STATISTICAL ANALYSIS OF WATER QUALITY DATA

AFFECTED BY LIMITS OF DETECTION

Submitted by

Paul Steven Porter

Department of Agricultural and Chemical Engineering

In partial fulfillment of the requirements

for the Degree of Doctor of Philosophy

Colorado State University

Fall 1986

COLORADO STATE UNIVERSITY

September 24, 1986

WE HEREBY RECOMMEND THAT THE THESIS PREPARED UNDER OUR SUPERVISION BY PAUL STEVEN PORTER ENTITLED:

STATISTICAL ANALYSIS OF WATER QUALITY DATA

AFFECTED BY LIMITS OF DETECTION

BE ACCEPTED AS FULFILLING IN PART REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY.

Committee on Graduate Work

Harry F. Bell

Jim C. Loftis

R. K. Hogeboom

Robert C. Ward

Adviser

Vincent J. Murphy

Department Head

ABSTRACT

STATISTICAL ANALYSIS OF WATER QUALITY DATA

AFFECTED BY LIMITS OF DETECTION

Many water quality problems are related to substances which are present at concentrations too low to be measured precisely. Obtaining information from a monitoring system which produces many results near the fringes of analytical capabilities is not straightforward. This thesis is a discussion of the concerns one should have when statistically analyzing water quality data from such a system. Two general approaches are discussed. The traditional approach is to regard all measurements as precise or imprecise. Precise results are simply numerical responses, for which statistical analysis may lead to valid and sound monitoring information. Imprecise results are reported as "ND", or not detected, with criteria for reporting based on categories of measurement precision.

Measurement error which leads to censoring is described. The impact of this error on the statistical characteristics of water quality data is illustrated using a model appropriate for analyte concentrations near the limit of detection. It is shown that the statistical properties of a set of measurements may not resemble the population from which samples were taken. This suggests the use of statistical methods which acknowledge observation error.

Loss of information due to censoring is demonstrated and it is proposed that a numerical result be reported for all measurements. It is also suggested that an estimate of data precision accompany all results. This would permit the data user to censor at levels of uncertainty chosen by the user, rather than having information censored by the measurement process.

When the number of results with significant observation error is small, or when data has been censored and no information is available regarding such error is available, statistical methods intended for censored data may appropriately be used. Such methods covering a variety of water quality problems are reviewed. Numerical examples of many methods are provided.

Paul Steven Porter
Department of Agricultural and Chemical Engineering
Colorado State University
Fort Collins, CO. 80523
Fall 1986

ACKNOWLEDGEMENTS

The author greatly appreciates the contributions of many of the faculty at Colorado State University. In particular, I wish to thank Dr. Robert Ward for the many hours he spent in consultation, planning, and review. This thesis is, to a large extent, a product of the sincere interest and effort he has taken.

I wish to thank Dr. Harry Bell for his practical insights and thought provoking questions. His concern about water quality problems forms the basis for this project.

I thank Dr. Jim Loftis for his help during the last several years, and for introducing me to this subject through many thoughtful discussions.

Thanks to Dr. Rod Skogerboe for the many conversations we had. As one who has produced a great deal of knowledge and information about this subject, his advice was invaluable.

I wish also to thank the Colorado State University Statistics Department, and Dr. Hari Iyer in particular, for introducing statistics to engineers.

The generous support of IBM Corporation for funding of this project is gratefully acknowledged.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

KEYWORDS AND SYMBOLS

LOD	limit of detection
LOQ	limit of quantification
LOG	concentration equal to the limit of guaranty of purity
RE	relative efficiency; for estimator Q relative to W, given by variance (W)/ variance (Q)
ARE	asymptotic relative efficiency of an estimator, equal to relative efficiency as sample size increases to infinity
MLE	maximum likelihood estimate
POE	propagation of errors estimate based on first order Taylor series approximation
FC	fraction censored, usually fraction below the LOD
FBQ	fraction below limit of quantification
CM	Cramer-von Mises test
CS	Chi-squared test
KS	Kolmogorov-Smirnov test
analyte	chemical substance whose concentration is sought
Cauchy	distribution with PDF is given by $f(x) = \alpha\pi\{1+[(x-a)/\alpha]^2\}^{-1}$
N(a,b)	normal random variable with mean a and variance b
T	random variable distributed as Student's T
χ^2	random variable distributed as chi-squared
Δ	non-central T random variable
$t(1-\alpha,f)$	value of t statistic when tail area $1-\alpha$ and f degrees of freedom
θ	parameter set

α	probability of type I error, size of critical area, size of hypothesis test, or "significance level" of test
$1-\alpha$	confidence level of test or size of acceptance region
β	probability of type II error
$1-\beta$	power of hypothesis test versus alternative
IUPAC	International Union of Pure and Applied Chemists
SIDF	method based on Satterthwaite's improved degrees of freedom
S	analytical signal
$\bar{S}$	average of analytical signals
$\underline{S}$	matrix of analytical signals
$S_{i,c}$	analytical signal produced by a standard (calibration) sample
S_p	analytical signal produced by a sample of a water quality random variable
S_{LOD}	analytical signal when analyte concentration is equal to the detection limit
S_{LOQ}	analytical signal when analyte concentration is equal to the limit of quantification
e_s	random signal error
k_o	coefficient of additive signal noise
k_1	coefficient of transitive signal noise
R	sensitivity of analytical signal = coefficient of independent noise/coefficient of dependent noise
X	analyte concentration
$\underline{X}$	matrix of analyte concentrations
X_p	concentration of sample taken from water quality population
X_m	measured concentration
X_B	measured concentration of analytical blank

$X_{i,c}$	concentration of calibration standard
X_{LOD}	concentration equal to the limit of detection
X_{LOQ}	concentration equal to the limit of quantification
X_{LOG}	concentration equal to the limit of guaranty of purity
$\bar{X}$	average concentration
ε	random measurement error present in concentration estimate
B_o	intercept of signal-concentration relationship
b_o	estimate of B_o
B_1	slope of signal-concentration relationship
b_1	estimate of B_1
$\sigma^2(\cdot)$	variance of $(\cdot)$
$\hat{\sigma}^2(\cdot)$	estimate of variance of $(\cdot)$
σ_p^2	variance of water quality population, X_p
β_p^2	variance of water quality population for X_p lognormally distributed
σ_m^2	variance of water quality measurements, X_m
σ_B^2	variance of standard blank
σ_{RSD}	relative standard deviation
$m_\cdot$	mean of $\cdot$
μ_p	population mean
α_p	population mean for X_p lognormally distributed
η_p	population coefficient of variation for X_p lognormally distributed
$f(\cdot)$	probability density function (PDF) of the random variable $(\cdot)$
$f(x/y)$	conditional probability density function (PDF) of the random variable X given Y

$F(\cdot)$	cumulative density function (CDF) of the random variable $(\cdot)$
MS	monitoring sensitivity equal to the ratio of population variance to the variance of observation error
Z	sensitivity of calibration slope = $B_1/\sigma^2(b_1)$
V	sum of squares for OLS = $\Sigma(X_{i,c}-\bar{X}_c)^2$
D	sum of squares for WLS = $\Sigma w_i \cdot \Sigma w_i X_{i,c}^2-(\Sigma w_i X_i)^2$
κ	number of replicate samples for each unknown
l	number of replicate measurements for each sample

I. INTRODUCTION

Water quality management in the United States, after receiving a major legal redirection in 1972 (PL 92-500), is evolving into an ongoing and routine management effort. Earlier, the emphasis seemed to be on defining problems and solving them as opposed to the type of continuous control mandated by this legislation. This evolution in management has resulted in a re-evaluation of information needs.

Water quality monitoring is managements' major source of information regarding progress toward national water quality goals, and perhaps its weakest link. Past inadequacies of monitoring for water quality management have been described (Council on Environment Quality, 1980; National Academy of Sciences,1977; General Accounting Office, 1981) and discussed in detail (U.S. House of Representatives, 1983). Simple questions such as, "is the environment cleaner than it was 15 years ago?", are, in fact, difficult to answer. A reason often cited for this is a lack of monitoring information. Witnesses at 1983 Congressional hearings on environmental monitoring have testified (U.S. House of Representatives, 1983):

- "Unquestionably there is a most serious and pervading need for knowledge . . . "
- ". . . current monitoring does not adequately serve the important purposes of evaluating the progress of national environmental programs."

- ". . . we don`t really know whether we are spending this ($50 billion per year) wisely..."

Statistical analysis of water quality data is an important aspect of water resource management. This analysis depends in part on the integrity of water quality measurements. An active quality assurance effort can often guarantee that such measurements reflect a true state of nature and not some property of the sampling and analysis system. However, many of today's water quality problems are associated with levels of chemicals which are too low to be measured precisely by a single analysis.

Advances in analytical technology can extend the range of precise measurement, but problems associated with data analysis will continue to arise. This is because safe levels for many substances are not known, or are believed to be zero. In addition, there is a desire for early detection of problems in both a temporal and spatial sense. The need for more information will continue to result in the monitoring of water with arbitrarily low analyte concentrations.

Measurement near the limit of detection is inherently imprecise. It is not possible to provide a response which meets the usual information expectations of water quality data users, namely a single number which adequately describes the contents of a sample. This creates a conflict between those who use water quality information and those who produce it. Water quality management has a need for information regarding trace level results, but the analyst

cannot produce it in a form that is easily interpreted by statistical methods now in use.

Currently, the analyst handles this conflict by defining a limit of detection (LOD) on the basis of analytical precision. Results below this limit are reported as "not detected" (ND) or "less than" the LOD. Though there is no uniformly accepted convention for the LOD, nearly every definition has the common objective of determining the smallest result which can be regarded as significantly different from zero.

Water quality managers have begun to investigate statistical methods developed for censored data and to adapt them for use in water quality. Developed for survival and failure time analysis, these methods presume that uncensored results (in water quality, those above the limit of detection), are known precisely and that nothing is known about results below this level. However, water quality data with ND observations is not truly censored in this sense. That is, useful information is recovered during chemical analysis that is not reported. This information includes the precision of a measurement and is useful in a decision making context. This applies to all results, not simply those near the LOD.

One can remedy the problem of imprecise results simply by raising the censoring level. However, this filters even more useful information from the data record. A more meaningful approach is to report with each analysis a result and

an estimate of precision. One could go a step further and describe the statistical structure of this error.

A. Objectives

The major purpose of this thesis is to present options for statistically analyzing imprecise data. This requires an understanding of the type of errors which lead to data censoring and statistical descriptions of measurement which convey as much useful information as possible. Specifically, the objectives are:

- describe the statistical nature of a measurement,
- illustrate the effect of random measurement error on the statistical properties of a set of data,
- present alternatives for the statistical analysis of data which acknowledge uncertainty, and
- for data which has previously been censored, and for statistical problems pertinent to water quality, summarize the literature of censored data methods.

The effect of measurement error on the statistical structure of water quality data is illustrated using a model for measurement error which is appropriate for low analyte concentrations. This is developed from theoretical consideration of random signal error in chemical analysis. The concept of a LOD is discussed in this context as well. It is asserted that any single definition of the LOD cannot meet all the information needs of a water quality management system.

Statistical methods which acknowledge monitoring system error are reviewed, and the concept of monitoring sensitivity is proposed as a means of characterizing the ability of a monitoring system to detect change. Finally, the literature of methods for censored data is reviewed. Modifications are made when necessary to make these methods usable for water quality data.

B. Scope

In order to provide a concise discussion of problems associated with analytical limits of detection, rather simple water quality statistical problems will be addressed. Estimation of the mean of a water quality population, and describing a confidence interval for the mean will be the primary examples. Given the discussion of the effect of measurement uncertainty on the statistical properties of data, it is straightforward to develop methods for trend detection, two sample comparisons and other types of problems in which measurement uncertainty is acknowledged.

To further simplify the discussion, water quality random variables will be considered to be distributed log-normally, and observations will be considered to be independent. In addition, sources of error will include only analytical error associated with detection limit problems. Extensions to other types of error will be mentioned.

The review of methods for censored data relates to the type of information sought by water quality monitoring. This frequently includes (Ward and McBride, 1986):

1) comparisons of water quality with criteria and standards,

2) spatial water quality (current water quality conditions and variation along a stream or in an aquifer), and

3) changes in water quality over time at a given point.

It is felt that these problems can be addressed to some extent by a fairly small set of statistical methods. The statistical questions this study will consider are:

- estimation of distribution parameters
- goodness of fit tests
- two sample tests of means
- tests for trend

The nomenclature described on page xiii is adhered to throughout the first four chapters. Chapter V, however, presents a problem in that a large number of methods from many references are described. Nomenclature for chapter V is defined separately for each method and is adhered to in Appendix A. In many cases, the symbols used are the same as the original references cited.

II. THE STATISTICAL NATURE OF THE MEASUREMENT PROCESS

Users of water quality data are often unaware of the statistical nature of numbers produced by a laboratory. Results are interpreted as arbitrarily accurate representations of reality. However, this type of confidence cannot always be justified. This chapter is a discussion of the uncertainty which arises from random analytical noise and leads to definitions of the limit of detection (LOD). Models for this type of error are suggested and used in subsequent chapters to simulate results of analyses.

This discussion does not explicitly include systematic error or errors due to sampling and sample preparation. It is assumed that systematic error can be detected and controlled by an active quality assurance effort. Uncertainty due to sample preparation can be estimated by replicate sample preparation and analysis. Similarly, replicate sampling techniques can be used to assess sampling error.

A. Objectives of measurement

Water quality monitoring may be viewed as an information system whose purpose is to supply knowledge and understanding about water quality conditions. The driving force

for the flow of information, as depicted in Figure II.1 (Ward and McBride, 1986) is the fundamental knowledge produced by the sampling and analysis portion of the monitoring system. The major objective of sampling is to provide representative portions of the water body for analysis, while the "aim of an analysis is to reduce the uncertainty with respect to the sample to be analyzed", which, it is noted, is equivalent to obtaining information (Massart et.al.,1978).

Reducing uncertainty with respect to a sample is hindered by imperfections in the sampling and measurement process. At each step, useful information as well as noise is introduced. Ideally the noise is removed without meaningful loss of information. In real situations, however, all results contain some noise, and no result preserves all of the information in a sample.

In order to assess the extent to which the major goal of measurement is accomplished, the system must be monitored. This effort of quality assurance has been defined as "those operations and procedures which are undertaken to provide measurement data of stated quality with a stated probability of being right. The measurement system must be in a state of statistical control in order to justify such a probability statement. This is attained by quality control procedures which reduce and maintain random and systematic errors within tolerable limits, and by quality assessment procedures which monitor the quality control procedures and

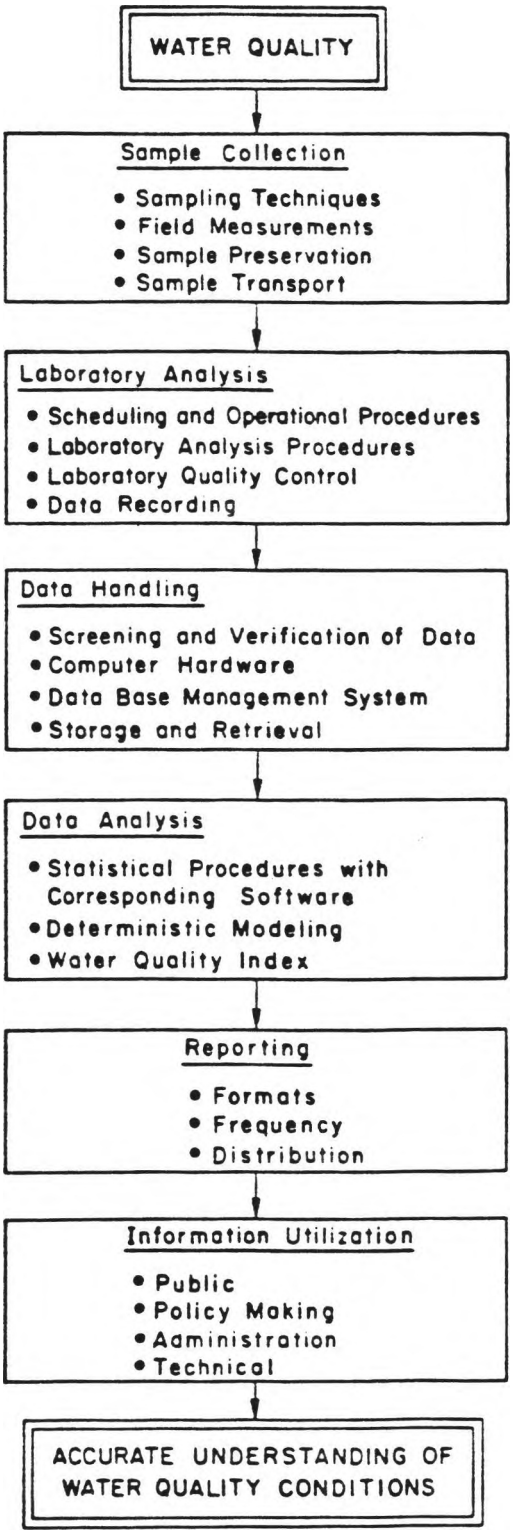

Figure II.1 The flow of information in a water quality monitoring system (After Ward and McBride, 1986).

evaluate the quality of the data produced" (Taylor and Stanley, 1985). "Statistical control" has been defined by Wilson (1970a) as meaning that "all the causes of errors remain the same".

Statistical control is accomplished by the regular monitoring of "quality parameters" (Kateman and Pijpers, 1981). Examples of quality parameters which are descriptions of data quality include:

- accuracy
- precision
- mean error
- calibration sensitivity
- limit of detection
- limit of quantification

Factors which directly affect data quality include the design of the calibration experiment, the quality of chemical reagents, the condition of the measuring apparatus and environment, the possibility of human error, and cost.

Under ideal circumstances, the objective of reducing uncertainty with respect to a sample can be satisfied by producing a single number, usually analyte concentration. In practice, however, it is more realistic to regard analytical results as "probability statements". Such statements require that one or more quality parameters be provided in addition to the result. In other words, data users should regard a result, at a minimum, as an estimate of the expected value of an analysis (the result) and its standard deviation (precision).

B. Random measurement error

1. Analytical signal noise

Figure II.2 after Mossotti (in Elving, 1984) is a schematic of the measurement process as an information model for a flame emission or flame atomic adsorption system. He considers the sample to be an "information generation and storage device in which the physical concentration of the analyte species is regarded as the informational symbol." Analysis is the conversion of chemical information to an output signal (say a digital voltage readout). A transducer converts the chemical signal into a related property such as light emission. A second transducer translates this quantity into a readout. The basic information produced by such a system is in the form of a property related to the concentration of analyte in the sample. Such properties include "emitted or absorbed light, electrical or thermal conductance, weight, volume and refractive index" (Skoog, 1985).

Because of imperfections in the measurement system, some portion of the observed property (signal) will not be related to the analyte, or some portion of the analyte may not produce a signal as expected. For example, in the flame atomic adsorption determination of sodium (Na) (a schematic of this measurement system is shown in Figure II.3), it is known that atomic Na absorbs light at a wavelength of 589.6 nm. Analysis is carried out by measuring the change in intensity of a light source emitting only at that

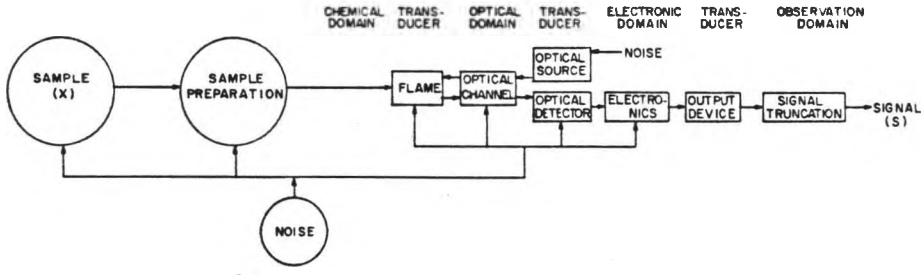

Figure II.2 The measurement process as an information system (After Mossotti in Elving, 1984).

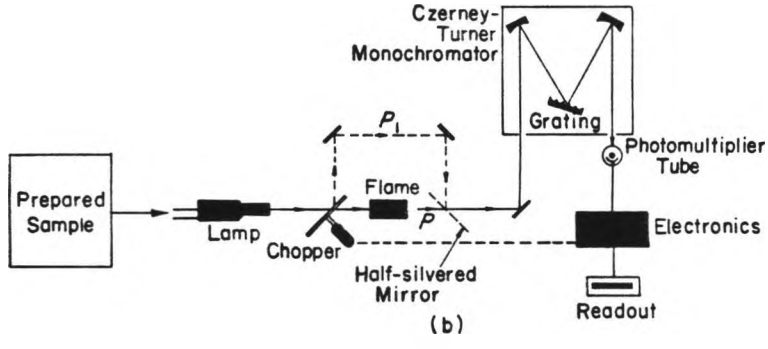

Figure II.3 Schematic of a flame adsorption measurement system (After Skoog, 1985).

wavelength, and its intensity after passing through the sample, which has been atomized in a flame. Changes in light intensity can be due, in part, to: imperfect detectors, imperfect light sources, species in the flame which adsorb or emit light within the range of the detector, ionization of Na in the flame, imperfect translation of light intensity into a readout, and imperfect knowledge of the relationship between light absorption and concentration.

Deviations from an ideal measurement process can be termed noise. Theoretical studies of noise have led to means by which it may be eliminated or reduced. Of particular interest are the statistical structures of various noise sources and the dependence of noise strength on concentration. Theoretical discussion may also be useful for explaining errors which are observed for a particular method even though the noise which is observed is a "complex composite, which usually cannot be fully observed" (Skoog, 1985).

2. Models for random signal noise

Mossotti (in Elving, 1984) presents a stochastic representation of analytical signal and noise using the example of a flame emission or atomic adsorption system. Noise produced by the system can be categorized based on dependence on analyte concentration. Noise which is statistically independent of concentration is termed additive. Noise dependent on concentration is termed multiplicative. Multiplicative noise is propagated through the measurement system along with the signal of interest. It can be further

categorized as "fundamental" or "nonfundamental" noise. Fundamental noise is usually considered to originate in the "chemical domain as a quantum characteristic of the chemical signal" (Mossotti in Elving, 1984). It is not a property introduced during recovery of the chemical signal. Nonfundamental noise "originates from temporal perturbations in the parameters of the transducing elements in the system" (Mossotti in Elving, 1984). Nonfundamental noise is characterized by nonstationary statistics, making it extremely difficult to model and remove.

Prudnikov(1981) applies theoretical discussion of noise to the flame emission determination of lithium in rocks and minerals. Prudnikov and Shapkina (1984) assert that the variance of an analytical signal [$\sigma^2(e_s)$] can be described by at most five terms:

$$(II.1)\quad \sigma^2(e_s) = \sigma^2_{add} + \sigma^2_{trans} + \sigma^2_{mult} + \sigma^2_{tnl} + \sigma^2_{nl}$$

The first term, σ^2_{add}, refers to additive noise, which is proportional to total instrument noise independent of signal strength. Mossotti (in Elving, 1984) terms this "additive channel noise", examples of which are Johnson noise in the electronics, "shot noise associated with optical detectors, power-line hum , and impulse noise."

Johnson noise is thermal noise resulting from the agitation of charge carriers in the components of an electronic system. Johnson noise is described by Skoog(1985):

$$(II.2)\quad v_{rms} = (4kTR\Delta f)^{1/2}$$

where v_{rms} is the root-mean-square noise voltage lying in a

frequency bandwidth of Δf Hz, k is Boltzmann's constant, T is absolute temperature, R is the resistance of the resistive element.

Shot noise is a result of a charge carrying particle moving across a junction (e.g., the movement of an electron between the anode and cathode in a vacuum tube) (Skoog, 1985). Shot noise is expressed (Skoog, 1985) as:

$$(II.3) \quad i_{rms} = (2Ie\Delta f)^{\frac{1}{2}}$$

where i_{rms} is the root-mean-square current fluctuation associated with the average direct current I, e is the charge of an electron, and Δf is as before. Both Johnson and shot noise are independent of the particular frequency of measurement, but not of band width being observed.

The term σ^2_{trans} refers to transitive noise associated with quantum-mechanical shot noise inherent in photoelectric detectors. Mossotti (in Elving,1984) terms this "Poisson" noise or a component of fundamental noise. It can be considered a quantum mechanical characteristic of the chemical signal. The variance component due to transitive noise is proportional to signal strength (Prudnikov and Shapkina, 1984).

The term σ^2_{mult} refers to multiplicative, or flicker noise, which is classical-mechanical in origin. Variance due to multiplicative noise is proportional to the square of the signal power. Methods most prone to this type of noise are those in which material transport and or sample atomization are required for signal generation (Mossotti in Elving,

1984). Skoog(1985) terms this "1/f" noise, as it is inversely proportional to frequency. Flicker, which is an example of nonfundamental noise, is a source of long term drift observed in some types of electrical measurement devices (Skoog, 1985).

