-2, N-r)$ is consistent with the
linearity of the regression equation. Since the data were
manipulated by transforming the amount values to obtain linearity,
i. e., to achieve the smallest lack of fit F^* statistic, the
significance level of this test is not reliable.
     Bickel and Doksum (32) discuss the influence of searching
for the proper transformation on the estimation of the regression

line. For cases where the error variances are small, such as we
have with calibration curves, they claim that the estimated slopes
and intercepts are much less accurate than the estimated variances
indicate. Carroll and Ruppert (33) also discuss this
phenomenon. Additional investigation of these type of data is
required to properly assess the effect on the prediction accuracy
in searching for the best transformation.

     The amount transformation process is illustrated with data
for chlorpyrifos in the flame photometric detector, phosphorus
mode, and shown in Table VI. Level 1 transformations were
calculated where the amount power was increased by 0.05 units for
each step. At an amount power of 0.20 the F statistic of 32.7
showed a minimum but at a confidence level of 95% did not satisfy
the F test for linearity. Power steps changed by only 0.01 and
0.001 units in the vicinity of the minimum were then calculated as
shown in levels 2 and 3. The best linearity was found in this
case at a power transformation of 0.182 although the F^*
statistic of 8.53 did not indicate linearity when compared with
the critical F of 2.99 at P=.95. Calculations at these second and
third levels were not always necessary and even when performed did
not always lead to a satisfactory condition of linearity.

     At any transformation level if the minimum F statistic were
less than or equal to the critical F value, our work was done and
the confidence band calculations began. Otherwise we either
accepted a lack of fit (and would note it in published results),
segmented the graph to shorter lengths, or sought a non-linear or
higher order model.

Table VI.   Convergence of the Optimal Amount Transformation for
                    the Determination of Data Linearity.
             Chlorpyrifos Data with Flame Photometric Detection.

Level 1 0.05 power steps		Level 2 0.01 power steps		Level 3 0.001 power steps	
Amount Power	F Statistic	Amount Power	F Statistic	Amount Power	F Statistic
0.30	932.	0.21	65.7		
0.25	330.	0.20	32.7	0.184	8.93
0.20	32.7	0.19	13.6	0.183	8.66
------->		------->		0.182	** 8.53
0.15	82.5	0.18	8.71	0.181	8.55
0.10	514	0.17	18.4	0.180	8.71
log	2580	0.16	42.9		

Table VII.   The Power for the Amount Transformation Achieving the
              Best Linearity on Chromatographic Data.

Amount Range on the Col, ng	Calib. Set r,n(a)	Transform. Power		Linear Prob. Level	Standard Deviation
		Response	Amount		
Fenvalerate					
0.05 - 100	6,6	0.15	0.16	0.00	0.014
Datasets					
A: 0.05 - 20	5,2	0.15	0.17	0.05	0.016
B: 0.05 - 20	5,5	0.15	0.17	0.00	0.018
C: 0.05 - 20	5,4	0.15	0.18	0.46	0.037
D: 0.05 - 20	5,4	0.15	0.19	0.50	0.045
E: 0.25 - 20	4,~4	0.15	0.16	0.47	0.013
F: 0.05 - 5	4,2	0.15	0.19	0.30	0.009
Chlorothalonil					
0.02 - 20	6,5	0.20	0.16	0.47	0.022
Chlorpyrifos, P-mode					
1.00 - 500	5,6	0.20	0.18	0.00	0.012

(a) r = No. of levels, n = No. of samples at each level.

Table VII displays the results of the test for linearity
after transforming the amount to a minimum F statistic.
Fenvalerate, Dataset B, and chlorpyrifos all demonstrated
significant linear lack of fit. This was due to relatively small
variations about the graph and a cubic or sigmoid shape to the
curve not removable by these transformations. The remaining
Datasets and chlorothalonil did not demonstrate significant lack
of linear fit due to larger variations about the graph, a shorter
range, or a combination of the two.
     The optimal power transformations on the amounts, when
examined in steps of 0.01, all fell between 0.16 and 0.19 for the
pesticides tested (Table VII). These values are very similar to
the powers required for the response transformation which varied
between .15 and .20.

Regression of Transformed Values. Step 3. The transformed
response values were regressed on the transformed amount data

using a simple linear regression model and least squares
estimation. Table VIII shows the estimated coefficients and their
standard errors (se) for each of the regressions performed.
Figure 2a-e shows the plot of the regression line with
superimposed data points for 5 of the data sets studied.

Table VIII.   Regression Coefficients and Standard Errors for
               Illustrated Data Sets Using the Optimally Transformed
               Data for X and Y.

$$Y' = a + b X' + error$$

Data Set	Y (a) Power	Intercept		Slope Coefficient		X (a) Power
		a		b		
Fenvalerate	0.15	0.033	0.0010	1.580	0.0008	0.16
Dataset A	0.15	0.005	0.0050	1.663	0.0044	0.17
Dataset B	0.15	0.022	0.0021	1.663	0.0019	0.17
Dataset C	0.15	0.136	0.0054	1.570	0.0046	0.18
Dataset D	0.15	0.256	0.0061	1.362	0.0052	0.19
Dataset E	0.15	-0.147	0.0039	1.743	0.0031	0.16
Dataset F	0.15	0.310	0.0036	1.418	0.0037	0.19
Chlorothalonil	0.20	-0.523	0.0022	3.205	0.0020	0.16
Chlorpyrifos	0.20	-0.135	0.0011	1.134	0.0005	0.182

(a) Y' corresponds to transformed response,
    X' to transformed amount

Examination of Data.   At this point, examination of the plot of
regression residuals verses transformed amount showed two
conditions: First, the condition of constant variance across the
graph was shown by observation of the spread of points at each of
the amount values. Second, the linearity was checked by verifying
that the clustering of the residuals was around zero and did not
show a normal or inverted "U" shape or an "S" shape throughout the
range of amount values. When the top end of the graph dropped off
(due to a saturation of the detector) the effect of forcing the
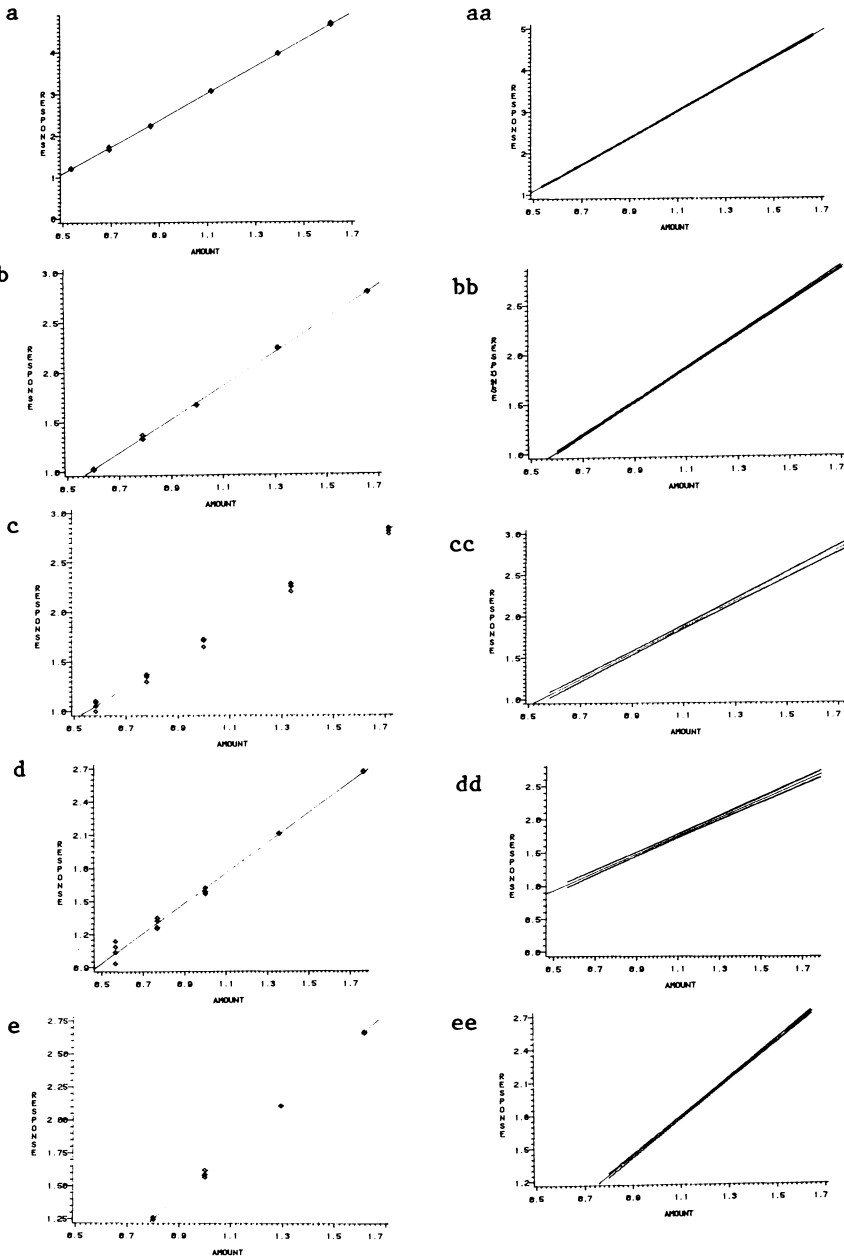
Figure 2. Series of plots of five cases showing linear regression
model with raw data points (a-e) and the 0.975 Working-Hotelling
confidence limits about the regressed line (aa-ee). The letter a
and aa refers to chlorothalonil and b-e and bb-ee refer to
Datasets B, C, D, and E, respectively.

data to linearity by transformation resulted in an "S" shaped
curve. Curvature at the top end of the graph was tolerated more
when the variance of the data points at each amount level was
higher because a sufficiently larger pure error term of the F
statistic lowered the value of the statistic. With long
chromatographic graphs and with data that had consistent and low
variance, the linearity tests became quite sensitive.

Confidence Band Calculation. Step 4.   Following the
determination of the proper regression graph from the transformed
data, the Working-Hotelling confidence limits were calculated.
These bands were calculated at an α =0.025 which, when combined
with the response error bounds α =0.025 (see Step 5), gave an
overall α =0.05 by the Bonferroni inequality. The confidence band
intervals calculated after inverse transformation for the data
sets cited in this work are presented in Table IX. It must be
emphasized that the intervals found in Table IX were calculated at
an α =0.025. The graphs for five of the data sets are shown in
Figure 2(aa-ee).

Table   IX.   Confidence Intervals for the Predicted Response from
              Inverse Transformed Data. α =0.025.

	Confidence Interval			
Data Set	At Min. Amount 0.05 ng(a)	Mean	Interval at Mean	At Max. Amount 20 ng(a)
Fenvalerate	1.01-1.16	102	100-105	550- 574 2990-3150(100)
Dataset A	0.85-1.24	49	45- 52	830- 970
Dataset B	1.03-1.28	52	50- 54	890- 980
Dataset C	1.08-1.80	57	52- 63	910-1140
Dataset D	0.87-1.61	41	36- 46	590- 780
Dataset E	4.00-4.90(.25)	101	98-105	660- 720
Dataset F	1.82-2.30	25	24- 26	200- 227(5.0)
Chlorothalonil	2.20-2.59(.02)	171	168-175	2130-2230
Chlorpyrifos, P-mode	0.95-1.05(1.0)	33	33- 34	434- 450(500)

(a) () gives different minimum and maximum amounts.

We chose to illustrate data at the mean point of the graph where the interval is the shortest and at the extreme ends where the interval is the longest. The low-scatter data sets, chlorothalonil and Dataset B, were found to have very narrow bands. Dataset C, chosen by combining data collected on two different days when the detector sensitivity was definitely different, produced a band wider than that of Set B. Datasets D and E are to be compared: The points in D at the low end of the graph contained various amounts of a contaminant so that many responses of the chromatograph in that region were too large resulting in a large variance. The width of the confidence band was obviously larger than that for Dataset B. When the data points representing the peaks containing the contaminant were properly removed from the dataset, the remaining points formed the Dataset E. This set had a shorter range and fewer points. It also had a narrower confidence band.

We represent the width of the confidence band numerically by the bandwidth (<u>1</u>) which is defined:

$$\text{Bandwidth} = \frac{\{(\text{Upper Limit} - \text{Lower Limit})/2\}100\%}{\text{Predicted Response}}$$

The Bandwidth is essentially a normalized half confidence band. The confidence interval bandwidths for 9 data sets using inverse transformed data are given in Table X. The bandwidths are approximately the vertical widths of response from the line to either band. The best band was found for chlorpyrifos, 1.5%, at the minimum width (located at the mean value of the response) and 4.9% at the minimum or lowest point on the graph. Values for fenvalerate and chlorothalonil were slightly higher, 2.1-2.2% at the mean level. The width at the lowest amount for the former was smaller due to a lower scatter of its points. The same reason explains the difference between fenvalerate and Dataset B. Similarly, the lack of points in Dataset A produced a band that was twice as wide when compared to Dataset B. Dataset C gave a much wider band when compared to Dataset B.

The widest bandwidth was found in Dataset D, the set containing artifacts. Its width was 11.5% at the mean and 31% at the lowest amount of the predicted response! When the chemical error was eliminated and the data recalculated as Dataset E, the bandwidths were only about 1/3 as much. Set E was now comparable to Set B despite having only 15 data points compared to the 25 for B.

Residue chemists would be most interested in comparable confidence bands or confidence bandwidths. These values become a performance characteristic for any detection system. The values indicate the precision of not only the prepared standards but also the precision of the overall operating detection system. It is ultimately envisioned that a given system of a separation column in a chromatograph with a certain type of detector should give bands of standardized values. If a chemist finds he has not met

Table X. Confidence Interval Bandwidths from the Regression of
           Transformed Data Sets.   Inverse Transformed Data.
                              α =0.025.

	Confidence Interval Bandwidth, % (a)		
Data Set	At Min. Amount 0.05 ng(b)	At Mean Amount	At Max. Amount 20 ng(b)
Fenvalerate	6.9	2.2	2.1
			2.6(100)
Dataset A	19.3	6.8	8.0
Dataset B	10.9	3.9	4.7
Dataset C	26.	9.2	11.1
Dataset D	31.	11.5	14.1
Dataset E	10.2(.25)	3.6	4.5
Dataset F	11.1	4.9	6.3(5.0)
Chlorothalonil	8.2(.02)	2.1	2.3
Chlorpyrifos, P-mode	4.9(1.0)	1.5	1.7(500)

(a) Bandwidth, % = $\dfrac{\{(\text{Upper C.I. } - \text{ Lower C.I.})/2\}\{100\%\}}{\text{Predicted Response}}$

(b) () gives different minimum and maximum amounts.

such values, then he must examine his system to find where it can
be improved. Hence it is recognized that there should be a lot of
data published in this regard in order that these standards are
developed. One caveat is necessary: The amount level at which
the bandwidth is calculated needs to be reported as part of the
performance characteristic.
     There is one final observation using the bandwidth informa-
tion: The data of Tables IX and X suggest that the flame photomet-
ric detector (chlorpyrifos) produces more consistent data than the
electron capture detector. The chlorpyrifos data clearly had the
narrowest bandwidth yet both the range and sample size of this set
were comparable to the others studied. The range of chlorpyrifos
was 500 to 1 whereas those of fenvalerate and chlorothalonil were
2000 to 1 and 1000 to 1, resp. Chlorpyrifos had 30 samples
whereas the other two had 36 and 30, resp. Chlorpyrifos had 5
analysis levels while the other two had 6 each. More data of this
sort is needed to compare various detector systems.

Calculation of Response Error Bounds.   The error bounds for each
step in the overall process have been individually calculated in
order that the reader appreciate the magnitude of the error at
that step. This paragraph will describe the error bounds in the
response error band. The bounds for a given data set are all
constant regardless of the magnitude of the response. They are,
however, dependent on the number of determinations of the unknown.
The bounds calculated in this paper were made under the assumption
of a single determination of a given unknown. Since the bounds
are inversely proportional to the square root of the number of
unknown determinations, utilizing a second determination of that
unknown will reduce the bounds by the square root of 2 to 70% of
the bounds for one determination. See Equation 5.
    The response error bounds and the response error bandwidths
for a single unknown determination are given in Tables XI and XII.
These values are all much larger than the regression confidence
bands because of the much smaller number of data points involved.
In comparing the response error bounds (Table XI) to the

Table XI.   Estimated Values of the Response Error Bounds from
             Inverse Transformed Data. α =0.025 where 95% of
             Response Unknowns Will Lie within the Response Error
                     Bounds of the True Response.

Data Set	At Min. Amount 0.05 ng(a)	Mean	Interval at Mean	At Max. Amount 20 ng(a)
	Estimated Response Error Bounds			
Fenvalerate	0.85-1.36	102	91-115	520- 620 2860-3300(100)
Dataset A	0.67-1.53	49	38- 61	770-1040
Dataset B	0.83-1.57	52	43- 62	830-1050
Dataset C	0.68-2.69	57	38- 84	790-1310
Dataset D	0.49-2.62	41	24- 65	490- 940
Dataset E	3.55-5.47(.25)	101	88-116	630- 770
Dataset F	1.62-2.55	25	21- 29	190- 240(5.0)
Chlorothalonil	1.85-3.05(.02)	171	154-190	2040-2320
Chlorpyrifos, P-mode	0.85-1.16(1.0)	33	31- 36	420- 460(500)

(a) () gives different minimum and maximum amounts.

confidence interval on the regression line at mean levels (Table
IX), values found were 24 to 5 response units for fenvalerate, 19
to 4 for Dataset B, 8 to 2 for Dataset F, and 5 to 1 for
chlorpyrifos. The comparison for all the datasets at the mean
were in the ranges of 4-5 to 1. The comparisons at the minimum
responses were 2-3 to 1 and those at the maximum response were
1.8-4 to 1. Since the response error bounds are constant for a
given experiment, the bandwidths get smaller and smaller as the
response gets larger.

Prediction of Unknowns. Step 5.   The point estimate of unknown
amounts is obtained from the intersection of the corresponding
response value and the regressed line which is then projected down
to the amount axis. The interval estimate for an unknown amount

Table XII.   Response Error Bound Bandwidths of Inverse Transformed
             Data. α =0.025 where 95% of Response Unknowns
             Will Lie within the Response Error Bounds of the True
                               Response.

	Estimated Response Error Bound Bandwidth, % (a)		
Data Set	At Min. Amount 0.05 ng(b)	At Mean Amount	At Max. Amount 20 ng(b)
Fenvalerate	24	11.9	9.2 7.0(100)
Dataset A	42	23	14.9
Dataset B	32	18	11.7
Dataset C	72	40	26
Dataset D	89	51	33
Dataset E	22(.25)	13.6	10.2
Dataset F	23	15.7	11.4(5.0)
Chlorothalonil	25(.02)	10.5	6.4
Chlorpyrifos, P-mode	16(1.0)	7.7	4.5(500)

(a) Bandwidth, % = {(Upper Bound - Lower Bound)/2}{100%}
                         ──────────────────────────────
                               Predicted Response

(b) () gives different minimum and maximum amounts.

is obtained from the intersections of the upper and lower response
error bounds with the lower and upper confidence bands which are
projected down to the amount axis. This calculation puts the
whole problem together as it is dependent on the length of the
calibration graph, the number of calibration points, and the
scatter of the points.   It is also dependent on the replication
of unknowns analyzed and the total number of future unknowns to be
determined.

<u>Estimated Amount Intervals.</u>   Estimated amount intervals,
calculated from the example data sets, are shown in Table XIII.
The bandwidths for this data are found in Table XIV.   Typical
estimated amount intervals are found in the analysis for
fenvalerate. At 4 ng this compound gave a range of 3.5 to 4.5 ng
at the confidence level described. This range was similar in the
analysis for fenvalerate in Dataset E, chlorothalonil, and
chlorpyrifos due to tight control of standards. These ranges
amounted to bandwidth percentages of from 10 to 14%. In more

Table XIII.   Estimated Amount Intervals from Inverse Transformed
Data. Overall α =0.05 where 95% of the Unknown
Amounts Will Lie within the Estimated Amount Interval
of the True Amount.

| | Estimated Amount Interval | | | |
Data Set	At Min. Amount 0.05 ng(a)	Mean	Interval at Mean	At Max. Amount 20 ng(a)
Fenvalerate	0.037-0.067	4.0	3.5 - 4.5	18. - 22.
				91. -110.(100)
Dataset A	0.028-0.083	1.5	1.16- 1.97	16.3- 25.
Dataset B	0.033-0.072	1.5	1.25- 1.85	17.3- 23.
Dataset C	0.018-0.113	1.6	0.98- 2.3	14.5- 28.
Dataset D	0.012-0.143	1.6	0.87- 2.7	13.4- 30.
Dataset E	0.190-0.32 (.25)	3.7	3.2 - 4.3	17.6- 23.
Dataset F	0.034-0.071	0.7	0.54- 0.80	4.3- 5.9(5.)
Chlorothalonil	0.015-0.026(.02)	1.3	1.09- 1.43	18.1- 22.
Chlorpyrifos, P-mode	0.82 -1.21 (1.0)	34.	31. -37.	470. -540.(500)

(a) () gives different minimum and maximum amounts.

routine analyses the bandwidths found in Datasets A, B, and F
would be indicative of typical error found. For these the
smallest bandwidth (at the mean of the graph) was in the range of
20 to 30%. These bandwidths were typical of electron capture and
phosphorus-mode flame photometric detection data utilizing packed
column chromatography with equipment described in the methods.
They were also typical when the regression model chosen was first
order. Mean-level bandwidths greater than 20-30% are probably
indicative that errors have been made in the analysis process that
should not be tolerated. In this case techniques would be
carefully scrutinized to find errors, outliers, or changing
chromatographic conditions. These should be remedied and the
analysis repeated whenever possible. Certain manipulation can be
done to reduce the bandwidth values. For example, they would be

Table XIV.   Estimated Amount Interval Bandwidth from Inverse
              Transformed Data. Overall α =0.05 where 95% of the
              Unknown Amounts Will Lie within the Estimated Amount
                      Interval of the True Amount.

	Estimated Amount Interval Bandwidth, % (a)		
Data Set	At Min. Amount 0.05 ng(b)	At Mean Amount	At Max. Amount 20 ng(b)
Fenvalerate	30	14	11 9.2(100)
Dataset A	55	27	21
Dataset B	39	20	15
Dataset C	95	45	33
Dataset D	131	58	42
Dataset E	27(.25)	15	13
Dataset F	37	20	16 (5.0)
Chlorothalonil	28(.02)	14	10
Chlorpyrifos, P-mode	20(1.0)	10	6.7(500)

(a) Bandwidth, % = $\frac{\{(\text{Upper Amount} - \text{Lower Amount})/2\}\{100\ \%\}}{\text{Amount Best Estimate}}$

(b) () gives different minimum and maximum amounts.

reduced if a higher order regression model were used to more
closely fit the data. Higher alpha values and lower probabilities
could also be chosen. However, data and methods should not be so
manipulated in order to produce an appearance of "improving" the
quality of results, i.e., the width of the estimated amount
interval.

Calibration Data Extrapolation Caveat.   Since extrapolations
cannot be done in performing proper regression line calculations,
it follows that there will be concern about interval estimates
from data that do occur at the extreme ends of the range of
standards. To avoid extrapolation we suggest that response values
be limited to those values corresponding to the range of amounts
of the calibration standards. At the extreme ends, however, one
end of the estimated amount interval would then extend into an
extrapolated region, the lower end at the minimum amount and the
upper end at the maximum amount.
    For determining the estimated amount interval at the lower or
minimum end of the regression line three cases arise: 1. The
lower end can be calculated and is positive. 2. The lower end can
be calculated and is negative and/or the interval is excessively
long. 3. The lower end cannot be calculated because it is
negatively infinite. A similar situation exists at the maximum
end of the regression line. In these cases the true uncertainty
is properly reflected by the calculated amount uncertainty. The
effects found in cases 2 and 3 could be studied to determine their
possibility with various types of calibration data. This paper,
however, will not delve into that aspect.

Inverse Transformation to Real Values. Step 6.   All of the
statistical steps are calculated with transformed data. With the
inverse transformation as a final calculation the original units
are recovered and evaluated as desired.
    The value of this work is to illustrate the importance of
including the estimated amount interval with every calculated
amount estimate in written reports. These can be calculated at
any response level. As an illustration of this process we can use
the data of Table XIII. If an analysis of fenvalerate were being
performed and the standards were those of Dataset A, an amount
estimated to be 1.5 ng would be reported as having a total error
range of 1.16 to 1.97 ng or in rounded figures 1.2 to 2.0 ng.
With the standards of Dataset B, however, the 1.5 ng value had a
range of 1.3 to 1.9 ng. This example shows that with more
standards a higher precision was obtained (assuming the quality of
the standards was equivalent). If the analyst used Dataset D
standards, those containing an artifact peak in the standards
would have to report that at 1.6 ng his range would be a much
larger 0.9 to 2.7 ng. Clearly a client would prefer the former
over the latter.

In using data of this sort regulatory agencies would have to redefine various limits important to their work. For example, if chlorothalonil had an "action" limit of 1.4 ng chlorothalonil (or appropriate concentration limit) and an analysis showed 1.3 ng present, what would they do with an amount interval of 1.1 to 1.4 ng for this determination? Then, also with fenvalerate if an "action" limit were 25 ng and the careless analyst using Dataset D found 20 ng present, its range of 13 to 30 might jeopardize the sale of food resulting from this determination. However, with some sharp observations and a little more chromatographic determinations, utilizing Dataset E, the same sample would be found to contain 20 ng with a range of only 18-23 ng. A different conclusion might be reached.

Model Selection. Various approaches can be used to select the best regression model, but most workers desire to use the simplest, the linear regression model. Higher order polynomials and spline methods have recently been studied and are reported in other papers of this volume (2,8). One common problem with chromatographic detectors is the loss of sensitivity at the top of the graph due to the inability of the detector to account for all of the molecules passing through. Attempting to linearize such data is often not possible. One solution for this problem is to segment the range needed into upper and lower regions so that a linear condition is achieved in each. The linear model is indeed the simplest with or without segmentation and the one that we used to study other aspects of the calibration problem, namely that of the calculation of the error bounds for unknowns.

Regression Accuracy

To demonstrate the accuracy of three data treatment methods we show the results of treating the same data in three ways: no transformation, log-log transformation, and the selected transformation as determined by this work.

Four datasets are taken as examples for discussion, those for chlorothalonil, Dataset B (a general fenvalerate data set), Dataset F (a data set with a fairly short range), and chlorpyrifos. The first three sets were obtained with an electron capture detector and the last with a flame photometric detector on the phosphorus mode. Comparisons of accuracy were made in amount accuracy, the standard deviation of the response about the regressed line, and the calculated R-squared value for the regression. Tables XV-XVIII give the comparison data calculated at half and twice the lowest response, at a median response, and at 0.8 and 1.2 times the highest response. The first and last conditions (data rows) in these tables from the standpoint of good laboratory practice were questionable because they would extend beyond the range of the data, but they were chosen to make a rigorous comparison.

The data were also hand plotted on log-log coordinates. The graph for chlorothalonil showed detector saturation at highest levels where increasing amounts gave slightly lower responses. Datasets B and F showed no saturation at high amounts but rather a decided leveling off at lowest amounts yielding higher responses. Chlorpyrifos showed a very slight saturation at the highest level.

Large differences were found at the lowest response values in all four data sets between the non-transformed data on one hand and the transformed data and the hand plotted data on the other hand. The amounts calculated by non-transformed linear regression at the lowest level for the four data sets were -2000, 860, 230, and -690%, respectively, different from the amount value determined by hand plotting. Those calculated by proper transformation regression as described by this paper were 9, 12, 0.0, and 8.5% different from the amount value determined by hand plotting. Those calculated by a log-log transformation, a transformation seen in Table I and in unpublished work to be too strong, were -15, 12, 12, and 0.0% different. Clearly the non-transformed regression gave erroneous results at the trace level. Also, the proper transformation as described in this work actually appears to have been more accurate than the log-log method.

On the other hand the non-transformed regression gave close control at the highest end of the regression which is expected since the large numbers control the direction of the regression

Table XV.   Regression Accuracy with Various Transformation
             Treatments of Data. Chlorothalonil Data Set.

Selected Response Values	Value of Unknown Amount			
	No Transform	Proper Transform	Log-Log Transform	Hand Plot
1.3	-0.21	0.012	0.0094	0.011
5.0	-0.18	0.038	0.037	0.039
57.	0.30	0.40	0.44	0.41
1700.	15.3	15.1	14.2	15.0
2530	23.0	23.7	21.3	24.5
Standard Deviation (a)	0.41	0.022	0.039	
Corr. Coef. Squared	99.7	100.	99.9	

(a) measured in transformed units; thus not directly comparable.

Table XVI.   Regression Accuracy with Various Transformation
                        Treatments of Data.   Dataset B.

Selected Response Values	Value of Unknown Amount			
	No Transform	Proper Transform	Log-Log Transform	Hand Plot
0.6	0.24	0.028	0.028	0.025
2.2	0.28	0.090	0.090	0.093
30.	0.88	0.93	0.93	1.00
730	16.0	16.1	16.2	15.7
1130	24.6	23.7	23.9	23.2
Standard Deviation (a)	0.21	0.018	0.029	
Corr. Coef. Squared	99.9	99.9	99.9	

(a) measured in transformed units; thus not directly comparable.

Table XVII.   Regression Accuracy with Various Transformation
                        Treatments of Data.   Dataset F.