Prudnikov and Shapkina(1984) describe σ^2_{tnl} (transitive non-linear) and σ^2_{nl} (non-linear) terms proportional to the 3^{rd} and 4^{th} power, respectively, of signal strength. These sources are due to non-linear quantum mechanical and non-linear classical-mechanical noise, respectively.

Noise reduction techniques include both hardware devices and signal processing software. Examples of hardware items include environmental controls (i.e., grounding of equipment, temperature control), low pass and high pass filters, signal modulators, and signal "choppers". Software methods include "box car" (moving average) filters, smoothing, and ensemble averages. For example, Johnson and shot noise can be reduced by lowering the temperature of a measuring device or reduction of the frequency bandwidth being observed. Signal averaging tends to cancel white noise (independent of time, zero mean), of which these are examples.

Equation (II.1) can be rewritten (Prudnikov and Shapkina,1984):

$$\text{(II.4)}\quad \sigma^2(e_s) = k_{add} + k_{trans}S + k_{mult}S^2 + k_{tnl}S^3 + k_{nl}S^4$$

where the k_i are coefficients related to the sources of

noise described above, and S is signal power. The k_i can be considered constant for a given method.

Considering the case of a linear response, corrected for background, $S = X\zeta$, where X is concentration and ζ is the change in signal produced by a unit change in analyte concentration,

$$(II.5)\quad \sigma^2(e_s) = k_{add} + k_{trans}X\zeta + k_{mult}X^2\zeta^2 + k_{tnl}X^3\zeta^3 + k_{nl}X^4\zeta^4$$

For the case of constant ζ, one can combine it with the k_i and obtain a single constant:

$$(II.6)\quad \sigma(e_s^2) = k_o + k_1X + k_2X^2 + k_3X^3 + k_4X^4$$

One can also define relative standard deviation (RSD) (the coefficient of variation) of a signal:

$$(II.7)\quad \sigma_{RSD} = \sigma(e_s)/S = (k_o/X^2 + k_1/X + k_2 + k_3X + k_4X^2)^{1/2}/\zeta$$

The signal(S) to noise(N) ratio (S/N) is defined as $1/\sigma_{RSD}$. Figure II.4 is a plot of σ_{RSD} vs. log X for arbitrary values of the k_i. (see also Figure 1 in Prudnikov and Shapkina(1984) and Prudnikov(1981)).

The values of the k_i depend on the analytical method and analyte matrix, but the shape of Figure II.4 holds for a wide range of methods. Models appearing in the literature, which take the form of a linear or squared dependence of noise variance on concentration, can be attributed to concentration ranges where particular noise sources are dominant. An analyst prefers to work in the range where σ_{RSD} is a minimum. In the range represented by the bottom of the

curve in Figure II.4, σ_{RSD} is approximately constant. Here one is dealing, for the most part, with transitive and multiplicative noise (Prudnikov and Shapkina, 1984). This is illustrated by Figure II.5, which shows the relative standard deviation for the determination of sodium in blood serum (Ross and Fraser (1977)), and Figure II.6 for the analysis of trichloroethylene in water by gas chromatography (Bell, 1986). Precision estimates for many of the methods described in Skougstad, et.al.(1979) also have a linear relationship between concentration and standard deviation. When relative error increases beyond what is optimal another method or concentration range should be selected. However, in trace analysis one is forced to work with less than optimal precision.

3. The distribution of random analytical noise

The frequency distribution of analytical noise has been the subject of much debate. Usually, it is assumed to be normal. Thompson and Howarth(1976 and 1980) describe several situations which might cause one to conclude that they were dealing with a skewed distribution.

(i) Repetition of a measurement until a positive result occurs is a practice which ignores the fact that although concentration cannot be negative, results can.

(ii) Assigning a constant to data below the detection limit has the effect of producing a "folded normal" distribution.

(iii) A lack of significant digits produces a discontinuous distribution of results, which makes identification of a distribution difficult.

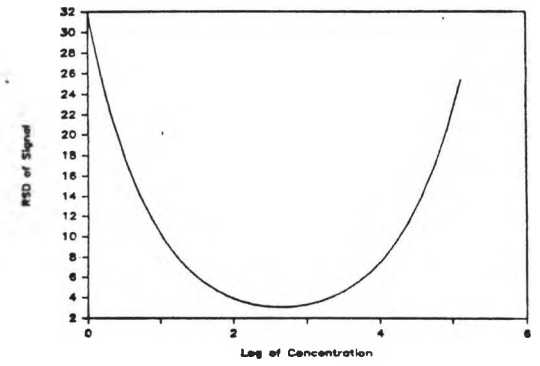

Figure II.4 Relative standard deviation of an analytical signal as a function of concentration.

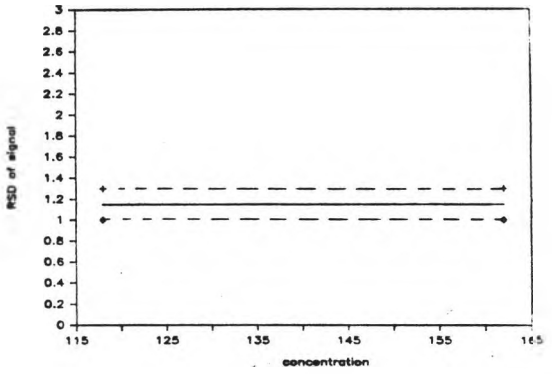

Figure II.5 Relative standard deviation of analytical signal as a function of concentration for the atomic adsorption determination of sodium - the high precision range (After Ross and Fraser, 1977). (Dashed lines are 95% confidence limits)

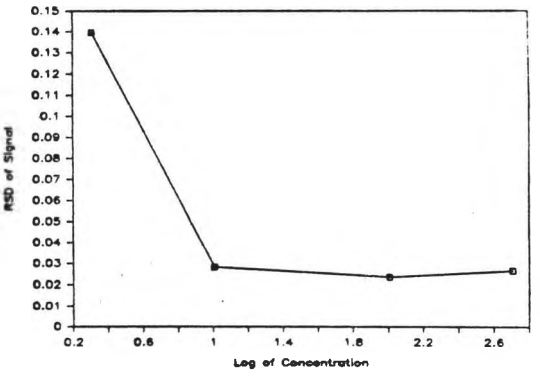

Figure II.6 Relative standard deviation of signal for the determination of trichloroethylene by gas chromatography.

(iv) Outliers due to a systematic error may cause an error distribution to appear skewed.

In addition, a large number of samples are required to distinguish nearly normal from normal distributions (see Shapiro et.al.,1968; and Kendall and Stuart,1979).

For purposes of this discussion, random signal error near the LOD will be assumed to be normally distributed with mean zero and variance k_o+k_1X. In standard notation,

$$(II.8) \quad e_s \approx N(0, k_o + k_1 \cdot X)$$

This model is consistent with theoretical descriptions of measurement error near the LOD. It also assumes that signal error is unbiased. This is reasonable for any measurement process with an active quality assurance effort that detects and removes systematic error.

Assumption of a normal distribution permits negative results. This does not imply that negative concentrations are possible, only that random fluctuations may lead to a negative signal. A similar model has been used by Skogerboe (1982) for describing the effect of blank contamination on trace level analysis. This model was used implicitly by Gilliom, et.al. (1984), to generate measurements near the LOD. Extension of this model to a more general form (higher order polynomial or power function) is straightforward.

C. Calibration

1. The calibration problem

The calibration problem refers to the transformation of observed analytical signals into estimates of concentration. In general, analytical signals can be described by:

$$S = g(X,\theta) \qquad (II.9)$$

where S is the analytical signal, X is concentration, θ is a parameter set, and g is a known function (e.g., for a linear calibration, $S = B_o + B_1X$ and $\theta = \{B_o, B_1\}$, the slope and intercept of the calibration curve).

The objective of calibration is to provide accurate estimates of concentration at a low cost, where these estimates are given by $X_m = h(\hat{\theta}, S_p)$, S_p is the signal from an unknown population, $\hat{\theta}$ is an estimate of θ, and h is some function of $\hat{\theta}$ and S_p.

Several approaches to calibration can be found in the literature. Inverse calibration considers the regression of concentration on signal. Estimates of unknown concentrations are provided by direct application of the regression equation. A method used more widely in chemical analysis is classical regression, where signal is regressed on concentration, and estimates of unknown concentration are found by inverting the calibration function. Other methods include the method of standard additions and Bayesian analysis. The method of standard additions is the regression of signals from replicate samples, which contain known amounts of added

analyte, on the concentration due only to the added analyte. Estimates of the unknown concentration are obtained by extrapolating the regression line to the abscissa, which represents the point of zero added standard. Bayesian classical analysis proceeds in a fashion similar to classical analysis, with added assumptions concerning the distribution of errors, presumably on the basis of prior knowledge or experience. This study will consider only classical calibration, as it is a very common technique used in water quality analytical work. There will also be limited discussion of a Bayesian calibration technique.

With regard to calibration, measurement error is influenced by knowledge of the functional relationship, $g(X,\theta)$, the number and spacing of standard samples, the number of replicate unknowns, and the nature of errors inherent in the measuring process. Cost considerations can be reduced to a question of the number of standard and unknown replicate samples. Naturally, a reduced error can be obtained at the price of more analyses.

A desirable property of an analytical method is that $g(X,\theta)$ be linear over a wide range (Skoog, 1985). For some methods, there is theoretical justification for a linear response model. For example, Beer's law, simply stated, holds that the attenuation of a light source of a given wavelength passing through a solution is proportional to the concentration of absorbing species:

(II.10) $A = \log(I_o/I) = \delta bX$

where A is absorbance, I_o is the initial energy of a light source, I the energy of a light source after attenuation by a sample, δ is a proportionality constant, b is the path length of the light source in the sample, and X is analyte concentration.

For such a relationship, equation (II.9) can be written:

(II.11) $S = B_o + B_1X + e_s$

where S is analytical signal (absorbance), B_o is the intercept (zero), and B_1 is the the slope (δb). B_o and B_1 are estimated by analyzing samples of known concentration (X) and observing signals (S) and e_s is random signal error.

For some methods, including inductively coupled plasma (ICP) analysis of metals, there is a linear relationship between the logs of signal and concentration.

(II.12) $\ln S = B_o + B_1 \ln X + e_s$, or

$$S = B_o' X_{B1} e',$$

where $B_o' = \exp(B_o)$, and $e_s' = \exp(e_s)$

2. Calibration design for the classical linear case

A number of techniques have been used to interpret signal-concentration relationships. Cardone (1986) lists many of these. Taylor and Iyer (1986) provide a general review of calibration. Buonoccorsi (1986a) reviews several approaches to optimizing a calibration experiment. The following is a description of classical calibration methods, including ordinary least squares (OLS) and weighted least

squares (WLS). Classical calibration refers parameter estimation by regressing signal on concentration.

Factors which influence the choice of the number and spacing of calibration standards include:

- the distribution of analytical signal error
- the chance that the response function will not be linear over the range of calibration standards; and
- cost.

For the case of constant signal variance, one has:

$$(II.13) \quad S = B_o + B_1 X + e_s,$$

where e_s is normally distributed with mean 0 and variance $\sigma^2(e_s)$. The least squares estimators of B_o, B_i, and $\sigma^2(e_s)$ are given by:

$$(II.14) \quad b_o = \left(\sum_{i=1}^{n} S_{i,c} \cdot \sum_{i=1}^{n} X_{i,c}^2 - \sum_{i=1}^{n} X_{i,c} \cdot \sum_{i=1}^{n} X_{i,c} S_{i,c} \right) / \left\{ n \cdot \sum_{i=1}^{n} X_{i,c}^2 - \left(\sum_{i=1}^{n} X_{i,c} \right)^2 \right\}$$

$$(II.15) \quad b_1 = \left(n \sum_{i=1}^{n} X_{i,c} S_{i,c} - \sum_{i=1}^{n} X_{i,c} \sum_{i=1}^{n} S_{i,c} \right) / \left\{ n \cdot \sum_{i=1}^{n} X_{i,c}^2 - \left(\sum_{i=1}^{n} X_{i,c} \right)^2 \right\}$$

$$(II.16) \quad \hat{\sigma}^2(e_s) = \sum_{i=1}^{n} (S_{i,c} - b_o - b_1 X_{i,c})^2 / (n-2)$$

The $X_{i,c}$ are the calibration standard concentrations and the sums are over the number of calibration standards (n). The properties of these estimators are:

$b_o \approx N[B_o, \sigma^2(b_o)]$

$b_1 \approx N[B_1, \sigma^2(b_1)]$

$S_{i,c}$, b_1, and $\hat{\sigma}^2(b_1)$ are mutually independent

$(n-2)\hat{\sigma}^2(e_s)/\sigma^2(e_s)$ is distributed as chi-squared with n-2 degrees of freedom

b_o and b_1 are uniformly minimum variance unbiased (UMVUE)

$$\hat{\sigma}^2(b_1) = \sigma^2(e_s) / \sum_{i=1}^{n} (X_{i,c} - \bar{X}_c)^2$$

$$\hat{\sigma}^2(b_o) = \sigma^2(e_s) \sum_{i=1}^{n} X_{i,c}^2 / \sum_{i=1}^{n} (X_{i,c} - \bar{X}_c)^2$$

Estimates (X_m) of the unknown, or population X (X_p) are given by:

$$(II.17) \quad X_m = (S_p - b_o)/b_1$$

where S_p is a signal from a sample with unknown concentration, and b_o and b_1 are the estimates of B_o and B_1. This estimate is also the maximum likelihood estimate (MLE). Given the distribution of S_p, b_o and b_1, it is noted that X_m is a ratio of normal distributions, which is distributed as Cauchy. The mean and variance of X_m do not exist because of the finite probability that b_1 will approach zero. Descriptions of X_m take the form of confidence interval construction or approximations to the distribution of X_m. Discussions of confidence intervals for the ratio of two normals are provided by Fieller(1932), Graybill(1976), Buonaccorsi and Iyer(1984), Mulrow(1986), and Satterthwaite(1946). Approximate distributions approaches are given by Marsaglia(1965), Hinkley(1969), and Hunter and Lamboy(1981). Buonaccorsi(1986a) summarizes many of their results.

From Buonaccorsi (1986a),

X_m is unimodal if

$$(X_p - \bar{X}_c)^2 < \sigma^2(e_s)(5.094)(1/l + 1/n)/B_1^2$$

and the median of $X_m = X_p$ if $\bar{X}_c = X_p$

$< X_p$ if $\bar{X}_c < X_p$

$> X_p$ if $\bar{X}_c > X_p$

where $\bar{X}_c$ is the mean concentration of the calibration standards and l is the number of measurements per standard. Therefore, X_m is median unbiased only when the unknown true X_p lies in the middle of the calibration range.

As the probability of b_1 approaching zero becomes small, X_m approaches normality. A guideline for judging whether this will occur may be given by the slope sensitivity, which is defined as:

$$(II.18) \quad Z = b_1/\sigma(b_1)$$

Unfortunately, there is not much information in the literature which helps one decide what values of Z should be interpreted as providing a normal X_m, or what estimators should be used for the mean and variance of X_m. Neither is there much information regarding "typical" values of Z for water analysis, though it is usually assumed to be large. For a particular gas chromatography analysis of toluene in water, Z was 190. For perchloroethylene analysis using the same calibration design and apparatus, the value of Z was about 50. The calibration experiment consisted of seven replicates each at 2, 10, 100, and 500 parts per billion

(ppb) (Bell, 1986). The value of Z can be expected to be smaller near the LOD.

Hinkley (1969) provides an approximation to the distribution of X_m given X_p and lower and upper bounds for the difference between the approximation and the exact distribution. A simpler approximation is given by Mood, et.al., (1974), where X_m is normally distributed with the mean and variance given by:

$$(II.19)\quad E(X_m/X_p) = X_p + (X_p - \bar{X}_c)/(VB_1^2)$$

$$(II.20)\quad Var(X_m/X_p) = \sigma^2(e_s)\{1/m + 1/l + (X_p-\bar{X}_c)^2/V\}/B_1^2$$

where V is the sum of squares of the calibration experiment $[= \sum_{i=1}^{n} (X_{i,c}-\bar{X}_c)^2]$, $X_{i,c}$ is a calibration standard concentration, $\bar{X}_c$ is the mean of standard concentrations, l is the number of analyses per standard, and $\sigma^2(e_s)$, B_1 and n are defined above.

A propagation of error approach, based on a first order Taylor series approximation to equation (II.17) results in:

$$(II.21)\quad E(X_m) = X_p$$

$$(II.22)\quad Var(X_m) = \sigma^2(e_s) + [\sigma^2(b_o)+2\sigma^2(b_o,b_1)X_p + \sigma^2(b_1)X_p^2]/B_1^2$$

Graybill(1976) presents a method for obtaining an approximate confidence interval for the unknown parameter, X_p, given κ replicate measurements of the unknown concentrations.

$$(II.23)\quad X_p = \bar{X}_c + b_1(\bar{S}_m-\bar{S}_c)/a \mp t(\alpha/2,n+\kappa-3)\hat{\sigma}(e_s) \cdot\{a(1/n+1/\kappa) + (\bar{S}_m - \bar{S}_c)^2/V\}^{1/2}/a$$

where $a = b_1^2 - \hat{\sigma}^2(e_s)t^2(\alpha/2,n+\kappa-3)/V$

$\bar{S}_m$ and $\bar{S}_c$ are the mean signals from κ replicate unknowns and n standard analyses, respectively. As before, $\bar{X}_c$ is the mean of the calibration concentrations. This is not a true $1-\alpha$ confidence interval for X_p. Instead, it may be said that there is a $1-\alpha$ probability that $\bar{X}_m$ satisfies equation (II.23). The confidence for this interval is in fact less than or equal to $1-\alpha$ (Graybill, 1976). (For virtually all calibration problems in analytical chemistry, this interval will exist.)

Ott and Myers (1968) used a criteria of minimum mean squared error (MSE), and Naszodi (1978) minimized bias as optimality criteria. Both result in equally weighted end point designs. This design also minimizes $\sigma^2(b_1)$. Other designs may be better if one knows in advance that the unknown X_p will lie in a particular region of the calibration range. This is usually impractical (Buonaccorsi, 1986a) because a calibration experiment is designed to cover as wide a range as possible of unknown concentrations. In the event that unknowns repeatedly fall in a small region of the calibration range, and the same calibration experiment is reused for all samples, significant bias may accumulate in the collection of results (Iyer, 1986).

In practice, few chemists use a design based on rigorous criteria. Often, rules of experience dictate a design with evenly spaced samples and the same number of replicate standards at each concentration. Two point (optimal end point) design is often shunned by chemists because

of a desire to detect a non-linear response. An approach to calibration design with power against non-linearity would assign some weight (cost) to power against a nonlinear (say a quadratic or power) response, and a weight to the relative error of X_m (Iyer, 1986). Buonaccorsi and Iyer (1984) compare methods for finding a confidence interval for X_m for a quadratic calibration model.

Another approach for overcoming the distributional problems of X_m is to modify the presumed distribution of signal error (e_s). For example, Williford, et.al.(1979) assume e_s to be a truncated normal distribution.

Mulrow (1986) studied the sampling distribution of X_m given an underlying normal population (unlike the other references in this section, in which X_p is treated as a known constant). It was assumed that signal errors were normally distributed with zero mean and constant variance. Several calibration designs were employed which resulted in a range of slope sensitivities (Z values). The methods were used to construct a confidence interval for the mean of X_p given a set of X_m's. They were denoted "uncorrected", propagation of errors (POE), and Satterthwaite's improved degrees of freedom (SIDF).

The "uncorrected" method is based on the application of the commonly used t statistic for computing a confidence interval for the mean of a normal population.

$$\text{(II.24)} \quad \mu_p = \hat{\mu}_p \pm t(\alpha/2, \kappa-1)/(\hat{\sigma}_m/\sqrt{\kappa})$$

Here, κ is the number of independent unknown samples from a

population with mean μ_p and variance σ_p^2. The sample estimates based on measurements are given by:

$$(II.25)\quad \hat{\mu}_m = \bar{X}_m = \Sigma X_{i,m}/\kappa, \quad \text{and}$$
$$\hat{\sigma}_m^2 = \Sigma(X_{i,m}-\bar{X}_m)^2/(\kappa-1).$$

This method proved to be very poor under a variety of conditions, providing much less than the nominal $1-\alpha$ coverage for Z values as large as 100. The exact coverage (γ), given by Mulrow(1986), is:

$$(II.26)\quad \gamma = P[|T| \leq t(\alpha/2,\kappa-1)/(1+\varepsilon)^{1/2}]$$

where T = a student's t random variable

$$= (\bar{X}_m-\mu_p)/(\hat{\sigma}_m/\sqrt{n})$$

$$\varepsilon = \kappa[\sigma^2(b_o)+2\sigma^2(b_o,b_1)\mu_p+\sigma^2(b_1)\mu_p^2]/[B_1^2\sigma_p^2+\sigma^2(e_s)]$$

and where $\sigma^2(e_s)$ is the signal variance, $\sigma^2(b_o)$ is the intercept variance, $\sigma^2(b_1)$ is the slope variance, and $\sigma^2(b_o,b_1)$ is the covariance of the slope and intercept. A typical result: [for α = 0.05, Z = 100, μ_p = 0.5, V = 1, B_1 = 1, σ_p = 0.005, κ = 35, $\sigma(e_s)$ = 0.01] γ(the true coverage) = 0.523 rather than 0.95.

Both POE and SIDF were found to behave better, and produced true coverage close to $1-\alpha$. The POE approach is based on the POE estimate of the variance of X_m given by equation (II.22), with an additional term due to X_p being a random variable.

$$(II.27)\quad \mathrm{Var}(\bar{X}_m) = \sigma_p^2/\kappa + \sigma^2(e_s)/(\kappa B_1^2) + [\sigma^2(b_o)+2\sigma^2(b_o,b_1)\mu_p+\sigma^2(b_1)\mu_p^2]/B_1^2$$

$$(II.28)\quad 1-\alpha \approx P[(\bar{X}_m-\mu_p)/\{\hat{\sigma}_m^2/\kappa + [\hat{\sigma}^2(b_o) + 2\hat{\sigma}^2(b_o,b_1)\hat{\mu}_m + \hat{\sigma}^2(b_1)\hat{\mu}_m^2]/B_1^2\}^{\frac{1}{2}} \leq t(\alpha/2,\kappa-1)]$$

SIDF is similar to a t test:

$$(II.29)\quad \gamma = P[(\bar{X}_m-\mu_p)/\{\hat{\sigma}_m^2/\kappa + [\hat{\sigma}^2(b_o) + 2\hat{\sigma}^2(b_o,b_1)\hat{\mu}_m + \hat{\sigma}^2(b_1)\hat{\mu}_m^2]/B_1^2\}^{\frac{1}{2}} \leq t(\alpha/2,f)]$$

an estimate of the degrees of freedom, f, is given by:

$$\hat{f} = \{\hat{\sigma}_m^2/\kappa + [\hat{\sigma}^2(b_o) + 2\hat{\sigma}^2(b_o,b_1)\hat{\mu}_m + \hat{\sigma}^2(b_1)\hat{\mu}_m^2]\}^2/\{\hat{\sigma}_m^4/[\kappa^2(\kappa-1)] + [\hat{\sigma}^2(b_o) + 2\hat{\sigma}^2(b_o,b_1)\hat{\mu}_m + \hat{\sigma}^2(b_1)\hat{\mu}_m^2]^2/(n-2)\}$$

The POE and SIDF simulated coverages for the parameters which provide a coverage of 0.523 with the uncorrected method were both 0.921. The properties of X_m when X_p is a random variable will be taken up in more detail in Chapter III.

Hunter and Lamboy (1981), with a Bayesian approach, suggest that the distribution of X_m is given approximately by:

$$(II.30)\quad X_m \approx N\{X_p,(s_{11}s_{22}+s_{33}s_{22}-s_{12}^2)/(s_{22}\cdot b_1^2)\}$$

where $\underline{S} = \begin{bmatrix} [\underline{XX}']_c^{-1} & 0 \\ 0 & \sigma^2/1 \end{bmatrix}$

$$\underline{S} = \begin{bmatrix} \begin{bmatrix} n & \Sigma X_{i,c} \\ \Sigma X_{i,c} & \Sigma X_{i,c}^2 \end{bmatrix}^{-1}\cdot\sigma^2 & 0 \\ 0 & \sigma^2/1 \end{bmatrix}$$

$$s_{11} = \sigma^2\Sigma X_{i,c}^2/d \quad s_{12} = -\sigma^2\Sigma X_{i,c}/d,$$

and $s_{22} = n\sigma^2\Sigma/d$,

$$d = n \cdot \Sigma w_i X_{i,c}^2 - (\Sigma X_{i,c})^2 \ .$$

where l is the number of unknown replicates and the sums are over the collection of calibration standards.

3. Weighted least squares calibration

For many analytical methods the variance of random signal error [$\sigma^2(e_s)$] is a function of concentration (X), or heteroscedastic.