Selected Response Values	Value of Unknown Amount			
	No Transform	Proper Transform	Log-Log Transform	Hand Plot
1.1	0.083	0.025	0.028	0.025
4.3	0.16	0.11	0.11	0.11
20.	0.52	0.53	0.50	0.53
170.	3.95	4.05	4.19	4.0
250.	5.78	5.77	6.14	5.7
Standard Deviation (a)	0.087	0.0088	0.034	
Corr. Coef. Squared	99.8	100.	99.8	

(a) measured in transformed units; thus not directly comparable.

Table XVIII.   Regression Accuracy with Various Transformation
                Treatments of Data.   Chlorpyrifos Data Set.

Selected Response Values	Value of Unknown Amount			
	No Transform	Proper Transform	Log-Log Transform	Hand Plot
0.5	-2.8	0.51	0.47	0.47
2.0	-1.0	1.97	1.96	1.95
20.	19.5	19.9	20.6	20.3
360.	408.	403.	394.	394.
540	614.	619.	597.	610.
Standard Deviation (a)	5.7	0.012	0.018	
Corr. Coef Squared	99.9	100.	100.	

(a) measured in transformed units; thus not directly comparable.

line more than small numbers. For the non-transformed data the
differences with the hand plotting amounted to only -6.1, 6.0,
1.4, and 0.7%. The proper transformation method gave even better
differences of only -3.3, 2.2, 1.2, and 1.5%.   The log-log
transformation method gave slightly less accurate differences of
-13, 3.0, 7.7, and -2.1%.

Discussion of the Regression Accuracy.   The inaccuracies found in
the above discussion for some data treatments are explained in two
ways. First, all the data sets have some curvature in them,
either at the top or at the bottom end of the graph.   Second,
least squares estimation puts greater influence on the large
values for non-transformed data and on small values for data sets
that have been transformed too much (i.e., log-log transformed).
For non-transformed data the large values have greater influence
on the line and the precision is best at the top of the curve.   On
the other hand a large error is expected at the bottom of the
curve.   For the log-log transformation the opposite is true since
the small numbers after transformation have a greater influence.
The general observation is that the non-transformed method uses
too weak a transformation and is dramatically in error at the
trace level.   On the other hand the log-log transformation is too
strong and some error is found at the highest level of amounts.
    If the data had been inherently linear, the estimation at
either end would have been fairly good with no transformation.

However, with improper transformation the calculation of confidence bands and amount interval estimates is erroneous because of the non-constant variance.

   We note that the use of the R-squared value in determining the quality of the regression line is dubious since almost all regressions using extended range of amounts in chromatographic processes will have R-squared values close to 100%. The standard deviation values for each of these regressions is given in the table merely for reference and cannot be compared since the units are affected by transformation.

Conclusions

The basic shortcoming of statistically determined calibration graphs as found in the literature has been the omission of confidence bands. When properly constructed, information is available about the error in the calibration process and the resulting uncertainty in the estimated unknown amounts.

   Although the results described in this paper have been known before, the combination of transforming the data to achieve both constant variance throughout the range of the standards and to provide linearity in the calibration graph reduces the calibration problem to the basic setting for which statistical procedures exist. We can then apply a statistical procedure which generates error bounds on our estimates with a precisely defined interpretation.

   It remains a problem requiring constant diligence to ensure that reported laboratory results are not only accurate, but that they are presented with an accurate assessment of their uncertainty. We hope this work helps laboratory workers meet this goal.

Acknowledgment

Literature Cited

1. Mitchell, D. G.; Mills, W. N.; Zdeb, M. Anal. Chem. 1977, 49 , 1655-9.
2. Mitchell, D. G., Garden, J. S. Talanta 1982, 29 ,921-9.
3. Willard, H. H.; Merritt, L. L.; Dean, J. A. "Instrumental Methods of Analysis"; 5th Ed.; D. Van Nostrand Co.: New York, 1974; p. 531.
4. Garden, J. S.; Mitchell, D. G.; Mills, W. N. Anal. Chem. 1980, 52 , 2310-15.
5. Schwartz, L. M. Anal. Chem. 1976, 48 , 2287-9.
6. Reekie, D.; Marshall, R. B.; Fleck, A. Clin. Chim. Acta 1973, 47 , 123-31.

7. Schwartz, L. M. Anal. Chem. 1977, 49 , 2062-8.
8. Wegscheider, W. Chapter 9 in this book.
9. Kurtz, D. A. Anal. Chim. Acta 1983, 150 , 105-14.
10. Agterdenbos. J Anal. Chim. Acta 1979, 108 , 315-23.
11. Thompson, M. Analyst 1982, 107 , 1169.
12. Schwartz, L. M. Anal. Chem. 1979, 51 , 723-7.
13. Bocek, P.; Novak, J. J. Chromatog. 1970, 51 , 375-83.
14. Cairns, T.; Rogers, W. M. Anal. Chem. 1983, 55 , 54A-57A.
15. Agterdenbos, J.; Maessen, Balke, J. Anal. Chim. Acta 1981, 132 , 127-37.
16. Schwartz, L. M. Anal. Chem. 1978, 50 , 980-5.
17. Tukey, J. W. Ann. Math. Stat. 1957, 28 , 602-32.
18. Box, G. E. P.; and Cox, D. R. J. Royal Stat. Soc. B. 1964, 26 , 211-252.
19. Neter, J.; Wasserman, W. "Applied Linear Statistical Models"; Richard D. Irwin, Inc.: Homewood, IL, 1974; p. 512-3.
20. Neter, J.; Wasserman, W. "Applied Linear Statistical Models"; Richard D. Irwin, Inc.: Homewood, IL, 1974; p. 149-153.
21. Lieberman, G. J.; Miller, R. G.; Hamilton, M. A. Biometrika 1967, 54 , 133-45.
22. Neter, J.; Wasserman, W. "Applied Linear Statistical Models"; Richard D. Irwin, Inc.: Homewood, IL, 1974; p. 153-4.
23. Scheffe, H. Annals of Stat. 1973, 1 , 1-37.
24. Trout, J. R.; Swallow, W. H. Technometrics 1979, 21 , 567-74.
25. Draper, N.; Smith, H. "Applied Regression Analysis"; John Wiley & Sons: New York, 1966.
26. Hunter, J. S. J. Assoc. Offic. Anal. Chem. 1981, 64 , 574-83.
27. Miller, R. G. "Simultaneous Statistical Inference"; McGraw Hill Book Co.: New York, 1966.
28. Conover, W. J.; Johnson, N.; Johnson, M. Technometrics 1981, 23 , 351-61.
29. Gottschalk, C. Fresenius Z. Anal. Chem. 1975, 276 , 81.
30. Maessen, F. J. M. J.; Elgersma, J. W.; Boumans, P. W. J. M. Spectrochim. Acta Part B, 1976, 31 , 179.
31. Neter, J.; Wasserman, W. "Applied Linear Statistical Models"; Richard D. Irwin, Inc.: Homewood, IL, 1974; p. 113-121.
32. Bickel, P. J.; Doksum, K. A. J. Am. Statist. Assoc. 1981, 76 , 296-311.
33. Carroll, R.; Ruppert, D. Biometrika 1981, 68 , 609-615.

Use of Cubic Spline Functions in Solving Calibration Problems

WOLFHARD WEGSCHEIDER

Institute for Analytical Chemistry, Micro- and Radiochemistry Technical University, Technikerstraße 4, A-8010-Graz, Austria

The analytical exploitation of the full dynamic range of a detection principle invariably encompasses non-linear portion of the concentration vs. response function. The use of cubic spline functions for the description of this relationship is discussed after a short introduction to the theoretical principles of spline approximations.
    Typical applications in calibration are reviewed and as practical example data from the determination of the pesticide fenvalerate by gas chromatography with electron capture detection are evaluated.

Non-linear concentration/response relationships are as common in pesticide residue analysis as in analytical chemistry in general. Although linear approximations have traditionally been helpful the complexity of physical phenomena is a prime reason that the limits of usefulness of such an approximation are frequently exceeded. In fact, it should be regarded the rule rather than the exception that calibration problems cannot be handled satisfactorily by linear relationships particularly as the dynamic range of analytical methods is fully exploited. This is true of principles as diverse as atomic absorption spectrometry (1), X-ray fluorescence spectrometry (2), radio-immunoassays (3), electron capture detection (4) and many more.
    One of the strategies for handling non-linear curves is the amending of the linear function by higher order terms. The resorting to polynomials of higher orders leads to success only in those instances where the shape can reasonably be represented by polynomial approximation. Other strategies include piecewise fitting of linear functions or the use of appropriate transformations with the aim of retaining

the readily accessible algebra of linear least squares al-
gorithms. In many instances there is, of course, another
driving force behind the search for transformations: that
of better understanding the nature of the processes involved.
    It is the aim of this contribution to review the principles
of the very flexible curve fitting approach by cubic spline
functions for construction of concentration/response curves,
to demonstrate the applicability of this methodology for
several data sets relevant in pesticide residue analysis
(5) and to make some general recommendations for the use
of cubic spline functions in calibration problems based on
experience in this laboratory with radio-immunoassay curves
(6) and atomic absorption curves (7). The reason for the
tremendous success of spline functions that also led to the
exploration of this possibility for solving calibration problems
can be summarized as follows (8, p. 123):
- Ordinary polynomials are inadequate in many sitations.
- Most functions arise for the experimenter from the physical
  world not from the mathematical world.
- Functions expressing physical relationship are frequently
  of a disjointed or disassociated nature; their behavior
  in one region may be totally unrelated to their behavior
  in another region. Polynomials and most other mathematical
  functions have just the opposite property. The same is
  not true for spline functions.

Theory

We are trying to determine the curve S, called <u>spline function</u>,
with

$$S(x) = E \, (y|x) \qquad\qquad Var \, S(x) = \sigma^2(x)$$

which is sufficiently smooth. The data are

$$(x_i, y_i), \quad i = 1, \ldots\ldots, n$$

The curve S consists of piecewise polynomials of degree m.
The pieces are linked together at the end of certain inter-
vals, the exact positions being termed <u>knots</u>. The coefficients
for each interval are chosen so that the first (m-1) - deriva-
tives of the functions are identical at the knots. The re-
sulting function is thus continuous with (m-1) continuous
derivatives. In a sense splines therefore can be regarded
as a compromise between a polygon and an interpolating polynom:
low order polynomials are assembled to a function that can
be differentiated in the entire fitting domain as often as
possible. This contribution will only treat cubic spline
functions (m = 3) whose representation can be given as

$$y = S(x) = P_i(x) = a_i + b_i x + c_i x^2 + d_i x^3 \qquad (1)$$

where the total curve $S(x)$ is composed of $(p-1)$ segments $P_i(x)$ leading to $(p-1)$ $(m+1)$ coefficients a_i, b_i, c_i, d_i; p being the number of knots. The choice of degree $m=3$ is a compromise between the too rigid $m=2$ that permits only the first derivative to be continuous at the knots and higher orders, $m>3$, that need the estimation of an excessive number of parameters without noticeable improvement at the knots as the continuity of higher order derivatives, $(m-1)>2$, cannot be sensed by the human eye/brain interface. The choice of the number and positions of the knots along with the selection of the degree m of the spline functions gives the experimenter a great flexibility that has been compared to the selection of functions type in ordinary curve fitting problems (9). Formal strategies (8,10) as well as heuristic arguments (9) are guiding the experimenter towards a useful selection of these parameters. This flexibility, however, leads to a multitude of solutions in calibration that can be a nuisance in practice. If every calibration point (with the exception of replicates that can be treated by averaging their response values) is treated as a separate knot, two different situations can be distinguished. In case of very precisely defined response values, y_i, obtained in practice by a high number of replicates in presence of small random errors, it is possible to use interpolating splines. Presumably, the more frequent case envisaged will be the one, where relatively few data points whose random errors are not negligible and/or that are not highly replicated span the concentration (or mass) domain. This needs to be considered in greater detail.

Calibration graphs defined by data with non-negligible error have to be constructed by some kind of smoothing operation. In cases, in which the form of the underlying curve is known a priori, the latter can be approximated by minimizing the squares of deviations. Otherwise a spline function can be used (11,12). The spline function $S(x)$ is constructed to minimize a measure of smoothness defined by

$$\int_{x_1}^{x_n} \left[S''(x) \right]^2 dx \qquad (2)$$

among all functions

$$\sum_{i=1}^{n} \left[\frac{S(x_i) - y_i}{\sigma y_i} \right]^2 \leqslant R \qquad (3)$$

$\delta y_i > 0$ and $R > 0$ are given numbers. Reinsch (11) suggests that for $\delta y_i \approx \sigma_i$ natural values of R lie in the confidence interval

$$n - (2n)^{1/2} \leqslant R \leqslant n + (2n)^{1/2} \qquad (4)$$

R = 0 leads back to the problem of interpolation by spline functions. It should be noted at this point that the condition stated by eq. (4) is not sufficient for the construction of calibration curves and additional considerations have to take effect. A reformulation of the problem stated in Equations (2) and (3) gives us with $\delta y_i = 1$ for all i calibration points another look at the problem that clarifies the role of the integral in Equation (2) as balanced against a value of R. Find S (x) to

$$\min \left\{ \frac{1}{n} \sum_{i=1}^{n} \left[S(x_i) - y_i \right]^2 + \lambda \int_{x_{min}}^{x_{max}} \left[S''(x) \right]^2 dx \right. \qquad (5)$$

where λ is a non-negative real number. The first term is a measure of fidelity to the data, and the second term is λ times the smoothness of S. It is important that for each λ 0 there exists an unique solution to S(x) (11). For $\lambda \to \infty$, S(x) becomes increasingly smooth and the limiting S(x) is a straight line. As $\lambda \to 0$ (equivalent to $R \to 0$) the spline function S(x) passes through each (possibly replicated) data point. Considerable research (e.g. refs. 13-17) has gone into choosing the appropriate degree of smoothing from the data without knowledge of errors σ_i^2 of the data y_i by cross validation. The intuitive idea of cross validation is quite simple and goes as follows (15): let S_n be the smoothing spline with a given value of λ calculated by using all (n-1) points except the k-th. The goodness of λ is now evaluated by predicting y_k by $S^k_{n,}$. Over all n values this gives

$$V_0(\lambda) = \frac{1}{n} \sum_{k=1}^{n} \left[S^k_n(x_k) - y_k \right]^2 \qquad (6)$$

The cross validation estimate of λ for equally spaced data is defined to the minimizer of $V_0(\lambda)$ (13,15). Results were also given for the "generalized" cross validation to compensate for unequally spaced data points (15-17) but will not be

discussed here for the following reason, founded in the par-
ticular problem treated in this contribution, namely cali-
bration: Only rarely sufficient number of data are available
for using cross validation. Wahba (18) does not recommend
generalized cross validation "for much fewer than 20 data
points". Excluding replicates this number is usually unavailable.
The appropriate degree of smoothing has thus to be defined
by other rules that are adapted to the particular calibration
situation. The various possibilities explored in literature
are now reviewed.

Past Development and Present Use of Cubic Spline Functions in Calibration

The first applications in calibration problems appear to
have been in the field of competitive protein binding assays.
Marschner et al. (19) have introduced the concept of smoothing
spline functions (11,12) to allow for the great variability
of radio-immunoassay (RIA) standard curves. It is advised
(19) that at least duplicates are run at each assay level
to adjust the relative weights δy_i of the data points.
Although no formal logic is given the algorithm was used
interactively and intermediate as well as the final spline
curves were plotted after appropriate adjustment of R in
eq. (3). This choice of Reinsch's approach seemingly is
in conflict with the advice given by Wold (9). He recommends
the method of Reinsch in cases where
"(i)   the number of data is sufficiently large,
  (ii)   the accuracy of the observed y-values is fairly well
           known,
 (iii)   no transformation of the data is needed,
 (IV)   the values are roughly equidistant ($x_k - x_{k-1}$ does
           does not vary with more than - say - 50 %)."

Upon closer examination none of the above requirements are
really met: one seldom has more than 5 - 7 calibration points
spread over the x-domain; occasional outliers do occur and
the accuracy estimate is far off in this instance; heteros-
cedascity of the data is commonplace; the range of the analytical
methods and the variation of sensitivity (expressed as first
derivative of the response vs. concentration graph) strongly
encourages the analyst to choose standards for calibration
that are unevenly spread over the x - axis (concentration,
mass) with the larger proportion of all data crowded at lower
x-values. It is not straightforwardly judged why the use
of smoothing splines has been successful in spite of the
fact given. One reason could be the operator interaction
that was inherent in the early approaches (19,20) and introduced
the past experience of human intellect to sort out outliers
and adapt smoothing factors. At the same time this must also

have helped to avoid drawbacks of overfitting. The use of
ß-splines with in some cases just a single knot was advocated
by Mosley and Bevan (21). However, the problem of how to
choose "correct" smoothing parameters is only substituted
by the problem of choosing the right number and position(s)
of the knot(s). Although no measures were taken to account
for the varying standard deviation, a logarithmic transformation
of the x - values served generally to yield roughly equidistant
knots. As the degree of automation of RIA procedures advanced
the tendency to substitute the operator interaction by some
objective and reproducible program logic was strong. The
regularities of the signal/concentration dependence served
as pivotal points for this accomplishment: no extreme points
are generally found in calibration, zero or one inflection
point is observed and the dependence of random errors as
a function of concentration is at least roughly known from
previous experience. These facts can be used in an interactive
procedure to adapt the smoothing parameters to yield the
smallest smoothness consistent with the known or assumed
characteristics of the curve. Smoothing parameter and position
of the inflection point can in turn be used to describe the
quality of a particular realization of the curve (22-24).
In spite of the great success of the spline functions for
radio-immunoassay standard curves caveats are voiced primarily
concerning the conscientious choice of the smoothing parameters
(25) and the overfitting (26). Both aspects deserve attention
in other applications as well.
    Another application that is closely related to the present
problem is the calibration of photographic emulsions for
atomic emission spectrochemical analysis (27,28). Here the
spline function serves to describe the complicated relationship
between incident intensity and consequent plate blackening
for which in spite of decades of research no fully satisfactory
linearization has been worked out. Although no exact strategy
for the optimization of smoothing parameters is described
it is also operational through successive increase to the
point where no theoretically unjustified "wiggles" are left
in the curve. Thus, the degree of smoothing is determined
empirically, but the smoothing parameters are adjusted auto-
matically and reproducibly. Examples of how this goal may
be accomplished are given in the next section.

Worked Examples

Testing the Accuracy of a Calibration Spline Function.
Of
primary concern in calibration is the freedom from systematic
errors introduced by fitting the wrong model. For judging
the accuracy of the cubic spline functions, it is therefore
desirable to start with a curve of known shape. Particularly
difficult to adapt by ordinary polynomial expressions are

curve shapes bounded by a limiting value. Among these the
model function selected here is

$$y = (1 - \frac{1}{x}) \qquad (7)$$

where y be signal and x be mass or concentration. The studied
range of x is from 1 to 10 and seven standards are placed
in this interval. The exact position of data along the x
- axis was determined arbitrarily but it was regarded important
to have a higher density of calibration points in those parts
of the calibration curve where the slope (sensitivity) is
relatively high. No noise was superimposed as the primary
goal of this exercise was to find the degree of correspondence
of the true and spline fitted curve between the knots. The
second point that can be demonstrated through this example
is the influence of grossly misplaced outliers. The raw data
are given in Table I. The misplaced point occurs for the
value of standard 4. The implementation of the smoothing

Table I. Raw Data for and Estimated Response of Fitting Cubic
         Splines to the Function in Equation (7)

number of standard	mass or concen- tration (a.u.)	true signal(a.u.)	fitted signal (a.u.) with all $\delta y = 1$	
			R = 1	R = 0.01
1	1	0	0.1235	0.0137
2	1.5	0.333	0.2961	0.3206
3	2	0.5	0.4142	0.4952
4	4	0.75	0.6739	0.7517
5	6	0.833	0.8101	0.8331
6	8	0.875	0.9018	0.8749
7	10	0.9	0.9721	0.9030

spline fitting program is given in ref. 29. First, it is
demonstrated how the smoothing parameters R and δy_i (Equation 3)
affect the shape of the spline function. For computation
the x - axis is transformed logarithmically and backtransformed
for the production of Figure 1. A smoothing parameter of
R = 1 (and all δy_i = 1) gives only a very rough approximation
while for R = 0.01 the difference between the true and the
calculated line cannot be represented graphically. Intermediate
values of R give curves that lie between the dotted and the
interrupted lines in Figure 1, but it is clear that as R → 0
the spline will run through the original data.

In case of gross deviations of the expected curve shape
it is possible to diagnose segments of the curve that are
suspected to be dominated by severly misplaced points. A
low reading for standard 4 (0.65 instead of 0.75) gives an
abnormal shape of the curve between x = 3.31 and x = 4.95,
depicted by the dotted line in Figure 2. If the neighboring
standards at x = 3,4 and 5 are given lower weights (δy =
100) because of this, the true shape is almost restored as
seen from the broken line in Figure 2. Alternatively, a high
reading at x = 4 (0.85 instead of 0.75) gives an anomalous
shape of the curve for R = 0.01 and all δy_i = 1 in the range
of x from 5.98 to 10. This can be seen in Figure 3 from the
dotted line that flattens out beyond x = 4 in an extraordinary
way and whose first derivative in effect increases towards
the end of the range. Here just the points at x = 4 and x = 6
were reweighted with δy = 10 and the resulting curve (Figure
3, broken line) is distinguishable, yet much closer to the
theoretical curve than if all standards are given equal weights.
It is clear that this rough check based on shape considerations
works only for grossly misplaced points. In practice replication
will permit a more powerful detection of errors.

Calculation of Spline Functions with Pesticide GC - ECD Data.
In considering the data from the determination of the pesticide
fenvalerate by GC - ECD it is evident that the variance in-
creases with mass. Although each of the Datasets points to
a somewhat different transformation (5) an exponential trans-
formation with the exponent being + 0.2 has been chosen by
standard procedures (30) as a compromise. Considerations
discussed in the last section are met by logarithmically
transforming the x-axis to achieve a close to even distribution
of points. The results of applying the cubic spline fitting
to Dataset A are shown in Figure 4. The entire curve is slightly
upward bent for R = 0.01 and all δy_i = 1. The choice of
these numbers assures that the curves satisfies two important
requirements of calibration graphs: (a) it hits within sta-
tistical limits ($s_i/\sqrt{n_i}$) the i-th calibration point replicated
n times with standard deviation of s_i; and (b) it shows no
undue oscillations or unevenness that can be observed if
the spline function is forced to run exactly through the
experimental mean response y_i. An overall decrease of the
precision of the measurements can be handled by increasing
R, an improvement of precision by decreasing R. Individual
points showing greater deviations between replicates or from
the "expected" curve can be given less weight by increasing
δy_i. For Dataset B with R = 0.01 and all δy_i = 1 the curve
shows the slightly sigmoidal shape depicted in Figure 5:
here an adjustment of R and/or selected δy_i's could lead
to a removal of the sigmoidal shape if so justified on theo-
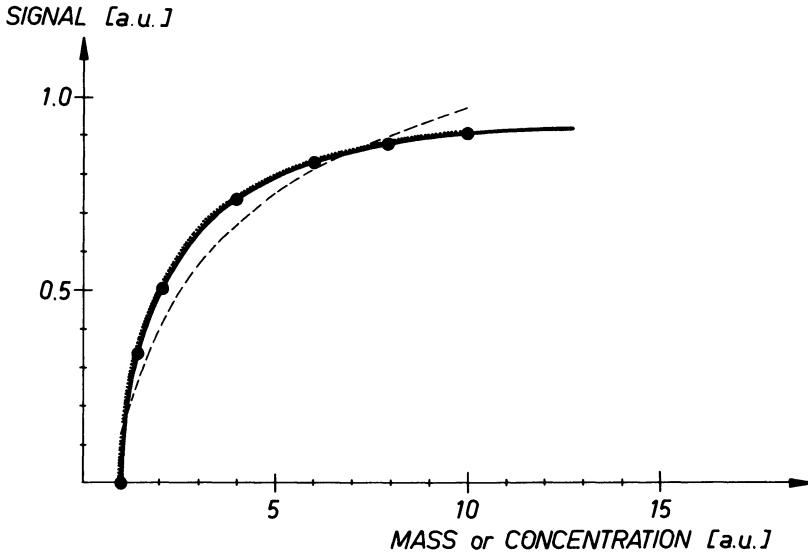
Figure 1. Graph of Equation (7) (full line) and spline approxi-
mations to Equation (7). All δy_i = 1
(R = 1: broken line; R = 0.01: dotted line)
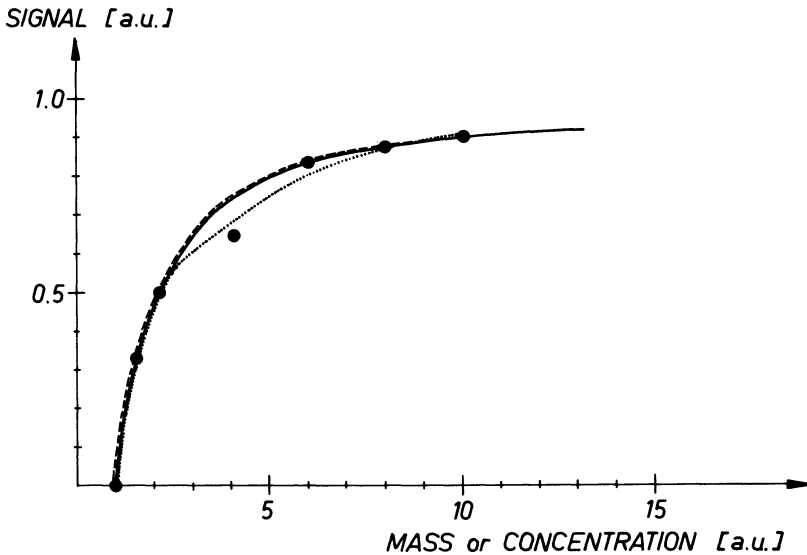
Figure 2. Spline approximation to Equation (7) with a low
reading at x = 4.
Equation (7): full line
all δy_i = 1: dotted line
$\delta y_1 = \delta y_2 = \delta y_6 = y_7 = 1$;
$\delta y_3 = y_4 = \delta y_5 = 10$: broken line

SIGNAL [a.u.]
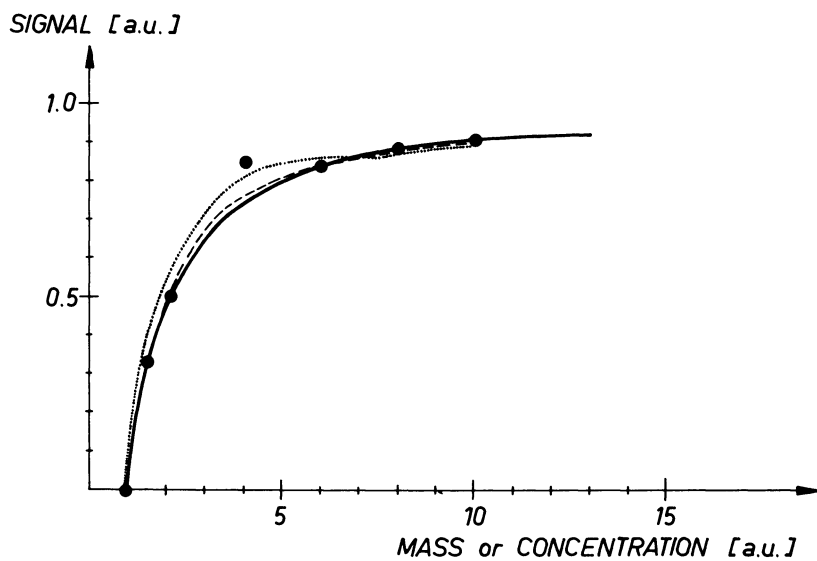
Figure 3. Spline approximation to Equation (7) with a high
reading at $x = 4$.
Equation (7): full line
all $\delta y_i = 1$: dotted line
$\delta y_1 = \delta y_2 = \delta y_3 = \delta y_6 = y_7 = 1$;
$\delta y_4 = \delta y_5 = 10$: broken line
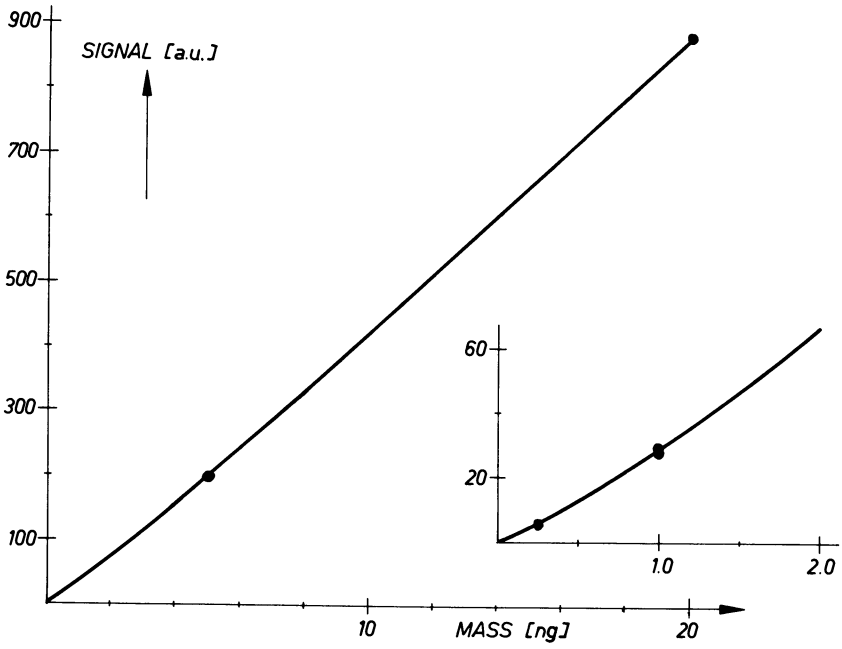
Figure 4. Spline approximation to Dataset A for the determination fenvalerate by GC-ECD
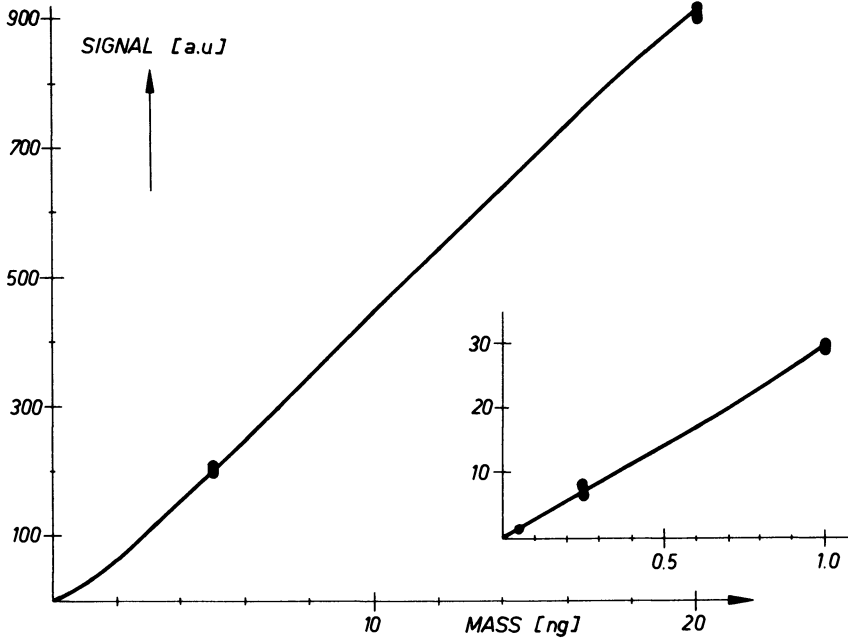
Figure 5. Spline approximation to Dataset B for the determination of fenvalerate by GC-ECD

retical or statistical grounds. Neither are giving this justifi-
cation for Dataset B so the parameters R and δy_i are given
the same values as with previous data sets.

Table II. Selected Values and Confidence Bands for the Determina-
tion of Fenvalerate Estimated by Cubic Spline Functions

| Dataset | Intensity (a.u.) | Central Value (ng) | Boundaries of Confidence Bands (ng) | | | |
| | | | 95 % | | 99 % | |
			lower	upper	lower	upper
A	1.2	0.054	-*	0.065	-*	0.073
A	30	1.04	0.95	1.13	0.89	1.20
A	440	10.2	9.6	10.8	9.2	11.2
A	850	19.2	18.2	20.3	17.5	21.1
B	1.2	0.051	-*	0.057	-*	0.060
B	2.2	0.081	0.072	0.092	0.068	0.097
B	30	1.13	1.05	1.21	1.02	1.24
B	460	9.98	9.51	10.48	9.34	10.67
B	730	15.7	14.9	16.5	14.7	16.8
C	5.0	0.173	0.104	0.267	0.084	0.311
C	34	1.00	0.76	1.29	0.67	1.42
C	227	5.00	4.13	6.00	3.83	6.42
C	500	10.1	8.5	12.0	8.0	12.8
C	1000	19.9	16.8	23.5	15.8	25.0
F	2.1	0.051	-*	0.060	-*	0.068
F	4.3	0.106	0.090	0.124	0.079	0.141
F	20	0.548	0.491	0.610	0.446	0.665
F	170	3.95	3.69	4.22	3.48	4.47

*not computed as the values are
 lying outside the calibration range

Construction of an Approximate Confidence Interval. An approxi-
mate confidence interval can be constructed for an assumed
class of distributions, if one is willing to neglect the
bias introduced by the spline approximation. This is accomplished
by estimation of the standard deviation in the transformed
domain of y-values from the replicates. The degrees of freedom
for this procedure is then diminished by one accounting for
the empirical search for the proper transformation. If one
accepts that the distribution of data can be approximated
by a normal distribution the Student t-distribution gives

the confidence bands required. Inserting these values into
the spline function gives the confidence bands in Table II.

Conclusions

Interpolating spline functions are well adaptable to non-linear
response/concentration data. To guard against the misgivings
of overfitting, checks for the appropriateness of the curve
shape have to be incorporated into the computer routines
that iteratively reweigh data points in the neighbourhood
of suspected outliers causing the shape anomalies. The resulting
curve is the least smooth approximation consistent with the
anticipated curve shape in terms of its first and second
derivatives. The particular requirements vary according to
the detection principle, but are frequently known from past
experience. The handling of data from the determination of
fenvalerate is performed quickly and reliably.

Acknowledgments

Literature Cited

1. Welz, B. "Atomabsorptionsspektrometrie", 2nd ed.,
    Verlag Chemie; Weinheim, 1978.
2. Tertain, R.; Claisse, F. "Principles of Quantitative
    X-ray Fluorescence Analysis", Heyden: London, 1982.
3. Finney, D.J. "Statistical Methods in Biological Assay",
    3rd. ed.; Griffin: London, 1978.
4. Wentworth, W.E.; Chen, E.C.M. in "Electron Capture:
    Theory and Practice in Chromatography"; Zlatkis, A.;
    Poole, C.F., eds.; Journal of Chromatography Library,
    Vol. 20, Elsevier; Amsterdam, 1981, p. 27-68.
5. Kurtz, D.A., personal communication; see the Appendix.
6. Wegscheider, W.; Knapp, G.; Frick, P.; Pittner, J. Ärztl.
    Lab. 1982, 28, 65-73.
7. Wegscheider, W., unpublished results.
8. Rice, J.R. "The Approximation of Functions", Vol. II,
    Addison-Wesley; Reading, MA, 1969.
9. Wold, S. Technometrics 1974, 16, 1-11.
10. Powell, M.J.D. "Curve Fitting by Cubic Splines", Rep.
    TP 307, Atomic Energy Establishment, Herwell; England, 1967.
11. Reinsch, C.H. Numer. Math. 1967, 10, 177-183.
12. Reinsch, C.H. Numer. Math. 1971, 16, 451-454.
13. Wahba, G.; Wold, S. Comm. Statist. 1975, 4, 1-17.
14. Wahba, G. Numer. Math. 1975, 24, 383-393.

15. Craven, P.; Wahba, G. Numer. Math. 1979, 31, 377-403.
16. Utreras, F. in "Smoothing Techniques for Curve Estimation"; Gasser, Th.; Rosenblatt, M., eds.; Lecture Notes in Mathematics, No. 757, Springer; Berlin, 1979, p. 196-232.
17. Utreras, F. "A Package for Smoothing Noisy Data With Splines"; Techn. Report, Nr. Ma-80-B-209, Departamento de Matematicas, Universidad de Chile, Santiago, 1980, 45 pp.
18. Wahba, G. personal communication.
19. Marschner, I.; Erhardt, F.; Scriba, P.C. in "IAEA Symp. Radioimmunoassay and Related Procedures in Clinical Medicine and Research", Istanbul, Sept. 1973, p. 111-122.
20. Marschner, I.; Dobry, I.H.; Erhardt, F.; Landesdorfer, T.; Popp, B.; Ringel, C.; Scriba, P.C. Ärztl. Lab. 1974, 20, 184-191.
21. Mosley, J.; Bevan, B.R. Ann. Clin. Biochem. 1977, 14, 16-21.
22. Nolte, H.; V.Z. Mühlen, A.; Hesch, R.D. J. Clin. Chem. Clin. Biochem. 1976, 14, 253-259.
23. Huber. W.; Mann, K.; Karl, H.J. J. Clin. Chem. Clin. Biochem. 1979, 17, 241-246.
24. Rawlins, T.G.R.; Yrjönen, T. Intern. Lab., Nov./Dec. 1978, 55-66.
25. Sandel, P.; Vogt, W.; W.; Popp, B.; Knedel, M. Fresenius'Z. Anal. Chem. 1976, 279, 126-127.
26. Finney, D.J. Clin. Chem. 1983, 29, 1762-1766.
27. Fischer, D.; Kreitz, N. Fresenius'Z. Anal. Chem. 1976, 281, 115-116.
28. Frigieri, P.F.; Rossi, F.B. Anal. Chem. 1979, 51, 54-57.
29. Späth, H. "Spline-Algorithmen zur Konstruktion glatter Kurven und Flächen"; Oldenburg; München, 1973, Chap. 6.
30. Box, G.E.P.; Hunter, W.G.; Hunter J.S. "Statistics for Experimenters", Wiley; New York, 1978, Chap. 7.

Comparison of Calibration Graph Amount and Estimated Amount Intervals

Calculated from Three Research Methods

DAVID A. KURTZ

Pesticide Research Laboratory, Department of Entomology, The Pennsylvania State University, University Park, PA 16802

Differences in calibration graph results were found
in amount and amount interval estimations in the use
of three common data sets of the chemical pesticide
fenvalerate by the individual methods of three
researchers. Differences in the methods included
constant variance treatments by weighting or
transforming response values. Linear single and
multiple curve functions and cubic spline functions
were used to fit the data. Amount differences were
found between three hand plotted methods and between
the hand plotted and three different statistical
regression line methods. Significant differences in
the calculated amount interval estimates were found
with the cubic spline function due to its limited
scope of inference. Smaller differences were
produced by the use of local versus global variance
estimators and a simple Bonferroni adjustment.

One of the opportunities that researchers rarely have is to be in
a position of a direct comparison of methods used by several
researchers that use the same data. Three of the fenvalerate
"unknown" Datasets described and used elsewhere in this volume
have been used as primary datasets by 3 research groups in the
solution of the calibration problem. Two aspects of the
calibration problem, namely, the accuracy of the calibration graph
and the description of statistical error as shown by the estimated
amount interval are examined here in comparing each of the
calibration methods.
    The methods used were those of Mitchell (1), Kurtz,
Rosenberger, and Tamayo (2), and Wegscheider (3). Mitchell
accounted for heteroscedastic error variance by using weighted
least squares regression. Mitchell fitted a curve either to all
or part of the calibration range, using either a linear or a
quadratic model. Kurtz, et al., achieved constant variance by a

power transformation and considered only a linear model in this
work. Wegscheider used a logarithmic transform to achieve
constant variance and used a cubic spline function to fit the
response standard means. Each of the methods has its purpose and
its limitations. In this comparison the reader will be able to
ascertain the effect on the calculated data when treated by each
of these three methods.

Data Description

Fenvalerate "unknown" data sets were used as examples in this
study. See Appendix A for the complete listing of the data.
Table I summarizes a description of the data sets. Datasets A and
B are both quality sets of the same amount range, 0.05 to 20 ng,
but Dataset A has only 2 replications/level while B has 5.
Dataset F has two replications/level and a much shorter range,
0.05 to 5 ng. Dataset A is barely linear at 95% level while
Dataset B shows curvature. Dataset F is an example of reasonably
good linear data with a minimum of range and data points.

Table I.   Description of Data Sets Studied.

Dataset	Samples Number	Levels Number	Amount Range ng	Linear Prob. Level(a)	Standard Deviation (a)
A	10	5	0.05-20.	.05	0.016
B	25	5	0.05-20.	.00	0.018
F	8	4	0.05- 5.	.30	0.009

(a) for transformed data.

Model Description

In Mitchell's work unequal variance of the response data was
compensated for by weighting the data by the variance at each
level. The regression parameters and the confidence band around
the regression line were estimated by least squares (4). The
overall level of uncertainty, α , was divided between the
variation in response values and the variance in the regression
estimation. His overall α was 0.05. The prediction interval was
estimated around a single response determination.

Mitchell's computer program first applied the method to a linear model and then calculated the amount values corresponding to response values of unknowns and the accompanying estimated amount interval ... calculated as a bandwidth. Bandwidth was defined as the percentage of half the difference of the upper and lower values of the estimated amount interval divided by the corresponding amount. The standards data was then shortened at the ends, always in such a way to maintain unknowns within the range, and the bandwidth recalculated. Narrower bandwidths were often found in this way. The method also allowed a further recalculation using a second order function model.

The method described by Kurtz, et al., (2) differs from the method described by Mitchell (1) in several minor respects. Kurtz, et al., used a power series transformation to stabilize the variance and to linearize the response. This approach can also be extended to include models with higher order terms. The prediction interval for a response was determined from standard normal statistics, making the assumption that a very large number of responses may be estimated from one calibration graph. The prediction interval around the average of a large number of responses will be narrower than the band around a single response. The overall α of 0.05 was split between the response variance and the variance in the regression estimation.

Wegscheider fitted a cubic spline function to the logarithmically transformed sample means of each level. This method obviates any lack of fit, and so it is not possible to calculate a confidence band about the fitted curve. Instead, the variance in response was estimated from the deviations of the calibration standards from their means at an α of 0.05. The intersection of this response interval with the fitted calibration line determined the estimated amount interval.

The Kurtz and Mitchell approaches both result in an amount interval which comes from the intersection of two confidence intervals: the confidence interval about the response and the confidence interval about the regression line. The Wegscheider approach will yield a narrower band, since the band is determined from the intersection of the response interval with the regression line itself. The three methods would be more directly comparable if there were a way to calculate a confidence band around the spline fit.

At any rate the practitioner must follow a two-step process in setting up a calibration graph: 1. Stabilize the response variance across the range needed and 2. choose an appropriate calculation function model. The response data is stabilized currently in two ways, either by weighting on a level-by-level basis or by applying some transformation function in the same manner to all the response values. The model chosen must approximate the data. It can be that a simple linear (as shown by a statistical test) function can serve this purpose adequately. The use of Mitchell's multiple linear function has been successfully

used by the author when the data have a slight curvature in them.
For data that does not conform well, a higher order model or an
alternate method should be chosen. Generally, then, the
practitioner should attempt to find some functional relationship
between the compound sought, such as the fenvalerate studied here,
and the readings on an instrument, such as a gas chromatograph.
The relationship takes into account that the day-to-day samples
determined are subject to error from several sources: <u>random
error</u> , instrument error, observer error, preparation error, etc.
This view is the basis of the process of fitting data to a model,
which results in confidence intervals based on the intrinsic lack
of fit and the random variation in the data.

Results

<u>Hand plotting of data.</u>   From the author's experience in
calculating the data for this work, it is very difficult to
determine what is the true or most accurate estimation of the
calibration graph. If hand plotting of the mean response at each
amount level is used, various estimations of unknown points can be
made depending on the methods used in plotting. Table II shows
amount estimations for various unknown response points using
Datasets A, B, and F when the data is plotted by three different
methods. Normally, analysts use simple rectangular coordinates.
To compile the data for this table with rectangular coordinates
the plotting was done in two scales, the lower values of responses
from 1 to 45 response area units and the overall scale for
responses greater than 45 area units. For extended range data,
however, analysts often use log-log coordinates. Data was thus
obtained using this method. Because difficulties were seen in
either of these methods, the response versus amount values were
plotted after they had been independently transformed as found in
Kurtz' work (<u>2</u>). Data from the use of all of these plotting
methods are shown in the table.
        The various plotting methods showed differences between them.
For Dataset A either the transformed plot was low at the low end
or the other two plots were high at the low end. Furthermore the
log-log plot appeared to be low at the high end as compared with
the other plotting methods. With Dataset B the log-log plot ap-
peared to be low at the high end. In Dataset F all three types of
plots gave similar amount values for the unknown responses listed.
        The main problem in plotting occurs in deciding where the
line should be placed when there is point scatter. A secondary
problem happens when the data has curvature and a curved line is
needed to describe the points. While the human mind can sense
relationships in plotting pretty well, there is difficulty in
estimating the proper plot position for these reasons. While two
of the data sets , A and F, are linear, corresponding to the lack
fit test at 95% probability, all of the sets show curvature in
log-log and rectangular coordinates.

Table II.   Estimated Amounts at Various Responses for Three
Regression Methods Using Datasets A, B, and F as Standards
and compared to Plotted Data.

| | Amounts by Method | | | | Amounts by Plotting | | |
| | Mitchell | | Kurtz, et al. | Weg-scheider | | | |
Response	Single	Multiple			Rect.	Transf.	Log-log
Dataset A (overall amount range is 0.05 to 20 ng)							
1.1	0.051	0.050(L)	0.053		0.050	0.046	0.050
1.2		(a)	0.057	0.054	0.054	0.050	0.054
48.6			1.52		1.57	1.49	1.58
223	6.4	5.4 (H)	5.9		5.6	5.8	5.6
440			10.7	10.2	10.3	10.4	9.6
883	19.6	20.5 (H)	19.8		20.0	19.6	18.0
Dataset B (overall amount range is 0.05 to 20 ng)							
0.6	0.33	0.31 (L)	0.028		0.025	0.026	0.025
1.1	0.048	0.047(L)	0.050		0.046	0.046	0.047
1.2			0.052	0.051	0.050	0.050	0.051
2.2	0.082	0.084(L)	0.090	0.081	0.090	0.089	0.90
30	0.92	0.99 (L)	0.93	1.13	1.01	0.97	0.99
234	6.3	5.4 (H)	5.8		5.5	6.0	5.9
460			10.7	10.0	10.3	10.7	10.6
730	16.4	16.2 (H)	16.1	15.7	16.0	16.0	15.7
920	19.6	20.3 (H)	20.0		20.0	19.8	19.2
1130	22.8	24.9 (H)	23.7		23.5	23.5	23.2
Dataset F (overall amount range is 0.05 to 5 ng)							
1.1	0.023	0.024(L)	0.025		0.027	0.025	0.025
2.1	0.051	0.051(L)	0.050	0.051	0.052	0.051	0.052
4.3	0.111	0.110(L)	0.111	0.106	0.113	0.112	0.110
20	0.54	same	0.53	0.55	0.54	0.53	0.53
66.3	1.18	same	1.68		1.70	1.68	1.68
170	4.1	same	4.1	4.0	4.0	4.0	4.0
217	5.0	same	5.0		5.0	5.0	5.0
250	5.7	same	5.8		5.7	5.7	5.7

(a) (L) is low range of 0.05-1 ng and (H) is high range of 1-20 ng

This part of the discussion is closed with the raising of a question. Is there an inherent truth that plotting itself should be done with respect to maintaining constant variance across the graph? The requirement of constant variance is invoked by statisticians when performing confidence band calculations. Thus, is it also a necessity in plotting the points themselves?

Differences in Estimated Amounts Determined from Three Methods.

Treatments of the data were compared at three points for Dataset A and 8 points each for Datasets B and F. While the points taken were actually points of the standards themselves, the described points should be thought of as unknown response points for analysis.

No data sets appear to conform to any of the hand plotted lines completely.

Of all of the three methods, Kurtz´ calculations appear to be the closest to its corresponding plot, the transformation plot. For these calculations only the lowest two points in both Datasets A and B appear to be greater than 4% different from the plotting of transformed data. These could even have been a result of slightly inappropriate plotting. The entire Dataset F falls quite closely to the plot.

For Wegscheider´s calculations the amount 10.2 in Dataset A appears to be a little high when compared with the log-log plot but very close to the other estimations. For Dataset B the 0.081 value appears to be 10% lower than the plotting figures and, strangely, similar to Mitchell´s single curve amount. At the same time the 1.13 amount appears to be 13% higher than the corresponding plotted amounts; it is even 12% high when compared to the mean of the original data points, 1.01, for that response level taken. (Adjusting a mean of 29.68 response units to 30.0 with a corresponding adjustment of amount from 1.00 to 1.01.). All the points for Dataset F are quite close to all other amount estimates shown.

Mitchell has scattered points that do not conform to plotted data. It was uncertain which plot should have been used to compare to; therefore the general consensus of all three was taken in this comparison and only obvious differences noted. For the single curve method the 6.4 amount in A was different from the plot. In B the 0.33, 0.082, 0.92, and 6.3 points were different. In F the 0.023 point was different and the 1.18 point appears to be in error. There was an improvement in precision when the multiple curve method was used. In A all three points conform. In B only the extreme points 0.31 and 24.9 do not appear to be similar. There were essentially no differences in F, save for the erroneous 1.18 point, over the single curve method.

The ultimate conclusion as to the method to be used for calibration graph usage is first to select the best calibration function. This may be linear or of higher order. Such a model must take into consideration instrumental and operator effects

which can be different at the ends of the calibration graph and
indeed even in the middle. Whatever the function, it should
reflect the real points. Then the response is transformed for
constant variance and the amount is transformed for conformity to
the proposed model. Other graphing methods may be chosen to more
closely fit the data. Finally statistical methods are chosen to
provide information about the error of the process.

Discussion on Model Selection for Accuracy.   The model selection
has a great deal to do with the calibration accuracy. We have
seen that some data sets can give close results regardless of the
model or statistical method chosen. For example, Dataset F data
were very close to being linear. The set tested, with power
transformed data, to show no evidence of lack of fit at the 30%
level with very little overall response dispersion. In this case
all methods of all authors closely agreed at all calculation
levels. Agreement was also obtained between all plotting methods
and between the plotting and statistical methods. In the cases of
Datasets A and B, both sets showed some curvature. With
transformed data, Dataset A did not show evidence for lack of fit
to the linear model at the 5% level and Dataset B did show
evidence of lack of fit even at the 1% level. In both these cases
we have found differences in accuracy between the methods that
utilized linear models and even with the cubic function. The
practitioner must verify the closeness of fit of the data to his
chosen model before proceeding to use his function.

Estimated Amount Interval and Bandwidth Data.   The bandwidths for
each of the points compared were calculated from Wegscheider's
data. They were markedly smaller than those calculated by either
of the other two workers. His bandwidths ranged from 4.9 to 16.0%
for all of the three data sets. Refer to Table III.
     The data of Mitchell and Kurtz, et al. are very similar as
compared both at given responses and across the whole range of
responses. For Mitchell's single curve results the bandwidths
ranged from 75 to 23% at the lowest response levels for the 3 data
sets. Note that the table contains bandwidth data at responses
both higher and lower than the range of standards for Datasets B
and F; these extrapolated data are not analytically appropriate
but were included to test the characteristics of the methods
tested. Kurtz' ranged from 54 to 37% for the same level. At the
top of the standards range, the bandwidths for Mitchell ranged
from 39 to 17% and Kurtz' ranged somewhat lower, 20 to 14%.
Mitchell's multiple curve method improved his bandwidth data
considerably. In shortening the range of standards he has halved
the bandwidths in Dataset A and achieved a half to one third
reduction in Dataset B.   Little improvement was found in
considering multiple curves for Dataset F since this set is quite
linear.

Table III.   Estimated Amount Bandwidths at Various Responses for
Three Regression Methods Using Datasets A, B, and F as Standards.

		Amount Bandwidth, %			
		Mitchell		Kurtz, et al.	Weg- scheider
Response	Amount Estimate	Single	Multiple		
Dataset A (overall amount range is 0.05 to 20 ng)					
1.1	0.05	75%	37% (L)	54%	-(a)
223		44	17 (H)	22%	-
440		-	-	21	5.9
883	20.0	39	18 (H)	20	-
Dataset B (overall amount range is 0.05 to 20 ng)					
0.6		60	38 (L)	44	-
1.1	0.05	40	24 (L)	39	-
2.2		34	20 (L)	35	12.3
30	1.00	33	17 (L)	22	7.1
234		25	7 (H)	16	-
460		-	-	15	4.9
730		21	7 (H)	15	4.9
920	20.0	20	7 (H)	14	-
1130		37(b)	6 (H)	14	-
Dataset F (overall amount range is 0.05 to 5 ng)					
1.1		56	51 (L)	44	-
2.1	0.05	23	23 (L)	37	-
4.3		20	20 (L)	30	16.0
20		20	20 same	21	10.9
66.3		18	18 same	17	-
170		17	17 same	16	6.7
217	5.00	17	17 same	16	-
250		18	18 same	16	-

(a) (L) is low range of 0.05-1 ng and (H) is high range of 1-20 ng
(b) appears to be in error; probably should be 17%.

The gradual shortening of the bandwidth from the lower end of
the standards to the upper end is normal due to the definition of
this quantity. From Figure 2 of Chapter 8 of this volume (2)
both the confidence band and the response interval are seen to be
similar in size across the total calibration range. This quantity
is then divided by the absolute calculated amount which is small
at the lower end giving relatively high bandwidths when compared
to the division of a large amount at the high end.

Discussion of Bandwidth Comparisons.   The narrowest bandwidths
found were those by Wegscheider. With the exception of the high
end of Dataset B in Mitchell´s multiple curve calculation
Wegscheider´s calculated bandwidths were one half to one third
those of the others. Several reasons could explain this effect:
    First, amount error estimations in Wegscheider´s work were
the result of only the response uncertainty with no regression
(confidence band) uncertainty about the spline. His spline
function knots were found from the means of the individual values
at each level. Hence the spline exactly followed the points and
there was no lack of fit in this method. Confidence intervals
around spline functions have not been calculated in the past but
are currently being explored (5).
    When I calculated the estimated amount interval from only the
response dispersion for the data using Kurtz´ methods, there was a
substantial reduction in the amount bandwidth from the total
bandwidth. This calculation was done by intersecting the bounds
of the response dispersion with the linear regressed line and
projecting these points to the amount axis. This reduction,
however, was not nearly enough to account for differences from
Wegscheider´s calculation to the others. In Table IV the data is

Table IV. Comparison of Overall and Response Dispersion portions
  of Amount Bandwidths of the Transformed Method (Kurtz, et al.)
with Bandwidths of the Spline Method (Wegscheider) for Dataset B.

Response	Calc. Amount	Kurtz, et al., Amount		Wegscheider Amount	
		Total Bandwidth Range	Response Dispersion Portion	Calc. Amount	Amount Bandwidth Range
2.2	0.090	35%	26%	0.081	12.3%
30.	0.93	22	18	1.13	7.1
730.	16.1	15	10.8	15.7	4.9

Table V.   Comparison of Overall and Response Dispersion portions
of Amount Ranges (in ng) of the Transformed Method (Kurtz, et al.)
   with Ranges of the Spline Method (Wegscheider) for Dataset B.

Re- sponse	Calc. Amount	Total Est. Amount Range	Response Dispersion Portion	Calc. Amount	Est. Amount Range
		Kurtz, et al., Amount		**Wegscheider Amount**	
2.2	0.090	0.063- 0.125	0.064- 0.112	0.081	0.072- 0.092
30.	0.93	0.75 - 1.15	0.79 - 1.13	1.13	1.05 - 1.21
730.	16.1	13.9 -18.6	14.2 -17.7	15.7	14.9 -16.5

shown for Dataset B. At 2.2 response units Kurtz´ overall amount
bandwidth of 35% has a contribution of 26% from the response
dispersion. This 26% is considerably larger than Wegscheider´s
calculation of 12.3%. At 30 response units the response
dispersion contribution from Kurtz´ work was 18% which is also
higher when compared to Wegscheider´s calculation of 7.1%. At 730
response units the difference between the two quantities has been
narrowed to about half that of Kurtz´.

     Table V includes the same comparison as Table IV between
Kurtz´ transformed method and Wegscheider´s spline method for
Dataset B and shows the actual amount units. It gives the reader
a clear idea of the actual sizes of estimated amount ranges
resulting from the uncertainty in response data.

     The substantial differences, when comparing response
dispersion alone, may be explained in two additional ways. The
first involves the size of the uncertainty when using the
Bonferroni inference. Kurtz, et al., and Mitchell use a
Bonferroni inference which splits the uncertainty in the
regression between the regression and the response portions.
Instead of a 95% probability there is a 97.5% probability for each
portion. The contribution to the uncertainty for a higher
probability thus has increased the width of the response
uncertainty band.

     A second reason is that the use of local variance versus
global variance can result in markedly different bands. The
separate calculations of variance at levels throughout the range
of standards produces a wider confidence interval at lower values
as seen in Kurtz´ method. If a common variance is used as the
variance estimate then a lower confidence interval is calculated
at each point as is probably the case in Wegscheider´s method.

It is not the purpose of this paper at this moment to investigate further for more detailed reasons for discrepancies in confidence bands or estimated amount intervals. That will be investigated fully at a later time. I do wish to point out that the assumptions one makes about the information he has and the statistical approaches he makes profoundly affect the resultant error calculations. Far from being a staid and dormant subject matter, statistical estimations of error are currently very actively being studied in order for scientific workers and citizens alike to be informed about the error in their work.

Acknowledgment

Literature Cited

1. Mitchell, D. G., Chapter 7 in this book.

2. Kurtz, D. A.; Rosenberger, J. R.; Tamayo, G., Chapter 8 in this book.

3. Wegscheider, W., Chapter 9 in this book.

4. Miller, R. G. "Simultaneous Statistical Inference", McGraw-Hill, New York (1966).

5. Cox, D. D. "Gaussian Approximation of Smoothing Splines," Technical Report #743, Department of Statistics, University of Wisconsin, Madison, WI, June 1984.

RECEIVED March 25, 1985

Application of Soft Independent Method of Class Analogy (SIMCA) in Isomer Specific Analysis of Polychlorinated Biphenyls

D. L. STALLING[1], W. J. DUNN III[2], T. R. SCHWARTZ[1], J. W. HOGAN[1], J. D. PETTY[1], E. JOHANSSON[3], and S. WOLD[3]

[1] Columbia National Fisheries Research Laboratory, U.S. Fish and Wildlife Service, Columbia, MO 65201
[2] University of Illinois Medical Center, Chicago, IL 60608
[3] Department of Organic Chemistry, University of Umea, Umea, Sweden

A principal components multivariate statistical approach (SIMCA) was evaluated and applied to interpretation of isomer specific analysis of polychlorinated biphenyls (PCBs) using both a microcomputer and a main frame computer. Capillary column gas chromatography was employed for separation and detection of 69 individual PCB isomers. Computer programs were written in ANSII MUMPS to provide a laboratory data base for data manipulation. This data base greatly assisted the analysts in calculating isomer concentrations and data management. Applications of SIMCA for quality control, classification, and estimation of the composition of multi-Aroclor mixtures are described for characterization and study of complex environmental residues.