(II.31) $$S = B_o + B_i X + e_s,$$

$$e_s/X \approx N(0,\ k_o + k_1 X + k_2 X^2 + k_3 X^3 + k_4 X^4)$$

where the k_i are as before. Near limits of detection, the variance of a signal given X can be modelled by k_o+k_1X. The best estimates for B_o, B_1, and $\sigma^2(e_s)$ are obtained from weighted least squares (WLS). Multiplying both sides by $\underline{W}^{-½}$,

(II.32) $$\underline{W}^{-½}\underline{S} = \underline{W}^{-½}\underline{X} \cdot \underline{B} + \underline{W}^{-½}\underline{e}_s, \quad \underline{W}^{-½}\underline{e}_s \approx N(\underline{0},\ I) \ .$$

where $\underline{X} = [\underline{1}\ \underline{X}_{i,c}]$, $\underline{k} = [k_o\ k_1]'$, $\underline{W}$ is diagonal with elements $k_o+k_1X_{i,c}$, and $\underline{I}$ is the identity matrix. Rewriting, with $\underline{W}^{-½}\underline{S} = \underline{S}^*$, etc.,

(II.33) $$\underline{S}^* = \underline{X}^*\underline{B} + \underline{e}_s^*, \quad \underline{e}_s^* \approx NID(\underline{0},\ I)$$

The least squares estimate of $\underline{B}$, based on $\underline{X}^*$ and $\underline{S}^*$ is referred to as weighted least squares (WLS). WLS produces unbiased parameter estimates with lower variances than OLS. Descriptions of generalized (or weighted) least squares are given by Graybill(1976), Draper and Smith(1981), and Box, et.al.,(1978). Garden, et.al.,(1980) present a WLS example for the atomic adsorption determination of lead and copper. The WLS estimates of B_o and B_1 are given by:

$$(II.34)\quad b_o = (\Sigma w_i S_{i,c} \cdot \Sigma w_i X_{i,c}^2 - \Sigma w_i X_{i,c} \cdot \Sigma w_i X_{i,c} S_{i,c}) / \{\Sigma w_i \cdot \Sigma w_i X_{i,c}^2 - (\Sigma w_i X_{i,c})^2\}$$

$$(II.35)\quad b_1 = (\Sigma w_i \Sigma w_i X_{i,c} S_{i,c} - \Sigma w_i X_{i,c} \Sigma w_i S_{i,c}) / \{\Sigma w_i \cdot \Sigma w_i X_{i,c}^2 - (\Sigma w_i X_{i,c})^2\}$$

$$(II.36)\quad \hat{\sigma}^2(e_s) = [\Sigma w_i S_i^2 - b_o \Sigma w_i S_i - b_1 \Sigma w_i S_{i,c} X_{i,c})]/(n-2)$$

where all the sums are from 1 to n, the number of standard samples. The value of Z, defined in equation (II.18) is calculated from:

$$(II.37)\quad Z = (\Sigma w_i \Sigma w_i X_{i,c} S_{i,c} - \Sigma w_i X_{i,c} \Sigma w_i S_{i,c}) / \Sigma w_i$$

D. Limits of Detection

1. Concepts of the detection limit

Data users have perceptions of the LOD which may differ from that of an analyst. Many mistakenly believe that a result of "ND" is equivalent to zero. Others believe that detection below the LOD cannot be accomplished. The "limit of detection" of a measurement process is a statistical concept which has meaning only in the context of the use to be made of a result. Massart et.al (1978) state that the LOD is "one of the most important performance-characteristics of an analytical process". "It should be emphasized that concentrations smaller than the limit of detection will sometimes be detected, i.e., the results will exceed the criterion of detection. In this sense, the term 'limit of detection' is rather misleading for it suggests that smaller concentrations will not be detected. However, without marked

expansion of the term, it is impossible to convey in one simple title all the statistical and analytical implications" (Wilson, 1970a).

Defining a LOD is the process of determining with reasonable certainty whether a signal from a measurement process is caused by analyte or an "uncontrolled chance perturbation", to use Kaiser's (1970b) words. It is a consequence of the presence of noise when useful signal due to analyte is relatively small. The determination of the LOD can also be stated in terms of a statistical hypothesis test:

(II.38) H_o: Analyte is present vs.

H_a: Analyte is not present

Wilson (1970c) states that the two important aspects of LOD determination are: 1) the criterion used to address the statistical hypothesis, and 2) the smallest concentration that an analyst can claim to be capable of detecting. The former aspect is subjective and includes (Wilson, 1970c): 1) the desired confidence level, 2) the distribution of the "random fluctuations", and 3) the standard deviation of this distribution. The latter pertains to the physical measurement apparatus and a specified measurement procedure. Kaiser (1970b) states this another way by stating that the LOD must be applied to a "complete analytical procedure" and requires a "criterion agreed upon by convention". Suffice to say that a criterion is needed, though it need not be agreed upon, only reported. "There is no reason why one confidence level

should be suitable for all analytical work; indeed, quite the opposite is true" (Wilson, 1970a). Nevertheless, there is a continual effort to standardize the definition of the LOD so that methods can be compared on this basis.

2. Definitions of the LOD

The International Union of Pure and Applied Chemists (IUPAC) definition of the LOD is:

$$(II.39) \quad X_{LOD} = 3 \cdot \sigma_B / B_1$$

where σ_B is the standard deviation of signals from repeated blank measurements and B_1 is the slope of a linear calibration function. In this discussion, a blank will be defined as identical to unknown samples except that no analyte is present.

If the random error is normally distributed and σ_B and B_1 are known, the confidence level provided by this definition is 0.9987. That is,

$$(II.40) \quad P(X_m < X_{LOD} \text{ given that } X_p = 0) = 0.9987$$

The probability of a type II error is, at the LOD, 0.50.

$$(II.41) \quad P(X_m < X_{LOD} \text{ given that } X = X_{LOD}) = 0.50$$

If $\sigma_{\bar{B}}$ is estimated from a (not large) set of replicate measurements, one must resort to the t distribution to calculate a confidence level:

$$P[S_B/(\hat{\sigma}_B/\sqrt{n}) \leq t(1-\alpha,n-1)] = 1-\alpha$$

and define $S_{LOD} = t(1-\alpha,n-1)\hat{\sigma}_B/\sqrt{\nu}$, the signal at X_{LOD}. Here $\bar{S}_B$ is the average signal from n blank measurements. If, for example, $n = 5$ and $t(1-\alpha,n-1)/\sqrt{n} = 3$, the confidence level is much larger than 0.9987 [$t(1-\alpha,4) = 6.71$]. To

achieve a confidence level of 0.995,

$$(II.42) \quad X_{LOD} = 4.604 \cdot \hat{\sigma}_B / \sqrt{(5B_1)}$$
$$= 2.06 \ \hat{\sigma}_B / B_1$$

Thus, the relatively simple IUPAC definition results in different confidence levels depending on how it is applied. One could remedy this by specifying that a large number of blank determinations be used (say n>30), then combined with the central limit theorem, the mean of the measurements would approach normality regardless of the true distribution of errors, and the confidence level would approach 0.9987. There would remain, however, several problems, including:

- low power,
- a confidence level of 0.9987 that is not appropriate for all purposes, and,
- for purposes of reporting a LOD in concentration units, the slope is assumed known.

Questions concerning power and confidence can be addressed only by considering the eventual use to be made of the data. Wilson (1970c) feels that in order for this decision to be made by the data user, the criterion for detection needs to be reported. This includes the "nature and standard deviation of the frequency distribution", plus the way in which this was estimated (number of blank samples, method, etc). This eliminates the need for explicitly including calibration information (slope and intercept and their variances), which is usually not obtained for a concentration region near the LOD. Calibration information depends on the number and spacing of calibration standards.

Usually, the lower limit of a calibration experiment is well above the LOD. It is an undesirable property of the definition of the LOD that it be influenced by the calibration design.

Another problem related to the use of a t statistic, as discussed above and in Long and Winefordner (1983), is that the hypothesis being tested does not relate directly to an analysis of an unknown sample, which is the ultimate use of a definition of a LOD. For the example used above, with n = 5, the confidence level calculated, 0.995, relates only to the mean of a population of blank measurements. In order to compare the set of blank measurements to a set of unknown replicate measurements, one would typically do a two sample t test for the means of the two groups, assuming that the variances of the two populations are the same. This is difficult if there is only a single unknown measurement. A t statistic can still be developed, however (Yao, 1984). Let $\{X_i\}_B$ be the set of n blank measurement results with mean $\mu_{x,B}$ and variance $\hat{\sigma}^2_{x,B}$, and X_m the result of a single unknown measurement using the same measurement process. For the sake of illustration, assume that measurement variance is constant. Then,

$$\text{(II.43)} \quad \bar{X}_B \approx N(\mu_{x,B}, \sigma^2_{x,B}/n)$$

$$X_m \approx N(\mu_p, \sigma^2_m)$$

$$(n-1)\hat{\sigma}^2_m/\sigma^2_m \approx \chi^2_{n-1}$$

Under the null hypothesis that analyte is not present, the means and variances of the two random variables are equal,

and (II.44) $T \approx (X_m - \bar{X}_B)/[\hat{\sigma}_{x,B}(1/n + 1)]^{\frac{1}{2}}$

and, (II.45) $X_{LOD} = [\hat{\sigma}_{x,B}(1/n + 1)]^{\frac{1}{2}} t(1-\alpha, n-1)$

Note that the LOD based on this concept can be arbitrarily lowered by increasing the number of unknown or blank replicates. If the measurement variance cannot be assumed constant one has the difficult problem of comparing two samples with different variances.

The EPA (Glaser, et.al., 1981) has endorsed a method which requires analyte to be present in the samples used to determine the LOD. The EPA definition of the LOD is:

(II.46) $X_{LOD} = \hat{\sigma} \cdot t(1-\alpha=0.99, n-1)$

where $\hat{\sigma}$ is the standard deviation of replicate sample concentrations which contain analyte concentrations near the LOD. This definition is equivalent to the development of the two sample t test described above with the 1/n term neglected. The assumptions of the IUPAC definition are inherent here. That is, the samples are field blanks, (but spiked with analyte concentration near X_{LOD}) and the slope of the calibration is known, so that $\hat{\sigma}$ can be calculated in terms of concentration units. However, there is the added complication of determining the proper spike concentration with a recursive method of repeated replicate analyses at different concentrations. This method is tedious. Moreover, Wilson(1970c) asserts: "the remaining discussion [concerning LOD's] rests on the basis that the standard deviation at zero concentration is the fundamental performance-characteristic requiring definition and evaluation."

Long and Winefordner(1983) consider the situation when the errors of the slope and intercept are included in the determination of the LOD. They also discuss the true statistical level when the distribution of the errors is not normal. A lower confidence level of 0.89 is given by consideration of Tschebyscheff's inequality (Larson and Marx, 1981). Hubaux and Vos(1970) define a LOD based heavily on the calibration experiment, using prediction intervals for a signal to project a confidence interval at a concentration of zero.

Winefordner and Ward(1980) add the constraints that for a LOD determination to be valid, the replicates must be analyzed at regular intervals, all analytical signals should be sampled for equal times, and the total signal sampling time should be reported. This is to account for low frequency noise and drift. They conclude that LOD's closer than a factor of three are not significantly different.

Skogerboe(1982) discusses the limiting effect of reagent contamination on the LOD. This applies to contamination of field blanks by analyte as well. His argument is based on consideration of the RSD and shows that the task of determining a LOD is basically a question of differentiating a blank from a low level measurement. If the "blank" contains analyte, significantly more analyte must be present in a sample in order to differentiate it from the "blank". It was shown for trace level measurements that this can be a significant problem.

Several concepts have been used to deal with the low power at the LOD associated with most definitions of the LOD. Menzies and Kaiser(1969) propose a "limit of guarantee of purity" (LOG) which is the purity (probability of absence of analyte) which can be assumed for a result below the LOD. Massert et.al.(1978) and Kaiser(1970b) both recommend using $6\sigma/B_1$ for this level. For a normal distribution with σ known:

(II.47) $P(X_m < X_{LOG}$ given that $X = X_{LOD}) = 0.9987$

Still another problem concerns the quality of measurements near the LOD. One can show, using the IUPAC definition of the LOD and the error model given in equation (II.8) that signals at the LOD have a precision no better than 33%.

(II.48) $\sigma_{RSD} = (1+ k_1X_{LOD}/k_o)^{\frac{1}{2}}/3$

In some cases, precision is poor for values well above the LOD. Thompson and Howarth(1976) illustrate this for a linear relationship between standard deviation and concentration. This has prompted a need for a "limit of quantification"(LOQ). ACS(1980) defines the LOQ as the "level above which quantitative results may be obtained with a specified degree of confidence". They divide the range of possible results into three regions.

Signal Region	Name	Reporting Convention
< LOD	Region of high uncertainty	ND
LOD < <LOQ	Region of less certain quantitation	"detected"
LOQ <	Region of quantitation	numerical result

They recommend that "quantitative interpretative decision making, and regulatory actions should be limited to data at or above the LOQ" and suggest that the LOQ be defined as $10 \cdot \sigma_B$.

Currie(1968) proposed this scheme, but included in his discussion the "critical level", which is defined by the type I error:

(II.49) $P(S < S_c \text{ given } X_p = 0) = 1 - \alpha$

Signals exceeding S_c indicate detection. He then proposes that the LOD be defined as the signal at which a desired power $(1-\beta)$ for detection is obtained.

(II.50) $P(S > S_{LOD} \text{ given } X_p = X_{LOD}) = 1-\beta$

This is equivalent to Kaiser's LOG.

In summary, there are several conventions for dividing results into reporting regions. They are all based on concepts of statistical confidence. There has been a great deal of effort to standardize the process so that LOD's or LOQ's reported for the same method by different laboratories, or at different times, would be comparable.

There are two major problems with this approach. One is that in order for a LOD definition to be standardized, a great deal more than the method and the computing formula must be specified. Particularly, the calibration experiment used to determine the slope, and the time of analysis of standards and unknowns should also be made standard. Ironically, if results less than the LOQ are to be reported simply as "detected" there is no need to determine the slope

in the region of the LOD except for purposes of calculating the LOD. This is implicit in the reporting recommendations of Currie (1968) who suggests reporting a quantitative result (a single number) only when $S > S_{LOQ}$.

The second problem is that specification of statistical criteria should be done by the data user. In water quality management, this is not usually the analyst. ACS(1980) recommendations that results below the LOQ not be used by data users is an infringement on the decision making process. The LOD has no meaning except in this context.

III. THE EFFECT OF MEASUREMENT ERROR ON WATER QUALITY DATA

System error, of which random signal error is a component, can obscure the statistical nature of a water quality variable. This can limit the amount of information on water quality behavior which can obtained from a monitoring system. More significantly perhaps, it may cause data users to believe they are observing a natural process, rather than a characteristic of the monitoring system. In this chapter, the effect of monitoring system error on the statistical characteristics of a set of data is illustrated. The error model appropriate for analyte concentrations near limits of detection will be used to generate "results" of analyses. The effect of calibration design and the distribution of the population being sampled on water quality data will be discussed. This approach can be extended to other sources of error.

The statistical properties of a set of measurements is influenced by system error. This can be illustrated in several ways. First, the moments of the conditional and unconditional distribution of measurements can be described using the propagation of errors (POE) approximation to a measurement (X_m). The POE approach provided accurate confidence intervals for X_m for classical calibration covering a wide range of calibration sensitivities (Mulrow, 1986). It

is shown by simulations that the POE approximation is also quite good for estimating other properties of X_m (mean, variance and skewness and kurtosis), for a particular weighted least squares (WLS) calibration design. It is also shown how the statistical properties of a set of data can be dominated by normally distributed measurement error.

A. Description of the measurement system

In this chapter, analytical signals (S) will be described by:

$$(III.1)\quad S = B_o + B_1X + e_s$$
$$e_s \approx N(0,\ k_o + k_1X)$$

where the terms are defined in Chapter II. Weighted least squares will be used to estimate B_o and B_1. The weights will be $(k_o + k_1X)^{-1}$, and k_o and k_1 will be assumed known. Estimates (X_m) of unknown concentrations (X_p) will be given by:

$$(III.2)\quad X_m = (S - b_o)/b_1,\quad S = B_o + B_1X_p + e_s$$

The POE approximation of the PDF of X_m given X_p is:

$$(III.3)\quad f(X_m/X_p) \approx N\{X_p,\ \sigma^2(e_s)/(\kappa B_1^2)$$
$$+ [\sigma^2(b_o)+2\sigma^2(b_ob_1)X_p + \sigma^2(b_1)X_p^2]/B_1^2)$$

where $\sigma^2(e_s)$ is the variance of random signal error, and b_o and b_1 are estimates of slope (B_o) and intercept (B_1) for a linear calibration. The $\sigma^2(\cdot)$ terms are the variances of the parameter estimates (variances of b_o, b_1 and the covariance

of b_o and b_1), X_m is the estimate of X_p (the true concentration) and κ is the number of replicate unknown samples.

A large number of variables are needed to describe the properties of a set of measurements. One must specify parameters of the measurement system (k_o, k_1, B_o, and B_1), the calibration experiment (number and spacing of standard samples), and the distribution of the population being sampled. The number of descriptors can be reduced by considering fundamental characteristics of the measurement system.

Fundamental characteristics should include measurement sensitivity and a description of the population being sampled. For this chapter, measurement sensitivity will be expressed by slope sensitivity (Z), given by equation (II.19):

$$(III.4) \quad Z = B_1/\sigma(b_1)$$

The object of the illustration is to study the properties of measurements which are inherently imprecise. Therefore, the calibration experiment should be designed to provide information about results near limits of detection. A calibration range of $[0, X_{LOQ}]$, where LOQ refers to "limit of quantification", covers the range in which results are not considered precise enough to report (ACS, 1980). The LOQ is defined here as the point at which the relative standard deviation (RSD) of a signal is 10%:

$$(III.5) \quad RSD = \sigma(e_s)/S$$

$$0.10 = (k_o + k_1 \cdot X_{LOQ})^{\frac{1}{2}}/S_{LOQ}$$

where S_{LOQ} is the signal produced by a sample with concentration X_{LOQ}. The calibration design should also reflect practical considerations. In practice, equally spaced, equally weighted designs are often employed. Such a design with 2 points has minimum variance, but no power to detect non-linear functions. Therefore, a 3 point design with points at [0, $X_{LOQ}/2$, X_{LOQ}] was chosen. Without loss of generality, the true calibration slope and intercept were set to 1 and 0, respectively, and the LOQ was set to 1. Each measurement was derived from a separate calibration experiment.

For this design and the error model of equation (III.1), slope sensitivity is described completely by a ratio of multiplicative to additive error, standardized by the slope of the calibration function.

$$R = k_1^2/(k_o B_1^2) \qquad \text{(III.6)}$$

Then,

$$Z^2 = [1250n/(50R+3)]\{[0.04\cdot(100R+1)\cdot(25R+1) + 0.2\cdot R^{1/2}\cdot(25R+1)^{1/2}(100R+3)]/[1+50\rho]\} \qquad \text{(III.7)}$$

where,

$$\rho = R + (R^2 + 0.04R)^{1/2}$$

For R large,

$$Z = (50n)^{1/2} \qquad \text{(III.8)}$$

The minimum value occurs at R = 0, where $Z = (50n/3)^{1/2}$. Figure III.1 shows the relationship between Z and R for three different values of n (number of calibration

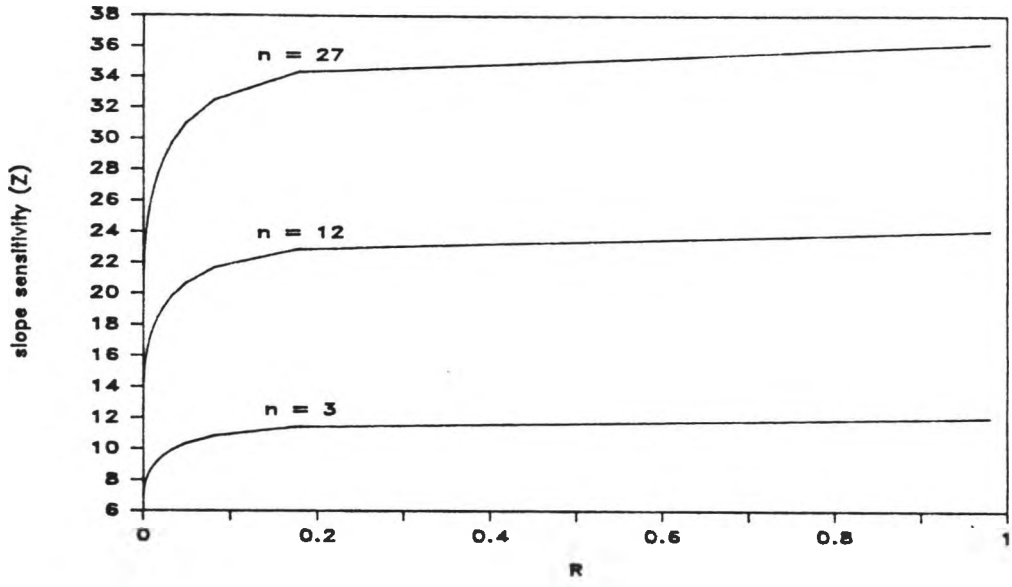

Figure III.1 Slope sensitivity (Z) vs. the ratio of multiplicative to additive error (R) for 3 point equal weight, equal spaced design

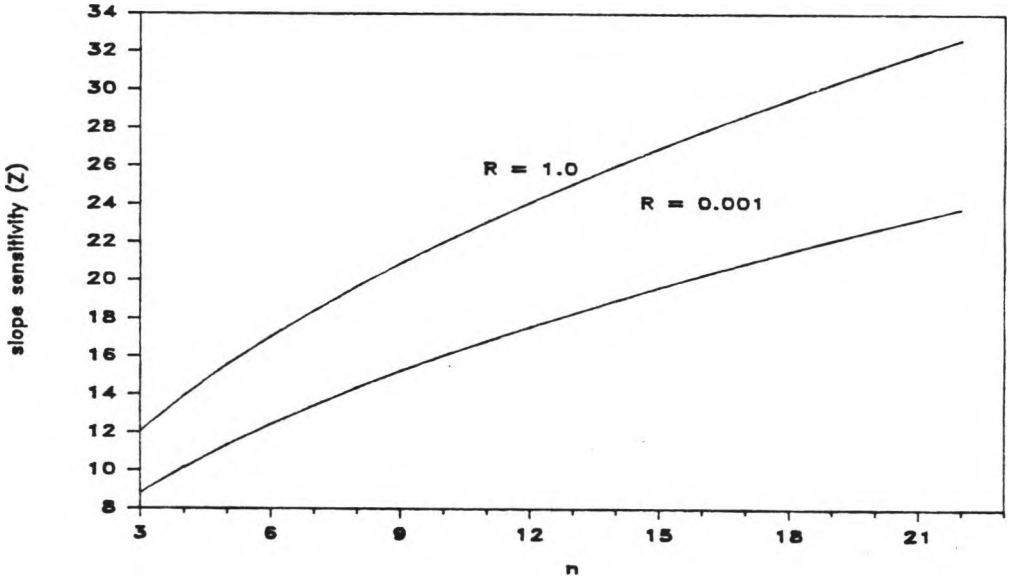

Figure III.2 Slope sensitivity (Z) vs. number of standard samples (n) for 3 point equal weight, equal spaced design

standards). Figure III.2 shows slope sensitivity as a function of the number of standard samples (n) for two different R values.

Lognormal distributions with skewness coefficients ranging from 0.1 to 5 were chosen to represent water quality random variables. (See chapter V for further discussion of the distribution of water quality random variables.) The mean and coefficient of skewness were chosen so that a variety of "censoring" conditions would exist and results could tentatively be compared with methods developed for censored data. Nominal censoring conditions were expressed as the fraction of true concentration below the LOD (FC) and the LOQ (FBQ). Most water quality data is currently censored at the LOD and it has been suggested in this thesis that the LOQ might be a more appropriate censoring level. The term "FC,FBQ" is used to describe the fraction of a sample which would be censored under each convention. For example, FC,FBQ = 0.5,1 refers to a distribution which would be approximately 50% ND's if censored at the LOD and approximately 100% ND's if censored at the LOQ. Specifying the FC, or FBQ, along with the population coefficient of skewness, fixes the population mean. Plots of the PDF's for the distributions used can be found in appendix C.

The measurement system can now be specified by describing the slope sensitivity (or, equivalently, R), the number of standard samples (n), the FC or FBQ of the population, and the population coefficient of skewness. Note again that

Z is completely determined for R and n. Therefore, a variety of simulation conditions can be described (using the POE approximation) as a function of R and n for a given X_{LOQ}. R may also be interpreted as the relative size of X_{LOD} compared to X_{LOQ}. For large R, X_{LOD}/X_{LOQ} approaches zero. For small R, this ratio approaches 0.3. R is a characteristic of the measurement system that is independent of calibration design. Values used for simulations are summarized in Table III.1. Parameters were chosen in order to provide a range of populations (coefficients of skewness), measurement systems (range of R values), and censoring conditions (FC,FBQ values).

B. The POE approximation to properties of measurements

The variance terms of equation III.3 are estimated as follows:

$$\text{(III.9)} \qquad \sigma^2(b_o) = \sum_{i=1}^{n} w_i X_{i,c}^2 / D$$

$$\sigma^2(b_1, b_o) = -\sum_{i=1}^{n} w_i X_{i,c} / D$$

$$\sigma^2(b_1) = \sum_{i=1}^{n} w_i / D$$

$$D = \sum_{i=1}^{n} w_i \sum_{i=1}^{n} w_i X_{i,c}^2 - \left(\sum_{i=1}^{n} w_i X_{i,c} \right)^2$$

Table III.1 Summary of simulation conditions

a. Simulation conditions

R	x	(FC,FBQ)	x	Skewness
1.0		1.0, 1.0		0.1
0.1		0.5, 1.0		0.5
0.01		0.0, 0.5		1.0
0.001				3.0
				5.0

b. Error Parameters

R	Z	X_{LOD}	k_o	k_1
1.0	24	0.0297	0.00009805	0.009902
0.1	22	0.0869	0.00083920	0.009161
0.01	18	0.1854	0.00382000	0.006180
0.001	15	0.2563	0.00729800	0.002702

c. Distribution parameters

FC,FBQ = 1,1; R = 1.0

R	skew.	σ^2	μ_p	α_p	β_p^2	η_p^2	std.dev. μ from ln of: LOQ	LOD
1.0	0.1	0.00111	-3.617	0.0269	8.03e-7	0.0011	108.6	3.0
	0.5	0.02691	-4.009	0.0184	9.24e-6	0.0273	24.4	3.0
	1	0.09877	-4.459	0.0122	1.53e-5	0.1038	14.2	3.0
	3	0.512	-5.663	0.0045	1.34e-5	0.6686	7.9	3.0
	5	0.8468	-6.277	0.0029	1.10e-5	1.3322	6.8	3.0
0.1	0.1	0.00111	-2.543	0.0787	6.87e-6	0.0011	76.3	3.0
	0.5	0.02691	-2.935	0.0538	7.9 e-5	0.0273	17.9	3.0
	1	0.09877	-3.386	0.0356	1.31e-4	0.1038	10.8	3.0
	3	0.512	-4.590	0.0131	1.15e-4	0.6686	6.4	3.0
	5	0.8468	-5.204	0.0084	9.3 e-5	1.3322	5.7	3.0
0.01	0.1	0.00111	-1.785	0.1679	3.1 e-5	0.0011	53.6	3.0
	0.5	0.02691	-2.177	0.1149	3.59e-4	0.0273	13.3	3.0
	1	0.09877	-2.628	0.0759	5.97e-4	0.1038	8.4	3.0
	3	0.512	-3.832	0.0280	5.23e-4	0.6686	5.4	3.0
	5	0.8468	-4.446	0.0179	4.27e-4	1.3322	4.8	3.0
0.001	0.1	0.00111	-1.461	0.2321	5.9 e-5	0.0011	43.9	3.0
	0.5	0.02691	-1.854	0.1588	6.87e-4	0.0273	11.3	3.0
	1	0.09877	-2.304	0.1049	1.14e-3	0.1038	7.3	3.0
	3	0.512	-3.508	0.0387	1.00e-3	0.6686	4.9	3.0
	5	0.8468	-4.122	0.0248	8.16e-4	1.3322	4.5	3.0

Table III.1 (continued)

FC,FBQ = 0.5,1;

R	skew.	σ^2	μ_p	α_p	β_p^2	η_p^2	std.dev. μ from ln of: LOQ	LOD
1.0	0.1	0.00111	-3.517	0.0297	9.80e-7	0.0011	105.6	0.0
	0.5	0.02691	-3.517	0.0301	2.4 e-5	0.0273	21.4	0.0
	1	0.09877	-3.517	0.0312	1.01e-4	0.1038	11.2	0.0
	3	0.512	-3.517	0.0384	9.84e-4	0.6686	4.9	0.0
	5	0.8468	-3.517	0.0454	2.74e-3	1.3322	3.8	0.0
0.1	0.1	0.00111	-2.443	0.0870	8.39e-6	0.0011	73.3	0.0
	0.5	0.02691	-2.443	0.0881	2.11e-4	0.0273	14.9	0.0
	1	0.09877	-2.443	0.0913	8.65e-4	0.1038	7.8	0.0
	3	0.512	-2.443	0.1123	8.43e-3	0.6686	3.4	0.0
	5	0.8468	-2.443	0.1327	2.35e-2	1.3322	2.7	0.0
0.01	0.1	0.00111	-1.685	0.1855	3.82e-5	0.0011	50.6	0.0
	0.5	0.02691	-1.685	0.1879	9.63e-4	0.0273	10.3	0.0
	1	0.09877	-1.685	0.1948	3.94e-3	0.1038	5.4	0.0
	3	0.512	-1.685	0.2395	3.83e-2	0.6686	2.4	0.0
	5	0.8468	-1.685	0.2831	1.07e-1	1.3322	1.8	0.0
0.001	0.1	0.00111	-1.361	0.2564	7.30e-5	0.0011	40.9	0.0
	0.5	0.02691	-1.361	0.2598	1.84e-3	0.0273	8.3	0.0
	1	0.09877	-1.361	0.2693	7.53e-3	0.1038	4.3	0.0
	3	0.512	-1.361	0.3311	7.33e-3	0.6686	1.9	0.0
	5	0.8468	-1.361	0.3914	2.04e-1	1.3322	1.5	0.0

FC,FBQ = 0, 0.5

1.0	0.1	0.00111	0	1.0006	1.11e-3	0.0011	0.0	105.6
	0.5	0.02691	0	1.0135	2.80e-2	0.0273	0.0	21.4
	1	0.09877	0	1.0506	1.15	0.1038	0.0	11.2
	3	0.512	0	1.2918	1.12	0.6686	0.0	4.9
	5	0.8468	0	1.5271	3.11	1.3322	0.0	3.8
0.1	0.1	0.00111	0	1.0006	1.11e-3	0.0011	0.0	73.3
	0.5	0.02691	0	1.0135	2.80e-2	0.0273	0.0	14.9
	1	0.09877	0	1.0506	1.15	0.1038	0.0	7.8
	3	0.512	0	1.2918	1.12	0.6686	0.0	3.4
	5	0.8468	0	1.5271	3.11	1.3322	0.0	2.7
0.01	0.1	0.00111	0	1.0006	1.11e-3	0.0011	0.0	50.6
	0.5	0.02691	0	1.0135	2.80e-2	0.0273	0.0	10.3
	1	0.09877	0	1.0506	1.15	0.1038	0.0	5.4
	3	0.512	0	1.2918	1.12	0.6686	0.0	2.4
	5	0.8468	0	1.5271	3.11	1.3322	0.0	1.8
0.001	0.1	0.00111	0	1.0006	1.11e-3	0.0011	0.0	40.9
	0.5	0.02691	0	1.0135	2.80e-2	0.0273	0.0	8.3
	1	0.09877	0	1.0506	1.15	0.1038	0.0	4.3
	3	0.512	0	1.2918	1.12	0.6686	0.0	1.9
	5	0.8468	0	1.5271	3.11	1.3322	0.0	1.5

$$(III.10)\quad \sum_{i=1}^{n} w_i = n[1/k_o+2/(2k_o+k_1X_c)+1/(k_o+k_1X_c)]/3$$

$$\sum_{i=1}^{n} w_iX_{i,c} = nX_c[1/(2k_o+k_1X_c) + 1/(k_o+k_1X_c)]/3$$

$$\sum_{i=1}^{n} w_iX_{i,c}^2 = nX_c^2[1/(4k_o+2k_1X_c) + 1/(k_o+k_1X_c)]/3$$

$$w_i = (k_o + k_1X_{i,c})^{-1}$$

where $X_{i,c}$ refers to the n calibration standards, and k_o and k_1 are related to the variance of an analytical signal, as described in Chapter II.

The variance of X_m given X_p can be written as a polynomial in X_p:

$$(III.11)\quad Var(X_m/X_p) = r + sX_p + tX_p^2$$

$$\text{where}\quad r = (k_o/\kappa + \Sigma w_iX_{i,c}^2/D)/B_1^2$$

$$s = (k_1/\kappa - 2\Sigma w_iX_{i,c}/D)/B_1^2$$

$$t = \sum_{i=1}^{n} w_i/(DB_1^2)$$

In terms of R, for $\kappa = 1$, $B_o = 0$, $B_1 = 1$, and $X_c = X_{LOQ} = 1$,

$$(III.12)\quad r = (k_o/B_1^2)\{1 + (2.5/n)\cdot[1 + 30\rho]/[1+25\rho]\}$$

$$s = (k_1/B_1^2)\{1 - [1/(25n)]\cdot[3 + 100\rho]$$

$$/[\rho(1+25\rho)]\}$$

$$t = 2\cdot[R/(625n)]\cdot[3 + 150\rho + 1250\rho^2]$$

$$/[\rho^2+25\rho^3]$$

$$\rho = R + (R^2 + 0.04R)^{1/2}$$

As n becomes large, $r \to k_o/B_1^2$, $s \to k_1/B_1^2$, $t \to 0$ and

$$(III.13)\quad Var(X_m \text{ given } X_p) \to (k_o+k_1X_p)/B_1^2$$

Figure III.3 shows the relative standard deviation of a measurement as a function of X_p for two different R values. Figure III.4 shows the relative standard deviation of X_m as

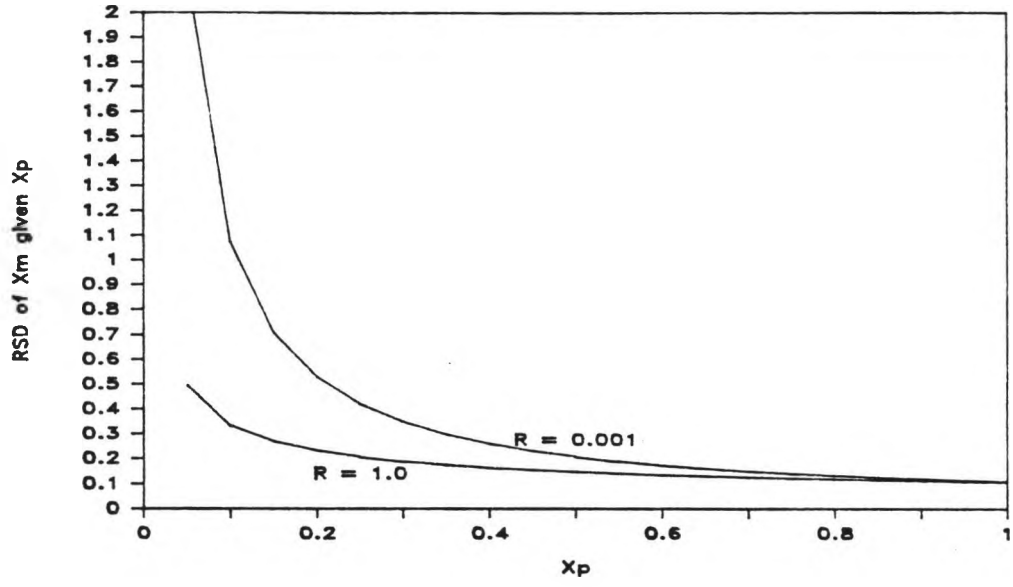

Figure III.3 Relative standard deviation of X_m as a function of X_p

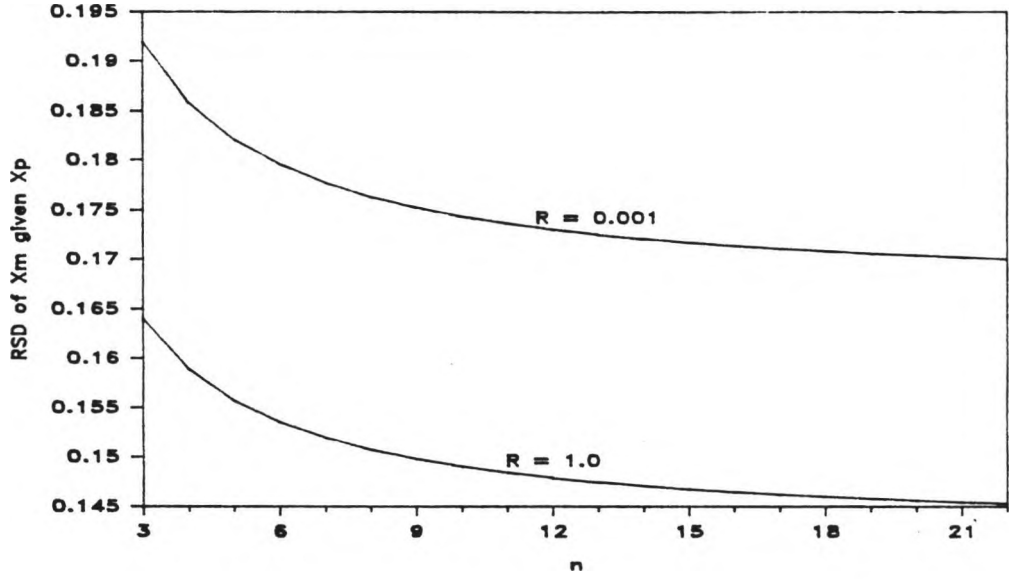

Figure III.4 Relative standard deviation of X_m as a function of the number of standard samples (n)

a function of n for X_p = 1/2 for the same two R values. These figures illustrate, for a limited number of cases, the influence of calibration design on data quality. Simulations were done, treating X_p as a constant, to evaluate the assumption of normality for X_m given X_p. A given value of X_m was 'measured' 1000 times for each measurement system. A Kolmogorov-Smirnov test for normality of X_m given X_p passed at the 80% confidence level.

The unconditional PDF of X_m can be expressed in integral form:

$$(III.14) \quad f(X_m) = \int_D f(X_m/X_p) \cdot f(X_p) dX_p$$

where D represents the domain of possible values for X_p. When X_p is distributed normally with mean μ_p and variance σ_p^2,

$$(III.15) \quad f(X_m) = \int [2\pi\sigma_m^2]^{-\frac{1}{2}} \exp\{-\tfrac{1}{2}[(X_m - X_p)/\sigma_m]^2\} \cdot (2\pi\sigma_p^2)^{-\frac{1}{2}} \exp\{-\tfrac{1}{2}[(X_p - \mu_p)/\sigma_p]^2\} dX_p$$

where the integration is from $-\infty$ to $+\infty$. For this and all but very simple distributions, the expression must be evaluated numerically. However, evaluation of the moments of $f(X_m)$ is relatively straightforward. The first moment about zero is simply the mean. The variance can be found by expressing $f(X_m)$ as a product of $f(X_m/X_p)$ and $f(X_p)$ and integrating.

$$\text{(III.16)}\quad \mathrm{Var}(X_m) = \sigma_m^2$$

$$= E(X_m^2) - [E(X_m)]^2$$

$$= \int\int X_m^2 f(X_m/X_p) f(X_p)\, dX_m dX_p - \left[\int\int X_m f(X_m/X_p) f(X_p)\, dX_m dX_p\right]^2$$

$$= \int f(X_p)\left[\int X_m^2 f(X_m/X_p)\, dX_m\right] dX_p - \left[\int f(X_p)\left[\int X_m f(X_m/X_p)\, dX_m\right] dX_p\right]^2$$

$$= \int f(X_p)\left[(E(X_m/X_p))^2\right] dX_p - \left[\int f(X_p) E(X_m/X_p)\, dX_p\right]^2$$

$$= \int f(X_p)\left[\mathrm{Var}(X_m/X_p) + (E(X_m/X_p))^2\right] dX_p - (E(X_p))^2$$

$$= \int f(X_p)\left[r + sX_p + tX_p^2 + X_p^2\right] dX_p - \mu_p^2$$

$$= r + s\mu_p + (t+1)\sigma_p^2 + t\cdot\mu_p^2$$

For X_p normally distributed with mean μ_p and variance σ_p^2, the coefficient of skewness is given by:

$$\text{(III.17)}\quad \text{Coef. of Skewness} = 3\sigma_p^2(s + 2\mu_p t)/\sigma_m^3$$

For X_p lognormally distributed with mean α_p and variance β_p^2,

$$\text{(III.18)}\quad \text{Coef. Skewness} = 3\eta_p^2\alpha_p^2[s + \eta_p^4\alpha_p(1/3+t) + \eta_p^2\alpha_p(1+t) + 2\alpha t]/\sigma_m^3$$

where $\eta_p = \beta_p/\alpha_p$. Similarly, the coefficient of kurtosis is given by:

$$\text{(III.19)}\quad \text{Coef. kurtosis} = \{[3\cdot(s^2+2rt+2r) + 6s(3t+1)\mu_p + 18t^2\mu_p^2]\sigma_p^2 + 3(3t^2+6t+1)\sigma_p^4 + 3t(t-2)\mu_p^4 + 6s(t-1)\mu_p^3 + 3(s^2+2rt-2r)\mu_p^2 + 6rs\mu_p + 3r^2)\}/\sigma_m^4$$

for X_p normally distributed, and

(III.20) Coef. kurtosis =

$$\{3\cdot\alpha_p^2\cdot(\eta_p^2+1)\cdot[s^2+2r(t+1)-4\alpha_p s+2(t+1)\alpha_p^2]$$
$$+\ 2\cdot\alpha_p^3\cdot(\eta_p^2+1)^3\cdot[3s(t+1)-2\alpha_p(3t+1)]$$
$$+\ \alpha_p^4\cdot(\eta_p^2+1)^6\cdot[3t^2+6t+1]$$
$$+\ 3(r^2+2rs\alpha_p+2s\alpha_p^3-2\alpha_p^2 r-\alpha_p^4)\}/\sigma_m^4$$

for X_p log-normally distributed.

Sample properties (mean, variance, and coefficients of skewness and kurtosis) calculated using the POE approximation were compared with those based on simulations. The unconditional properties of X_m were determined by "sampling" and "analyzing" from a given distribution 1000 times. The calibration design described above with n = 12 was used in all cases.

Figures III.5 through III.13 compare POE approximations with simulated results, for the coefficient of variation, skewness, and kurtosis. Figures III.5 through III.7 compare the POE estimate of the coefficient of variation of X_m (the expression for standard deviation given by equation III.16 divided by the population mean) with that from simulations. Figures III.8 through III.10 compare skewness coefficient calculations from equation III.18 (based on the POE estimate) with the simulated coefficient of skewness of the measurements. For large amounts of censoring (FB:FBQ = 1,1), the POE approximation is not too good, but skewness values are small. For all other conditions examined, the POE based

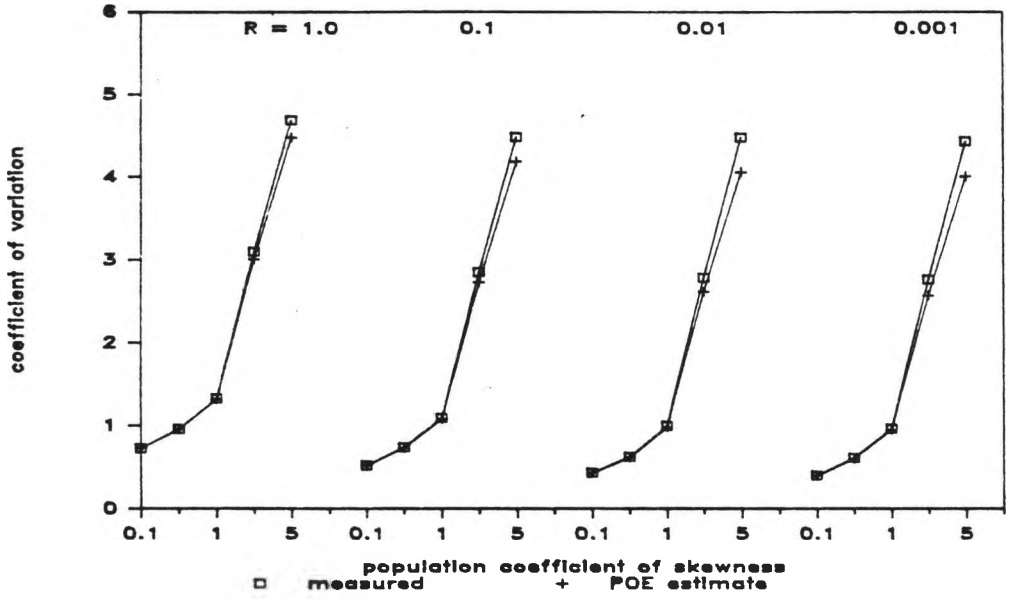

Figure III.5 Comparison of the POE estimate of the measured coefficient of variation with simulated values for FC, FBQ = 1,1

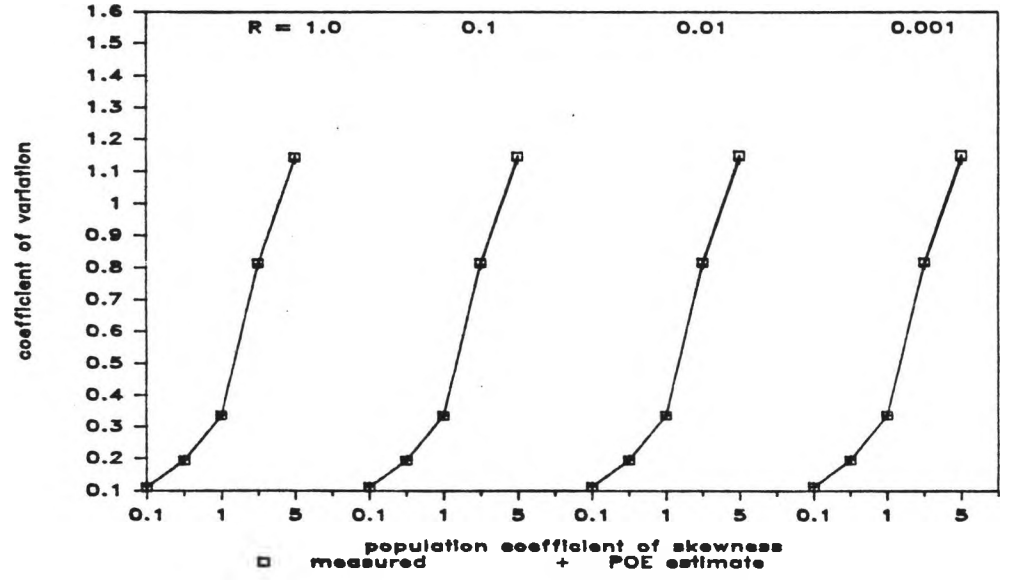

Figure III.6 Comparison of the POE estimate of the measured coefficient of variation with simulated values for FC, FBQ = 0.5,1

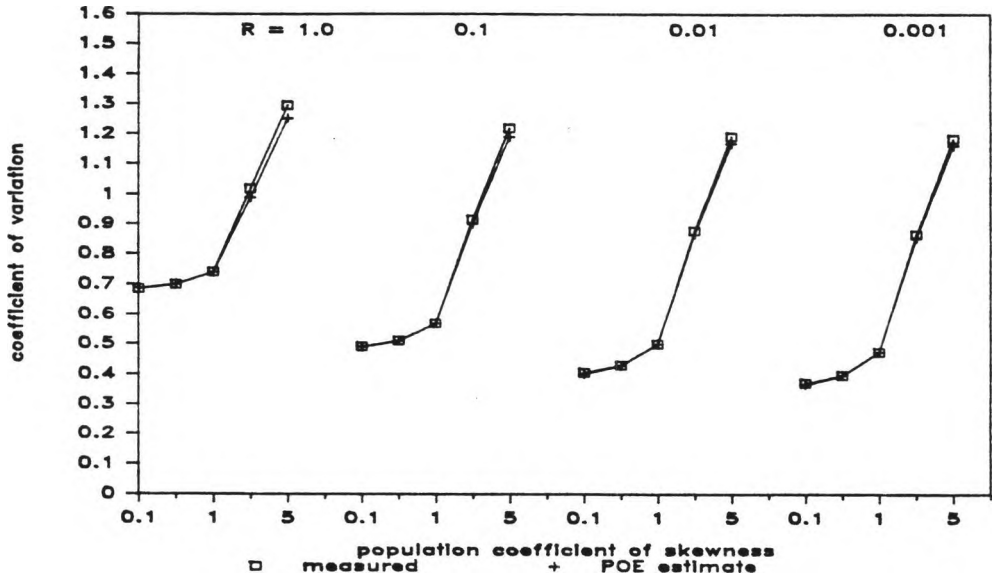

Figure III.7 Comparison of the POE estimate of the measured coefficient of variation with simulated values for FC, FBQ = 0,0.5