Polychlorinated biphenyls (PCBs) are a class of synthetic chlorinated compounds with a total of 10 possible positions for chlorine attachment and 209 possible isomers, although the number of constituents observed in technical formulations is much smaller (1-3). PCBs have been produced by several industries worldwide in the form of technical formulations (4). Most PCBs produced in the U.S. originated as one of several products designated as Aroclors and were previously manufactured by the Monsanto Chemical Company (5). The major Aroclors produced were designated Aroclors 1242, 1248, 1254, and 1260, where the last two digits designate the percentage of chlorine by weight in the material. Each Aroclor is characterized by a somewhat different distribution of constituents having a characteristic chromatographic profile of about 70 to 100 chromatographic peaks (1,3). After PCBs were identified in 1966 as pollutants in fish and wildlife (6), they were soon recognized as global pollutants

widely distributed in the biosphere and in tissues of most forms
of life, including man (7-13). Concern about the occurrence of
PCBs in the environment centers on their persistence,
bioaccumulaton in organisms and toxicity (14, 15).

Of additional concern is the thermal conversion of PCBs to
polychlorinated dibenzofurans, especially when PCB--filled
electrical transformers are involved (16). Further, PCB isomers
with no chlorine atoms substituted in the ortho, ortho'-biphenyl
positions are particularly potent as inducers of aryl hydrocarbon
hydroxylase activity. The potency of certain dibenzofurans and
non-ortho, ortho' chlorine substituted PCBs in inducing activity
of these enzymes is similar to that of the highly toxic 2,3,7,8-
tetrachloro-dibenzo-p-dioxin (15, 17, 18).

Residues of PCBs exist as complex mixtures in almost every
segment of the environment, and characterization of these
residues poses a difficult challenge to the analyst. Routinely,
PCBs have been quantitated by comparing selected peak areas
observed in samples with those in one of several Aroclor mixtures
(19, 20). Packed column gas chromatography has usually been used
in these analyses, even though this technique provides poor
resolution of individual isomers and congener groups (21). The
problems associated with characterizing metabolically altered or
weathered PCBs is a formidable task that requires an analytical
approach with enhanced resolution.

Both the degree and position of chlorine substitution on the
biphenyl rings influence the physical-chemical and toxicological
behavior of the individual PCB constituents (15). When PCBs are
released into the environment, the original isomer distribution
pattern of any PCB formulation may be altered as a result of
specific interactions with the environment (14, 22, 23).

Further complications are encountered in describing the
residue profiles when more than one Aroclor mixture is
encountered in an ecosystem. Thus, it is important to consider
not only the total PCB concentration in a sample, but also to
characterize the distribution of individual PCB isomers present
in a sample.

Because information defining the distribution of the various
PCB isomers among the compartments of lotic aquatic ecosystems
(streams) is limited, the Columbia National Fisheries Research
Laboratory began a detailed analysis of the distribution of PCBs
in a segment of the mainstream and of a pool of the Upper
Mississippi River. The objectives of this study were to model
and understand the hydrological and biological distribution of
PCBs, and to aid in assessing the environmental dynamics of PCBs
in the Upper Mississippi River. Ecologically, Lake Onalaska, the
river pool being studied, provides a major staging and feeding
area for migration of the canvasback (Aythya valisineria) in the
Mississippi River Flyway. About 1200 samples were analyzed by
capillary gas chromatography for 69 PCB isomers. Sample types
included water, suspended sediments, sediments, fish, benthos,
and plants.

The amount and complexity of data resulting from these analyses prompted us to search for an improved method for characterizing and comparing information gathered from multi-component analyses of large numbers of samples. Multivariate statistics were applied in the process of characterization of large numbers of complex residues. Such methods have been referred to as Chemometrics (24).

Soft Independent Method of Class Analogy (SIMCA), a pattern recognition technique based on principal components (25) was selected to evaluate and apply to the problems of establishing similarities among sample residue profiles. The development of a laboratory data management system to assist in the calculation and organization of results greatly enhanced the feasibility of this approach (26).

Materials and Methods

Gas Chromatographic Analysis. We used temperature programmed glass capillary gas chromatography to separate PCB residues. Use of an electron capture detector required an efficaceous sample cleanup for isomer quantitation (27). These combined techniques offered enhanced separations and enabled us to identify and quantitate individual PCB constituents (1, 27). Schwartz (27) separated more than 100 constituents from a 1:1:1:1 mixture of Aroclors 1242, 1248, 1254, and 1260.

A Varian Model 3700 gas chromatograph (Varian Associates, Palo Alto, CA) equipped with a [63]Ni electron capture detector was used for sample analysis. The column was a glass capillary coated with a hydrocarbon stationary phase having the formula $C_{87}H_{176}$ and structure (24,24-diethyl-19,29-dioctadecylheptatetra-contane) (Quadrex Corp., New Haven, CT). The column was a wall coated open tube (WCOT) (0.25 mm x 100 m) having a film thickness of approximately 0.3 microns and 3670 theoretical plates per meter. A 30 m section of the column was conditioned at 260 °C overnight and the temperature program rate, linear gas velocity, and detector makeup gas were determined by peak resolution of PCB isomers.

Chromatographic conditions were similar for all sample analyses. The temperature program was as follows: initial temperature 120 °C programmed at 2 °C/min to a final temperature of 255 °C; injector 220 °C; and detector temperature 320 °C. The carrier gas, hydrogen, was maintained at a column head pressure of 20 psi. The electron capture detector makeup gas was nitrogen with a flow rate into the detector of 17 mL/min.

The same chromatographic parameters were used in determining the molar response of the individual PCB isomers except, that the area responses were determined with a flame ionization detector. The flow rates of the hydrogen and air combustion gases were 30 and 300 mL/min, respectively. Response factors needed to

calculate individual congener concentrations were calculated by
using the ratios of flame ionization to electron capture response
for each standard peak (28, 29).

To integrate the digitized electrical signals from the gas
chromatograph, we used a PDP 11/34 computer (Digital Equipment
Corp., Maynard, MA) and the computer program PEAK-11, supplied
from that corporation. An operational amplifier with variable
gain was interfaced to the the gas chromatographic amplifier
output to amplify the signal from the electron capture detector.
Data were acquired at the rate of 20 points per sec and the
integration and gas chromatographic operations were controlled by
a Varian Autosampler (Model 8000), which also delivered a
calibrated amount of sample to the gas chromatographic injection
port. Extensive quality control and verification of analyses
resulted in reproducible results (26).

Laboratory Data Base System. Data representing integrated peak
areas were transfered to a second PDP-11/34 computer by magnetic
tape, where programs written in Digital Standard MUMPs (DSM-11)
created and maintained the chromatography data base. MUMPs was
selected for development of this data base because this language
is structured to efficiently deal with sparse arrays. The
computer was configured with 96k words of memory, 20M words of
disk storage, and a 9 track magnetic tape.

The data base programs accomplished the following tasks: (1)
generation of retention values for each PCB isomer peak (relative
to p,p'-DDE) in every chromatogram; (2) comparison of an Aroclor
standard, or any mixture of Aroclors, to a master file of isomers
containing structural identification, number of ortho, ortho'-
chlorine substituents, and flame ionization response factors; (3)
generation of linear and logarithmic (to accomodate non-linear
detector response) regression equations for each of the
constituents (69, 92, or 105 isomers) on the basis of Aroclor
concentration and area of each component; (4) assistance in
resolving ambiguities in the process of matching standard peaks
to sample constituents; (5) generation of numerous reports on
sample processing, including status of sample data, i.e., samples
logged in, raw chromatograpic data matched to standards,
concentration of constituents calculated from regression curve,
data archived to magnetic tape, and (6) preparation and magnetic
tape storage of final reports on concentrations of up to 105
isomers in samples, with notations on whether the measured
concentration fell within, above, or below the calibration range
for that constituent. A flow chart for the data analysis scheme
is outlined in Figure 1.

In addition to these functions, other data base programs
provided output formating and retrieval of concentration data
from completed analytical reports and transfer of these data onto
magnetic tape for subsequent examination by SIMCA programs.
Additional features of the pattern recognition data management
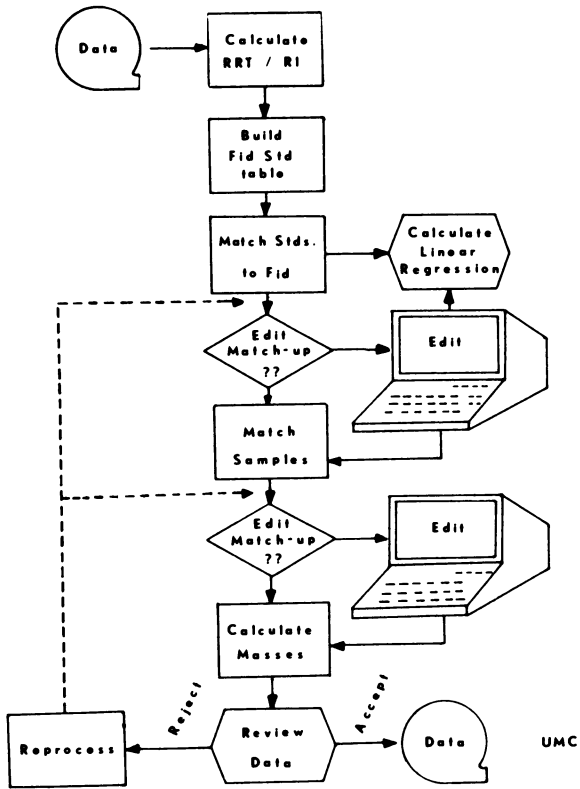
Figure 1. Flow Diagram of PCB Data Base.

subprograms provided peak frequency and ranking summaries within
each selected sample set. Programs were written for graphic
plots of concentrations of PCB isomer components and related
statistics.

A summary of these functions follows:
>   1. Build the matrix of samples to output
>   2. Output the summary matrix of sample identity
>   3. Edit the summary matrix
>   4. Build the set of sample concentrations
>   5. Output the set of sample concentrations
>   6. Build the non-zero concentration matrices
>   7. Output the non-zero concentration matrices
>   8. Plot concentration histogram of a sample
>   9. Plot means and standard deviations
>   10. Generate principal components plots of data blocks

Results and Discussion

When large numbers of chromatograms are analyzed visually from
chart paper or tabular data, it is difficult to detect minor (or
perhaps major) significant differences that may be present in
samples from different locations or differences that could be
attributed to changing physiochemical processes. A typical gas
chromatogram of a PCB mixture (Figure 2) represents a 1:1:1:1
mixture of Aroclors 1242, 1248, 1254, and 1260. It contains more
than 100 component peaks, of which 69 were selected for use in
these analyses.

An abreviated report from the analysis of the mixture of
Aroclor standards just described is shown in Table I. The report
indicates the individual isomer structure and degree of
chlorination, ortho-, ortho-' substitution, isomer concentration,
total concentration.

To illustrate the problems associated with evaluating such
data, we conducted several studies with Aroclor standards and
mixtures of these standards in an effort to determine what
information could be readily obtained with the SIMCA method of
pattern recognition (30-32). The following discussion
illustrates some of the features of this approach and describes
how the SIMCA method works when applied to Aroclor mixtures.

General Discussion of SIMCA.

Consider a series of gas
chromatograms obtained on a large number of samples (N). These
samples can all contain the same number of constituents (P)--in
the present study, 69 PCB isomers. Such data can be tabulated in
matrix form as in Figure 3. Each row of the matrix is a
chromatogram with the elements of the matrix, x_{ki}, representing
concentration of peak i in sample k.

We first consider the hypothetical case when there is no
variation in composition of the samples, either in relation to
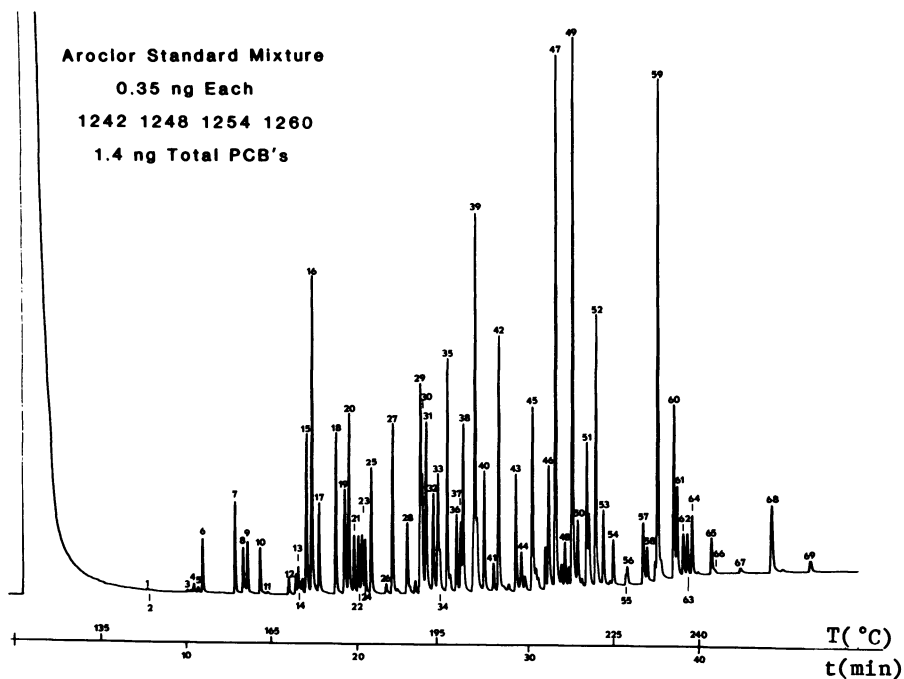
Figure 2. Capillary Gas Chromatogram of 1:1:1:1 Aroclor 1242:1248:1254:1260 Mixture.

the relative concentrations of constituents or in the absolute
concentration of isomers.   If our $\underline{N}$ samples are projected into
$\underline{P}$ space, due to our prior requirement of identity, they would
appear as a tight cluster, the only variation being due to
measurement error.

Table I. Isomer Structure Assignment, Response Factors, and
Concentration in 1:1:1:1 Mixture of Aroclor 1242;1248;1254
and 1260.

Peak	RRT[1]	Chlorines Total	Ortho	Structure	F_i[2]	Conc.$_i$ ug/g
1	.2681	2	2	2,2'	.0098	2.53E-02
2	.2706	2	2	2,6	3.00E-5	9.00E-05
3	.3599	2	1	2,5	.0012	3.31E-03
4	.3695	2	1	2,4	.0046	1.46E-02
5	.3776	3	–	3CL	.0029	9.18E-03
6	.3903	2	1	2,4'	.0206	6.86E-02
7	.4615	3	2	2,2',5	.0468	1.53E-01
8	.4810	3	2	2,2',4	.0150	4.91E-02
9	.4873	3	2	2,2',3	.0152	5.05E-02
10	.5173	3	2	2,4',6	.0096	3.10E-02
11	.5352	2	0	4,4'	.0039	1.03E-02
12	.5803	4	3	2,2',5,6'	.0062	2.01E-02
13	.6031	3	1	2,3',5	.0076	1.93E-02
14	.6107	4	3	2,2',4,6'	.0040	1.39E-02
15	.6252	3	1	2,4',5	.0377	1.21E-01
16	.6390	3	1	2,4,4'	.0313	8.57E-02
17	.6525	3	1	2,3',4	.0106	3.42E-02
18	.6933	4	2	2,2',5,5'	.0623	2.00E-01
19	.7145	4	2	2,2',4,5'	.0226	7.33E-02
20	.7201	4	2	2,2',3,5'	.0574	1.83E-01
21	.7387	4	2	2,2',4,4'	.0080	2.72E-02
22	.7441	4	2	2,2',3,4'	.0085	2.75E-02
23	.7502	4	2	2,2',3,3'	.0125	4.12E-02
24	.7596	4	2	2,2',3,4	.0095	3.08E-02
25	.7758	4	2	2,3,4',6	.0159	5.10E-02
26	.8069	3	0	3,4,4'	.0034	1.16E-02
27	.8237	5	3	2,2',3,5,6	.0429	1.37E-01
28	.8547	4	3	2,2',4,5,6'	.0107	3,48E-02
29	.8897	4	1	2,3',4',5	.0351	1.10E-01
30	.8990	5	3	2,2',3,4,6'	.0100	3.14E-02
31	.9043	5	–	5CL	.0252	7.80E-02
32	.9174	4	1	2,3,3',4'	.0112	3.66E-02
33	.9320	4	1	2,4,4',5	.0236	8.01E-02
34	.9353	6	4	2,2',3,3',6,6'	.0002	6.28E-04
35	.9596	5	2	2,2',4,5,5'	.0560	1.77E-01

Table I. Continued

Peak	RRT[1]	Chlorines		Structure	F_i[2]	Conc.$_i$ ug/g
		Total	Ortho			
36	.9832	5	2	2,2',4,4',5	.0176	5.78E-02
38	.9913	5	2	2,2',3,3',5	.0122	3.98E-02
39	1.0205	5	2	2,2',3,3',4	.0571	1.83E-01
40	1.0499	6	3	2,2',3,5,5',6	.0192	6.10E-02
41	1.0726	6	3	2,2',3,4',5,6	.0021	6.59E-03
42	1.0815	6	3	2,2',3,4',5,6'	.0485	1.54E-01
43	1.1181	6	3	2,2',3,4,5',6	.0144	4.62E-02
44	1.1385	5	1	2,3',4,5,5'	.0044	1.40E-02
45	1.1613	5	1	2,3,3',4,4'	.0274	8.50E-02
46	1.1938	6	–	6CL	.0062	2.00E-02
47	1.2218	6	2	2,2',4,4',5,5'	.0425	1.45E-01
48	1.2382	6	2	2,2',3,3',5,5'	.0004	1.84E-03
49	1.2561	6	2	2,2',3,4,4',5	.0273	8.79E-02
50	1.2730	7	3	2,2',3,4',5,5',6	.0033	1.01E-02
51	1.2998	7	3	2,2',3,3',4,5,6	.0145	4.52E-02
52	1.3146	7	3	2,2',3,3',4,5,6'	.0080	2.48E-02
53	1.3228	7	3	2,2',3,4,4',5',6	.0070	2.18E-02
54	1.3358	7	3	2,2',3,3',4,5',6	.0024	7.49E-03
55	1.3604	7	3	2,2',3,3',4,4',6	.0048	1.51E-02
56	1.3936	8	3	2,2',3,4,4',5,5',6	7.0E-5	2.24E-04
57	1.4280	6	1	2,3,3',4,4',5	.0028	8.73E-03
58	1.4444	7	2	2,3,3',4',5,5',6	8.0E-5	2.51E-04
59	1.4693	7	2	2,2',3,4,4',5,5'	.0250	9.45E-02
60	1.5032	7	2	2,2',3,3',4,4',5	.0096	2.99E-02
61	1.5208	8	3	2,2',3,3',4,5,5',6	.0025	7.84E-03
62	1.5295	8	3	2,2',3,3',4',5,5',6	.0013	4.13E-03
63	1.5454	8	–	8CL	.0020	6.28E-03
64	1.5575	8	–	8CL	.0025	7.73E-03
65	1.5948	8	3	2,2',3,3',4,4',5,6	.0016	5.05E-03
66	1.6119	9	4	2,2',3,3',4,4',5,6,6'	5.0E-5	1.25E-04
67	1.6502	8	–	8CL	1.0E-5	3.13E-05
68	1.7155	8	2	2,2',3,3',4,4',5,5'	.0011	3.58E-03
69	1.7380	9	4	2,2',3,3',4,4',5,5',6	2.0E-5	6.71E-05
					Total	3.24E+00

[1] Retention Time Relative to p-,p'-DDE
[2] Ratio of Electron Capture to Flame Ionization Response

Peak Number

Sample number	1	2	3	.	.	i	.	$\underline{P}$
1						.		
2						.		
3						.		
.						.		
$\underline{k}$	.	.	.	.	.	$\underline{x}_{ki}$		
.								
$\underline{N}$								

Figure 3. Matrix Representation of Sample Analysis for $\underline{P}$ Peaks and $\underline{N}$ Samples (Chromatography Data Matrix).

Each sample can be modeled by Equation 1, where $\underline{m}_i$

$$\underline{x}_{ki} = \underline{m}_i + \underline{e}_{ki} \qquad (1)$$

is the mean concenration of each peak in the matrix and $\underline{e}_{ki}$ contains the error of measurement and error associated with the mathematical modeling of the data array (model error).
We now relax our requirement for identity and allow the samples to be _similar._ This is analogous to having a group of $\underline{N}$ samples of the same species, type of Aroclor, or mixture of Aroclors. Such data can be shown to be modeled by Equation 2,

$$\underline{x}_{ki} = \underline{m}_i + \sum_{a=1}^{A} \underline{t}_{ka}\, b_{ai} + \underline{e}_{ki} \qquad (2)$$

in which A-product terms have been added to account for the variation in the data (30). This is shown in 3-dimensions below in Figure 4.
This is a principal components model in which $\underline{b}_{ai}$ is the loading of peak $\underline{i}$ in term $\underline{a}$, and $\underline{t}_{ka}$ is the score of object $\underline{k}$ in term $\underline{a}$; $\underline{b}$ is a peak specific term and $\underline{t}$ is an object or sample specific term. The variation about the mean, $\underline{m}_i$, can be random or systematic. If random variation is observed it can be due to measurement error, and this variation can be used in quality

assurance of the data. If the measured variation is systematic, it can be the result of class specific internal variation, which can be used in classification or correlation studies.

Conc. Peak 3

o
ooo $\underline{N}$-points as a cluster
o

Conc. Peak 2

Conc. Peak 1
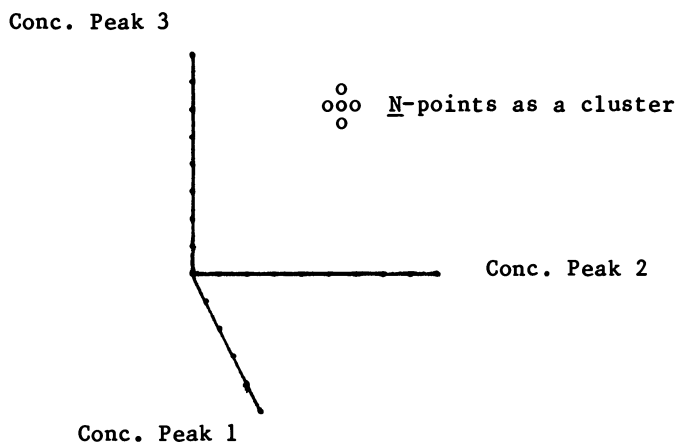
Figure 4.   Clustering of a Class of Identical Samples.

    Figure 5 illustrates the cases in which the data are represented by a point ($\underline{A}$=0), line ($\underline{A}$=1) or plane ($\underline{A}$=2). $\underline{A}$ is the number of product terms in Equation 2. Samples clustered in a point represent replicate analyses of a single sample in which there is no variation other than measurement error, and the product term in Equation (2) is 0. In these last two situations, the data vary about the mean, $\underline{m}_i$, and the position of each object on the line or plane given by the peak coordinates. An example of data that would form a line are those based on an analysis of a range of concentrations of a single Aroclor ($\underline{A}$=1). Data that could be represented in a plane result from the analysis of the fractional composition of two (or more) Aroclor mixtures ($\underline{A}$=2). In Fiqure 5, $\underline{q}$ designates the class number of the these hypothetical samples.
    One can use principal components plots to visually inspect higher dimensional data. Their use is equivalent to projecting the higher dimensional data onto a two-dimensional plane. Such plots are helpful in interpreting chromatographic or other scientific data composed of many measurements (peaks or dimensions).
    If it is known from the history of the samples illustrated in Figure 3 that the items represent $\underline{q}$-distinctly different groups or classes, a classification problem can be formulated. These classes may result from samples being derived from different locations or of different species, etc. In this study,

sample data from the analysis of five Aroclors were used (Aroclor 1242, 1248, 1254, 1260 and a 1:1:1:1 mixture of these four Aroclors). Together, these sample data compose five classes.

Using SIMCA, one can derive q-disjoint principal components models for distinct classes. This derivation is given in Equation 3 and a hypothetical set of data representing three classes (q=3) of data of increasing complexity is shown in Figure 5. The axes in Figure 5 may represent either the coordinates of the original data or those resulting from a transformation of the data such as will be discussed later in the section on data pretreatment. In the examples used in this paper, data have not been transformed. Disjoint principal components models are bounded by limits derived from the data used to calibrate the class model.

$$x_{ki}{}^q = m_i{}^q + \sum_{a=1}^{A} t_{ka}{}^q \, b_{ai}{}^q + e_{ki}{}^q \qquad (3)$$

From the $e_{ki}{}^q$-values in Equation 3, a standard deviation for each object can be calculated and from these a residual standard deviation for each class (q) can be obtained. Twice the residual standard deviation around the line or above and below the plane for a class defines a volume in P-space where classes have a high probability of occurrence. Classification of an unknown can be based on its projection into P-space and determination of its position. It may be inside one or none of the defined classes. This method has a number of other features that have previously been treated (32, 33) and we present only a limited discussion here. Principal components analysis and the closely related factor analysis, as applied to scientific data, have been well reviewed (25, 30).

Another feature of SIMCA that is of considerable utility lies in the assistance the technique provides in selecting relevant variables. Information contained in the residuals, e_{ki}, can be used to select variables relevant to the classification objective. If the residuals for a variable are not well predicted by the model, the standard deviation is large. An expression defined as modeling power has been defined to quantitatively express this relationship. The modeling power (MPOW) is defined as:

$$MPOW = 1 - s_i / s_{i,y}$$

where s_i is the standard deviation of a variable after being fitted to a model and $s_{i,y}$ is the standard deviation before it is fitted to a class model. As MPOW approaches 1, the variable contributes strongly to class description.

Principal Components Plots. To obtain a visual representation of the data structure for the class or classes being examined, one can project the data onto a two-dimensional plane. A convenient plane for mapping is that described by the two first principal components. These plots are derived by plotting the t_{1i} and t_{2i} values for each sample and designating the location of the sample with the sample number. The axes are designated in these principal components plots as Theta 1 and Theta 2. Information about the similarity of samples is derived in most cases, by their close proximity in the principal components plot. Usually, separation of different classes are clear. By examining the clusters of samples in the graphical projections, the analyst gain insight into sample outliers from the principal components plots.
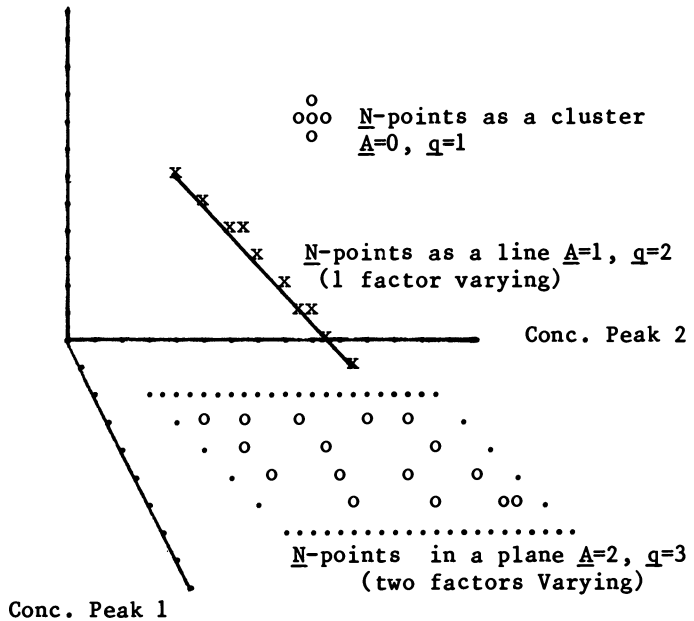
Figure 5. Clustering of Three Classes of Samples. Samples that are identical ($A=0$, $q=1$); or have one factor ($A=1$, $q=2$); or two factors varying ($A=2$, $q=3$). Class number = q and A = number or product terms in Equation 2.

Information about relation among variables in samples belonging to a data set can be obtained from an examination of plots of the sample loading terms (b_{1i} vs b_{2i}). In these plots the axes are designated Beta 1 and Beta 2. Variables responsible for certain samples being different from the main cluster of

samples are readily discerned. Three dimensional plots derived
from the first, second, and third components can sometimes
clarify additional relations among samples. Such a case is
discussed later in discussions of modeling differences both in
concentration and composition.

The similarity of samples can be evaluated by using
geometrical constructs based on the standard deviation of the
objects modeled by SIMCA. By enclosing classes in volume
elements in descriptor space, the SIMCA method provides
information about the existence of similarities among the members
of the defined classes. Relations among samples, when visualized
in this way, increase one's ability to formulate questions or
hypotheses about the data being examined. The selection of
variables on the basis of MPOW also provides clues as to how
samples within a class are similar, and the derived class model
describes how the objects are similar, with regard to the
internal variation of these variables.

One must consider the number of product terms that should be
included in a model. For chromatography data obtained from
similar samples, it can be expected that the data will contain a
high degree of correlation. In our experiments, two- or three-
component models usually accounted for >90% of the variance in
the data for a class of similar samples. Results from
crossvalidation should be considered as the primary criteria in
selecting the number of principal components to be extracted from
a given data set (34).

In the principal components plots presented in this paper,
the number plotted corresponds to the sample identification
number given in the appendix. If more than one sample has the
same locus in the score (Theta's) or loading plots (Beta's), the
letter M is plotted. The values for the sample coordinates in
the principal components plots can be listed by the SIMCA-3B
program.