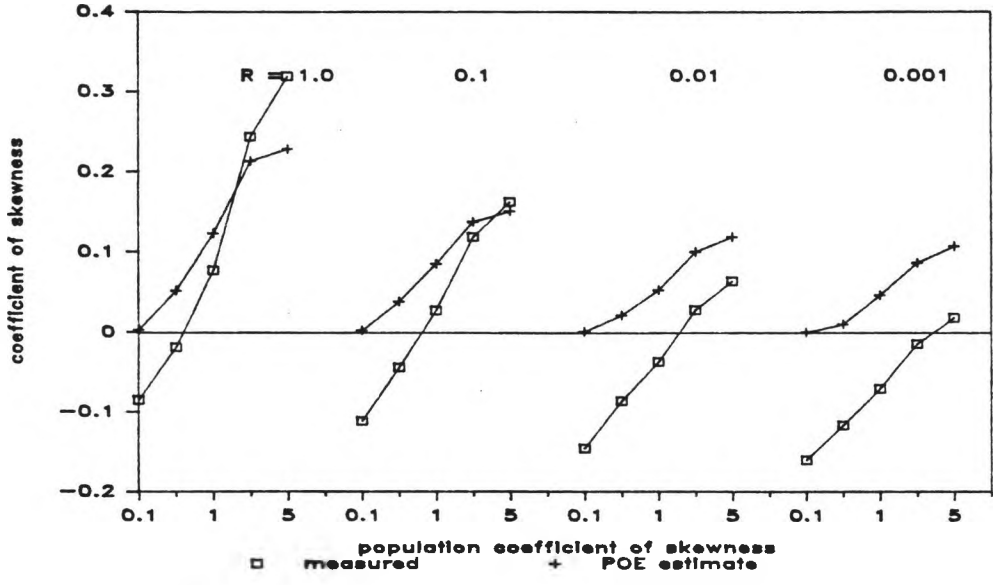

Figure III.8 Comparison of the POE estimate of the measured coefficient of skewness with simulated values for FC,FBQ = 1.0, 1.0

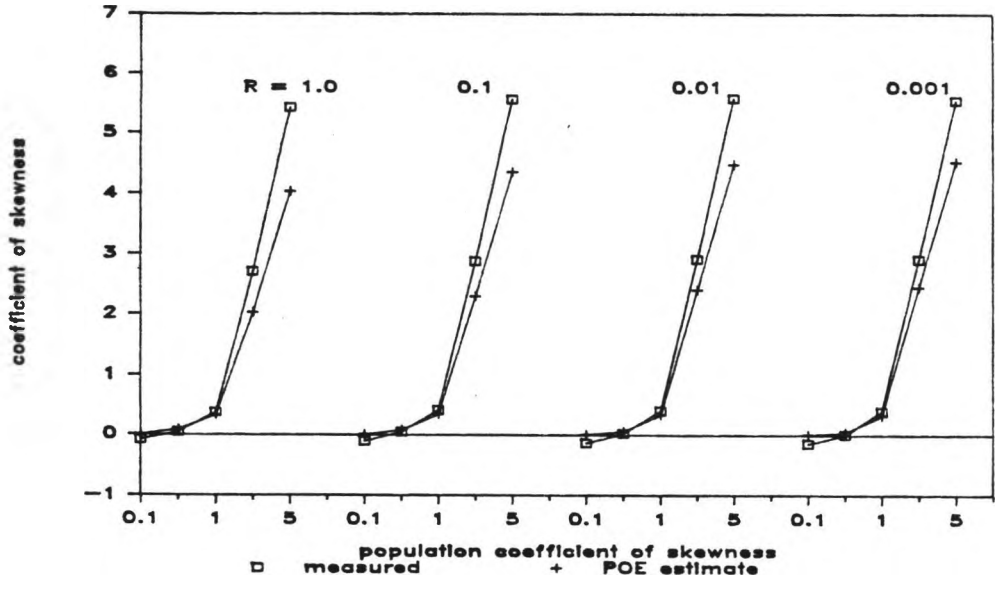

Figure III.9 Comparison of the POE estimate of the measured coefficient of skewness with simulated values for FC,FBQ = 0.5, 1.0

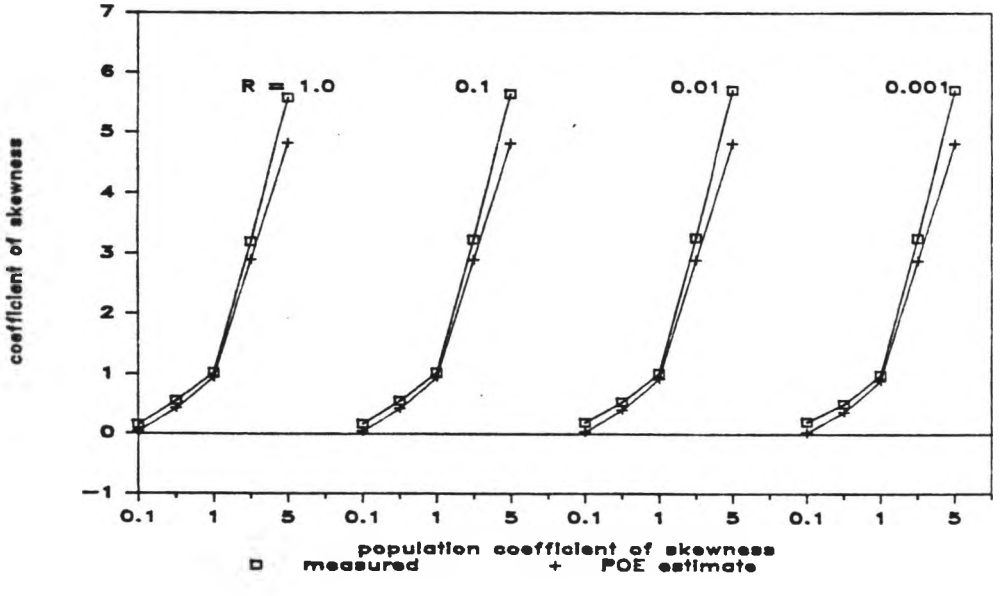

Figure III.10 Comparison of the POE estimate of the measured coefficient of skewness with simulated values for FC,FBQ = 0.0, 0.5

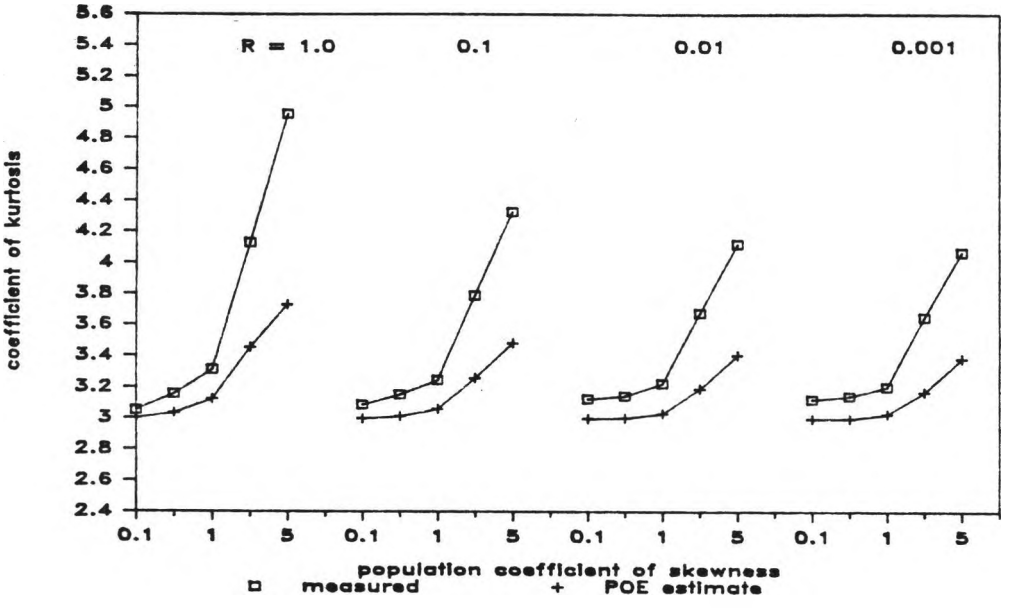

Figure III.11 Comparison of the POE estimate of the measured coefficient of kurtosis with simulated values for FC,FBQ = 1.0, 1.0

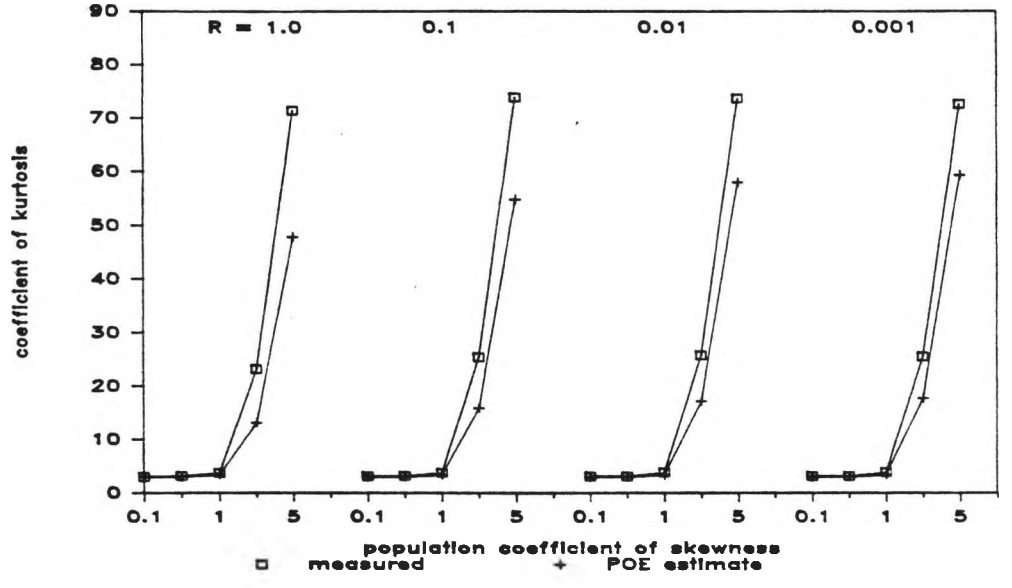

Figure III.12 Comparison of the POE estimate of the measured coefficient of kurtosis with simulated values for FC,FBQ = 0.5, 1.0

approximation was quite good. Figures III.11 through III.13 compare the same information for the coefficient of kurtosis.

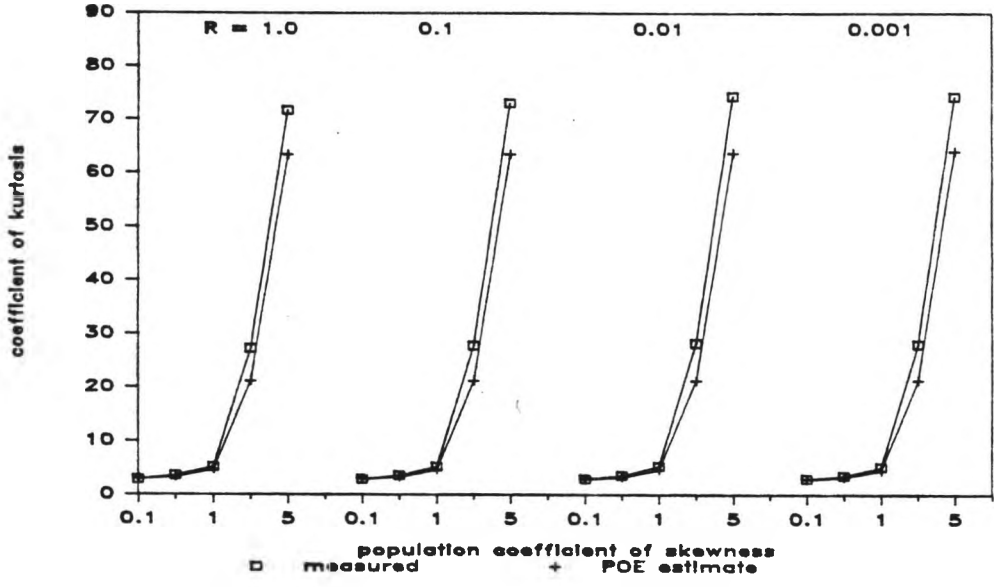

Figure III.13 Comparison of the POE estimate of the measured coefficient of kurtosis with simulated values for FC,FBQ = 0.0, 0.5

C. Comparing measurement and population properties

The simulated properties of X_m were also compared to X_p. These results are summarized in Figures III.14 - III.20. The effect of normally distributed measurement error on the observed, or measured data can be striking. For example, in the absence of measurement error, the unconditional variance becomes, simply, the population variance. Skewness becomes zero for the normal case, and $\eta_p^3+3\eta_p$ for the lognormal case, as expected. However, when measurement error dominates, the properties of X_m approach those for a normal distribution regardless of the population distribution.

Figure III.14 shows the fractional bias in the mean as a fraction of the true mean for each of the simulation conditions.

(III.21) Frac. Bias = (measured - true mean)/true mean

Significant negative bias occurred for populations with skewness coefficients of 3 and 5. Bias increases as the population median moves further from the center of the calibration curve. Positive bias occurs for the population with the largest mean because most samples are from the high end of the calibration curve. A quantitative expression of bias is given by equation II.20.

Figures III.15 and III.16 show the coefficient of variation of X_m as a function of the population coefficient of variation. Figure III.16 shows populations with different 'FC,FBQ' values. Solid lines indicate where measured values

would equal population values. Figures III.17 through III.20 show similar information for the coefficients of skewness and kurtosis, respectively.

Measurements made near limits of detection do not have the same statistical properties as the population. These figures can be used to classify systems (combinations of measurement error and water quality distributions) as being dominated by measurement error (moment estimates approximate those for a normal distribution) or by population fluctuations (moment estimates approximate those of the population).

One begins to see the problem with inferences regarding a population when measurement error becomes very much larger than population variance. In some cases numbers would be reported and used even though what is being observed bears little resemblance to the water quality variable.

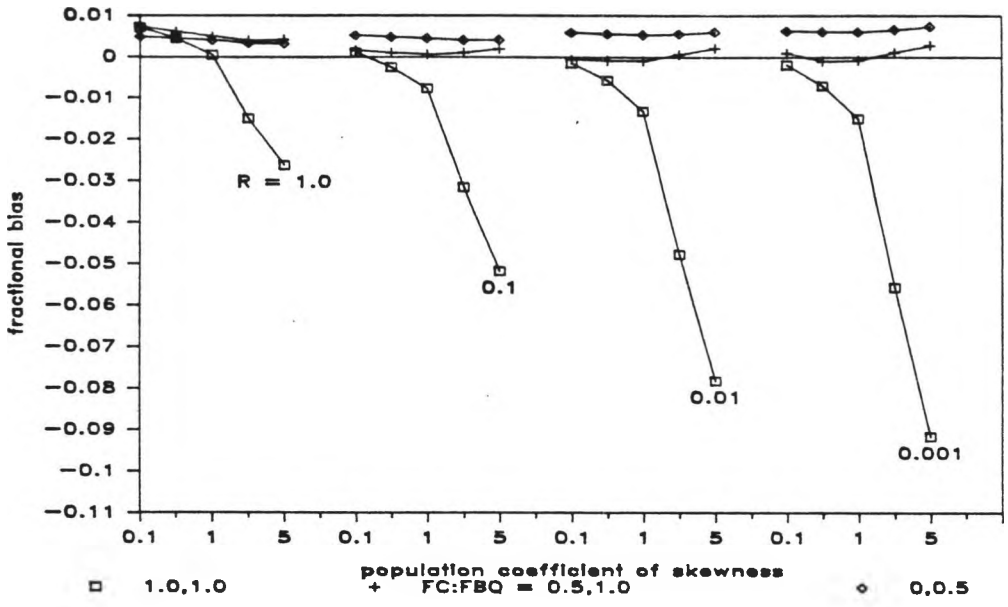

Figure III.14 Fractional bias of the estimate of the mean for the conditions described in Table III.1

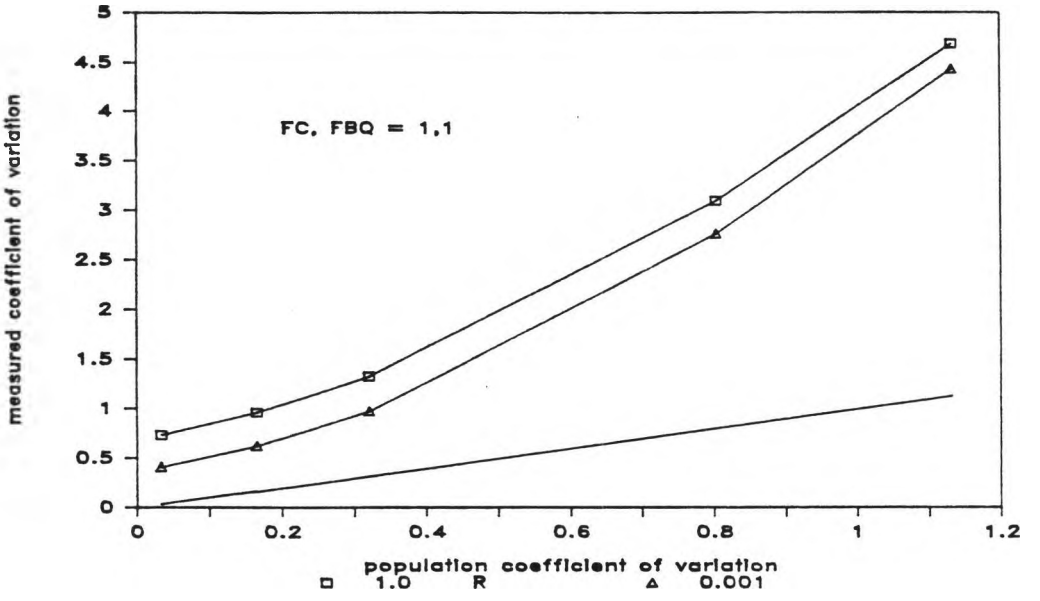

Figure III.15 The coefficient of variation of measurements as a function of population coefficient of variation for several R values.

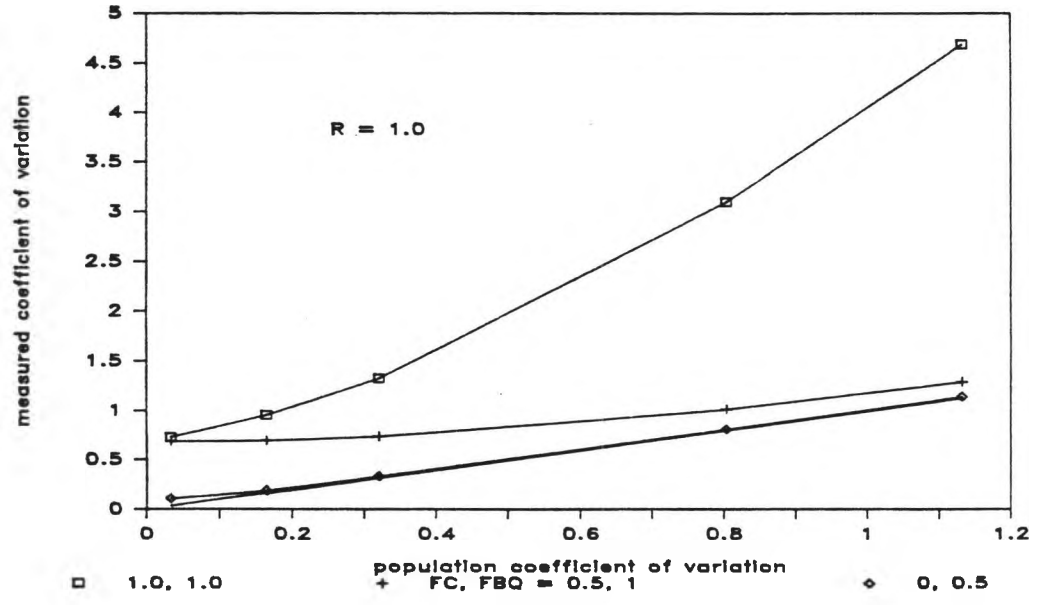

Figure III.16 The coefficient of variation of measurements as a function of population coefficient of variation for several FC,FBQ values.

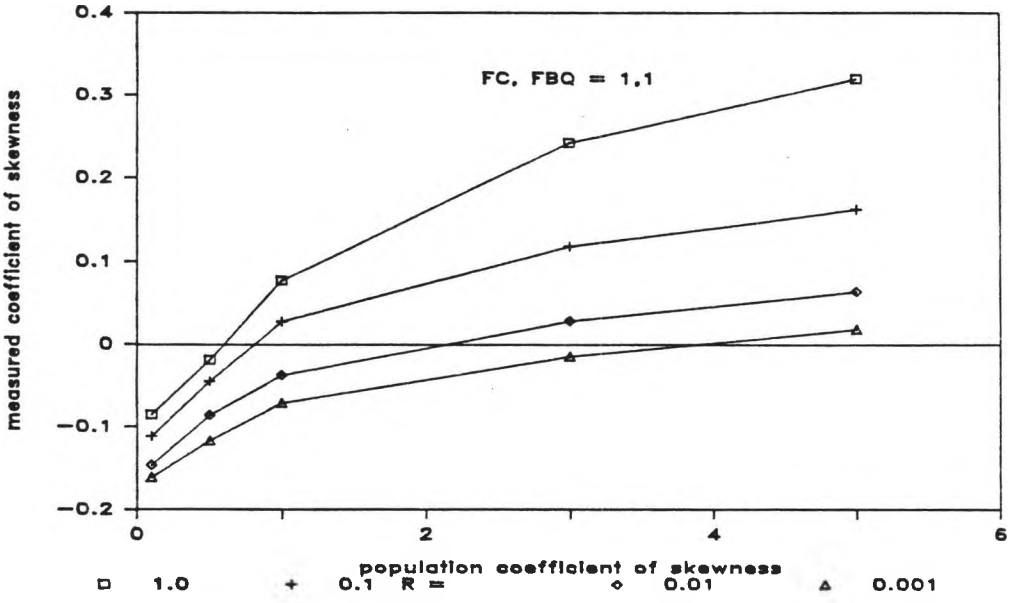

Figure III.17 The coefficient of skewness of measurements as a function of population coefficient of variation for several R values.

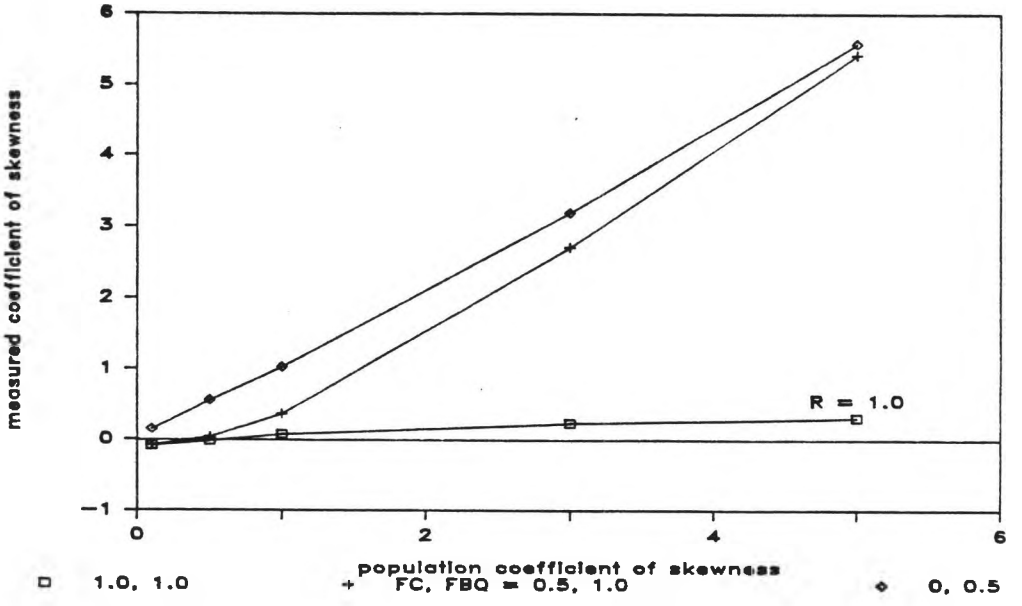

Figure III.18 The coefficient of skewness of measurements as a function of population coefficient of variation for several FC,FBQ values.

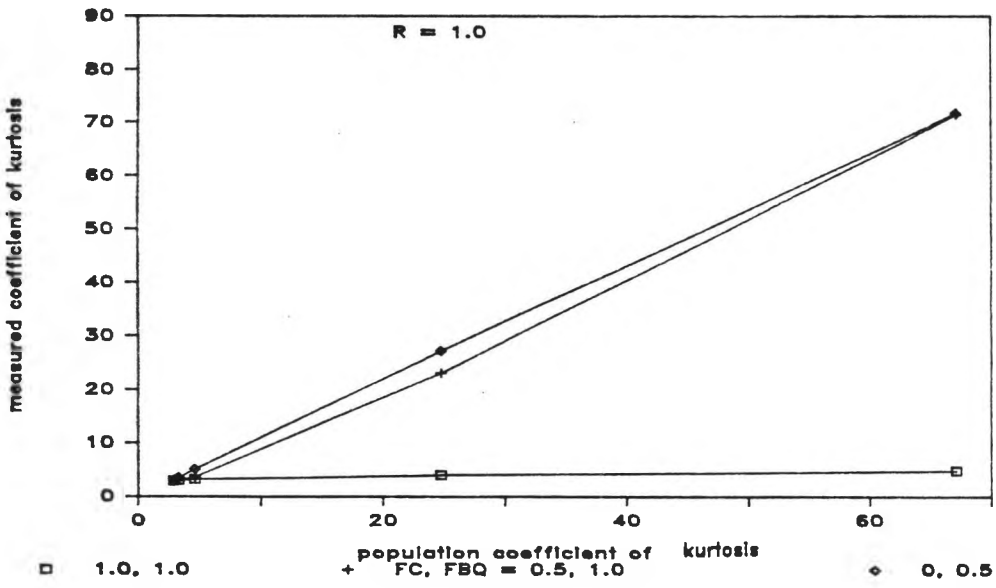

Figure III.19 The coefficient of kurtosis of measurements as a function of population coefficient of variation for several R values.