Data Pretreatment

A number of data pretreatments (scalings and transformations) are
possible with SIMCA (and other methods). The appropriate scaling
for a given data set is difficult to anticipate and depends on
the nature of the problem being addressed as well as on the
structure of the data being examined. This aspect of pattern
recognition has not been been adequately investigated for gas
chromatography data. In the applications discussed here, we have
used two techniques--normalization and regularization.

Normalization. Each peak is weighted to its fractional (or
percentage) composition. This weighting prevents samples with
large total concentrations from heavily influencing an analysis.
Also, this scaling technique focuses the classification on
questions about variations in relative composition of samples
independently of the absolute concentration measured. Although

instrumental limits of detection can strongly influence this data pretreatment, this feature, when used with caution, can be effective in characterization of the range of total concentrations that can be analyzed.

Regularization. Regularization, the autoscaling of Kowalski, (35) and scaling of Massart, (36) transforms the data so that the data set has a zero mean and a variance of one for each variable. This method equalizes the influence of peaks or measurements.
   A risk with this approach is that if there are a large number of constant non-zero entries in the data matrix, they can act as binary variables and perhaps weight the analysis toward yielding trivial results. When cross validation is used, however, this risk is reduced.

Calculation of Composition of Mixed Aroclors

SIMCA can be applied to the problem of classification when attempting to correlate measurable effect variables with composition of the classified samples. In correlation analyses one may wish to determine how other sample variables, such as sediment composition, organic content, lipid concentration, etc., influence the composition of measured residues or concentrations of PCBs.
   This determination cannot be made by using multiple regression techniques because there are too many independent variables, x_{kj}, relative to dependent variables, y_{kj}. The SIMCA-3B program, PLS-2, uses the partial least squares (PLS) method. This method has been proposed by H. Wold (37) and was discussed by S. Wold (25). In such a problem there are two blocks of data, Y and X. It is assumed that Y is related to X by latent variables u and t; t is derived from the X block and u is derived from the Y block.
   These relations are mathematically stated in Equation 4 and 5.

$$Y \longleftarrow u \longleftarrow t \longleftarrow X \tag{4}$$

With the PLS method the X-block is modeled by the equation

$$x_{ki} = m_i + \sum_{a=1}^{A} t_{ka} \; b_{ka} + e_{ki} \tag{5}$$

and the Y-block by the equation

$$y_{kj} = m_j + \sum_{a=1}^{A} u_{ka} \; c_{aj} + f_{kj} \tag{6}$$

Both resemble principal components models, but are derived so as to <u>simultaneously</u> minimize e_{ki} and f_{kj} in the least squares sense <u>while</u> yielding $\underline{t}_{ka}$'s and $\underline{u}_{ka}$'s that optimize the correlation between the x's and y's. The $\underline{u}_{ka}$'s and $\underline{t}_{ka}$'s are analogous to the components in principal components analysis whereas the <u>b</u> and <u>c</u> terms correspond to loading values. The loadings reflect the importance or influence of the individual peaks in the PLS model.

A predictive relation between the latent variables is constructed as:

$$\underline{U}_{ka} \quad = \quad \underline{d}_a \; \underline{t}_{ka} \; + \; \underline{e}_k \qquad\qquad (7)$$

and can be used to estimate $\underline{u}$ (and the y's) from the X-block for PCB compositions similar to those modeled.

Applications of SIMCA to Chromatography Data

In the discussion that follows, the SIMCA method is illustrated by applying it to three problems: (1) quality assurance of chromatography data, (2) classification of unknowns, and (3) predicting the composition of unknown samples. This third problem is one of deconvolution of a mixture and calculation of the relative concentration of the constituents (<u>25</u>, <u>38</u>).

The data modeled are from gas chromatograms obtained for Aroclors 1242, 1248, 1254 and 1260. The unknown samples are from the anaysis of used transformer oil obtained from a waste dump in New Jersey. The concentration of individual isomers in selected Aroclor and transformer oil samples are given in Appendix I. The data are organized in a matrix in which the first four data entries for each sample in row 1 of the data array (Table 2, Apendix I) designate the composition of the sample. For standards, these four variables represent the fractional parts of Aroclor 1242, 1248, 1254, or 1260, respectively, that were combined. Results from the analysis of transformer oil (samples 21-23) are of unknown fractional composition and variables 1 through 4 are null entries. In the examples that follow data from samples analyzed (Table 1, Appendix I) were used in part or in total to illustrate the PLS method.

As in many such problems, some form of pretreatment of the data is warranted. In all applications discussed here, the analytical data either have been untreated or have been normalized to relative concentration of each peak in the sample. <u>Quality Assurance.</u> Principal components analysis can be used to detect large sample differences that may be due to instrument error, noise, etc. This is illustrated by using samples 17-20 in Appendix I (Figure 6). These samples are replicate assays of a 1:1:1:1 mixture of the standard Aroclors. Fitting these data for the four samples to a 2-component model and plotting the two first principal components (Theta 1 and Theta 2 [scores] in

Equation 3 for the samples shows that samples 19 and 20 are different from samples 17 and 18 (Figure 7).

The plot of the loadings for each chromatographic peak (Beta 1 vs. Beta 2; b_{1i} and b_{2i} Equation [3]) reveals information about the sources of the variance in the four samples (Figure 8). Information in Table II confirms these findings, as it is seen that the variability is largely the result of the failure to detect two peaks (peak 1 [variable 5], and peak 4 [variable 9]). Also, peak 63 (variable 67) in sample 20 exceeds its average concentration measured in all samples by 9%.

The data show that peak 1 is zero in sample #19 and #20, that peak 5 is zero in sample 19, and that there is about 12% variation in peak 63 in samples 19 and 20 (Table II). Only after a careful examination of the data plotted for each analysis are these deviations apparent in the chromatograms. These results illustrate the utility of principal components analysis (PCA) for checking the internal consistency of complex chromatograms.

Table II. Partial Peak Summary of Replicate Analysis of an Aroclor 1242:1248:1254:1260 Mixture

Sample #	Concentration (ng) of Peak Number					ng Total
	1 (5)[1]	4 (8)	5 (9)	6 (10)	63 (67)	
17	0.054	0.020	0.014	0.106	.0933	4.38
18	0.058	0.020	0.014	0.101	.0956	4.19
19	0[2]	0.016	0[2]	0.104	.0929	4.22
20	0[2]	0.020	0.020	0.101	.1030	4.10

[1](SIMCA Variable Number)
[2]Not Detected by PEAK-11 Integration Program

To illustrate the influence of changes in sample composition resulting from failure to detect constituents because of decreasing concentration or sensitivity, we offer a hypothetical set of data in histogram form (Figure 9). This histogram illustrates how the results of analyzing of a mixture of five compounds would change as a result of varying the sensitivity of the detection system or from sample dilution at a fixed sensivity.

The changes in fractional composition resulting from analyzing the sample under different sensitivities in this
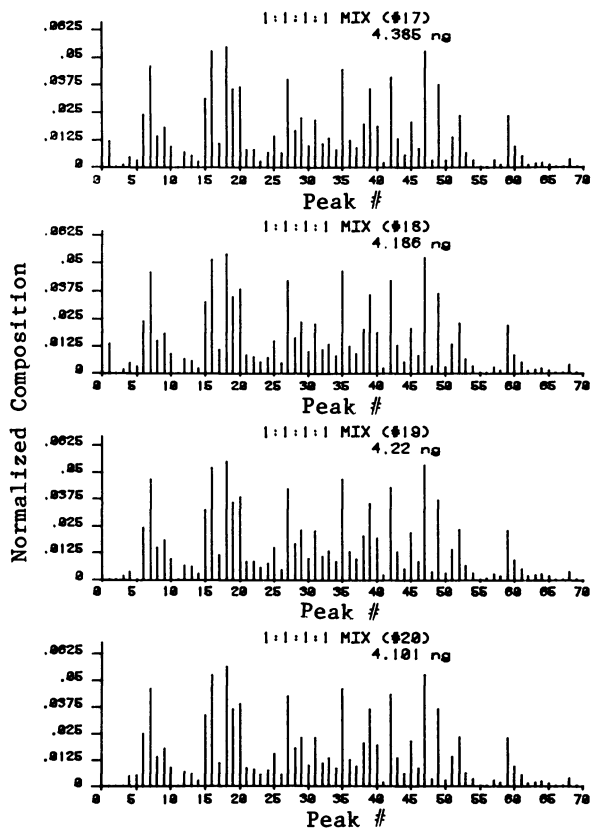
Figure 6.   Fractional Composition Histograms from
Four Replicate Analyses of Aroclor Mixture.

Sample # in TRS 3/6/82

Theta-2

18
17

20

19

Theta-1
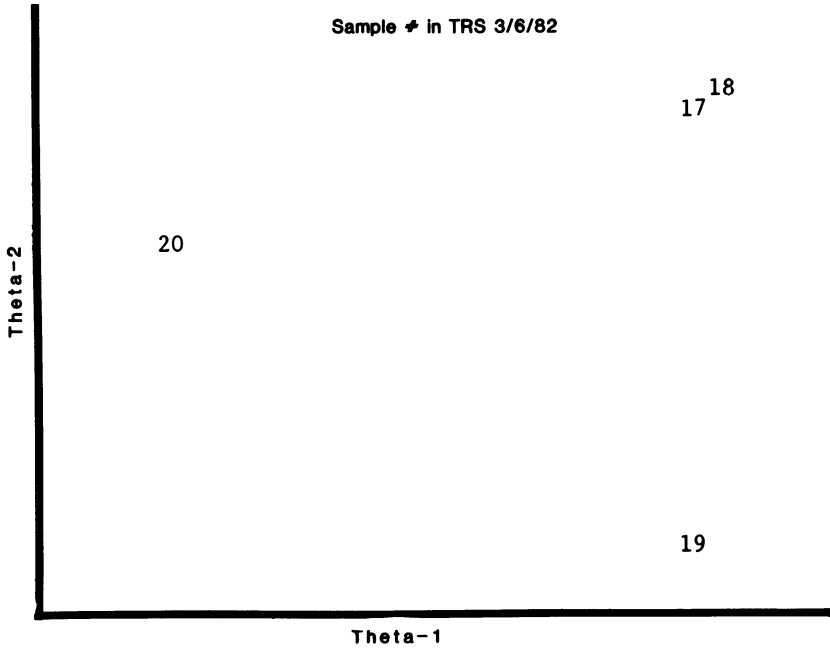
Figure 7. Principal Components Plot Derived from Fractional Composition Replicate Aroclor Analysis (Figure 2).
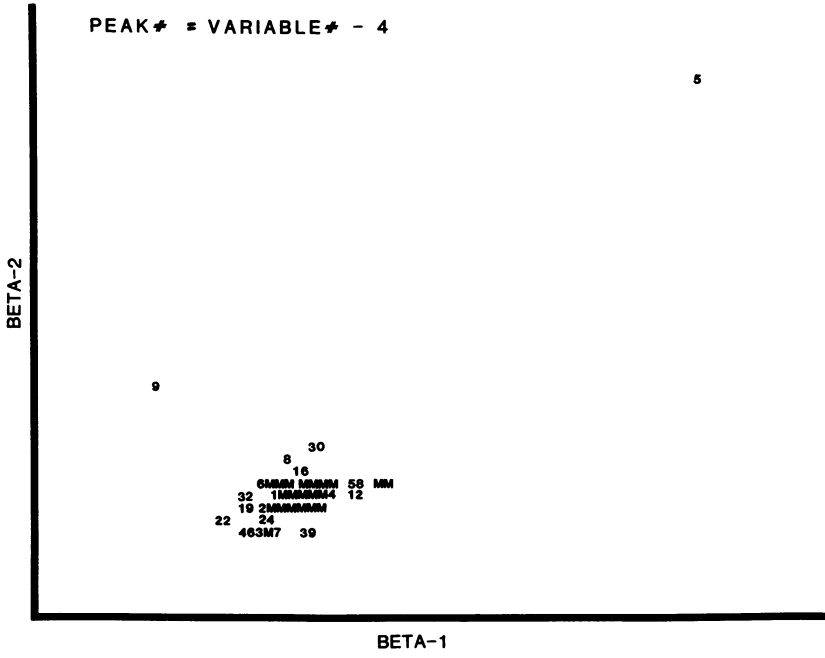
Figure 8.   Plot of Variable Loadings (Betas) in
Aroclor Mixtures.

example are summarized in Table III. The hypothetical total of
the constituents is 30 arbitary units and correct results are
obtained at sensitivity level I. Results determined at each of
the lower sensitities are presented in brackets. Sensitivity
decreases from level I to IV and the dashed lines represent the
baseline at the different sensitivity levels. These results are
commonly encountered when concentration data are expressed as
proportions and this example serves to point out the importance
of controlling this problem when the examining fractional
composition data. It is shown later that SIMCA is able to
accomodate for this type of change, if the change is a consistent
function of concentration.

```
                                   *
                                   *
     SENSITIVITY              *    *
       LEVEL     IV      --*----*------------------- [ 6]
                              *    *    *
                 III     --*----*----*-------------- [12]
                              *    *    *    *
                 II      --*----*----*----*--------- [20]
                              *    *    *    *    *
                 I         *    *    *    *    *    [30]
           COMPONENT #     1    2    3    4    5
       COMPONENT CONC.     8   10    6    4    2
```
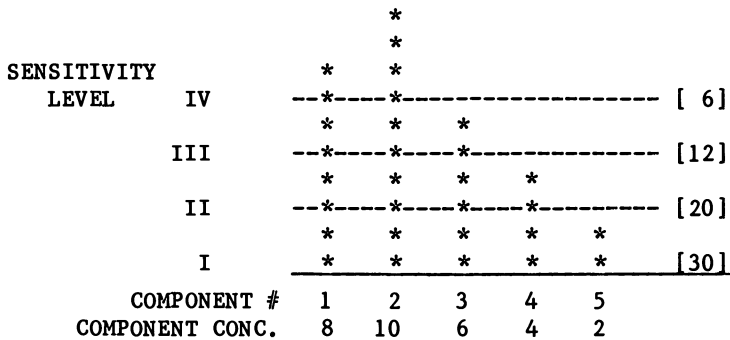
Figure 9. Example of changes in composition resulting from
decreased sensitivity.

Table III. Changes in fraction composition of a hypothetical
mixture due to decreasing concentration or decreasing
sensitivity.

SENSITIVITY LEVEL	FRACTIONAL COMPOSITION OF COMPONENT #					TOTAL
	1	2	3	4	5	
I	.266	.333	.24	.133	.067	30
II	.333	.400	.20	.10	ND[1]	20
III	.333	.50	.166	ND	ND	12
IV	.333	.667	ND	ND	ND	6

[1]ND=Assumed to be not detected due to limited sensitivity

To further illustrate the utility of the SIMCA method we

examined an even more complex selection of Aroclor samples in
which both composition and concentration of samples were modeled.
Data for this illustration were from a group of 27 samples of
four Aroclor types, where three Aroclors were analyzed over a
range of concentrations (Table IV).

   In this set of samples, the concentration data for the
Aroclor 1254 concentration series were from another set of data
in which there were 105 constituents instead of the usual 69. In
this study, the concentration of the first 69 peaks of the total
of 105 were selected. A sample of Aroclor 1254 having the same
69 constituents as Aroclor 1248 and 1260 was included to test the
discrimination ability when concentration and composition were
being modeled.
   The data were modeled by a principal components model with
three components. The statistical results method ($\underline{25}$, $\underline{31}$) are
presented in Table IV and V. In addition, the measured total PCB
concentration is included in Table IV. One of the three sets of
two-dimension plots (Theta 1 vs Theta 2) is presented in Figure
10. Individual samples of a given Aroclor were distributed
regularly in these plots and samples were ordered according to
concentration. The sums of squares decreased from 4,360 to 52.4
(Table V.) and approximately 88 percent of the standard deviation
was explained by the three term component model.
   Because of this structure in the plots, three dimensional
representation of the data was investigated. The three principal
components score terms ($\underline{t}_{k1}$, $\underline{t}_{k2}$, and $\underline{t}_{k3}$) from the statistical
analysis of this data set were used as plotting coordinates for
each sample (Table IV). The data were displayed and rotated
about the axes in 3-D using a Dazzler TV graphics board
(Cromemco, Inc., Mountain View, CA). Following this display of
the data, the coordinates for each sample were used to generate a
3-D plot with a Texas Instruments Plotter driven by a MUMPS
program in the laboratory data base.
   Two graphics viewpoints were selected that allowed us to
discern the three clustering of Aroclors as lines (Figure 11).
The single Aroclor 1254 sample (point "3") that was composed of
data from an alternate set of data is readily observed in these
3-D plots as not being similar to any of the other sample types.

Classification To illustrate the use of SIMCA in classification
problems, we applied the method to the data for 23 samples of
Aroclors and their mixtures (samples 1-23 in Appendix I).   In
this example, the Aroclor content of the three samples of
transformer oil was unknown. Samples 1-4, 5-8, 9-12 and 13-16,
were Aroclors 1242, 1248, 1254, and 1260, respectively. Samples
17-20 were 1:1:1:1 mixtures of the Aroclors. Application of
SIMCA to these data generated a principal components score plot
(Figure 12) that shows the transformer oil is similar, but not
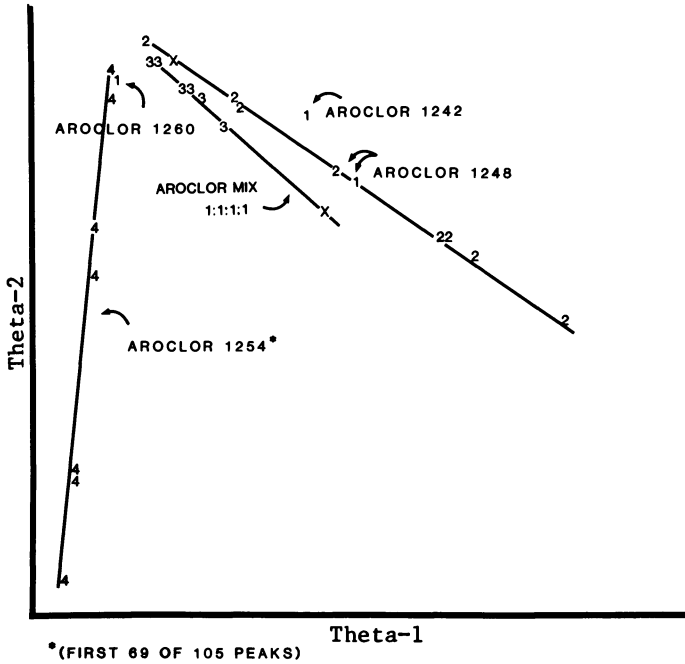
Figure 10. Principal Components Plot (Theta 1 vs. Theta 2) from Aroclor Classes (Table IX).

* Aroclor 1248 Series
@ 1:1:1:1 Aroclor Mix
# Aroclor 1254 Series (69/105)
2 1248 Test
3 1254 (69) Test
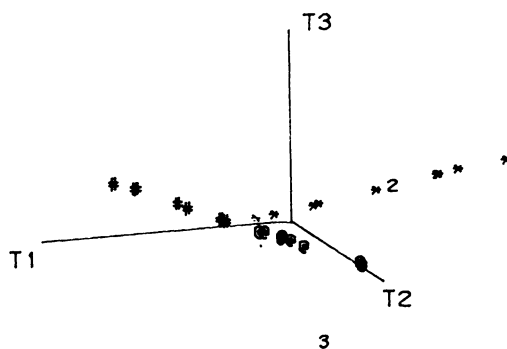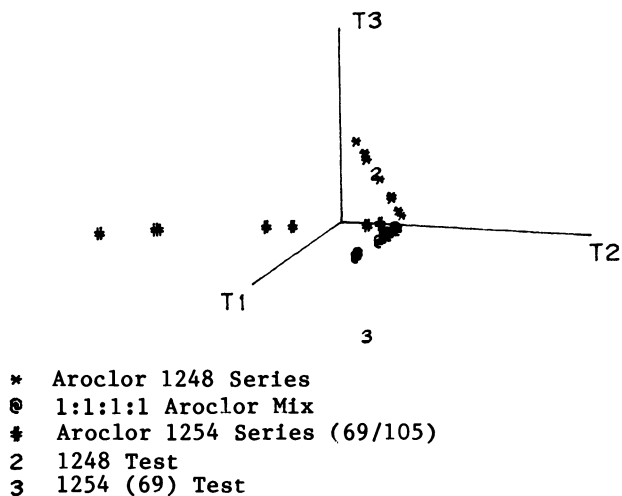
Figure 11. 3-D Views of Theta Values Derived from
PC Analysis of Aroclors. Key: * = Aroclor 1248;
@ = 1:1:1:1 Mixture; # = Aroclor 1254 (69/105 pks);
2 = Aroclor 1248 (test); and 3 = Aroclor 1254.
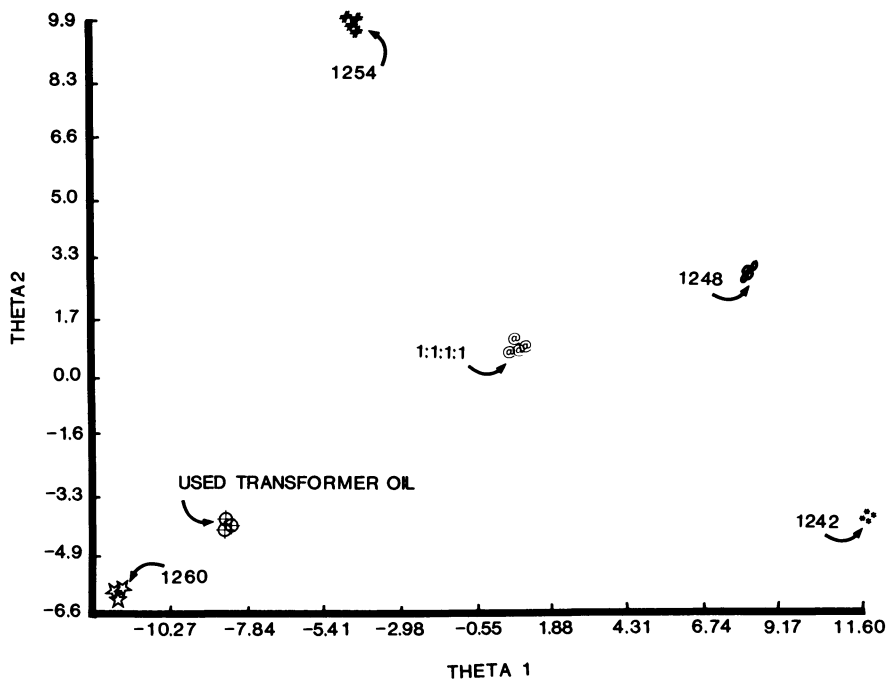
Figure 12. Principal Components Plot from Five Aroclors Classes and a Used Transformer Fluid (most similar to Aroclor 1260).

identical, to Aroclor 1260. A more detailed discussion of classification using these data is presented by Dunn et al. (39).

Table IV. Theta Values from a Three Component Principal Components Class Model (A=3) and Total Measured Total Concentration of PCBs.

Aroclor and SAMPLE #	THETA 1	THETA 2	THETA 3	CONC.
1248				
1	-0.813	-0.003	0.227	6
1254				
2	-0.180	0.11	-1.39	6
1248				
3	0.524	0.621	0.21	0.76
4	0.351	0.541	0.216	1.58
5	0.352	0.541	0.211	1.61
6	-0.0173	0.369	0.211	3.46
7	-0.0641	0.352	0.225	5.63
8	-0.645	0.077	0.22	6.56
10	-1.32	-0.246	0.231	9.85
11	-1.29	-0.236	0.231	9.74
9	-1.52	-0.334	0.245	10.9
12	-2.07	-0.592	0.240	13.7
1:1:1:1 mixture				
19	0.486	0.570	0.048	1.48
14	0.442	0.546	0.038	1.76
13	0.255	0.418	-0.075	3.22
18	0.264	0.430	-0.089	3.24
17	0.167	0.370	-0.14	3.89
15	0.0237	0.273	-0.253	5.11
20	-0.587	-0.099	-0.629	9.7
16	-0.603	-0.122	-0.67	9.99
1254[1]				
24	0.721	0.513	0.173	0.80
26	0.733	0.41	0.159	1.38
25	0.839	-0.139	0.126	3.82
23	0.866	-0.353	0.112	4.76
27	1.08	-1.64	0.000	10.7

[1]Data Set Obtained from first 69 of 105 Isomers Quantitated

Prediction of Composition of Unknown Samples PLS.

Because many samples are analzyed in which the analyst is interested in determining which Aroclor mixtures are present, we applied the PLS method to the data obtained from the analysis of

the used transformer fluid previously discussed. In order to estimate the Aroclor content of the previously classified the used transformer fluid previously discussed. In order to estimate the Aroclor content of the previously classified the used transformer fluid previously discussed. In order to estimate the Aroclor content of the previously classified transformer oil samples, we obtained additional data from the analysis of Aroclors of varying proportions. In Appendix I, the data are ordered in an array and the first four variables designate the fractional part of each Aroclor composing the sample in the order 1242, 1248, 1254, and 1260. This composition data represents the $\underline{Y}$-block and the 69 peaks represent the $\underline{X}$-block of data analyzed with the PLS-2 program.

Table V. Statistical Summary for A=3 Principal Components SIMCA Analysis of Aroclor Samples.

Step	Par.	A	NDF	SS	SD	1-SD/SDY	SS(TETA)	ITET
0		0	1862	4.36E+03	1.53E+00	0.000	0.	0
1	Alpha	0	1794	2.25E+03	1.12E+00	0.268	0.	
2	Beta-	1	1700	5.83E+02	5.86E-01	0.617	1.7E+03	14
3	Beta-	2	1608	1.36E+02	2.91E-01	0.810	4.5E+02	9
4	Beta-	3	1518	5.24E+01	1.86E-01	0.879	8.4E+01	17

The samples of unknown composition--21-23 and samples 1-20, 24-34 (Appendix I) were those of Aroclors of variable composition. Variables 5-73 are isomer concentrations (Variable 74, the total PCB concentration in ppm was not included in the analysis). Variables 5-73 represent the fractional composition or isomer proportional concentration values. Representative concentration histograms of the data set are presented in Figure 13. Four PLS components were extracted and then used to estimate the Aroclor content of the unknowns and of a standard sample (No. 24). The Aroclor standard is a mixture of three Aroclors in the ratio of 0.33:0.33:0:0.33. Chromatograms of the samples for which the PLS estimates were made (Table VI) were similar when compared to a chromatogram of a similar mixture of standards.

The partial least squares (PLS) method has been applied to structure activity problems by Wold $\underline{et}$ $\underline{al.}$ ($\underline{38}$). Recently, Lindberg $\underline{et}$ $\underline{al.}$ ($\underline{40}$) employed this approach to resolve mixtures of humic acid and ligninsulfonate on the basic of fluorescence spectra.

This example demonstrates that the PLS method gives a stable estimate of the $\underline{Y}$-block, even though there are many more $\underline{X}$-variables than samples, a condition that removes the possibility of applying multiple regression. Another advantage of the method
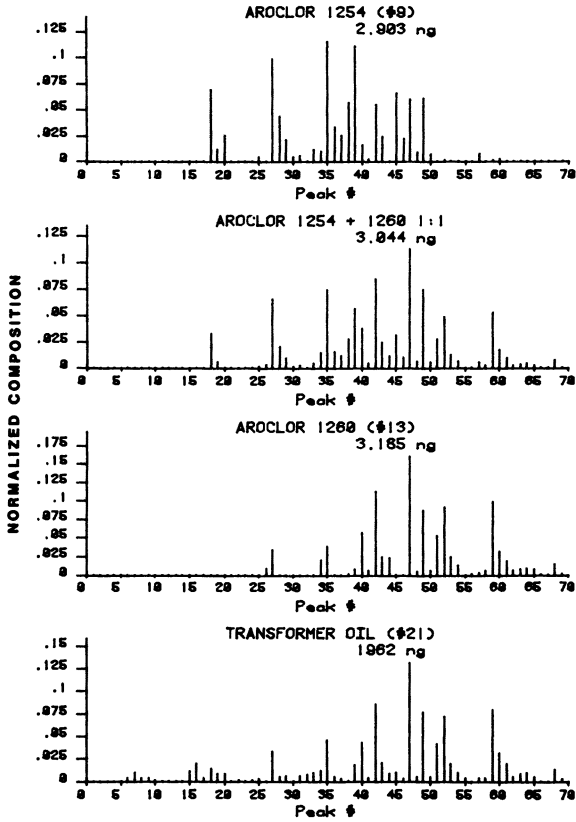
Figure 13. Fractional Composition Histograms of Used Transformer Fluid and Aroclors.

that makes it attractive for use in analytical problems is its computational efficiency and simplicity, which makes it possible to use microcomputers and minicomputers to carry out such calculations.

The SIMCA software is available in two forms, both developed by Wold (25, 31): 1) an interactive, Fortran version which runs on Control Data Corporation (CDC) machines, and 2) an interactive version, SIMCA-3B. Additional information on these programs is contained in Appendix I. Only the SIMCA-3B version contains the CPLS-2 program used for PLS analysis.

Table VI.   Fractional Composition of Aroclors in Transformer Oils Estimated by partial least squares.

Sample number	Aroclor			
	1242	1248	1254	1260
21	.03	.03	.08	.84
22	.03	.03	.08	.84
23	.03	.03	.08	.84
25	.37 (.33)[1]	.36 (.33)	.05 (.00)	.24 (.33)

[1] Actual composition

Environmental Applications

To illustrate the environmental application of the SIMCA method we examined a set of isomer specific analyses of sediment samples. The data examined were derived from more than 200 sediment samples taken from a study site on the Upper Mississippi River (41). These analytical data were transferred via magnetic tape from the laboratory data base to the Cyber 175 computer where principal component analysis were conducted on the isomer concentration data (ug/g each isomer).