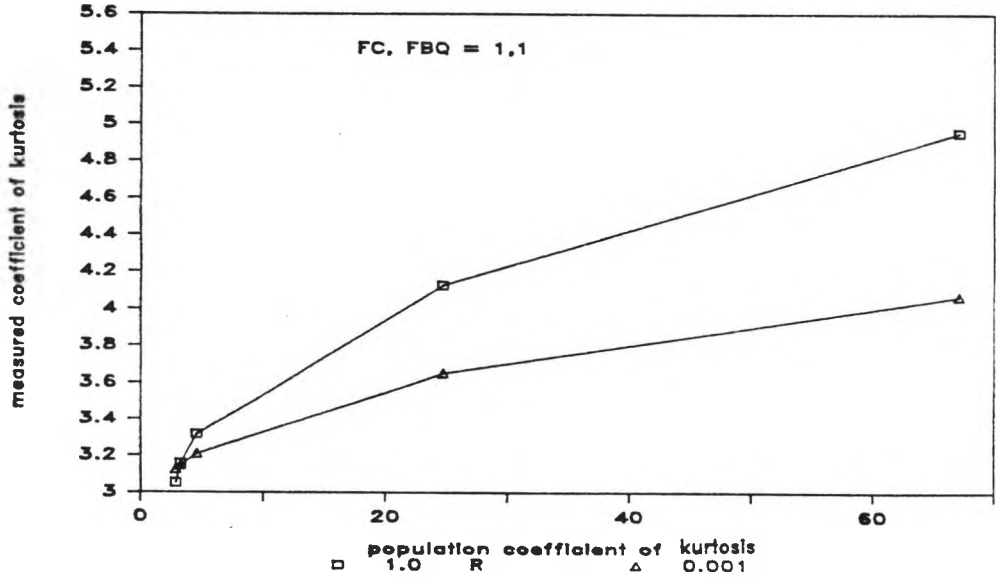

Figure III.20 The coefficient of kurtosis of measurements as a function of population coefficient of variation for several FC,FBQ values.

IV. STATISTICAL ANALYSIS WHICH INCLUDES MEASUREMENT ERROR

It has been suggested that statistical data analysis should in some way acknowledge observation error. This chapter uses an example of estimating a population mean from a set of observations to illustrate this.

A. Estimating the mean of a normal distribution

The most simple problem considered in this section is the construction of a confidence interval about the mean of a normally distributed population given normally distributed independent observation error. The problem is made more complex by acknowledging that measurements are calibrated and as such are not normally distributed. Still another approach to this problem is to estimate and remove components of variance due to the observation error. The object of this is to draw information from the data which pertains to the underlying population. This is a departure from solutions which do not distinguish measurements from true population values.

1. Confidence intervals when system error is included

Assume that a water quality random variable is normally distributed [$X_p \approx N(\mu_p, \sigma_p^2)$]. Were it possible to make observations without error,

$$(IV.1) \quad \mu_m = \mu_p \quad \text{and} \quad \sigma_m^2 = \sigma_p^2$$

A $1-\alpha$ confidence interval for μ_p is given by:

$$(IV.2) \quad CI = \bar{X}_m \pm t(1-\alpha, n-1) \cdot \hat{\sigma}_m / \sqrt{n}$$

where $\hat{\sigma}_m^2 = \sum_{i=1}^{n} (X_{i,m} - \bar{X}_m)^2 / (n-1)$, and n is the sample size.

The information content of the sample with regard to this problem is related to the size of the confidence interval. Observation error decreases information by increasing the variance of X_m and therefore $\hat{\sigma}_m$. For example, if one admits to an unbiased error (ε) independent of X_p and normally distributed,

$$(IV.3) \quad X_m = X_p + \varepsilon,$$

$$\varepsilon \approx N[0, \sigma^2(\varepsilon)]$$

then, $X_m \approx N(\mu_p, \sigma_m^2)$

where $\sigma_m^2 = \sigma_p^2 + \sigma^2(\varepsilon)$ $\hat{\mu}_p = \bar{X}_m$, and $\hat{\sigma}_m^2$ is as before.

A $1-\alpha$ confidence interval given by equation (IV.2) would be larger than before.

A more quantitative expression of information content for this problem may be given by the relative efficiency (RE) of $\hat{\mu}_p$. The efficiency of the sample mean of error-free observations relative to those which contain error is given by:

$$(IV.4) \quad RE = \sigma_p^2 / \sigma_m^2$$

Water quality observations, however, are calibrated results. The distribution of X_m given X_p is Cauchy. Not only will the confidence interval for μ_p using equation (IV.2) be larger than for error free data, it will not be with confidence $1-\alpha$ but something less. Mulrow (1986) showed that for $\hat{\mu}_p$ and signal error normally distributed, it is more accurate to calculate a confidence interval using the POE approximation or SIDF. These are given by equations (II.29) and (II.30), respectively. For the POE approximation, and a single unknown measurement per sample, the efficiency of $\hat{\mu}_p$ relative to error free estimates is given by:

$$(IV.5) \quad RE = \sigma_p^2/\{\sigma_p^2+[\sigma^2(b_o)+2\ \sigma^2(b_o,b_1)\mu_p+\sigma^2(b_1)\mu_p^2]/B_1^2\}$$

2. Removing components of variance due to system error

Using SIDF or POE helps to eliminate the problem caused by assuming the wrong distribution for X_m. However, one is left with a larger confidence interval and a low relative efficiency for $\hat{X}_p$ as compared with error free observations. One could alleviate this problem by estimating and subtracting components of variance due to observation error. One can write:

$$(IV.6) \quad \sigma_m^2 = \sigma_p^2 + \sigma^2(\varepsilon)$$

then,
$$\sigma_p^2 = \sigma_m^2 - \sigma^2(\varepsilon)$$

Satterthwaite (1946) and Gaylor and Hopper (1969) discuss the distribution of linear combinations of variance estimates (sums of squares). This forms the basis for the SIDF approach. Satterthwaite shows that a linear combination

of sample variance estimates can be approximated as chi-squared (χ^2) with "improved" degrees of freedom. e.g.,

$$(IV.7) \quad \hat{\sigma}^2 = a_1\hat{\sigma}_1^2 + a_2\hat{\sigma}_2^2 \approx \sigma^2\chi_f^2/f$$

where $f = (a_1\hat{\sigma}_1^2 + a_2\hat{\sigma}_2^2)^2/[(a_1\hat{\sigma}_1^2)^2/f_1 + (a_2\hat{\sigma}_2^2)^2/f_2]$ and f_1 and f_2 are the degrees of freedom of the estimates $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$, respectively. This estimate is valid for positive a_i. (See Satterthwaite (1946), Mulrow (1986), and Welch (1947)).

Satterthwaite (1946) cautions against using this expression when some of the a_i are negative. Gaylor and Hopper (1969) examine this in more detail. Consider:

$$(IV.8) \quad \hat{\sigma}^2 = \hat{\sigma}_1^2 - \hat{\sigma}_2^2$$

The object is to approximate $\hat{\sigma}^2$ by:

$$(IV.9) \quad \hat{\sigma}^2 \approx \sigma^2\chi_f^2/f \approx \sigma_1^2\chi_{f1}^2/f1 - \sigma_2^2\chi_{f2}^2/f2$$

and to approximate f by:

$$(IV.10) \quad \hat{f} = (Q-1)^2/(Q^2/f_1 + 1/f_2) \quad \text{where } Q = \hat{\sigma}_1^2/\hat{\sigma}_2^2$$

On the basis of a simulations study, Gaylor and Hopper (1969) suggest that equations (IV.9) and (IV.10) are good approximations, with fewer than 2.5% negative $\hat{\sigma}^2$ results, if:

$$(IV.11) \quad Q \geq F(f_2,f_1,0.975)\cdot F(f_1,f_2,0.50)$$

for $f_1 \leq 100$ and $f_2 \geq f_1/2$, where F(a,b,p) is p^{th} percentile of the F distribution with a and b degrees of freedom. Note also that $0 \leq \hat{f} \leq f_1$.

If the above conditions are met, one can construct a new confidence interval for μ_p:

$$(IV.12)\quad CI = \hat{X} \pm t(1-\alpha, \hat{f}) \cdot \hat{\sigma}_p/(n-1)^{\frac{1}{2}},$$

where $\hat{\sigma}_p^2 = \hat{\sigma}_m^2 - \hat{\sigma}^2(\varepsilon)$

This will, in general, be smaller than the confidence interval given by the POE or SIDF approach described by Mulrow(1986). If data fail this criteria, they suggest increasing f_1 and/or f_2 (by performing more analyses).

Gaylor and Hopper (1969) also investigated the distribution of $\hat{f}$, providing estimates of the coefficient of variation of $\hat{f}$ for $2 \leq (\hat{f}_1, \hat{f}_2) \leq 500$. In general, the coefficient of variation of $\hat{f}$ is rather large. However, it is still useful for calculating a confidence interval using the t statistic. As $\hat{f}$ increases the size of a confidence interval decreases, but $\hat{\sigma}_p^2$ will also tend to increase with larger $\hat{f}$. Gaylor and Hopper (1969) consequently demonstrate that the confidence interval remains about the same size, given widely varying estimates of $\hat{f}$.

This approach is valid when both variance estimates are distributed as chi-square (χ^2). (That is, X_p and ε are normally distributed.) One can estimate observation error by making κ replicate measurements of each unknown sample. Then,

$$(IV.13)\quad \sigma_m^2 = \sigma_p^2 + \sigma^2(\varepsilon)/\kappa$$

$$\hat{\sigma}_p^2 = \sum_{i=1}^{n} (X_i. - X..)^2/(n-1)$$

$$\hat{\sigma}^2(\varepsilon) = \sum_{i=1}^{n} \Sigma_{j=1}^{\kappa} (X_{i,j} - X_i.)^2/[n(\kappa-1)]$$

where $X_{i\cdot} = \sum_{j=1}^{\kappa} X_{i,j}/\kappa$

$$X_{\cdot\cdot} = \sum_{i=1}^{n} \sum_{j=1}^{\kappa} X_{i,j}/(\kappa \cdot n)$$

and $f_1 = n - 1$

$f_2 = n(\kappa - 1)$

This problem has been extended to the construction of tolerance limits by Hahn (1982), Mee (1984), Jaech (1984), and Mee, et.al., (1986). The problem they address is analogous to problems in water quality. One wishes to estimate the probability that a product is out of specification, but must base the probability estimate on a set of measurements. The crucial element is the desire to estimate the true probability of a product out of specification (or, in the context of water quality monitoring, a violation of a standard), not the probability of a measurement exceeding a specification. This is depicted in Figure IV.1, which shows a normally distributed population with observation error causing an increase in the variance of measurements, and the true Cauchy distribution of measurements. This figure exaggerates the difference between these distributions which one might expect, but even small differences between the true and assumed distribution can cause a large difference in the size of a confidence interval.

Jaech (1984) considers criteria for using the SIDF method of estimating $\hat{f}$ for tolerance intervals. The two criteria are similar. Mulrow (1986) evaluated tolerance

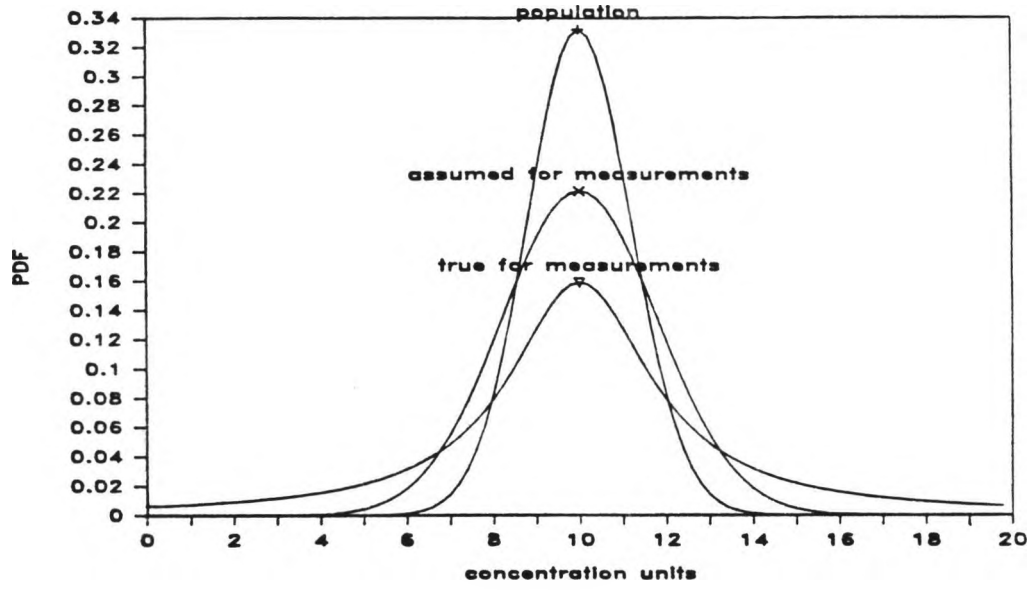

Figure IV.1 The PDF of a water quality random variable compared with the PDF commonly assumed for a measurement and the true PDF of a measurement.

intervals using SIDF when data were calibrated and found that they were conservative as compared to the nominal confidence levels. This is the opposite of what occurred when forming confidence intervals.

A numerical example of the use of Satterthwaite's method for computing the confidence interval of the mean can be found in appendix B.

B. Estimating the mean and variance of a water quality random variable

Unlike the random variables used as examples in the foregoing discussion, water quality random variables tend to be non-normal and have observation error which is dependent on concentration. This section provides small sample and asymptotic efficiencies for estimates of the mean given a lognormally distributed population and the error model used in Chapter III. Asymptotic results are compared with methods intended for censored data. The removal of components of variance due to measurement error is also illustrated. One objective of this is to show some of the difficulties with extending results from the previous section to data which more closely resembles water quality data.

Small sample properties of estimates for the mean were studied by taking samples of size 10, 20, and 40 from several of the population-measurement systems summarized in Table III.1. Coefficients of skewness values of 0.1, 1.0, and 5.0, and R = 0.001 were chosen.

Results are summarized in Table IV.1 and Figures IV.2 through IV.5. Figures IV.2-IV.4 show the efficiency of estimates for the mean relative to observations made without error. When data has a large relative error (FC, FBQ = 1,1), efficiencies are low for all sample sizes. Efficiencies are also small for all sample sizes for coefficients of skewness = 0.1. This is because this population has a large relative observation error (variance of observations/variance of

Table IV.1 Summary of simulation results for small samples (R = 0.001)

n = 10

population

FC,FBQ	coef. skew.	mean	var.mean	variance	var.var.
1.0,1.0	0.1	0.232	6.109E-06	6.041E-05	8.401E-10
	1.0	0.105	1.168E-04	1.169E-03	6.012E-07
	5.0	0.025	8.879E-05	9.176E-04	6.667E-06
0.5,1.0	0.1	0.257	7.461E-06	7.378E-05	1.253E-09
	1.0	0.270	7.700E-04	7.706E-03	2.611E-05
	5.0	0.398	2.219E-02	2.294E-01	4.166E-01
0,0.5	0.1	1.001	1.136E-04	1.123E-03	2.904E-07
	1.0	1.054	1.172E-02	1.173E-01	6.051E-03
	5.0	1.551	3.379E-01	3.492E+00	9.654E+01

measurements

FC,FBQ	coef. skew.	mean	var.mean	variance	var.var.	RMSE mean
1.0,1.0	0.1	0.232	1.064E-03	8.920E-03	1.850E-05	0.128
	1.0	0.105	1.554E-03	9.591E-03	2.405E-05	0.464
	5.0	0.024	1.378E-03	9.586E-03	3.246E-05	1.278
0.5,1.0	0.1	0.257	1.079E-03	8.941E-03	1.863E-05	0.128
	1.0	0.270	3.026E-03	1.534E-02	8.692E-05	0.153
	5.0	0.398	2.902E-02	2.327E-01	3.652E-01	0.385
0,0.5	0.1	1.004	1.928E-03	1.262E-02	4.128E-05	0.037
	1.0	1.057	1.877E-02	1.245E-01	7.215E-03	0.110
	5.0	1.555	3.670E-01	3.478E+00	8.233E+01	0.378

Table IV.1 (continued)

n = 20

population

FC,FBQ	coef. skew.	mean	var.mean	variance	var.var.
1.0,1.0	0.1	0.232	2.868E-06	6.059E-05	3.648E-10
	1.0	0.105	5.483E-05	1.157E-03	2.456E-07
	5.0	0.025	3.896E-05	8.322E-04	2.091E-06
0.5,1.0	0.1	0.257	3.503E-06	7.400E-05	5.440E-10
	1.0	0.270	3.614E-04	7.625E-03	1.067E-05
	5.0	0.393	9.738E-03	2.080E-01	1.307E-01
0,0.5	0.1	1.001	5.333E-05	1.126E-03	1.261E-07
	1.0	1.051	5.501E-03	1.161E-01	2.472E-03
	5.0	1.534	1.482E-01	3.167E+00	3.028E+01

measurements

FC,FBQ	coef. skew.	mean	var.mean	variance	var.var.	RMSE mean
1.0,1.0	0.1	0.231	5.286E-04	9.017E-03	8.746E-06	0.099
	1.0	0.104	7.848E-04	9.891E-03	1.201E-05	0.267
	5.0	0.023	6.900E-04	9.691E-03	1.449E-05	1.059
0.5,1.0	0.1	0.256	5.370E-04	9.049E-03	8.840E-06	0.914
	1.0	0.269	1.533E-03	1.610E-02	4.459E-05	0.145
	5.0	0.393	1.344E-02	2.168E-01	1.227E-01	0.295
0,0.5	0.1	1.004	6.333E-04	1.349E-02	1.975E-05	0.025
	1.0	1.055	6.037E-03	1.302E-01	3.114E-03	0.074
	5.0	1.539	1.509E-01	3.231E+00	2.923E+01	0.253

Table IV.1 (continued)

n = 40

population

FC,FBQ	coef. skew.	mean	var.mean	variance	var.var.
1.0,1.0	0.1	0.232	1.475E-06	6.047E-05	1.880E-10
	1.0	0.105	2.839E-05	1.154E-03	1.254E-07
	5.0	0.025	2.001E-05	8.303E-04	1.031E-06
0.5,1.0	0.1	0.257	1.802E-06	7.385E-05	2.804E-10
	1.0	0.270	1.871E-04	7.608E-03	5.447E-06
	5.0	0.393	5.003E-03	2.075E-01	6.445E-02
0,0.5	0.1	1.001	2.743E-05	1.124E-03	6.499E-08
	1.0	1.051	2.848E-03	1.158E-01	1.262E-03
	5.0	1.534	7.616E-02	3.159E+00	1.493E+01

measurements

FC,FBQ	coef. skew.	mean	var.mean	variance	var.var.	RMSE mean
1.0,1.0	0.1	0.232	2.270E-04	9.142E-03	4.088E-06	0.065
	1.0	0.104	2.606E-04	1.019E-02	4.764E-06	0.154
	5.0	0.024	2.506E-04	9.895E-03	5.329E-06	0.638
0.5,1.0	0.1	0.256	2.274E-04	9.179E-03	4.112E-06	0.059
	1.0	0.269	4.267E-04	1.676E-02	1.583E-05	0.077
	5.0	0.394	5.325E-03	2.194E-01	6.240E-02	0.186
0,0.5	0.1	1.004	3.203E-04	1.346E-02	8.807E-06	0.018
	1.0	1.055	3.212E-03	1.297E-01	1.563E-03	0.054
	5.0	1.539	7.792E-02	3.223E+00	1.443E+01	0.182

population). Also, as censoring decreases, the efficiency improves.

Solid lines on these figures indicate asymptotic efficiencies. The distribution of X_m when X_p is lognormally distributed is given by equation (III.15). This distribution is approximately normal for large observation error and approximately lognormal for small observation error. The most efficient estimator for the mean of a sample from a normally distributed population [$N(\mu_p,\sigma_p^2)$] is simply the sample mean, which has variance σ_p^2/n. The sample mean for a lognormally distributed population [$LN(\mu_p, \sigma_p^2)$ with mean α_p and variance β_p^2] has variance β_p^2/n, but this is not minimum variance. The MLE estimate, given by Finney (1941) and discussed by Aitchison and Brown (1957) has minimum variance given by:

(IV.14) $\text{Variance}(\hat{\alpha}_p) = \alpha_p^2(\sigma_p^2+\sigma_p^4/2)/n$

For the simulations, sample means were used, not MLE. As a set of values becomes more skewed some loss of efficiency could be expected because of the use of sample means. Asymptotic efficiency of the sample mean of observations relative to an error free sample mean is given by:

(IV.15) $\text{ARE} = \beta_p^2/(k_o+k_1\alpha_p+\beta_p^2)$

Note from the figures that the small sample efficiency rapidly approaches the asymptotic result. Figure IV.5 shows the root mean squared error (RMSE) of the sample mean.

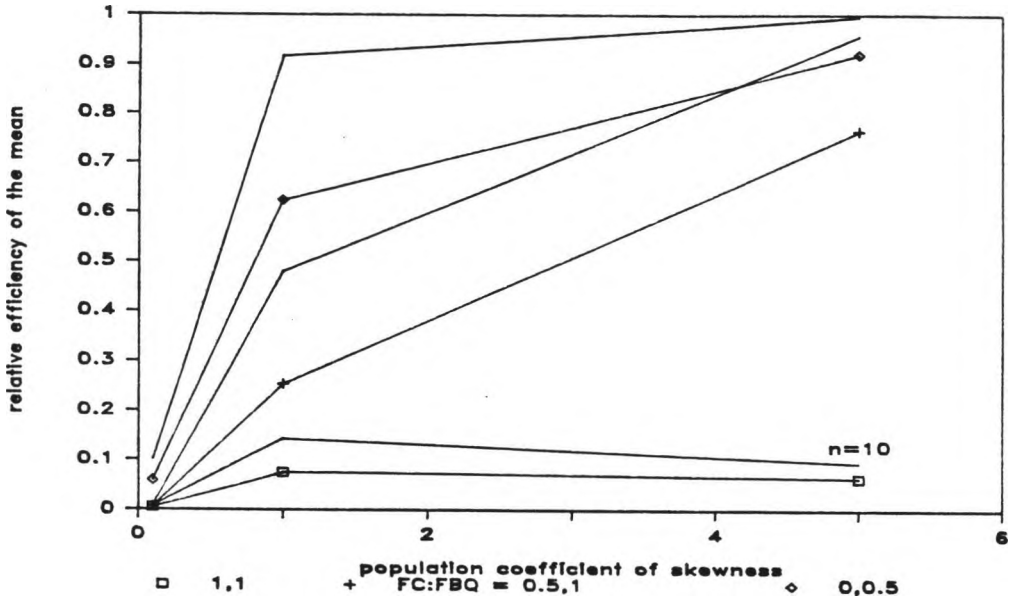

Figure IV.2 The efficiency of the sample mean when error is present relative to an error free sample mean for n=10. Solid lines are asymptotic results.

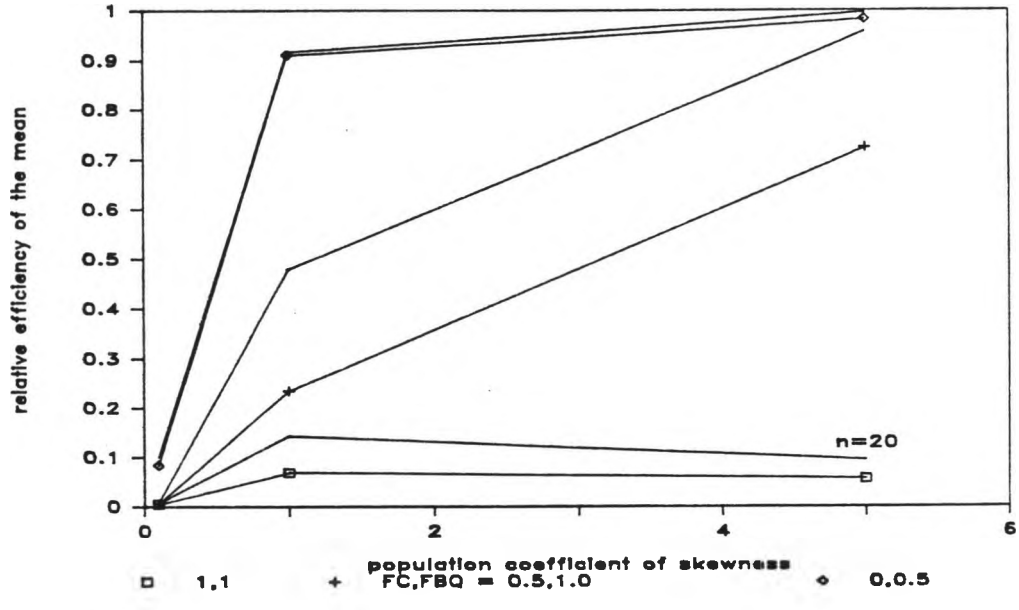

Figure IV.3 The efficiency of the sample mean when error is present relative to an error free sample mean for n=20. Solid lines are asymptotic results.

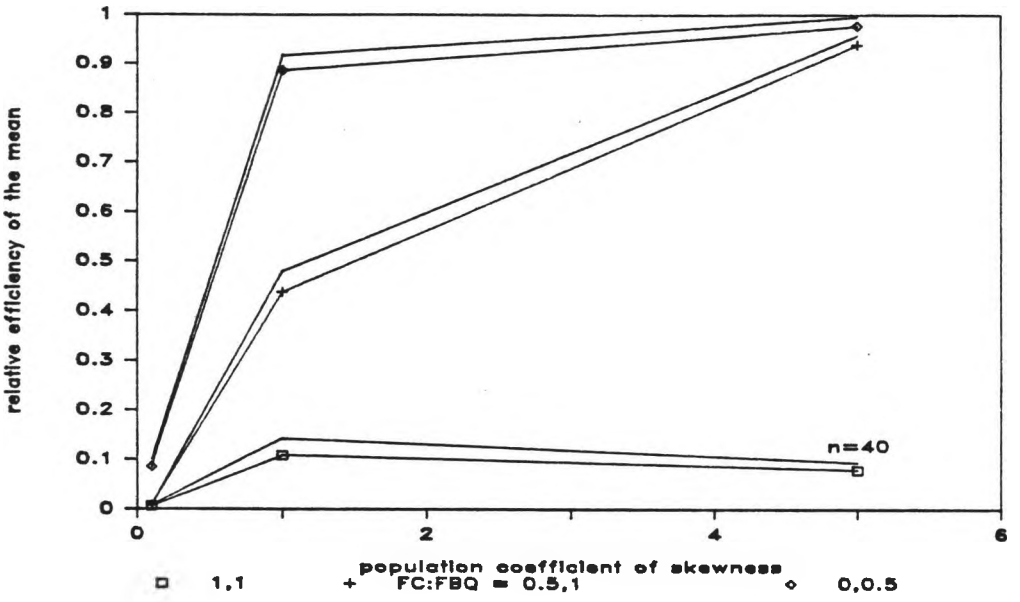

Figure IV.4 The efficiency of the sample mean when error is present relative to an error free sample mean for n=40. Solid lines are asymptotic results.

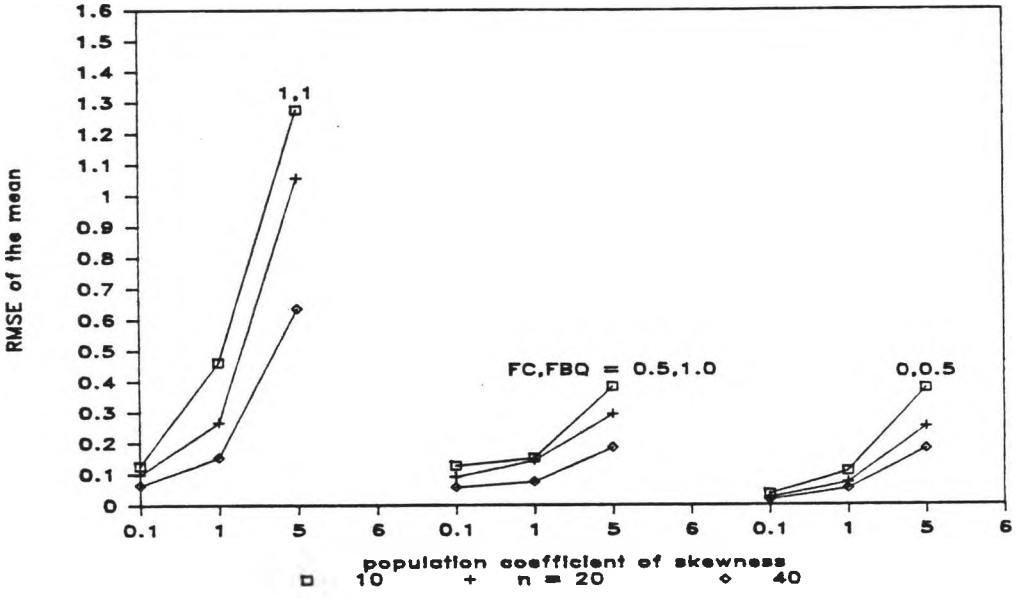

Figure IV.5 The RMSE of the sample mean when error is present and n=10. Solid lines are asymptotic results.

The RMSE is given by :

$$(IV.16) \quad RMSE = \left[\sum_{i=1}^{n} (\bar{X}_m - \alpha_p)^2 / 1000 \right]^{1/2}$$

where n is the sample size.

RMSE is an absolute indication of the performance of an estimator, while efficiency is comparative. For FC,FBQ = 1,1, when data would be 100% censored, RMSE in some cases is not too bad. For example, when the coefficient of skewness = 0.1 and n = 10, RMSE is about 13%. This means that one knows fairly precisely what the mean is given this sample, but not nearly so well as when observations are made without error (RE about 0.001). An ad hoc 95% confidence interval given by $\hat{\mu}_p \pm t(\alpha-1,n-1)\hat{\sigma}_p/\sqrt{n}$ gives [0.161 0.303] (where $\hat{\mu}_p$ and $\hat{\sigma}_p$ are the sample mean and standard deviation respectively, of the measurements). Had the data been 100% censored, which is likely given the simulation conditions, the most one could say about the mean would be that it was less than the detection limit of 0.256. Moreover, increasing the sample size doesn't help much in this case when data is censored, but for n = 40, a similar 95% confidence interval when results are reported gives [0.201 0.263], a significant improvement.

This point can also be illustrated by directly comparing censored methods with reporting methods recommended here. The asymptotic relative efficiency (ARE) based on MLE when all values are reported can be approximated:

$$(IV.17)\quad ARE = [\alpha_p(\sigma_p^2+\sigma_p^4/2)]/[\alpha_m(\sigma_m^2+\sigma_m^4/2)]$$

For the POE approximation to X_m, $\alpha_p = \alpha_m$ and $\beta_m^2 = k_o + k_1\alpha_p + \beta_p^2$. The ARE can be determined given a mean and coefficient of variation for the population and a value for R.

For censored samples, it is assumed that μ and σ are estimated using MLE and converted to an estimate of α using $\hat{\alpha} = \exp(\hat{\mu} + \hat{\sigma}^2/2)$. The variances of $\hat{\mu}$ and $\hat{\sigma}$ are given by Cohen(1961b).

$$(IV.18)\quad \text{Variance of } \hat{\mu}_m = K(\mu_m)\cdot\sigma_m^2/n$$

$$\text{Variance of } \hat{\sigma}_m = K(\sigma_m)\cdot\sigma_m^2/n$$

$$\text{Covariance of } (\hat{\sigma}_m,\hat{\mu}_m) = K(\mu_m\cdot\sigma_m)\cdot\sigma_m^2/n$$

where the values of the $K(\cdot)$ found in Cohen (1961b) depend on the degree of censoring. The variance of $\hat{\alpha}_p$ can be approximated by a propagation of errors argument based on the expression for $\hat{\alpha}$ given above.

$$(IV.19)\quad \text{Variance of } \hat{\alpha}_o \approx \alpha_o^2\cdot\sigma_\mu^2[K(\mu_\mu)+\sigma_\mu^2K(\sigma_\mu) + 2\sigma_m K(\mu_m\cdot\sigma_m)]/n$$

Figures (IV.6) and (IV.7) show the ARE of the estimate for $\hat{\alpha}$ given above as a function of the fraction which would have been censored. Also shown is the ARE of an estimate of the mean for a censored sample from a lognormally distributed population in which the sample contains observation error. These are based on MLE for both the censored and uncensored samples. These figures show that estimates of the mean based on uncensored observations are better than those based on censored samples, reinforce what is intuitively obvious, that censoring filters information from data.

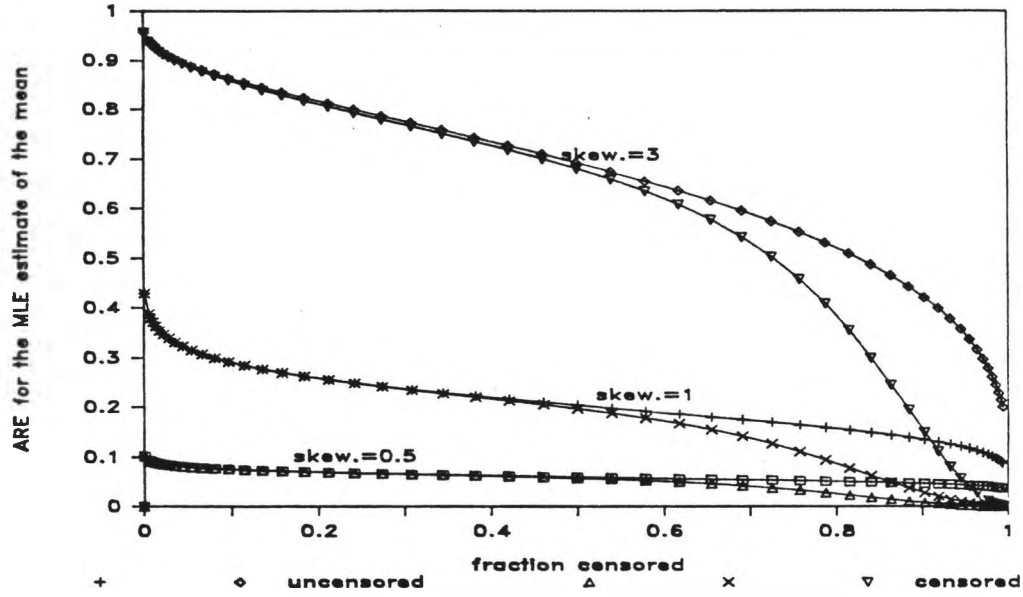

Figure IV.6 The asymptotic relative efficiency for the MLE estimate of the mean when error is present, comparing censored and uncensored data when R = 0.001

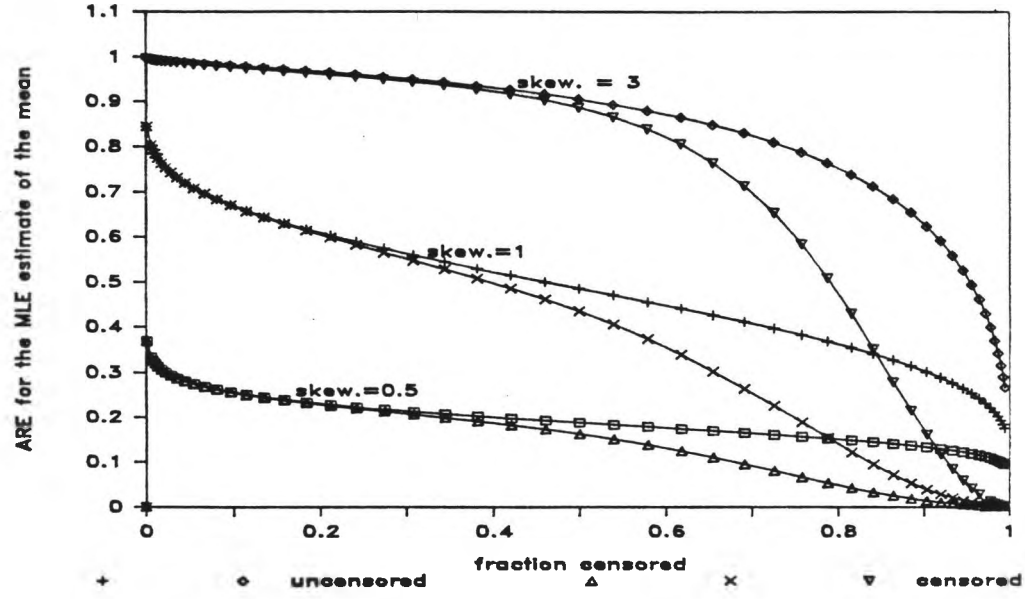

Figure IV.7 The asymptotic relative efficiency for the MLE estimate of the mean when error is present, comparing censored and uncensored data when R = 1.0.

As for removing components of variance due to observation error, the POE expression for the variance of a water quality measurement, when solved for β_p^2, suggests an estimator:

$$(IV.20) \quad \beta_p^2 = [\sigma_m^2-(r+s\alpha_p+t\alpha_p^2)]/(t+1)$$

$$(IV.21) \quad \hat{\beta}_p^2 = [\hat{\sigma}_m^2-(r+s\hat{X}_m+t\hat{X}_m)]/(t+1)$$

The usefulness of (IV.20) as an estimator was studied in the small sample simulations. Results are summarized in Table IV.2. The problem with this estimator, as is true for SIDF for differences of variance estimators, is the occurrence of negative estimates. The table shows that although $\hat{\beta}_p^2$ is unbiased (for large numbers of estimates), a large fraction may be negative. Note also the large variance of $\hat{\beta}_p^2$.

C. Monitoring sensitivity

A value of estimating observation error is that it allows one to consider the limitations, or sensitivity of a monitoring system. Monitoring system sensitivity (MS) could be defined simply as the ratio of population variance to the variance of observation error:

$$(IV.22) \quad MS = \sigma_p^2/\sigma^2(\varepsilon)$$

For X_p normally distributed with mean μ_p and variance σ_p^2, and using the POE expression for the variance of X_m,

$$(IV.23) \quad MS = \sigma_p^2/(\sigma_m^2-\sigma_p^2)$$
$$= \sigma_p^2/[r + s\mu_p + t(\sigma_p^2+\mu_p^2)]$$

Large values of MS imply that small changes in the mean of a

Table IV.2 Use of $\hat{\beta}^2$ as an estimator

n = 10

FC,FBQ	coef. skew.	estimate pop.var.	variance of est.	frac. est.>0	var.ratio (meas/pop)	effic. of $\hat{\beta}^2$
1.0,1.0	0.1	7.32E-05	1.82E-05	0.49	1.48E+02	4.62E-05
	1.0	7.78E-04	2.37E-05	0.49	8.20E+00	2.54E-02
	5.0	7.27E-04	3.20E-05	0.48	1.05E+01	2.08E-01
0.5,1.0	0.1	7.00E-05	1.83E-05	0.45	1.21E+02	6.84E-05
	1.0	6.42E-03	8.56E-05	0.76	1.99E+00	3.05E-01
	5.0	2.23E-01	3.62E-01	0.99	1.01E+00	1.15E+00
0,0.5	0.1	5.43E-04	4.06E-05	0.46	1.12E+01	7.15E-03
	1.0	1.11E-01	7.06E-03	1.00	1.06E+00	8.57E-01
	5.0	3.44E+00	8.15E+01	1.00	9.96E-01	1.19E+00

n = 20

FC,FBQ	coef. skew.	estimate pop.var.	variance of est.	frac. est.>0	var.ratio (meas/pop)	effic. of $\hat{\beta}^2$
1.0,1.0	0.1	1.73E-04	8.53E-06	0.48	1.49E+02	4.28E-05
	1.0	1.08E-03	1.18E-05	0.58	8.55E+00	2.09E-02
	5.0	8.32E-04	1.43E-05	0.54	1.16E+01	1.47E-01
0.5,1.0	0.1	1.80E-04	8.62E-06	0.48	1.22E+02	6.31E-05
	1.0	7.19E-03	4.38E-05	0.89	2.11E+00	2.43E-01
	5.0	2.07E-01	1.22E-01	1.00	1.04E+00	1.08E+00
0,0.5	0.1	1.42E-03	1.92E-05	0.58	1.20E+01	6.57E-03
	1.0	1.17E-01	3.05E-03	1.00	1.12E+00	8.09E-01
	5.0	3.20E+00	2.89E+01	1.00	1.02E+00	1.05E+00

n = 40

FC,FBQ	coef. skew.	estimate pop.var.	variance of est.	frac. est.>0	var.ratio (meas/pop)	effic. of $\hat{\beta}^2$
1.0,1.0	0.1	3.02E-04	4.02E-06	0.52	1.51E+02	4.68E-05
	1.0	1.38E-03	4.66E-06	0.72	8.83E+00	2.69E-02
	5.0	1.04E-03	5.22E-06	0.67	1.19E+01	1.98E-01
0.5,1.0	0.1	3.15E-04	4.05E-06	0.53	1.24E+02	6.93E-05
	1.0	7.85E-03	1.55E-05	0.99	2.20E+00	3.51E-01
	5.0	2.09E-01	6.18E-02	1.00	1.06E+00	1.04E+00
0,0.5	0.1	1.40E-03	8.54E-06	0.67	1.20E+01	7.61E-03
	1.0	1.17E-01	1.53E-03	1.00	1.12E+00	8.24E-01
	5.0	3.19E+00	1.43E+01	1.00	1.02E+00	1.05E+00

water quality random variable will be detected with high confidence. Low values of MS mean that large changes in the mean could occur without detection.

This process is easily visualized by considering estimates of sample means and variances from the simulation conditions described above. Typical simulated estimates of α_p and β_p^2 (the mean and variance of a lognormal distribution) are shown in Figures IV.9-IV.12. Figures IV.9 and IV.10 show insensitive systems, while IV.11 and IV.12 show sensitive systems.

Sensitivity which is too poor for a particular management objective would presumably result in a shift of monitoring resources in an effort to raise the the value of MS. This could be done by optimizing the calibration design and analyzing more standards and, or unknowns. Specific objectives could be related to particular values of MS.

For example, suppose the objective is to state the change in the mean of X_p which could be detected with a particular power $(1-\beta)$ and confidence $(1-\alpha)$ given a sample size of n. where C is the "critical value" used to reject the null hypothesis that the mean has not changed.

(IV.24) $P(\hat{\mu}_p > C$ given that μ_p has shifted to some value $k \cdot \mu_p)$
$= 1 - \beta$

(IV.25) $P(\hat{\mu}_p < C$ given that $\hat{\mu}_p$ has not changed$) = 1-\alpha$

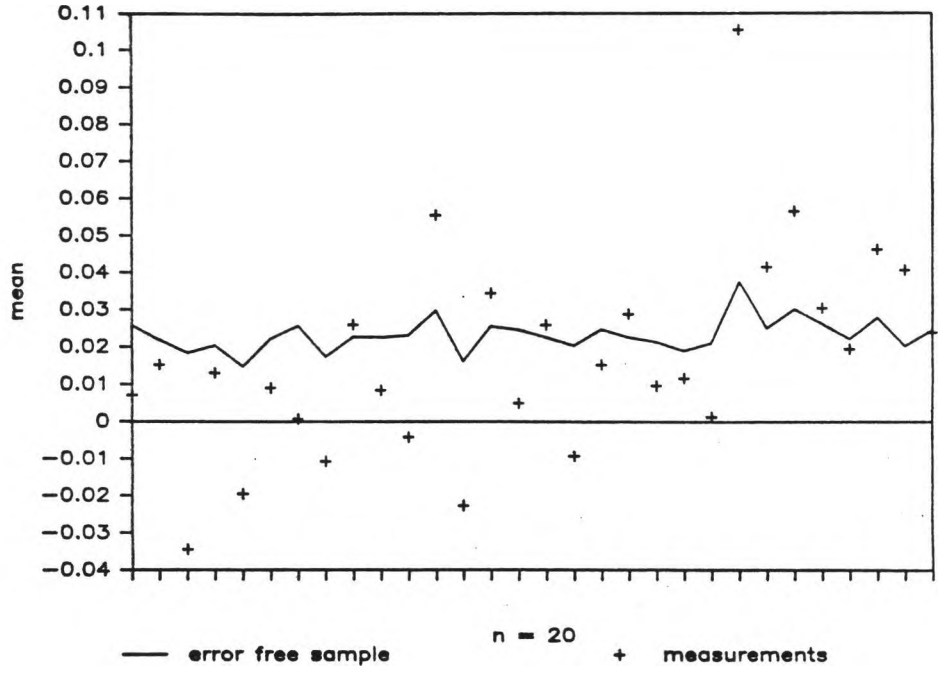

Figure IV.8 Simulation result comparing error free mean with that based on measurements. n = 20, pop. coef.skew. = 0.1, R = 0.001, and FC,FBQ = 1,1.

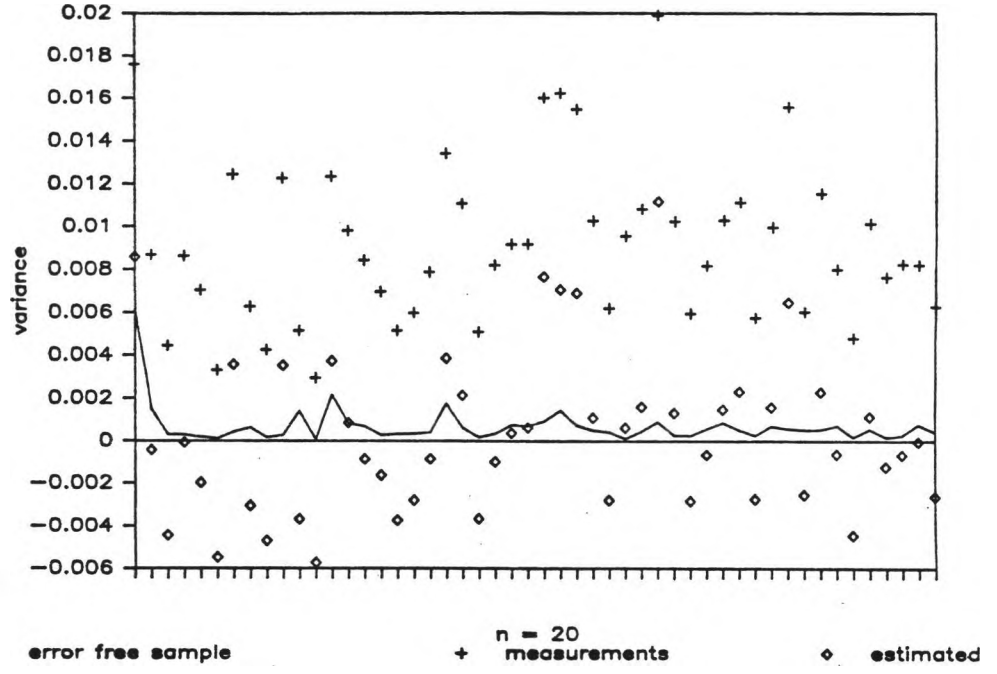

Figure IV.9 Simulation result comparing error free variance with that based on measurements. n=20, pop.coef.skew. = 0.1, R=0.001, and FC,FBQ=1,1.

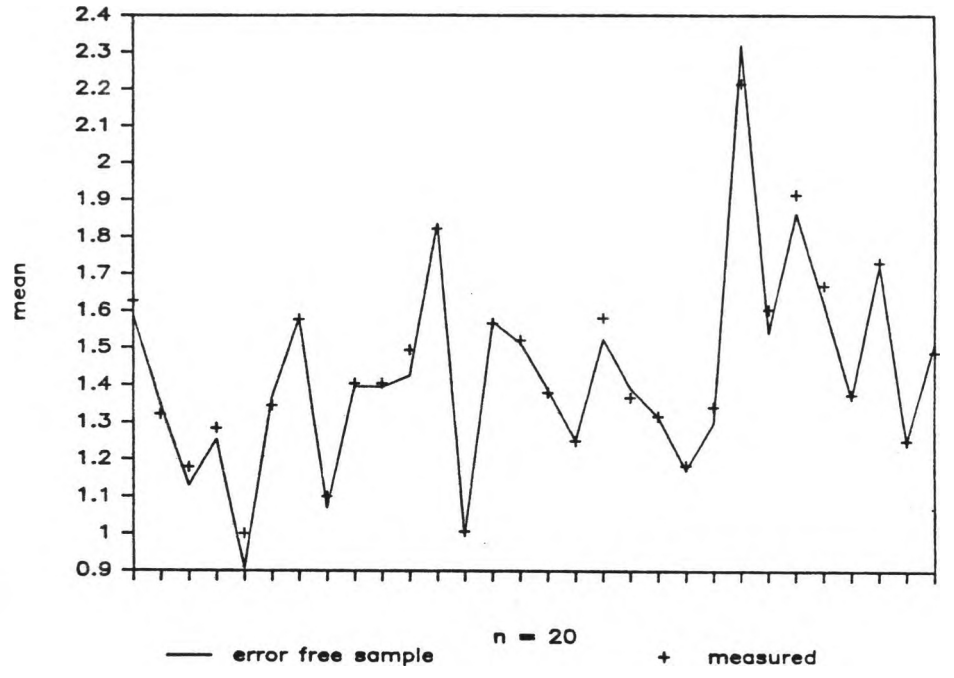

Figure IV.10 Simulation result comparing error free mean with that based on measurements. n = 20, pop. coef.skew. = 5, R = 0.001, and FC,FBQ = 1,1.