The first principal component values (Theta 1) for each sample were determined and these values were correlated with the total PCB concentration (Figure 14) recorded for each sample in a separate computer data base that contained other environmental data such as hydrology and sediment texture. The results indicated that certain samples deviated by factors of about two. Upon examining the sample records, the recorded dilution values
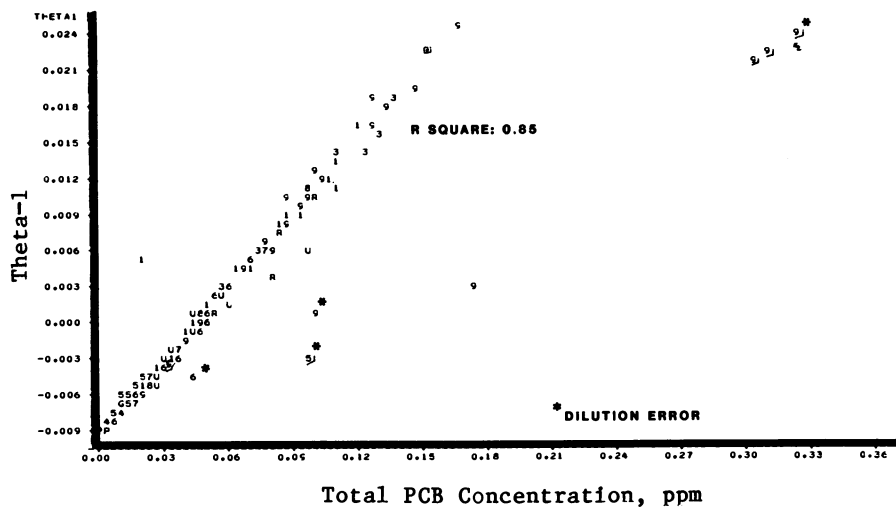
Figure 14. Correlation of Theta 1 vs. Total PCB
Concentration Measured in 201 Lake Onalaska Sediment
Samples. Plot symbol is site number.

for samples designated with an asterisk in Figure 14 were found to be in error by a factor of 2. When the values were corrected, these samples fell on the correlaton line. Little or no correlation of the second principal component value existed with total composition (Figure 15).

An examination of the sample distributions observed in principal components projections using isomer concentration data expressed as fractional composition, as well as the clustering of samples by similar values of their second principal component score term, revealed consistent differences existed in sample profiles. The next step in this data evaluation is to statistically analyze correlations of the PLS components from analyses with the external variables such as percent sand, clay and silt, and total organic matter in samples. These correlations may play an important role in identifying factors resulting in changes in PCB composition and enable one to more clearly understand the forces determining the distribution and fate of PCB in a complex ecosystem.

Total PCB Concentration, ppm
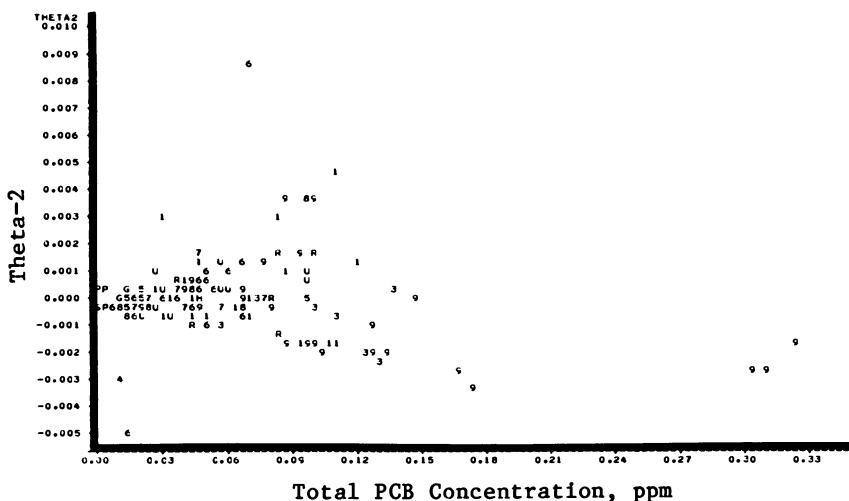
Figure 15. Plot of Theta 2 vs. Total PCB Concentration measured in 201 Lake Onalaska Sediment Samples.

Appendix I.

This appendix contains some of the data generated for the SIMCA
and PLS analyses. The complete data set is available from the
authors. Upon request the data will be provided on 8" single
density single sided floppy disks in IBM 3740 format for CP/M
based systems or on 5 1/4" double sided double density floppy
disk for the IBM/PC or other MS/DOS based computers.   The
requestor, however, must supply a properly formated floppy disk.

Software Availability. The SIMCA software is available in two
forms, both developed by Wold (25): 1) an interactive, Fortran
version which runs on Control Data Corporation (CDC) machines.
The second set of programs are an interactive microcomputer
version, SIMCA-3B, are available from Principal Data Components,
2505 Shepard Blvd., Columbia, MO 65201. The SIMCA-3B pattern
recognition programs includes the CPLS-2 program used for PLS
analysis and are available for CP/M (Digital Research, Pacific
Grove, CA) and MS-DOS (Microsoft Corporation, Bellueve, WA) for
8088 or 8086 based microcomputers.
     The Fortran version used in this study was located at the
Computer Center at the University of Illinois at
Champaign/Urbana.   The Fortran version is useful for analysis of
very large data sets, i.e. 400 x 70 matrices.   The SIMCA-3B
version for microcomputer systems is interactive, menu driven,
and is applicable to intermediate sized data sets and runs under
CPM or MS-DOS.   In this study, the SIMCA-3B program--CPLS-2, was
used to obtain the results in the PLS examples discussed.
     An earlier Fortran version of SIMCA is available for use in
the ARTHUR package available from Chemical Information Systems,
Box 2227, Falls Church ,VA. Recently, the operating system was
changed on the CDC Cyber computer system at the University of
Illinois.   The new operating system does not allow the earlier
SIMCA-2T version used to perform the environmental analyses to
operate correctly.   The authors expect that a new version of
SIMCA will be installed that will function with the current
operating system in use on the CDC Cyber computer.

Partial Summary of Data from the Gas Chromatographic Analysis of
Aroclor, Aroclor Mixtures, and Transformer Oil analyses.

Table 1. Identity of Samples Analyzed—Aroclor 1242, 1248, 1254,
1260, Their Mixtures and a Transformer Oil. Data are included in
this appendix for sample numbers designated with an asterik.

SIMCA ID	Description
1*	Aroclor 1242 Replicate
2	Aroclor 1242 Replicate
3	Aroclor 1242 Replicate
4	Aroclor 1242 Replicate
5*	Aroclor 1248 Replicate
6	Aroclor 1248 Replicate
7	Aroclor 1248 Replicate
8	Aroclor 1248 Replicate
9*	Aroclor 1254 Replicate
10	Aroclor 1254 Replicate
11	Aroclor 1254 Replicate
12	Aroclor 1254 Replicate
13*	Aroclor 1260 Replicate
14	Aroclor 1260 Replicate
15	Aroclor 1260 Replicate
16	Aroclor 1260 Replicate
17*	Aroclor42:48:54:60 1:1:1:1
18*	Aroclor42:48:54:60 1:1:1:1
19*	Aroclor42:48:54:60 1:1:1:1
20*	Aroclor42:48:54:60 1:1:1:1
21*	Used Transformer Oil Replicate
22*	Used Transformer Oil Replicate
23*	Used Transformer Oil Replicate
24	Aroclor 42:48:54:60 1:1:0:1
25	Aroclor 42:48:54:60 1:0:1:1
26	Aroclor 42:48:54:60 0:1:1:1
27	Aroclor 42:48:54:60 1:1:0:0
28	Aroclor 42:48:54:60 1:0:1:0
29	Aroclor 42:48:54:60 1:0:0:1
30	Aroclor 42:48:54:60 0:1:1:0
31	Aroclor 42:48:54:60 0:0:1:1
32	Aroclor 42:48:54:60 0:1:0:1
33	Aroclor 42:48:54:60 1:1:1:1

Table II. Data Matrix Organization for Aroclors and Samples.

Data Matrix – Variable #

Sample #	ID code			
1(1242)[1]	2(1248)	3(1254)	4(1260)	1
6(peak 2)[2]	7(peak 3)	8(peak 4)	9	10
11	.	.	.	.
16	.	.	.	.
21	.	.	.	.
26	.	.	.	.
31	.	.	.	.
35	.	.	.	.
41	.	.	.	.
46	.	.	.	.
51	.	.	.	.
56	.	.	.	.
61	.	.	.	.
66	.	.	.	.
71	72	73(peak 69)	74[3](Total conc.)	

[1]Weight fraction each Aroclor in sample variable 1-4
[2]Variables 5-73 are fractional concentration of each PCB isomer
[3]Variable 74 designates total PCB concentrationin sample

Table III. Representative Analyses of Aroclors, Their Mixtures and a Transformer Oil Sample. Refer to Table 2 for key to data organization.

Sample #	ID code			
	1 1 42			
100	0	0	0	.3318
0	.03722	.1324	.1501	.6295
.8347	.3023	.3691	.1762	0
.07662	.07048	.02987	.5528	1.14
.2134	.3437	.3375	.3169	.09121
.08603	.02791	.07272	.1301	.1613
.04002	.0355	.1661	.08603	.2138
.1051	.09709	0	.03421	.02422
.01593	.02423	.02633	0	0
0	0	0	.01582	.01318
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0
0	0	0	7.511	

Table III. Continued

5   5 48

0	100	0	0	.02185
0	0	0	.02243	.04143
.3688	.09205	.1105	.07285	0
.1108	.08407	.03458	.3342	.4488
.09049	.6112	.4933	.5239	.1222
.1164	.04218	.1092	.2177	.05379
.1572	.1294	.3208	.1477	.3446
.1657	.1605	0	.1479	.09148
.06115	.09849	.1533	0	0
0	0	0	.09543	.05943
0	0	0	0	0
0	0	0	0	0
.0009057	0	0	0	.0003501
0	0	1.463E-05	0	0
0	6.426E-05	0	6.257	

9   9 54

0	0	100	0	0
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0
0	.2006	.03684	.07303	0
.001386	0	0	.01534	0
.2877	.1281	.06263	.01397	.01731
.003889	.03496	.03043	.3352	.09876
.07539	.1657	.3222	.04839	.007758
.1595	.07247	0	.1939	.06609
.1752	.02599	.1777	.01992	0
.005103	.002069	.001326	0	.002222
.02406	0	.007326	.0102	0
.001136	0	0	0	0
0	0	0	2.903	

13   13 60

0	0	0	100	0
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0
0	0	0	0	.02807
.1134	0	0	0	0
0	.0004755	.06472	.1251	0
0	.006675	.02628	.1837	.02169
.3606	.07934	.07441	.0009664	0
.5124	.01567	.2802	.01874	.1711
.2944	.08182	.04352	0	.006068
.01198	.02128	.3146	.1039	.06354
.01951	.02727	.03142	.02323	.002931
0	.04784	.009139	3.185	

Table III. Continued

17 17 M4

25	25	25	25	.05351
0	.007108	.02049	.01411	.1062
.2019	.06352	.0797	.04155	0
.03203	.02606	.01205	.1385	.2308
.04948	.2394	.156	.1621	.03699
.03584	.01305	.02929	.06368	.03058
.1765	.07449	.09968	.04474	.09625
.04801	.05987	.03558	.1951	.05449
.04109	.08646	.1563	.08263	.01062
.1798	.05681	.02459	.09232	.03886
.2308	.01501	.1661	.01465	.06061
.1034	.03089	.01593	0	.002988
.01464	.007318	.103	.04175	.02351
.007217	.009673	.01155	.007575	.0004781
0	.01765	.002694	4.385	

18 18 M4

25	25	25	25	.0582
0	.009318	.02033	.01366	.1012
.1923	.06195	.07673	.03757	0
.02852	.0246	.01163	.1362	.2158
.04677	.2254	.1456	.1597	.03718
.03311	.02247	.03049	.06241	.02041
.1754	.06777	.0975	.04221	.09346
.04581	.0564	.03511	.1943	.05288
.03923	.08372	.1494	.07889	.01055
.1755	.05492	.02173	.0869	.03402
.2197	.01502	.1514	.01265	.05624
.09693	.02829	.01499	0	.002718
.01202	.006532	.09292	.03756	.02181
.006168	.008577	.01005	.007134	0
0	.01566	.002565	4.186	

19 19 M4

25	25	25	25	0
0	.00758	.01591	0	.1035
.1979	.06332	.07692	.03998	0
.02768	.02566	.01181	.139	.2205
.04811	.2343	.1521	.1627	.0366
.03671	.02365	.03228	.06273	.01992
.1798	.07212	.09911	.04306	.0963
.04658	.05797	.03613	.198	.05478
.04075	.08714	.1496	.08243	.01029
.1816	.05516	.02203	.09154	.03546
.2255	.01544	.1564	.0134	.05881
.09807	.02874	.01439	0	.002419
.01253	.00762	.09556	.03932	.02241
.006722	.009386	.01079	.006948	0
0	.0158	.001706	4.22	

Table III. Continued

20 20 M4

25	25	25	25	0
0	0	.01953	.02258	.1012
.1888	.05618	.07373	.03659	0
.029	.02383	.01114	.1375	.2149
.04563	.2313	.1486	.159	.03436
.03259	.0219	.03021	.06214	.0217
.1742	.07348	.09438	.04086	.09305
.04368	.05558	.03355	.1881	.05164
.03871	.08279	.1486	.07911	.008697
.1776	.05508	.02429	.08819	.03461
.2167	.0147	.1513	.01244	.05699
.09493	.0272	.01298	0	.002228
.01184	.007262	.09327	.03842	.02261
.006812	.008346	.01053	.006725	0
0	.01538	.001846	4.101	

21 21 TO

0	0	0	0	0
0	0	0	0	9.761
21.21	9.022	7.668	4.709	0
0	0	0	22.01	37.65
9.456	29.72	18.87	16.88	0
5.23	0	3.79	6.906	0
67.07	11.2	13.59	0	13.89
16.32	19.48	25.79	90.13	10.21
7.393	2.339	36.12	83.84	9.548
169.9	41.68	17.97	19.92	7.416
259.6	10.67	150.3	11.58	82.64
141.5	39.53	20.73	9.535	0
9.323	8.306	156.3	62.34	39.39
10.12	18.16	20.98	10.67	0
0	27.11	6.732	1962	

022 22 TO

0	0	0	0	0
0.	0	0	0	7.651
18.84	6.861	6.402	4.596	0
0	0	0	20.97	36.33
9.354	28.37	16.44	16.38	0
4.054	0	0	7.090	0
68.15	11.62	13.79	0	13.65
15.17	17.98	24.69	89.59	8.900
6.871	2.331	36.60	82.98	9.401
167.6	42.73	18.34	18.26	69.10
258.6	10.65	150.1	11.48	81.87
140.7	39.08	20.85	9.035	0
8.846	8.124	156.7	61.77	39.50
10.08	17.99	20.90	10.63	0
0	26.88	6.672	1929.	

Table III. Continued

	023	23 TO		
0	0	0	0	0
0	0	0	0	9.198
20.07	8.285	5.973	4.554	0
0	0	0	21.05	33.38
8.918	27.98	15.53	15.96	4.408
0	0	0	6.221	0
63.37	10.25	13.64	0	12.98
15.78	19.00	24.82	86.27	8.967
0	2.185	37.18	79.38	9.282
161.7	39.50	17.65	20.34	6.783
244.7	9.602	141.9	10.87	77.59
131.9	36.90	19.59	8.435	0
8.368	7.868	146.8	56.95	36.46
9.241	16.73	19.50	9.806	0
0	24.92	6.579	1835	

Literature Cited

1. Ballschmiter, K.; Zell, M. Freshenius Z. Anal Chem. 1980, 302, 20.

2. Albro, P. W.; Corbett, J. T.; Schroeder, J. L. J. Chromatogr. 1981, 205, 103.

3. Bush, B.; Connor S.; Snow, J. J. Assoc. Off. Anal. Chem. 1982, 65, 555.

4. Hutzinger, O.; Safe, S.; Zitko, V. in "The Chemistry of PCB's," CRC Press, Cleveland, OH, 1974.

5. Brinkman, U. A. Th. de Kok, A. in "Topics in Environmental Health, Halogenated Biphenyls, Terphenyls, Naphthalenes, Dibenzodioxins and Related Products, Kimbrough, R. D., Ed. ; Elsevier/North Holland Biomedical Press: New York , 1980; 2-4.

6. Jensen, S. New Sci. 1966, 32, 612.

7. Zell, M.; Ballschmiter, K. Fresenius Z. Anal. Chem. 1980, 304., 337

8. Koeman, J. H.; Debrouw, M.C.; De Vos, R.H. Nature (Lond.) 1969, 221, 1126.

9. Biros, F.J.; Walker, A.C.; Medbery, A.; Bull. Environ. Contam. Toxicol. 1970, 5, 317.

10. Fishbein, L. J. Chromatogr. 1972, 68, p 345.

11. Hammond, P. B.; Nisbet, I.C.T.; Sarofim, A.F.; Drurry, W.H.; Nelson, W.; Rall, D. P. Environmental Impact. Environ. Res. 1972, 5, 249.

12. Bush, B.; Tumasonis, C.F.; Baker, F.D. Arch. Environ. Contam. Toxicol. 1982, 28, 97.

13. Mes, J.; Davies, D.J.; Turton, D. Bull. Environ. Contam. Toxicol. 1982, 28, 97.

14. Bandera, S.; Sawyer, T.; Campbell, M.A.; Robertson, L.W.; Safe, S. Life Sciences 1982, 31, 517.

15. Safe, S.; Robertson, L.W.; Safe, L.; Parkinson, A.; Bndera, S.; Sawyer, T.; Campbell, M. A.   Can. J. Pharmacol., 1982, 60, 1057.
16. Rappe, C.; Buser, H.-R. in "Topics in Environmental Health, Halogenated Biphenyls, Terphenyls, Naphthalenes, Dibenzo dioxins and Related Products, Kimbrough, R. D., Ed.; Elsevier/North Holland Biomedical Press: New York, 1980; Chap. 2.
17. Poland, A.; Glover, E. Mol. Pharmacol. 1977, 13, 924.
18. Trotter, ; Young, W.J.; Casterline, S.J.V., Jr.; Bradlaw, J.L.; Kamps, L.R. J. Assoc. Off. Anal. Chem. 1982, 65, 838.
19. Webb, R.G.; McCall, A.C. J. Chromatogr. Sci. 1973, 11, 366.
20. Environmental Protection Agency, Washington, D. C., Method 625, Fed. Reg. 1979, 44, 69540.
21. Duinker, J.C.; Hillebrand, J.I.J.; Palmark, K.H.; Wilhemsen, S. Bull. Environ. Contam. Toxicol. 1980, 25, 956.
22. Zell, M.: Ballschmiter, K. Fresenius Z. Anal. Chem. 1980, 304, 337.
23. Bopp, R.F.; Simpson, J.; Olsen, C.R.; Kostyk, N. Environ. Sci. and Technol. 15, 1981, 210.
24. Kowalski, B.R. Anal. Chem. 1980, 52, 112R.
25. Wold, S. Kemia-Kemi 1982, 9, 401.
26. Schwartz, T.R.; Stalling, D.L.; Petty, J. D.; Hogan, J.W.; Marlow, B. K.; Campbell, R.D.; Little, R.L. 184th National Meeting of the American Chemical Society, Environmental 1982, Papers 20, 21, Kansas City, MO.
27. Schwartz, T.R. M.S. Thesis, University of Missouri-Columbia 1982.
28. Albro, P. W.; Fishbein, L. J. Chromatogr. 1972, 69, 273.
29. Ugawa, M.; Nakamura, A.; Kashimota, C. in "New Methods in Enviromental Chemistry and Toxicology," Proceedings of the International Symposium, Susonso, Japan, International Academic Printing Co., Tokyo, Japan 1973.
30. Wold, S. Pattern Recognition 1976, 8, 127.
31. Wold, S.; Albano, C.; Dunn, W. D.,III; Esbensen, E.; Helberg, S.; Johansson, E.; Sjostrom, M.; "Pattern Recognition: Finding and Using Regularities in Multivariate Data in Food Research and Data Analysis," Eds. Martens, H. and H. Russwurm, Jr.; Applied Science Publishers, New York, 1983, pp. 147-188.
32. Wold, S.; Sjostrom, M. in ACS Symposium Series No. 52, American Chemical Society: Washington, D.C., 1977, pp. 243.
33. Joreskog, K.G.; Klovan, J.E.; Reyment, R.A. "Geological Factor Analysis," Elsevier, Amsterdam, 1976.
34. Wold, S. Technometrics 1978, 20, 397.
35. Kowalski, B. J. Amer. Chem. Soc. 1973, 95, 686.
36. Massart, D.L.; Dijkstra, A.; Kaufman, L.; in "Evaluation and Optimization of Laboratory Methods and Analytical Procedures"; Elsevier, Amsterdam 1978.

37. Wold, H. Soft Modeling by Latent Variables: the Nonlinear
     Iterative Partial Least Squares Approach," Ed. J. Gani,
     in Perspective in Probability and Statistics - Papers in
     Honor of M. S. Bartlett, Academic Press, London 1975,
     pp. 117-142
38. Wold, S.; Dunn, W.J.,III; J. Chem. Inf. Comput. Sci.
     1983, 23, 6.

39. Dunn, W.J.,III.; Stalling, D.L.; Schwartz, T.R.; Hogan,
     J.W.;   Petty, J.D.; Johansson, E.; Wold, S.; "Pattern
     Recognition for Classification and Determination of Poly-
     chlorinated Biphenyls in Environmental Samples," Anal.
     Chem. 1984, in press.
40. Lindberg, W.; Persson, J. A.; Wold, S. Anal. Chem. 1983,
     55, 643.
41. Dexter, R. N.; Pavlou, S. P.; Hines, W. G.; Anderson,
     C.; 1978, "Dynamics of Polychlorinated Biphenyls in the
     Upper Mississippi River: Final Report. Phase I, Task 1:
     Compilation of Information. U. S. Fish and Wildlife
     Service, Columbia, MO Contract No. 14-16-009 78-026.

From Data to Information to Knowledge

The Problems of Metamorphosis

C. ZERVOS

Pharmaceutical Research and Testing, National Center for Drugs and Biologics, U.S. Food and Drug Administration, Washington, DC 20204

Rational control of health and
environmental risks from technical
development requires scientific
knowledge which must be acquired through
the orderly process of the scientific
method of inquiry. Contrary to widely
held opinions the latter is no less
subjective than other rational human
endeavors which require decisions under
uncertainty. Indeed, to be applied, the
method requires a value system which in
ordinary research is supplied by the
various scientific disciplines. Because
of differences among the disciplinary
value systems problems often arise in
the interdisciplinary settings of
efforts to control risks from technical
development. Metrics, the concepts,
theory, and practice of measurement is
suggested here as a way to deal with
such problems.

The terms data, information, and knowledge are often
used interchangeably for diverse purposes by
researchers in all scientific disciplines. In the
scientific enterprise, however, they are not
interchangeable, despite arguments to the contrary. As
a matter of fact, Chemometrics, the subject of this
Symposium, may spring at least in part from the real

differences among these terms. The International
Chemometrics Society, for instance, declares that:

"Chemometrics is the chemical discipline that
uses mathematical and statistical methods (a) to
design and select optimal measurement procedures
and experiments, and (b) to provide maximum
chemical information by analyzing chemical
data."(1)

The differences among the three terms are not
just of theoretical interest. They touch nearly every
aspect of our daily lives because they are central to
the functions of the agencies that protect the public
health and the environment. Specifically, to be
credible, these agencies must base their actions on
"accepted" scientific knowledge. Consequently, they
collect enormous amounts of experimental data. These
data, however, are of little use until they are first
converted to scientific information and then placed in
the context of other relevant scientific information
and thereby become knowledge.

The standards for converting data to information
are short-cut or "economy" solutions to the universal
problem of having to decide under uncertainty; these
standards are based on convention, not on science.
They vary from discipline to discipline and from time
to time. As might be expected, although appropriate in
the context of their development, such standards are
often likely to be incomplete or otherwise
inappropriate for universal application because they
are value-laden rules for making choices (vide infra).
Through use, however, they become valuable to those
who use them. Often these standards also become a
cause of contention when, in interdisciplinary
settings, practical knowledge must be extracted from
experimental data. Thus, controversies often arise
when scientists trained in different disciplines
influence public policies or make decisions based on
the conversion standards in which they were trained.

Here I will examine at some length the problems
with one such standard, namely, the odds for deciding
"gating" hypotheses (vide infra) in the life science-
related disciplines. This standard plays a pivotal
role in the assessment or management of technological
risks and thus is at the root of many controversies of
the genre. I will also suggest that this and similar
problems can be overcome by renewed emphasis on the
proper use of the scientific method of inquiry and by
focusing attention on Metrics, i.e., the concepts, the

theory and the practice of measurement. The following
examination of these problems will include a brief
description of the scientific method of inquiry; an
analysis of its value foundations; a description of
representative examples of subjective choices in
science; an analysis of the clash of values during
interdisciplinary investigations of societally
important topics; and a recommendation to develop and
expand the uses of Metrics to overcome the
difficulties of making decisions under uncertainty.

The Scientific Method of Inquiry; An Overview.

Different authors describe the scientific method of
inquiry differently depending on what they wish to
emphasize(2-7). The following description
serves the purpose of this paper.

   The scientific research method consists of a
cyclic sequence of experimental and logical steps
designed to yield knowledge about the world by using
conjecture and appeals to experience. The first step
in the sequence is to divide the world into a
conveniently sized part called the "study system" and
a complementary part called the "environment." A study
system and its complement are created by and for the
convenience of the investigator only. A study system
can be of any size and type as long as it is known
where in the universe it comes from and exactly where
it fits in it.

   The second step in the sequence eludes adequate
description but is important. It is an understatement
to describe it as a review, analysis and resynthesis
of accepted knowledge about a particular study system
and about its relation to its environment. Yet, that
is all that can be said to characterize this second
step besides acknowledging a central role for
intangibles such as genius, intuition, serendipity,
insights, etc.

   Reviews and analyses of accepted knowledge are
performed in order to accommodate and assimilate
information. Most epistemologists seem to agree that
attempts at resyntheses aim at ever more general
conjectures about the nature of the universe and the
interactions between its parts.

   The properties of the conjectures that result
from successful resyntheses have been discussed
(6,7). Although opinions may differ about the
role of such conjectures and their contribution to
scientific research, there is agreement that they lead
to development and articulation of statistical

hypotheses about the specific aspects of study
systems.(2,4,8)
     All scientific conjectures are speculations but
not idle speculations. Rather, they are rooted in
accepted scientific knowledge. This is one of the
properties that distinguish scientific conjectures
from metaphysical schemes of knowledge. Another is the
requirement that, in principle, they be subject to
refutation by experimental observations
(2-4,6).
     Experiments designed to test specific statistical
hypotheses comprise the third step of the scientific
method of inquiry. They constitute appeals to
experience regarding the validity of a specific
scientific conjecture. They provide for the collection
of data, i.e., numbers that refer to, characterize, or
specify the attributes of a study system of interest.
     Data are the raw product of the scientific method
of inquiry. By analysis, refinement and reduction
which collectively constitute the fourth step in the
sequence, data are converted to information about the
nature of study systems. The conversion is
accomplished by the Neymann-Pearson process of
statistical hypotheses testing(8). If the
collected data are sufficient and pertinent enough to
support rejecting or accepting the statistical
hypothesis under test, a measurable(9,10)
quantity of information about the study system has
been extracted. If not, the data cannot be converted
to information and therefore cannot contribute to the
pool of accepted scientific knowledge.
     The cycle begins again with new reviews, analyses
and attempts at resynthesis. The continuing purpose is
to assimilate all available information including that
which was newly acquired and to incorporate it into a
more consistent and more general scheme, an
overarching conjecture of ever expanding scope.
     Attempts at resynthesis take the form of
articulation and testing of new theories by the
scientific community. Theory testing procedures are
not as rigorous or as easy to describe as the
Neymann-Pearson procedures for testing statistical
hypothesis. Rather, they are unstructured and they
reserve a large role for "scientific intuition" and
other similar intangibles(11). Buck(6,7)
has discussed the criteria for having "meshed"
established or accepted knowledge and newly acquired
information, i.e., for having converted the latter to
the former. These criteria are (a) consistency of
information from different sources and (b) the

emergence or furtherance of general principles
which describe the observed behavior of the study
system and predict how it will behave under a
specified set of circumstances. Buck has also noted
that theory testing has the following properties: it
can further substantiate what is considered accepted
knowledge about the study system; it can raise
questions about it; it always gives rise to new
hypotheses which commence a new cycle of research;
often, it necessitates a redefinition of the study
system with its boundaries drawn either more clearly
or less; and on occasion, it can change the
perspectives of most scientists about accepted
knowledge(12).

Because of the nature of the scientific method,
Metrics is an indispensable tool of scientific
research. It can provide rigorous indices of the
internal consistency and the predictive power of
"accepted knowledge" about study systems. Thus, it can
aid with theory testing. It can also provide rigorous
indices of the strength of correlations between the
attributes of the study system and the external
factors that might influence it. Thus, it can assist
with statistical hypothesis formulation and testing.

Although recognition of its importance in
research is still growing, Metrics is not new. Indeed,
over the past 400 years, it has evolved into an
indispensable element of the scientific method and is
now an integral part of it. Thus, some of the concepts
and the theories of Metrics are old. They were
developed hand-in-hand with and for the benefit of the
older scientific disciplines, principally physics. In
earlier times systematic application of Metrics in the
less mature disciplines was infrequent and
controversial(13) and for the most part intuition
and "professional judgement" substituted for Metrics
in research. Presently, Metrics is used more and more
by researchers in all disciplines. As a result, it has
become part of our daily lives and activities and is
manifested in a variety of unexpected disciplinary
apparel (viz. biometrics, chemometrics, econometrics,
psychometrics, sociometrics, anthropometrics, etc.).
Two factors are responsible for this blossoming: the
advent of large computers that can collect, store, and
manipulate unlimited amounts of data; and the
increasing societal significance of scientific data.
The first factor and its effects on science have been
discussed elsewhere recently(1). It is the second
that is of interest here, especially as it relates to
the potential contribution of Metrics to the

standards, techniques, and theories of conversion of
data into information, often a troublesome
undertaking.

The value foundations of the scientific method of inquiry.

It is a commonly held view that the scientific method
of inquiry is, or at least ought to be, objective in
the sense that it should transcend the personal value
system of the researcher. The terms bad science,
better science or best science are often used to
connote the perceived degree of objectivity of a
particular scientific activity(14). Further,
"trans-science" is a term which was coined to sort and
label putative encroachments upon "objective" science
by issues that can be posed as scientific questions
but cannot be answered by the available means of
scientific experimentation(15). The term was
introduced because it was thought that the division of
technologically important issues into scientific and
trans-scientific would significantly reduce the
problems of converting data into useful iformation.
      As will be shown below, the notion that science
is objective is mistaken and the mistake contributes
to many of the problems encountered when scientific
reseach has immediate economic or political
consequences(16-21). It should not be
surprising that years after the introduction of the
concept of trans-science, we still observe difficulty,
conflict and general misunderstanding not of what
science can or cannot do, but of what science does do
and how. The reason is a general lack of appreciation
that the scientific method of inquiry is inherently
and specifically subjective and that it requires a
value system without which it simply cannot be
applied.
      This is not a restatement of the old observation
that science is influenced diffusely and subtly by the
prevailing value system of society(22). It is to
say that value-laden judgments are specific, readily
identifiable and necessary parts of the cyclic
sequence of activities that make up the scientific
research method. Such judgments are the scientists'
means of coping with experimental uncertainty, i.e.,
the decision rules for testing statistical
hypotheses(8).
      Uncertainty in measurement is a universal feature
of all experimental work(2,3,8) and must
be dealt with if data are to be converted to

information. Decision theory, a branch of Metrics,
provides the method of maximum expected
utilities(23) for handling uncertainty in
statistical hypotheses testing. But this method
requires the construction of an explicit utility
scale. As expected, the method finds use when and
where it is advantageous to construct such a scale,
e.g., engineering, product development research,
business, economics research, etc. In such cases, it
is both possible and generally acceptable to assign
monetary values to the consequences of correct or
incorrect choices on the basis of experimental data.
The approach is difficult and not always preferable
for other scientific disciplines, which consequently
opt to deal with experimental uncertainty in less
transparent ways. In general toxicology, cancer
toxicology, analytical chemistry, biology,
diagnostics, clinical medicine, etc., for instance, it
is not always advantageous to state openly and
explicitly the utility scale used by researchers.
Consequently, the cultural dimensions of using the
scientific method are not obvious.

Examples of Subjectives Choices.

Discussion of the following three examples, namely,
general toxicology, cancer toxicology, and analytical
chemistry will be used to illustrate the point. They
have been selected because, among the life
science-related disciplines, they cover the spectrum
of needs for scientific "judgements."

General Toxicology and its special branch cancer
toxicology or chemical oncology are concerned with the
adverse effects of chemicals on living systems. A
major force behind current toxicological research is
society's need to protect the public health and the
environment from the toxic effects of chemicals
without unnecessarily inhibiting the introduction of
new products and technologies. Accordingly, a major
part of present day toxicological research consists in
using groups of experimental animals as devices to
measure the toxic potential of commercially important
chemicals.
     The measurement begins with Neyman-Pearson type
of testing(8) of the following null hypothesis:

    Given a defined set of experimental conditions
    of observation, the incidence of adverse effects
    in a sample population of experimental animals

is not affected by a specific rate of exposure
to the substance under study.

    The test is repeated for all identifiable effects
and for a number of different rates of exposure.
Standard 2x2 contingency tables(24) are used to
determine the odds of observing the bioassay outcome
if the null hypothesis is true using Fisher's exact
test. For all adverse effects, including cancer, the
first rule of the strategy for deciding whether a
substance is toxic is to:

    Assume that the null hypothesis is true and if
    the odds against the observed outcome are equal
    to or greater than 19 to 1 to reject the
    assumption; otherwise to accept it.

    The decision rules that follow vary depending on
whether the observed effects include cancer or not. If
they do not, the strategy rules are clear(25),
even if not clearly justified(26). If they do, the
stategy rules are unsettled and still the subject of
much controversy(27-32). Significantly, they
depend on the uses or potential uses of the substance
under study. As a matter of fact, as a discipline,
chemical oncology does not seem to have evolved a
clearly identifiable and accepted set of hypotheses
and test rules for deciding whether a substance will
be considered a carcinogen and how potent it may
be(27). The U.S. Congress on the other hand has
made several attempts to specify a utility scale so
that decisions concerning the carcinogenic potential
of societally important chemicals can be made by
maximizing the expected utilities. The outstanding
problem has been that the utility scale outlined by
the Congress is constantly in either general or
specific conflict with one or more decision rules
devised by several scientific groups with an interest
in the matter(30).

In Analytical Chemistry, one of the oldest and
most "objective" scientific disciplines, the current
impetus for research comes from the needs of other
disciplines and from society's need to protect itself
and the environment from noxious chemicals. Analytical
chemistry uses a large number of physical, chemical
and biochemical principles to determine whether a
particular, potentially noxious substance, the
analyte, is part of specific, commercially useful and
societally important matrices of substances (e.g.,

foods, water, air, soil, etc.) and at what proportion.
    Analytical methods are not ordinarily associated
with the Neyman-Pearson theory of hypothesis testing.
Yet, statistical hypothesis tests are an indispensable
part of method development, validation, and use . Such
tests are used to construct analytical curves, to
decide the "minimum significant measured" quantity,
and the "minimum detectable true" quantity
(33,34) of a method, and in handling the
"outlier value problem"(35,36).
    Two somewhat different types of null hypotheses
are tested, one during the development and validation
of an analytical method and the other each time the
method is used for one purpose or another. They are
stated here in general form but they can be made
suitably specific for experimentation and testing
after review and specification of the physical,
chemical and biochemical properties of the analyte,
the matrix, and any probable interfering substances
likely to be in the same matrix. Further, the null
hypotheses of analytical chemistry are cast and tested
in terms of electronic signal to noise ratios because
modern analytical chemistry is overwhelmingly
dependent on electronic instrument responses which are
characterized by noise.
    The basic null hypothesis of method development
is:

    Given a defined set of experimental conditions
    suitable for observing the effects of the
    analyte on the properties of the matrix, the
    probability of observing a signal to noise ratio
    greater than, or equal to, a predetermined
    number is not affected by the presence of a
    specified number of units of the analyte in a
    specified number of units of the matrix.

    This null hypothesis is tested repeatedly for
different analyte concentrations in each matrix of
interest. The overall testing strategy is frequently
complicated by special rules that depend on the
reasons why the method is being developed in the first
place. Always, however, the first rule is:

    For each analyte concentration, assume that the
    null hypothesis is true. If the odds against the
    observed signal to noise ratio are equal to or
    greater than 19 to 1 reject the assumption;
    otherwise accept it.

    The basic null hypothesis of method use is of

course predicated on the availability of a validated
method, i.e., on the availability of a difined set of
experimental conditions suitable for measuring analyte
concentrations in a suitable range and on the
expectation of an appropriately large (i.e., p>0.95)
probability that under the specified conditions, the
signal to noise ratio due to an analyte concentration
in the appropriate range will be greater than or equal
to a predeternimed number. Under such circumstances
the null hypothesis is:

> The observed signal to noise ratio is less than
> that expected from the minimum detectable
> concentration of analyte.

The first rule of the test strategy is:

> Assume that the null hypothesis is true, and if
> the odds against the observed signal to noise
> ratio, are 19 to 1 or greater reject the
> assumtion, otherwise, accept it.

It is unnecessary to delve into hypothesis
testing strategies further to discern the role of
values in the choice of decision rules. The following
are immediately noteworthy. First, whether one
observes analytical chemistry, an old and well
established discipline, or chemical oncology, a
relatively recent one, the observer is struck by the
fact that "making odds" is an indispensable step of
the process of converting data to information. This
fact alone establishes beyond reasonable doubt that
science is no less subjective than other rational
human endeavors.
       Second, all three disciplines in the chosen
sample begin the process of converting data to
information by using what might be called a "gating"
hypothesis which establishes a standard of strength
for the correlation between the stimulus applied to
the study system and the subsequent system changes
that will be considered a response.
       Third, implicitly or explicitly, all three
disciplines decide the "gating" hypothesis by using a
rule which rejects it only if the observed odds
against it are overwhelming, i.e., no less than 19 to
1 and often much higher. However, very rarely, if
ever, is the choice justified explicitly and
analytically. This is a common and important feature
of decision strategies.
       Undoubtedly the choice of odds in each specific

case is based on some sort of a utility scale. It is generally difficult if not impossible, however, to tell whose utility scale it is and how and why it has the form it does. For instance, the particular choices of the gating odds (commonly known as the significance level, the alpha value or the acceptable error of the first kind) are a measure of individual and collective concern about false-positive decisions. The conspicuous absence from the decision rules of, or the perfunctory concern about, the limitations of all methods of measurement for errors of the second kind is a measure of individual or collective indifference toward false-negative decisions. If such attitudes are indeed the foundation of the decision rules, they manifest the implicit utility scales of the concerned disciplines. The difficulty of identifying the architects of these scales and the reasons for their characteristic features is an adverse commentary on science. Knowing more about them is important to all, not simply a matter of "science" for toxicologists or analytical chemists only.

But where have these attitudes come from and what is their justification? Why should there be strong and pervasive concern among scientists about errors of the first kind (false-positive decisions) while little concern and only perfunctory thought is given to errors of the second kind?

Undoubtedly, some justification for this attitude stems from the dynamics among scientists and from the way scientists perceive themselves and their work individually and collectively. Scientists are generally a very conservative group(37). They reserve harsh penalties for those among them who take chances and announce conclusions that might not be confirmed subsequently. In terms of peer recognition and standing among one's fellow researchers, the consequences of losing a Neyman-Pearson bet are potentially devastating and the betting odds that scientists require are therefore adjusted accordingly.

To justify the selection of overwhelming odds against the null hypothesis many will also argue that a lax standard for errors of the first kind would promote inefficiency in research and would therefore be detrimental to the scientific enterprise as a whole. This argument is behind the often-heard assertion that scientists need to be certain about the positive results they accept "....because they are used to construct new hypotheses and theories and will be incorporated into the body of assumed scientific knowledge."(37,38)

The argument sounds impressive. But all
scientific theories and knowledge are temporary and
incomplete descriptions of physical reality; they are
forever subject to change. Thus, it is at least
arguable and perhaps impossible to substantiate
whether science stands to advance more efficiently by
being overly cautious about errors of the first kind
and essentially indifferent to those of the second.
      Finally, a third type of argument is possible
especially for the choices of disciplines that use and
depend on costly experimental techniques (e.g., the
2-year bioassays of toxicology). The concern about
false positive decisions implied by the requirement
for small values of alpha regardless of the
consequences probably reflects where toxicologists
collectively strike a balance between the need for a
broad front advance and scarcity of resources.
      Whatever the true merits of the reasons and the
justifications for the conservative attitudes of most
scientists, it is true that one seldom hears arguments
for avoiding errors of the second kind, especially for
small differences between means(8). It appears
instead that there is a convention in the life
science-related disciplines which automatically sets
at 0.05 the maximum acceptable value for errors of the
first kind without critical consideration of all that
that might entail.

Decision Rules for Societally Important Study
Systems: Clashing Values

It would be preferable, of course, if individually and
collectively, scientists were more forthcoming with
explanations and justifications of their choices of
values. But it is difficult to advance strong
arguments against their choices if research is
practiced only as the asymptotic pursuit of "pure
truth" and if the choices are not misrepresented as
objective and scientific. However, as it is funded
and practiced currently, scientifc research deals
almost exclusively with societally important study
systems and therefore has practical as well as "purely
scientific" objectives. Thus, it is proper to require
that decision rules be adequately explained and
justified in their societal context and even to look
with a jaundiced eye at misrepresentations of value
choices as "objective" or "scientific" for whatever
reason and in any research context.
      Examination of a sample of three disciplines
revealed that the requirement for 19 to 1 or greater

odds against gating hypotheses is quite general. The unsuspecting may indeed run the danger of mistaking this requirement for a scientific principle. It is not. It is a gambler's rule. In common jargon, when using it in science, the investigator, the data analyst, etc. behave like gamblers with a betting strategy.

Most scientists are taught and understand this aspect of science. But the literature contains many admonitions against granting a special status to this convention, and many indications that investigators assume it is a scientifically derived discriminant between what has been proved "scientifically" and chance occurences.

Ackoff(23), for example, states:

The value of alpha is conventionally taken to be 0.05 but this choice generally has no rational basis."

Also, in discussing matters of statistical significance and its role in FDA decisions a former Commissioner of Food and Drugs stated(38):

"Although the 0.05 confidence level has often been used in the scientific literature to determine whether a result is positive, there is no fixed convention for the matter."

And finally in discussing the procedures for testing statistical hypotheses Lehmann stated that(8):

"The choice of a level of significance, alpha, will usually be somewhat arbitrary since in most situations there is no precise limit to the probability of an error of the first kind that can be tolerated. It has become customary to choose for alpha one of a number of standard values such as .005, .01, or .05. There is some convenience in such standardization since it permits a reduction in certain tables needed for carrying out various tests. Otherwise, there appears to be no particular reason for selecting these values. In fact, when choosing a level of significance one should also consider the power that the test will achieve against various alternatives. If the power is too low one may wish to use much higher values of alpha than the customary ones, for example, .1 or .2."

On the other side of the issue, many scientists
explicitly classify the results of hypothesis testing
as positive or negative by using the conventional
values of p<.05 as the discriminant. Others are not so
explicit but show similar predilections by
identifying, for example, "significant" results
according to this rule.

In what is perhaps the most explicit case of this
type of practice it has been argued(38) that if
the convention is not followed, "...science is done a
disservice; any attempt to use the levels of
significance otherwise is an exercise in futility;"
and that decisions that are not based on the special
place that must be reserved for p<0.05 are:
"...subjective and arbitrary and have not been
established by practice."

Even though usually less explicit, this practice
is very common whether one considers bona fide, peer
reviewed, scientific reports or official public health
protection documents(39).   The scientific
literature is full of subtle practices of the genre.
Tables, for instance, contain entries which are
footnoted or otherwise marked to indicate significance
based on the "magic" p<.05(40-44).

It is perhaps futile to hope that scientists can
continuously focus on the true meaning of their
logical tools and that they will use them properly.
Their daily tasks are perhaps too complex to permit
constant fine-tuning of the decision rules of their
disciplines. Thus, they prefer to adopt conventions
based on what others have experienced and described or
on what appears to have served their disciplines well
in the past(37).

Such practices might or might not be sanctioned
by custom. Their chief effect, however, is that they
give the appearance of a law of nature to the p<0.05
and other such conventions. And what is more, they
give the scientific enterprise a cloak of objectivity
which is unecessary, unbecoming and, what is most
important, potentially damaging to the societal role
of science.

In the context of health and environmental
protection the choice of the value of alpha is a
critical issue of the decision making processe. It
involves much more than the protection of personal or
disciplinary reputations and there is continuous need
to reexamine it along with other similar conventions.
If hypothesis testing decisions reduce to making bets
about the future status of the the public health and

the environment, the public has the right to know the
true nature of the bet and to decide the weight it
will accord to rules apparently designed to protect
professional reputations and standings among
scientific peers.

It is noteworthy that in this broader context of
decision making, the conservative behavior of
scientists who use conventions as if they were rules
of nature constitutes a paradox. On the one hand, they
usually demand odds overwhelmingly against the null
hypothesis before concluding it is false, and
therefore behave as a risk-aversive group reluctant to
gamble personal and collective reputations. On the
other, many behave as risk-takers and are often
willing, collectively or individually, to loose major
health and environmental risks upon the public.
Concern about this troublesome aspect of the behavior
of scientists acting as technical advisers and
decision makers was expressed by the same former
Commissioner of Food and Drugs in the context of
regulating carcinogenic substances in the environment.
He stated(38):

> "In the same context we are especially troubled
> by the prospect of mistakenly declaring that the
> results of a study are negative, i.e., of
> mistakenly concluding that a study demonstrates
> safety. Such a decision, if incorrect, could
> result in widespread marketing of a carcinogen.
> A regulatory agency may therefore have less
> reason than the scientists to insist on a very
> high degree of cerainty before concluding that a
> study is positive, i.e., before rejecting the
> null-hypothesis. Similarly, there may be reason
> for a regulatory agency to require a greater
> stringency than other scientists require before
> concluding that a study is negative."

However, it is not unusual for some scientists to
raise the straw-man of a putative need for better
science(45). They do this to publicly impeach
policy decisions that they consider unmindful of
hypothesis-testing rules for whatever reasons. Yet,
to a large measure, these rules are designed to
protect the interests of specific scientific
disciplines and usually have little to do with the
regulatory function. It should be kept in mind that
the ends to be served by the disciplines and the
agencies, although both legitimate, are entirely
different. Their difference can be gauged by a simple

appraisal of the consequences of mistaken decisions in
the two cases.

A Recommendation

Thus, how to deal effectively and openly with
experimental uncertainty remains a problem of modern
industrial society if it is admitted that use of the
scientific method of inquiry has aspects and
consequences which are societally too important to be
the exclusive preserve of the scientific disciplines
to deal with in terms of unstated assumptions or
conventions.
     The expanded use of Metrics suggests itself as
the most rigorous means to overcome this problem which
arises in the most stressful form when converting data
to information. The excellent start made by the
chemists who established the discipline of
Chemometrics should now be followed by others who must
deal with the societal implications of our
continuously expanding ability to collect data.
Specifically, life science-related disciplines other
than chemistry should now focus on Metrics which
should become a generalized discipline defined in
analogy to Chemometrics as the "discipline that uses
mathematical and statistical methods (a) to design or
select optimal measurement procedures and experiments,
and (b) to provide maximum information by analyzing
collected data."

----

*/The views expressed herein are those of the author
only and do not represent official FDA positions.

Literature Cited

1. Frank, I.E.; Kowalski, B.R. Anal. Chem.
1982, 54(5), 233 R.
2. Braithwaite, R.B. "Scientific Explaination: A Study
of the Function of Theory, Probability and Law in
Science"; Cambridge University Press: 1964; pp. 9-21.
3. Russell, B. "The Scientific Outlook"; W.W. Norton &
Company, Inc.: New York, 1936; p. 57.
4. Dolby, G.R.: Biometrics 1982, 38, 1069.
5. Bray, J. Nature 1984, 308, 9.
6. Buck, C. Int. J. Epidemiol. 1975,
5(1), 97.
7. Buck, C. Int. J. Epidemiol. 1975,
4(3), 159.
8. Lehmann, E.L. "Testing Statistical Hypotheses";
John Wiley & Sons: New York, 1975; p. 61.

9. Kaiser, H. **Pure   Appl.   Chem**.   1973,
34(1),   35.

10.Brilluin, L. "Science and Information Theory"
2[d] ed.; Academic: New York, 1963; pp. 8 and 302.

11.Wilder, R.L. **Science** 1976, **156**, 605.

12.Reingold, M. **Soc**. **Stud**. **Sci**. 1980,
10, 475.

13.Armitage, P. **J**. **R**. **Statist**. **Soc**.
(**A**) 1983, **146**(4), 321.

14.Moss, T.; Lubin, R. in "Health Risk Analysis:
Proceedings of the 3[d] Life Science Symposium";
Richmond, C.R.; Walsh, P.J.; Coppenhaver E.D., Eds.;
The Franklin Institute Press: Philadelphia, 1981;
p.29.

15.Weinberg, A.M. **Minerva** 1972, **10**, 209.

16.Gori, G.B. **Chem**. **Eng**. **News** 1982,
60(36), 25.

17.Hattis, D.; Ashford, N.A.; Hollomon, J.E. **Chem**.
Eng. **News** 1982, **60**(50), 35.

18.Hunter, W.G. **Am**. **Stat**. 1983, **37**(4),
360.

19.Yelling, J. **Am**. **Stat**. 1983, **37**(4), 362.

20.Martin, J.A. **Am**. **Stat**. 1983, **37**(4),
367.

21.Whittemore, A. **Risk   Anal**. 1984, **3**(1),
23.

22.Russell, C. **Nature** 1984, **306**, 777.

23.Ackoff, R.L.; Gupta, S.K.; Minas, J.S. "Scientific
Method: Optimizing Applied Research Decisions"; John
Wiley and Sons, Inc.: New York and London, 1962; p.
295.

24.Haight, F. A.: "International Encyclopedia of
Statistics"; Kruskal, W.H.; Tanur, J.M., Eds.;
Macmillan and Free Press: London and New York, 1978;
p. 155.

25.Lehmann, A.J.; Fitzhugh, O.G. **Assoc**. **Food**
Drug Offic.**US** 1954, **XVII**(1), 33.

26."Risk Assessment in the Federal Government:
Managing the Process," National Academy of Science,
1984; p 9.

27.Salsburg, D.S. **J**. **Toxicol**. **Environ**.
Health 1977, **3**, 611.

28.Salsburg, D. **Fund**. **Appl**. **Toxicol**. 1983,
3, 63.

29.Haseman, J.K. **J**. **Toxicol**. **Environm**.
Health 1977, **3**, 633.

30."Oversight of Food Safety, 1983", Hearings,
Committee on Labor and Human Resources, U.S. Senate,
98th Congress, 1st Session; U.S.G.P.O.:
Washington, D.C. 1983; pp. 536-666.

31. Marshall, E. Science 1982, 218, 975.
32. Ashforth, N.A; Ryan, W.W.; Caldart, C.C.
Science 1983, 222, 894.
33. Currie, L.A. Anal. Chem. 1968, 40(3),
586.
34. Altshuler, B.; Pasternack, B. Health
Phys. 1963, 9, 293.
35. Dixon, W.J. Biometrics 1953, 9, 74.
36. Youden, W.J.: "Statistical Techniques for
Collaborative Tests"; Association of Official
Analytical Chemists: Washington, D.C., 1969; p. 30.
37. Sagan, C. The Washington Post
Parade Magazine; Sunday, October 30, 1983; p.
5.
38. Fed. Reg. 1980, 45(181), 61474; Section
III.A.(3).
39. Fed. Reg. 1983, 48(25), 5252.
40. National Toxicology Program: Fiscal Year 1984
Annual Plan: U.S Department of Health and Human
Services, Washington, D.C., pp. 99-103.
41. Zychlinski, L.; Montgomery, M. R.; Shamblin, P. B.;
Reasor, M. J. Fund. Appl. Toxicol. 1983,
3(3), 192.
42. Revis, N.W.; Major, T.C.; Horton, C.Y. J.
Environ. Pathol. Toxicol. 1980,
4(2,3), 293.
43. Rudman, D. "Insulins, Growth Hormone and
Recombinant DNA Technology"; Gueriguian, J.L., Ed.;
Raven Press: New York, 1981; p. 167.
44. Andrews. J.W.; Hames, C.G.; Metts, J.C.; Waters,
L.; Davis, J.M.; Carpenter, R. J. Environ.
Pathol. Toxicol. 1980, 4(2,3), 313.
45. Ref. 30, p. 246.

Panel Discussion
Advice to Analytical Chemists

Edited by DAVID A. KURTZ

A panel discussion of symposium speakers with
audience participation was held to discuss means and
methods of instituting chemometric (statistical,
computer) methods into general use in analytical
problems. Discussion centered around three topics:
analysts in their own work, analysts in the
educational process, and analysts in the political
and social scene.

A panel of all of the speakers in the symposium was set up at its
conclusion to discuss the subject, "Advice to Analytical Chemists"
for the utilization of chemometric techniques in analysis. The
problem as we saw it, is that we have to overcome this fear of
statistics and especially of this word chemometrics. We have to
learn what it means, how to understand it, and how to use it.
When the proposal to the ACS books committee was first made to
publish this symposium of papers, the comment was made by their
editorial board that the information was far ahead of its time.
Of course, the counter response of our people was that we are
behind the times in the use of these concepts. It is with this
idea in mind that we embarked on the panel discussion.
   This chapter discussion is not a verbatim listing in order of
comment. Quotes of speakers, panel members or not, are placed in
an order designed for clarity to the reader. There are three
major sections: 1. Analysts in their own work, 2. Analysts in the
educational process, and 3. Analysts in the political/social
scene.
   The first area covers laboratory analysis. Analysts can
express results from data in a number of ways...mean, standard
deviation, and range. The analysts can separate the error
involved in various phases of an analytical problem. They can
work on a single detection technique, which is univariate, or they
can utilize more than one measuring technique, which is

multivariate. The latter often yields much more information
relating the sample to the subject area than the extra work needed
for the accumulation of the additional data. Statistical
processes also allow data base handling, number crunching, and
discovery of the structure of data. All of these are covered in
this first section.

The second area discusses the educational process. The
statements were made as to how analysts can bring about or educate
the general society into a greater appreciation and use of such
methods. These can be done through educating at the analyst,
user, or manager level. It could be brought about at the
manufacturers level in the design and construction of analytical
instruments, the black boxes, that are used to get the data.
Finally, the need for legislative education is presented.

Questions relating the political/social area of our society
with analytical information are found in the third section. These
include an evaluation of risk in giving numerical results, the
interaction of the analyst and submitter of samples, and the
problem of receiving data from divergent sources. This section
also includes looking at different motives in the use of the data.

We hope by presenting this forum that ideas can flow from
readers to aid in a fuller understanding and use of statistical
methods. We hope we can use such methods in our laboratories and
spread to other parts of society a greater appreciation and
understanding of them.

Analysts in their Own Work

Describing Data. The use of the mean, standard deviation, and
range are all well known methods of describing data. Problems
arise in such descriptions, however, when individual values are
given as "less than." Dr. Currie warns: "the reporting of
experimental results simply as upper limits or 'less than numbers
will frequently result in a subsequent rounding off and expressing
as zero."

One person suggested that maybe the best thing in calculating
a mean is to exclude 'less than' figures. Dr. Mitchell concurs:
"That's correct. In a majority of cases the 'less than' values
are beyond the limit of determination and are statistically
supposedly invalid."

But when pushed by a requester who wants a number and an
average, Dr. Currie again responds,

"An average obviously can only be deduced by going back to
the original estimates. If one of the numbers in your data set is
'less than' a certain value and you round it to zero or some upper
limit, then your result will be biased. There are, however, some
robust statistical techniques that can be applied in these cases.

"For example, if you wish to estimate a mean for a series of
results, including a number of upper limits, you always have to
make an assumption. You could in certain cases use a median as an

estimate that would at least be less biased than trying some kind of average ... but the best is preserving all of the original information."

"Another way of presenting data involving 'less than' data points," said Dr. Kurtz, "is to present the mean as a range. The lower limit of the range is the value of the mean assuming all 'less than' data points are zero or as a defined point. The upper limit of the range is the value of the mean when all the 'less than' data points are at the maximum level within that uncertainty range. Thus the limits of the uncertainty range of the mean can be ascertained."

The Confidence Interval. The use of a confidence band around the data was of concern to the panel. Dr. Mitchell: "The confidence band is the first thing the analytical chemist should stick around the number he reports. That will not be a relative standard deviation or a linearity statement. The confidence band is the final output on data statements. It is something we should get in the habit of doing. The meaning of the band that I use is a band which encloses the 'true value' with a certain probability, typically 90%. They are computations concerned only with precision. You would need, in addition, some sort of accuracy statement on that as necessary. The confidence band is typically symmetrical: If the most likely concentration is 3 and the 90% confidence range is 1 to 7, or -3 to +8 as the case might be, the range encloses 90% of the values obtained. The users perhaps should get used to that type of information. People will also get a feel for the spectacular difference in data quality between the optimum range for your machine and right down the bottom."

In the description of a calibration graph Dr. Kurtz adds, "It may be helpful to give the confidence bandwidth at three different points, the maximum and minimum points of the calibration graph and the mean point (where the band is of minimum width). At the same time you may wish to give the number of calibration points or something like this."

And he continues, "The confidence bandwidth of a calibration problem does not adequately describe the solution of unknowns, but it could give information about the status of the analytical instrument, such as a gas chromatograph. It could tell how efficient it was running on a particular day. For the solution of unknowns, one needs, in addition to the confidence bandwidth, information about the response bandwidth which provides the more important portion of the error estimations in a calibration situation."

In reporting data, Dr. Kurtz suggests "Maybe we should report not the point estimate but rather just the band. We all know the point estimate is located about the middle."

Dr. Mitchell adds, "You could give a rather amusing example of that and that would be discussions of the 1986 budget deficit, which, of course, is the difference between two very large numbers each of which have a considerable amount of uncertainty on it.

You don't get that impression from the people we see on
television."

Pictures.   If experimental results are to be described by
multiple points, such as the mean, range, and distribution, Dr.
Mitchell suggested, "Your output will be pictures rather than
numbers."
        Pictures as graphs or other forms do convey more information
about a subject area that numbers alone can do. This was
suggested in numerous places of the discussion, but, of course,
few hard suggestions could be made in this area as aids to the
analytical chemist. Dr Stalling presented a strong argument for
pictures:
        "I would like to respond to one aspect of this that seems
very important and that is to try to present the pictorial image
of the information. Too many times a collection of samples what
goes out of our lab is presented as a tabular report. Included is
no histogram of frequency, no distribution as to function of
sample size; there is no real attempt to try to pictorially
represent what was done. I have tried to deal with this by using
the graphics presentation that would encompass the information.
Now, it gets to be a real pain if you have to do that by hand; it
takes as much time sometimes to put that together as it did to put
the rest of the report together. But I really think we would be
ahead of the game if we didn't leave the data in tables. If you
are dealing with a lay person or an administrator, the best
approach I have had is to try to pictorially present the
information. Then you can say "Here is the problem and here is
the distribution. It doesn't take 2% of the effort that it could
take to square with what is going to happen if he misunderstands
the problem. Television news does that sort of thing very well.
I don't care what you use, pie charts or whatever, but make some
attempt to convey a pictorial perspective of the information you
have gathered. And this may make men translate some information."
        The comment was then made from the audience that your
distribution may not be a common distribution about a point at
all. It may be limited to one side or the other. By using the
distribution pictorial, you can clearly show the true situation.

Univariate Verses Multivariate.   The problem of working in
univariate or multivariate environment was addressed in only one
comment even though a whole book could be written on this topic
alone. "Analysts should change their direction, wherever they
can, to work in a multivariate area." Dr. Stalling said, "And the
thing that impresses me so much about the chemometrics potential,
is the capability of using multivariate statistics. How many
problems can you define in the real world better in a univariate
way? Name me one!"

Design and the Measurement Process.   The final and concluding
advice on describing data is that numerical results should be
accompanied by statements telling about the method used in
obtaining the number.   Dr. Currie stressed the importance of
including an adequate number of blanks and reference samples, an
adequate experimental design and description of the measurement
process, and in continuing quality control.

In answer to this an unidentified analyst in the audience
comments: "Not only that, but in some instances you are doing
analysis in an indirect way.   You can't receive all impulses in
the measurement system, but you are seeing a group of molecules
that you are humping together... you are lumping that together and
the mixture of these can change results a little bit.   Whatever
the combination happened in that particular sample, you are not
endeavoring to find out what the real mixture is and make a
general statement of that type.   Knowing those kind of
experiments, knowing   the   method   and   reflecting   that
in   the   answer   is awfully important.   Only then, meaning is
given to the answer in relation to the situation where you are
very specific to a particular molecule or to a particular isomer,
whatever the case may be.   This is particularly important in trace
analysis, and that is one of the reasons why I've always harped on
the importance of describing or referring to the method ...
somebody later on could come long, take a look at the result, and
say 'I know where that result came from and I know what it means.'
If he just takes the number, he doesn't know unless he got some
sort of reference or method comment on technique as to what and
where the method came from."

Systematic Error in Laboratory Analysis.   Dr. Currie: "I would
like to raise a nasty issue that is pertinent to some of the
comments I am hearing with respect to complex measurements and
environmental systems and that is, it is rare for me to see much
reference to possible systematic error.   I just want to raise the
fact that the systematic error, as we heard yesterday, in
sampling, can be overwhelming."

Dr. Stalling adds to that: "I have participated in quite a
number of dioxin analyses for TCDD down to the parts per trillion
range, and I have been extremely impressed by the fact that at
15-20 ppt in fish residues that the agreement falls within 20-30%
of these values.   This may all come about with the fact that we
are using C-13 labels as an internal standard.   Some reflection on
this situation has lead me to think that perhaps what we need is
element 205, that doesn't occur in our universe, which we always
add as an internal standard to the sample at the beginning of the
analysis.   No matter what we do earlier, when we get to the final
output for the sample, we can compare the expected value for that
compound to see if it falls within a range that we expect for the
point.   We then can look at the measurement ratio and compare
that, too.   Now that doesn't mean it goes through the same

procedure. There are a lot of problems with that. The only way I
have seen that we have been able to manage at all from a
calibration viewpoint on this dioxin problem has been through the
use of the internal standard."

The measurement of uncertainties is always a special problem
and is usually indirectly determined. One direct measurement was
done in Dr. Currie's laboratory. As he explained, "We recently
ran through some interesting exercises where we created data sets
to simulate the behavior of pollutants in the environment. In
generating the data matrix, we included elemental and isotopic
compositions, as well as stochastic behavior. In any case we knew
the truth as well as the measures of precision. (We had three
different kinds of imprecision working here.) Now, there were
brave souls out there who operated on these data knowing the truth
was known to somebody. They presented us with their results and,
more importantly, their estimates of uncertainty. We have learned
a tremendous amount from this, as have the participants involved,
in using alternate ways of operating on the data and discovering
the problems that are needed to be solved in each case. While I
can't go into detail here, both those who were using factor
analysis techniques and those using least squares methods
identified elements of scientific judgment, such as how many and
what type of source components are present in a given sample. I
just wanted to state that this kind of exercise speaks to the
question of real uncertainties in chemometric operations and I
think it can help determine them."

Number Crunching.   We poke into a general discussion and hear "I
gather data over periods of time. Later, I do my analysis, not
with one sample, but with a dozen or a hundred samples. Then I
put it all together. What you might think of in your
interpretation may not be the interpretation you want to rest on a
year from now. When you go back and put it all together, you may
want to go back to the raw data."

"To respond to that point," says Dr. Stalling, "I think there
is a terrible loss of time and efficiency by the fact that
information isn't projected into the data that has already been
gathered as the system goes long. We have used that approach as
sort of a feedback loop. Every time a new block of data of 40
samples or so comes in, it gets looked at. It may be that you
have 3 samples in there that don't even relate to the rest. We
end up with outliers, and this is a real problem. There are no
possible means of going back out into the field or doing anything
in retrospect to understand why we had that outlier.

"The first step in all this is the capacity to transfer
measured data," he continued. "That's the first hurdle that I see
you get over for any reasonable number of samples. I certainly
feel that 50 samples is a reasonable size of information. It
could be 1000 or it could be less. The cost to do these kind of
things is a cost more in time than money. Time wise, I think it's

impossible to do the data digestion and manipulation in less than
10% of the time...20% is a more reasonable guess. You should see
some of the gymnastics that I do routinely. The data base that we
have lives on a PDP-11 34. It has a capacity for a magnetic tape,
9 track output. I can easily go in and in 5 minutes dump 100
samples to the tape. Federal express will get that to Urbana, IL
overnight where the CDC-Cyber system is. It takes 2 days, unless
it is real urgent, if I am lucky to get the consultant in the CDC
center to dump the data file in my directory so I can get at it.
Now, the short cut to this is to call Ma Bell up with a little
RS-232 floppy disk and run over to the terminal and yank out the
plug and stick it in the back of the floppy disk, capture the
data, and hope that the operator doesn't come in real quick while
I have his terminal unplugged. I then turn right around, call the
data system up, and transfer it over the line. You can transfer
10K of data in 15 minutes. You lose 2 to 3 lines into the editor,
decode, and take another 15 minutes to straighten it up and get it
back in. In the second way I can get 50 samples into the machine
in under 30 minutes. That data transporting problem is a real
nasty mess. Now, in the lab we have, we go though one PDP-11
integrator, the PDP data base, the modem or data tape output, and
the CDC-Cyber for analysis. The data comes back to my micro so I
can pick up the data values from it. I may have to go home and
put it up on my micro to get a three dimensional plot if I want to
do that. So, right now it is a difficult thing to put these
operations in place. It needn't be that complicated, but I just
try to respond as to how to manage the information. It is
possible to do the measurement, it is possible to quantify it, and
it is possible to process the data with very little more effort
than what you are currently doing."

"I think you have identified the general core of the transfer
information problem," says an audience respondent. "At the moment
there are not adequate conventions for data communication and
transportability ... in terms of information processing standards
and software protocols for hardware devices. These have to be
agreed upon in order to allow hierarchal and distributed type
databases to exist through exchange of data. I think that the
answer to that is that the analytical chemist and those people who
are interested in that sort of data exchange will have to become
extremely active within the voluntary consensus standards process
to reflect the needs of the user. I think that that's one of the
deficiencies in why these standards don't exist because the users
have not been represented, and, more, the vendors and suppliers of
equipment have been involved in setting these standards. The
various professional disciplines are going to have to participate
if in fact they are going to be responsive to the need of the
user. Otherwise, who is going to know who those users are."

Database Questions.   Because of his heavy involvement in
databases, Dr. Stalling now asks questions about the form of data:

"How readily is the information that is generated by the
analyst in machine-readable format? Better, what is the state of
affairs in terms of being able to get data from one system to
another without having to type or write it, or physically do
things with the information? Instrument companies, it seems to
me, are completely ignoring the information management aspects of
analytical chemistry. If I am wrong in that perception, I would
like to know about it. I think that it seems to me to be the
major impediment in doing something more with the information.
There is data out there, and we shuffle information back and
forth, and reports, and secretaries are kept busy, etc. How many
iterations does one have to go through before you get data into
the final database. And then, are people even using databases for
decisions?"

From the audience in a different regard, "Different analysts
are producing different data sets which are pooled in standard
databases. For example, the the database of constituents and
nutrients in food is one the Department of Agriculture maintains.
Different people submit various analytical levels of these
quantities along with various pollutants like pesticide residues
and that sort of thing. The question is how can these databases
identify method bias which they do not now do? How would you
identify the data elements that should exist in pooled data bases
that would allow bias adjustment to be made?" You readers must
search for that answer.

Structure of Data.   There is another factor in the analysis of
materials that provide complicated results, materials containing
hundreds of compounds, for example. The data has a structure that
is very hard to describe in a simple way. Dr. Stalling puts it
this way:
"The chemometricians' tools and resolution power has a
tremendous role to play in expanding and clarifying the
information that we get. There is another side to the question
and that is that it is critical for the analytical chemist to
involve himself in helping understand the structure of the data.
If I could try to define what I mean by the structure of the data,
I might say this: 'In the contamination of fish, it is not widely
recognized that the contamination has a size-concentration
dependency. The concentration of PCB, a persistent and
complicated group of compounds, in a fish certainly has a size
factor in it. As a random sample, it is important to place it in
the context for which the range of values fall, i. e., the mean
and distribution. If we put that in place of a population
estimate, we see that there is a factor involved in those
measurements.'"

Stalling goes on to say: "In regards to PCB analyses there is
no problem in showing the structure of 209 possible isomers in a
sample if there is minimal interference from other components. We
can at least model this data structure. That's clear. What

samples have similar composition and concentration dependency can
be very clearly described. No problem in saying that this group
of samples has some spatial relationship if they are related to a
process. That you can find very quickly. The problem in taking
that piece of data obtained from the environment and coming up
with how it relates to a manufactured Aroclor mixture, however,
can only be done, as far as I know, through this PLS block
method."

Analysts in the Educational Process

Towards The Education of Analysts.   Dr. Stalling expressed the
view, "I think there are two critical factors which are the error
in sampling estimates verses repeatability estimates. Those are
two different universes as far as I can see. Without preserving
that information in the final report that says the analytical
uncertainty is such and such, a good deal of information is lost."

Dr. Stalling went on to strongly state that the role of the
internal standard has been very important in reducing the range of
agreement to 20-30% for state-of-the-art residue analyses of
dioxin in the environment.

Dr. Stalling goes on to say, "'A picture is worth a thousand
words.' I would make a real plea to present as much information
as possible in graphic format. I have come to the conclusion that
understanding comes only by seeing and even less in hearing. A
mental image of the discussion brings it much clearer in the case
of discussing the environmental facts. When you begin to see
variations in the data, over several graphical transformations,
the analyst will put up the flag in 99% of the cases. He sees
some sort of bias or another component. In the case of our PCB
modelling work we used 70 component matrix in which we summed each
of the 4,5, and 6 chlorine-containing compounds into a linear
matrix for each sample for our model. Looking at that matrix in a
particular analysis case, we found there were four components,
four major peaks, there that were not PCB's that fell completely
outside of the distribution of the rest of the PCB's. In some
cases that was 20% of the sum. Now, you know, there is the
beginning of a significantly wrong model if we proceed without
getting a better grasp of the data. Well, I want to express the
thought that we must find means and tools that are convenient to
express the information in complex analyses. And, personally, I
am an overzealous advocate of the tools I have found. I don't
care what you use as long as you can defend the picture and it
conveys the information."

Education of the Manufacturer.   Raw data needs to be preserved,
and, as one said, "Sometimes I emphasize to the instrument makers
that today's microprocessors cannot be depended upon to stashing
away raw data; they can't archive it very well, there, on the
instrument. Besides you are stuck with the instrument that the
makers supply us...we should have better control."

The preservation of raw data saved the day (or saved a week or a month) when an error in software was discovered by Dr. Stalling: "We were about 3 man-months into this block of sample data; some 250 to 300 samples had been analyzed through a routine procedure, calibration. I had begun to do this pattern recognition work by looking at the data with SIMCA. Suddenly I realized that there was a bug in the data base. Those numbers were wrong. The coefficients that were being extracted from the slope to the line, related to the last 3 samples, had reasonable slopes, looked pretty good, and were on good floppy disks. But they were wrong. The archival factor we built into the data base said to preserve the original and every piece of data as well as the calculation algorithm. Well, after about a week they had corrected the bug in the program, and we were able to go back after 4-5 days to reprocess all the data. Just anticipate going to the director and saying you know we just lost 3 man-months when you are already 9 days behind schedule."

Other problems with instruments include "many of the newer ones convert negative numbers to zeros", "you are losing information when you use the instruments", and "the necessity to get the raw material out of the instrument into your own calculating device is very important." "The instrument manufacturer has his own biases, data biases. If the instrument interprets and if you read the interpretation off the instrument, you are stuck with the manufacturer's algorithms and attitude."

And then there is the charge by Dr. Kurtz, "In a different way, instead of a point estimate, can an instrument handle a range estimate?"

From the general audience came this comment: "Instruments today are not too compatible to transfer information. There is the RS-232 port with its 8-bit code and all sorts of hand shaking and lines to be hooked up properly. It is just difficult to set up." And as we have already said in this panel discussion, "There are not adequate conventions for data communication and transportability. I think that the answer to that is that the analytical chemist and these people who are interested in that sort of data exchange will have to become extremely active in this area to reflect the needs of the user."

Often lost in the discussions are the enabling methods needed for bringing new methods into common use. Dr. Kowalski offered an unusual but effective suggestion: "I believe the first use is by education of the instrument companys' directors and not the general public. By placing these techniques that process the actual measurements before them, they will make better products. Those products will be able to do things that other products won't, and they will sell better. Then, also, we should go along and say that we should provide the uncertainty of these estimates. All right, but the acceptance of these estimates takes the education of the user you get. Now, it's difficult to educate all

of the people that use analytical data; it is very, very difficult
to do that. It is a very slow process. Do you want to know how
it is going to be done? It will be done, it may take 50 years.
The first step is going to be lot easier; there are far fewer
instrument manufacturers than end users of instruments."

Education of Managers.   The insight of Dr. Stalling is seen:
"...We make no effort to understand the continuum of information
until we give it to the administrator or until we give it to the
guy whose job it is to prepare a model independent from me (the
analyst) when the questions about the importance of the data are
asked. And I see the chemometrics tool as a bridge between the
people who are having to make the information judgments and
providing a mechanism of understanding the data."

        "In about 2-3 management steps the number no longer is 15
+/-5, but simply 15. They tend to drop reference to the spread.
Perhaps it is only a matter of education. If we could bring about
a change so that a number has any meaning at all only when the
uncertainties are attached to it..."

        Dr. Currie capped this discussion: "Put the uncertainty
statement first so the other part may be forgotten."

Education of the Legislators and Regulators.   From one
participant: "The analytical chemist is asked to make a
measurement level of a compound in the environment. He provides
that to another professional, such as an industrial toxicologist,
who then has to interpret that result, make certain policy
decisions on it, and explain it to the public. And yet the
implications of the initial question are that the analytical
chemist has a contribution to make if he is aware of that initial
question. I am impressed in the manner in which the dialogue is
carried out in the public ... in an atmosphere of total informa-
tional blackout. The implications of the data that the analytical
chemist really possesses is not brought into the dialogue."

        Dr. Mitchell: "That is often the truth and I guess it is a
political question. I ought to comment on it. We've done studies
in our place which couldn't, in effect, possibly provide useful
information, and it didn't. So we should challenge these results
to be truthful."

        He continues: "We generate our numbers all right. We feed
our numbers to people who are trying to relate them to health
effects. So, no matter how bad it is what we have, they are in a
much worse situation, and most of the time they don't know quite
what they want...

        "There are a couple of motives there. The persons making
that decision may well have different motives from yours. You
must stick that confidence statement in because someone will then
proceed and make a decision. At some later period he may well
discover that the decision was stupid, and they are going to look
for someone to blame. I have some real beauties."

Another commentator adds, "Lots of time the decision-making process is really clear ... it is not the person you talk to but two managers above or a fellow in the legislature who starts quoting data. They believe the scientist is some sort of manageable block and all of a sudden your number is going to be God's word. You are going to have to be protected in what you report."

A different aspect was also brought out: "...All your data has to go someplace, though. When you leave your spot, and a similar professional has to go back to your data and look at it, he has to find all the information. "That part has to be retained somewhere."

Analysts in the Political/Social Scene

Interaction of the Analyst with the Submitter of Samples.   Dr. Mitchell continues: "The analytical chemist cannot operate in a vacuum. The requester has to convey what he intends to use that information for and that goes into public policy sorts of things. I think that begs the question as to what is the role and what sorts of things should the analytical chemist and chemometrician do to convey knowledge or intelligence to the generally unsophisticated public or the nontechnological public, if you will."

A comment from the audience set the stage for the need to interact with the submitter of samples. He said, "Your report should also reflect somewhat the experimental design from which your data was obtained, for example, by a certain instrumental method. It should be reflected in it because without that information it is quite often misleading information for someone who is going to use it but doesn't know how it was obtained. If I don't know how it was obtained, I don't know if it is of any importance. Even though I was given the experimental range, I don't know the meaning of it without having an idea from where it came."

"There isn't enough involvement of the analytical chemist with the people providing the sample, those who are designing a solution or who want the answer to a particular question. We are often answering a different question from what is to be answered. We get the sample, we analyze it carefully, and we put error limits, probability, and confidence interval on it. That confidence interval may not be the confidence interval on the population you are looking at ... the patient, for example, or the environment. What we need to do, I think, is to get chemometrics to be more involved in the design of the proper question and in formulating how we can best answer that. Quite often I don't think our analytical method needs to be as precise perhaps, or as expensive as we have used. We can often use a cheaper, less expensive, and maybe less precise method and still answer the question that you need to answer."

"Yes, they are handing you a bucket and saying, 'Tell me what's in this.' 'I'll tell you what's in it when you tell me what your going to do with it.'" ... Dr. Mitchell.

And put a little less bluntly was the comment: "In clinical chemistry one requests a test to be done and a result to come back. Yet, the interpretation of that result and the confidence in it depends an awful lot on the goals and objectives of the requester of the test. The proposal has been made somehow we should be exchanging goals and objectives between the requester and performer."

Dr. Kowalski put the problem in another way. He said, "If I need to be 19 out of 20 right or 99 out of 100 right, I should give my answer either as a 'yes' or 'I can't say'. The tighter you put the confidence, the more I am going to say 'I can't say'."

Another way of phrasing that, is, "If you are going to have surgery and you want to have such or such a percentage of surviving, it may be that you have a tradeoff if you don't do the surgery. How do you give the patient a picture of what the tradeoffs are?"

Interaction of Analyst with Other User. A number of statements have already been made on this topic. One aspect yet to be addressed is the desire for results to be expressed in a single number. "I think it is a serious problem. I have lots of people who say they want an average; they don't want three numbers nor ten numbers. They want one number they call average."

To that it was brought out that "we have been attached to this idea for so long, we can't think of any other way of expressing a result. Yet, we are now opening our minds to realize that it is a distribution. In my field there is a reluctance to do this, that somehow if you have a large variance, your data is bad. I see people running around saying that they want to express their result as a standard error or a standard deviation because, if it is larger or smaller, it would look better."

Significant digits are often forgotten. "A lot of data we generate gets passed out politically. If we have a tendency to provide too much data, there is an urge for people to use 3 or 4 significant figures from experiments that really should not allow you more than one or two. It can get back to you at a later time in an embarrassing way as if you were originally accurate to that place."

Where Data is Used. Data is used in such a wide variety of places that there may be difficulty in even putting data in a single form. These comments speak to that point:

"One breakdown in terms of goals and objectives that has been brought out is that one uses test result information for diagnostic purposes to determine the problem, another for developing a prognosis of the eventual outcome, and a third for management of a course of therapy. For example, if you are

interested in management of a course of therapy, you may be
interested in knowing if a patient is taking his medicine, and if
he is taking his medicine, the results may be fairly gross, either
up or down. In this case the analyst really doesn't have to go
through large contortions to deal with very exact measurements.
So that, convey that kind of information that allows the analyst
to say that the particular result in hand can indicate that the
patient is or is not in the therapeutic range without having to go
through agreements. That's really what the requester is
interested in. In diagnosis it may be much more critical because
you may be combining a number of things, propagation of error and
this sort of thing, and may be very critical as to the outcome."

In a completely different area, the setting of minimum
detectable levels, the politically expedient method may be counter
productive to the expansion of knowledge. Dr. Kurtz related an
example of this: "We had a local manufacturer dribble amounts of
kepone and mirex in the local streams from waste disposal on their
site. At one point a few wells of homeowners along the stream
were found to contain kepone at levels in the low parts per
trillion. Later in annual state testing programs these wells were
found not to contain any kepone. Was there a lowering of
contamination? What had probably happened was that the minimum
detectable level was set above most of these early levels, and the
later samples were less than this minimum level. It took on a
suspicious air that the state had purposely set this level to
avoid future controversy in this area."

There are always these problems of getting socially involved.
"How can not the analytical chemist get involved since he knows so
much about the sample that others simply do not know. Whether a
person puts a charcoal filter on his water supply or not puts an
entirely different quality in the understanding of the problem.
In clinical chemistry every method is different. Even the
instrument used has certain values for controls such that
different values result from the analytical process."

Then, Dr. Mitchell hits on another problem: "There is a
practical problem for a commercial lab. This is that you wouldn't
want to be the first boy in the block to stand up and admit the
[poor] quality of your data because no one else is doing that, and
you are going to look a lot worse if you are honest. So, I guess
that is a real psychological block for developing data. The
transition period is going to be rough."

Another person brought out the problems where the community
user is separate and distinct from the analyst. "Again, you are
getting back to the different levels of interpretation that the
analyst who does the measurement, the intermediate, and the top
level user community all have different forms of the answer. Each
one has to be placed in their frame of reference. That is a very
difficult problem. The state department director of health, for
example, is interested in sort of simplified or bulletized
information, if you will."

The whole issue of needed data, as Dr. Currie pointed out, was addressed in a recently published report by the American Chemical Society. "Improving the Reliability and Acceptability of Analytical Chemical Data Used for Public Purposes," published by the Joint Board/Council Committee on Science of the ACS in May 1982 (see also C&EN 1982, 60 ,(23) 44. It tried to relate chemical analytical measurements to socio-political and regulatory needs.

Assessing Risk.   There is always "the risk of being wrong.   If you are wrong, and you proclaim it, people are going to do something that probably costs money."
    "So, I think it is interesting in that the analyst cannot really operate in a vacuum; the requester of the information for the eventual consumer user of that information has to somehow convey what he intends to use that information for. That goes into public policy sorts of things to sort of pick up on the theme of Dr. Zervos. I think that begs the question as to what is the role and what sorts of things should the analytical chemist and chemometrician do to convey knowledge or intelligence to the generally unsophisticated public or the nontechnological public, if you will."
    "That is a general problem that we all follow ... relative risk," concludes Dr. Mitchell.   "... a massive problem in the environmental field. I think that if you present a particular result, you have to sort of prove that context. I think people have to understand that there isn't just one side. You can't have your cake and eat it, too."

Listing of Data Sets

Appendix A

Fenvalerate "Unknown" Data Sets (Electron Capture Data):

Dataset A.   This data set represents good data but only in
duplicate at each level.

Response	Amount,ng	Response	Amount,ng
1.15	0.05	1.07	0.05
6.30	0.25	5.64	0.25
29.9	1.00	27.5	1.00
197.	5.0	198.	5.0
881.	20.	884.	20.

Dataset B.   This data set represents good data taken over a two
day period with environmental samples and standards interspersed.
Compare with Dataset C.

Response	Amount,ng	Response	Amount,ng
1.13	0.05	212.	5.0
6.55	0.25	922.	20.
29.7	1.00	1.20	0.05
211.	5.	6.37	0.25
929.	20.	29.5	1.00
1.23	0.05	213.	5.0
7.98	0.25	928.	20.
30.0	1.00	1.12	0.05
204.	5.0	7.96	0.25
905.	20.	29.1	1.00
1.22	0.05	205.	5.0
6.54	0.05	919.	20.
30.1	1.00		

Fenvalerate "Unknown" Data Sets - Continued

Dataset C.   This data set is illustrative of data obtained with
an equipment malfunction such that the response on one day was
significantly different from the earlier day. Day one data
regresses to a line parallel to day two data. Compare with
Dataset B.

Response	Amount,ng	Response	Amount,ng
1.43	0.05	1.79	0.05
7.25	0.25	7.96	0.25
35.2	1.00	36.4	1.00
225.	5.0	242.	5.0
986.	20.	1051.	20.
1.00	0.05	1.97	0.05
5.57	0.25	8.12	0.25
27.5	1.00	37.2	1.00
192.	5.0	247.	5.0
914.	20.	1072.	20.

Dataset D.   In this data set the compound sought had super-
imposed near it an artifact peak such that at at lower levels
the response values were significantly inflated. Compare with
Dataset E.

Response	Amount,ng	Response	Amount,ng
2.34	0.05	1.71	0.05
6.39	0.25	7.36	0.25
21.6	1.00	25.0	1.00
146.	5.0	145.	5.0
692.	20.	687.	20.
1.26	0.05	0.63	0.05
4.76	0.25	4.41	0.25
20.0	1.00	21.9	1.00
146.	5.0	146.	5.0
694.	20.	704.	20.

Fenvalerate "Unknown" Data Sets - Continued

Dataset E.   This data set is the same as Dataset D except that
 artifact peak data was removed.   Compare with Dataset D.

Response	Amount,ng	Response	Amount,ng
-	-	-	-
-	-	-	-
21.6	1.00	25.0	1.00
146.	5.0	145.	5.0
692.	20.	687.	20.
-	-	0.63	0.05
4.76	0.25	4.41	0.25
20.0	1.00	21.9	1.00
146.	5.0	146.	5.0
694.	20.	704.	20.

Dataset F.   This short range set of data is an example of
reasonably good routine data with minimal sample size.

Response	Amount,ng	Response	Amount,ng
2.17	0.05	1.99	0.05
9.26	0.25	9.67	0.25
36.9	1.00	37.4	1.00
212.	5.0	221.	5.00

Fenvalerate Data Set (Electron Capture Data):

Response	Amount,ng	Response	Amount,ng
0.95	0.05	1.03	0.05
6.11	0.25	5.87	0.25
24.5	1.00	24.3	1.00
122.	5.0	127.	5.0
753.	25.	755.	25.
2970.	100.	3106.	100.
1.15	0.05	0.97	0.05
6.21	0.25	6.30	0.25
24.7	1.00	23.7	1.00
124.	5.0	129.	5.0
755.	25.	745.	25.0
3017.	100.	3077.	100.
1.05	0.05	1.02	0.05
6.17	0.25	5.82	0.25
24.3	1.0	24.1	1.0
125.	5.0	123.	5.0
732.	25.	720.	25.
2935.	100.	3028.	100.

Chlorothalonil Data Set (Electron Capture Data):

Response	Amount,ng	Response	Amount,ng
2.56	0.02	269.	2.0
15.6	0.10	977.	8.0
56.8	0.40	2143.	20.
265.	2.0	2.40	0.02
949.	8.0	12.2	0.10
2113.	20.	54.6	0.40
2.42	0.02	266.	2.0
15.4	0.10	955.	8.0
57.5	0.40	2181.	20.
261.	2.0	2.68	0.02
972.	8.0	13.2	0.10
2149.	20.	56.2	0.40
2.49	0.02	274.	2.0
12.1	0.10	988.	8.0
53.6	0.40	2277.	20.

Chlorpyrifos Data Set (Flame Photometric Data):

Phosphorus mode with no electronic filtering.

Response	Amount,ng	Response	Amount,ng
1.06	1	0.91	1
4.89	5	5.14	5
19.5	20	19.6	20
98.9	100	98.1	100
451.	500	441.	500
1.06	1	1.02	1
4.85	5	5.01	5
19.5	20	19.7	20
97.9	100	95.9	100
437.	500	430.	500
1.02	1	1.09	1
4.98	5	5.07	5
19.4	20	20.3	20
97.3	100	99.1	100
438.	500	434.	500

Kepone Data Set (Electron Capture Data):

Response	Amount,ng	Response	Amount,ng
0.73	0.01	24.9	0.375
2.59	0.05	61.8	1.0
12.1	0.20	0.79	0.01
24.0	0.375	2.86	0.05
60.0	1.0	12.5	0.20
0.82	0.01	25.5	0.375
2.90	0.05	60.6	1.0
11.5	0.20		

For additional raw data on Arochlor PCB analyses of transformer oils, see Chapter 12 in this book.

RECEIVED March 25, 1985

Author Index

Subject Index

277

Author Index

Subject Index

277