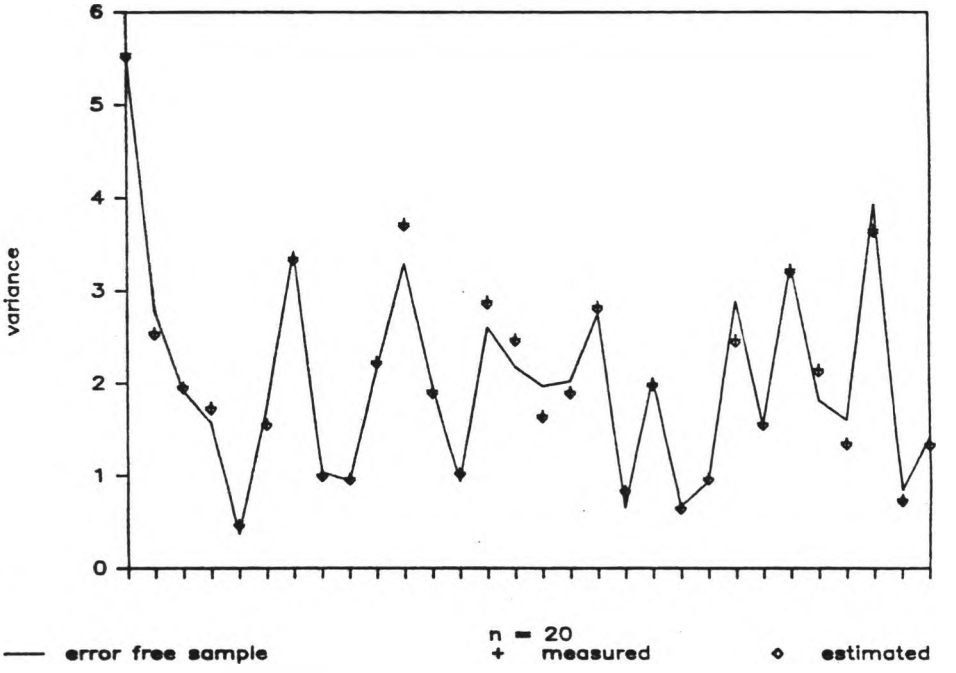

Figure IV.11 Simulation result comparing error free variance with that based on measurements. n=20, pop.coef.skew. = 5, R=0.001, and FC,FBQ=1,1.

The non-central t distribution can be used to relate C and k for a given n, α, and β. The non-central distribution is given by:

$$(IV.26) \quad \Delta = (\bar{X}_m - \mu_p)/(\hat{\sigma}_m/\sqrt{n})$$

The variable Δ is distributed as a non-central t with n-1 degrees of freedom and noncentrality parameter λ, where

$$(IV.27) \quad \lambda = \sqrt{n}(k\cdot\mu_p-\mu_p)/\sigma_m$$
$$= \sqrt{n}(k-1)\cdot\mu_p/\sigma_m$$

Tables (for example see Chemical and Rubber Company, 1985) provide values of:

$$(IV.28) \quad \phi = (n/2)^{1/2}\mu_p(k-1)/\sigma_m$$

For the sake of illustration, assume that a water quality random variable (X_p) is normally distributed with mean μ_p = 1.0 and variance σ_p = 0.2, and that observations are made without error. In this case, MS = ∞. Let α = 0.01, β = 0.10, and n = 10. Tables of noncentral t show ϕ = 3.3. Solving equation (IV.26) for k gives k = 1.66.

In other words, given the statistical criteria, α and β, and a sample size of 10, the mean would need to change from 1.0 to 1.66 before this test would detect it with the stated confidence and power. With observation error, things would have to change even more. For MS = 1.0, $\sigma_m^2 = 2\cdot\sigma_p^2$. In this case, k = 1.93. Similarly, for MS = 0.1, k = 3.19. In other words, the mean would have to shift by a factor of 3.19 to be detected. Figure IV.12 shows the percent change in mean needed [(k-1)·100] vs. MS for μ_p = 1.0, σ_p = 0.2, α = 0.01, β = 0.10, and various values of n.

This is a simple and very specific example, but it is not hard to see how this procedure could be extended to other monitoring objectives. For example, one could similarly determine the trend magnitude a monitoring system is capable of detecting, given n, α, $1-\beta$, and MS.

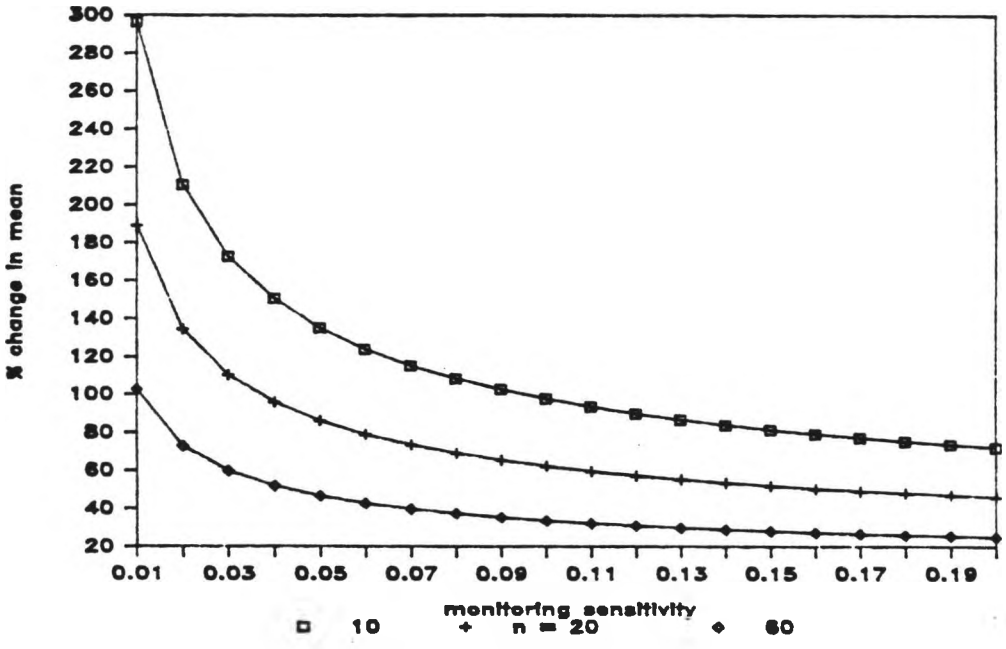

Figure IV.12 Change in mean necessary before detection by a two sample t test with confidence 0.99 and power 0.90 as a function of monitoring sensitivity.

V. STATISTICAL ANALYSIS OF CENSORED WATER QUALITY DATA

The previous chapters have presented a case for not censoring data. However, there are many existing data records which have previously been censored. This chapter presents methods for analyzing such data. It should be emphasized that if future data is not censored, the type of data analysis procedures presented here will not be necessary. Example problems for many of these methods are provided in Appendix A. These examples should not be used to infer properties of the methods, but serve only to demonstrate numerical applications.

A. The censored data problem

When observations fall below (or above) a certain limiting level and the actual values below (or above) this "limit" are unknown, the sample is said to be type I censored. When the values of a fixed fraction of a sample are not known, it is said to be type II censored. In the former case, the censoring level is known and the fraction of sample censored is random, while in the latter, the censoring level is random and the fraction censored is known, usually in advance of taking the sample. It is routinely

assumed by the data user that uncensored observations are made with high precision.

Type II censoring occurs frequently in life testing, where the experiment is terminated after a fixed number of deaths. This is important in medical research, and has consequently received a great deal of attention. In this case, methods derived from order statistics are appropriate, and there is a temptation to use those results for type I censoring as well. Many methods for type I and type II censoring are asymptotically equivalent. However, small sample properties of these methods may be quite different. Also, the knowledge of the censoring level cannot be used in type II methods.

Water quality data with "not detected" (ND) results are type I censored on the left. The level of censoring is determined by the analytical chemist and is based on the confidence with which analytical signal can be discerned from noise. Samples taken over time may be censored at different levels as changes in analytical technology alter the precision of a method. This type of sample is referred to as multiply censored.

The statistical literature contains censored data analysis procedures useful for virtually every problem found in water quality. Many are summarized in several books (Cox and Oakes, 1984; Miller, 1981; and Kalbfleisch and Prentice, 1980) which include worked examples. These methods were developed for failure time and survival analysis, which

produces data that is type II censored on the right. Some work may be necessary to modify particular methods so that they can be used for water quality data. Also, many distributions found in survival analysis are not commonly used for water quality data analysis. The following is a survey of censored data methods pertinent to water quality data analysis. Modifications for left type I censored data have been made. Numerical examples of many, using samples from normally and lognormally distributed populations, are provided in appendix A. Portions of tables necessary for the completion of examples are also provided.

B. Estimation of the mean and variance

1. Assigning a value to censored observations

"Assigning a value" refers to using a number in place of ND, including 0, the detection limit, or a random number. The data is then treated as if it were not censored. For example, the sample mean and variance includeing these assigned values are used to estimate the mean and variance. Gilbert and Kinnison(1981) have pointed out that assigning a fixed constant, such as zero or the detection limit, produces biased estimates. Gilliom and Helsel(1984), in extensive Monte Carlo simulations using synthetic water quality samples, found assigning a constant less suitable than several other estimators. It should not be used unless censoring is insignificant; that is, unless the fraction

censored is small, or the censoring level is much less than the mean. An example of this method is provided in appendix A.1.

Gilliom and Helsel(1984) investigated assigning a random number to the censored results. Assignment of normal, log-normal, and uniform random numbers were simulated with a variety of distributions. For the normal case, normal scores are calculated for results above the LOD and plotted on normal probability paper. A least squares fit to the normal plot is extrapolated into the "less than" region. Values were assigned to censored results based on the plotting position.

$$\text{(V.1)} \quad \text{Normal Score} = F^{-1}[r/(n+1)]$$

where F^{-1} is the inverse normal CDF, and r/(n+1) is the plotting position, with r the rank of the given observation. Extrapolated values of less than zero are readjusted to zero. The method for lognormal numbers is the same except that data is log transformed prior to plotting. The method for uniform numbers is similar. Random uniform numbers from the range of zero to the detection limit are assigned to the censored data. Examples of assigning normal, lognormal, and uniform random numbers appear in appendix A.1.

2. Estimation of the mean and variance for samples from normal distributions with type I censoring

a. Maximum likelihood estimation (MLE)

Maximum likelihood estimation (MLE) of the mean and variance of a type I censored sample from a normal population has been studied extensively (Cohen 1957,1959,1961a,1961b; Gupta, 1952; Hald, 1949; Halperin ,1952). Properties which make MLE desirable are discussed in Kendall and Stuart(1979). To summarize, if a single sufficient statistic exists, the MLE is a function of it. If a minimum variance bound(MVB) estimator exists, and the MLE exists, then the MLE is unique and MVB. For type I censoring, no single sufficient statistic exists. In this case, the MLE is not unique, but is asymptotically normal, efficient, and consistent.

The likelihood function (L) for a random sample with type I (possibly multiple) censoring on the left is:

$$(V.2) \quad L \propto \prod_{i=1}^{r}\{F(a_i;\theta)\}\prod_{i=1}^{n-r} f(x,\theta)$$

where the a_i are the set of censoring levels (fixed constants), $F(x,\theta)$ is the population cumulative density function (CDF) of X with parameter set θ, $f(x,\theta)$ is the population probability density function (PDF), and r of n observations are censored. When the sample is from a normal population,

$$(V.3)\quad f(x;\theta) = f(x;\mu,\sigma^2) = (2\pi\sigma^2)^{-\frac{1}{2}}\exp\{-\tfrac{1}{2}[(x-\mu)/\sigma]^2\}]$$

then,

$$(V.4)\quad L \propto \prod_{i=1}^{r}\{F(a_i;\theta)\}\cdot(2\pi\sigma^2)^{-\frac{1}{2}(n-r)} \cdot\exp\{-\sum_{i=1}^{n-r}\tfrac{1}{2}[(x_i-\mu)/\sigma]^2\}$$

$$(V.5)\quad \ln L \propto \sum_{i=1}^{r}\ln\{F[(a_i-\mu)/\sigma]\} - \tfrac{1}{2}(n-r)\ln(2\pi\sigma^2) - \tfrac{1}{2}\sum_{i=1}^{n-r}[(x_i-\mu)/\sigma]^2$$

$$(V.6)\quad \partial(\ln L)/\partial\mu = -\sum_{i=1}^{r}f[(a_i-\mu)/\sigma]/(\sigma F[(a_i-\mu)/\sigma]) + \sigma^{-2}\sum_{i=1}^{n-r}(x_i-\mu) = 0$$

$$(V.7)\quad \partial(\ln L)/\partial\sigma = -\sum_{i=1}^{r}(a_i-\mu)\sigma^{-2}f[(a_i-\mu)/\sigma] /F[(a_i-\mu)/\sigma] - (n-r)/\sigma + \sum_{i=1}^{n-r}(x_i-\mu)^2/\sigma^3 = 0$$

Rearranging equation (V.6) and substituting the sample estimates $\hat{\mu}$ and $\hat{\sigma}$ for μ and σ gives:

$$(V.8)\quad \hat{\mu} = \bar{X}_o - [\hat{\sigma}/(n-r)]\Sigma f[(a_i-\hat{\mu})/\hat{\sigma}]/F[(a_i-\hat{\mu})/\hat{\sigma}]$$

Similarly, equation (V.7) gives:

$$(V.9)\quad \hat{\sigma}_2 = [S_o^2 + (\bar{X}_o-\hat{\mu})^2]\{1 + [1/(n-r)]\Sigma[(a_i-\hat{\mu})/\sigma]f[(a_i-\hat{\mu})/\hat{\sigma}]/F[(a_i-\hat{\mu})/\hat{\sigma}]\}^{-1}$$

Where Σ_c is the sum over all censored values, $\bar{X}_o$ is the mean of the observed values and S_o^2 is the sample variance of the observed values:

$$(V.10)\quad \bar{X}_o = \Sigma x_i/(n-r)$$

$$(V.11)\quad S_o^2 = \Sigma(x_i - \bar{X}_o)^2/(n-r)$$

Because they are not linear functions of the sample, finding the solution to (V.8) and (V.9) is an iterative process. MLE methods found in the literature may differ because of the manner of solution to those equations. Some

include tables which obviate the need for an iterative approach (see Cohen, 1961b). Currently, tables exist only for small samples ($n \leq 20$). Wolynetz(1979a) has provided a computerized solution.

Cohen(1961a, 1961b) has computed the asymptotic covariance matrix and efficiency of the MLE estimators. Part of his results are reproduced as Figures V.1 and V.2. The efficiencies of these estimators are relative to uncensored samples. The efficiency of the variance estimate deteriorates more rapidly than that for the mean. This is because the most important observations for the variance, one tail, are missing (Saw,1961). Also, $\hat{\mu}$ and $\hat{\sigma}$ become progressively more correlated as censoring increases.

The small sample properties of MLE have not been studied extensively. A misleading characteristic with regard to small sample estimation is that the variance of the estimator actually decreases as the sample becomes progressively (> 50%) censored. This is because there are fewer observations to work with and the number of likely outcomes actually decreases. Based on a small number of simulations, Saw(1961) noted that the MLE was biased for small samples. An example of MLE estimation for normal data can be found in appendix A.

The covariance matrix of MLE of the mean and variance estimates is given by the inverse of the matrix of negative second derivatives of the likelihood function with respect to each estimated parameter. From Wolynetz(1979a):

$$(V.12)\quad \mathrm{Var}\,[\hat{\mu}\ \ \hat{\sigma}]' = \begin{bmatrix} -\partial^2 \ln(L)/\partial\mu^2 & -\partial^2 \ln(L)/\partial\mu\partial\sigma \\ -\partial^2 \ln(L)/\partial\sigma\partial\mu & -\partial^2 \ln(L)/\partial\sigma^2 \end{bmatrix}^{-1}$$

$$(V.13)\quad \partial^2 \ln(L)/\partial\mu^2 = -(n-r)/\hat{\sigma}^2 - \Sigma_c[(\cdot)f(\cdot)/F(\cdot) + f^2(\cdot)/F^2(\cdot)]/\hat{\sigma}^2$$

$$(V.14)\quad \partial^2 \ln(L)/\partial\sigma^2 = -\Sigma_u(X_i-\hat{\mu})^2/\hat{\sigma}^4 - (n-r)/\hat{\sigma}^2 - \Sigma_c\{[(\cdot)f(\cdot)/F(\cdot)]^2 + (\cdot)^3 f(\cdot)/F(\cdot)\}/\hat{\sigma}^2$$

$$(V.15)\quad \partial^2 \ln(L)/\partial\mu\partial\sigma = -\Sigma_u(X_i-\hat{\mu})/\hat{\sigma}^3 - \Sigma_c\{(\cdot)^2 f(\cdot)/F(\cdot) + (\cdot)f^2(\cdot)/F^2(,)\}/\hat{\sigma}^2$$

when evaluated at the MLE for μ and σ^2, and where $(\cdot) = (a_i-\hat{\mu})/\hat{\sigma}$.

A class of estimators which are linear functions of the sample were developed to overcome the computational difficulties of the MLE estimates. Some of these have quite high efficiencies in comparison with MLE estimators. For samples censored at the same fixed point, there is Saw(1961):

$$(V.16)\quad \hat{\mu} = \{X_o - af(\cdot)/[1-F(\cdot)(f^{-1}(\cdot)]\} \cdot \{1-f(\cdot)/[(1-F(\cdot)f^{-1}(\cdot)\}^{-1}$$

where a is the censoring level, $(\cdot) = r/n$, $f^{-1}(\cdot)$ is the standard normal value corresponding to the area r/n, and $F(\cdot)$ is approximated by r/n, and r is the number censored out of a sample size of n. This is based on a least squares approach. The asymptotic efficiency of this estimate is within 95% of the asymptotic efficiency of athe MLE estimate.

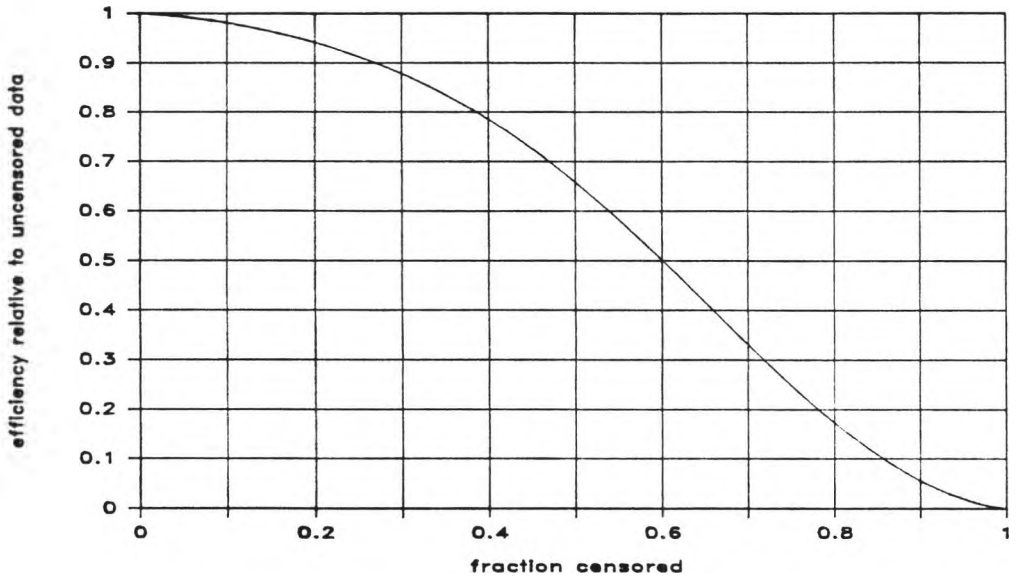

Figure V.1 Asymptotic efficiency relative to uncensored data for MLE estimation of the sample mean of a normal distribution.

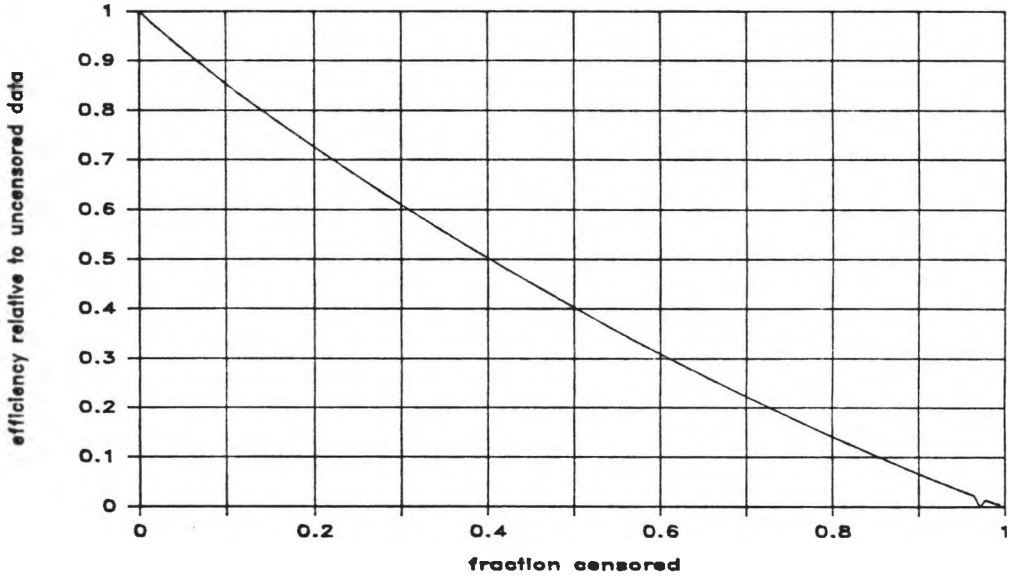

Figure V.2 Asymptotic efficiency relative to uncensored data for MLE estimation of the sample standard deviation of a normal distribution.

b. Linear estimators

Persson and Rootzen(1977) proposed estimators for μ and σ based on a combination of MLE and the method of moments. Starting with equation (V.3) above, approximate $F[(a-\mu)/\sigma]$ with r/n, then letting $\tau = F^{-1}(r/n)$,

$$(V.17)\quad \ln L = \ln\binom{n}{r} + r\ln[F(\tau)] - \tfrac{1}{2}(n-r)\ln(2\pi\sigma^2) - \tfrac{1}{2}\Sigma[(x_i-\mu)/\sigma]^2$$

Let $x_i - \mu = (x_i - a) + (a - \mu)$

Then approximate $(a-\mu)/\sigma$ with τ:

$$(V.18)\quad \ln L = \ln\binom{n}{r} + r\ln F(\tau) - \tfrac{1}{2}(n-r)\ln(2\pi\sigma^2) - \tfrac{1}{2}\Sigma[(x_i - a + \tau\sigma)/\sigma]^2$$

Maximizing ln(L) then results in:

$$(V.19)\quad \sigma_{rml} = \tfrac{1}{2}\{[\tau/(n-r)]\Sigma(x_i-a) + \{[\tau/(n-r)]^2[\Sigma(x_i-a)]^2+[4/(n-r)]\Sigma(x_i-a)_2\}^{1/2}\}$$

$$= \tfrac{1}{2}\{F^{-1}(r/n)(X_o-a) + \{[F^{-1}(r/n)(X_o-a)]^2 + [4/(n-r)]\Sigma(x_i-a)^2\}^{1/2}\}$$

where rml refers to "reduced maximum likelihood". By definition, $(a-\mu)/\sigma = \tau$, so approximate the mean by : $a - \tau\sigma_{rml}$

$$(V.20)\quad \mu_{rml} = a - F^{-1}(r/n)\sigma_{rml}$$

Here, μ_{rml} and σ_{rml} are not asymptotically unbiased therefore Persson and Rootzen(1977) derived estimators which are corrected for bias:

$$(V.21)\quad \hat{\sigma}^2 = S_o^2 - \{(n/(n-r))f(r/n)F^{-1}(r/n) - [n/(n-r)]^2[f(r/n)]^2\}\sigma_{rml}^2$$

$$(V.22)\quad \hat{\mu} = X_o - [n/(n-r)]f(r/n)\sigma_{rml}$$

The asymptotic efficiency of $\hat{\mu}$ and $\hat{\sigma}$ are shown in Figures (V.1) and (V.2). An example of the use of Persson and Rootzen's method can be found in appendix A.1.

4. Estimation of the mean and variance for lognormally distributed samples with type I censoring

The two parameter lognormal (LN) distribution has the PDF:

$$(V.23)\quad f(x;\mu,\sigma) = (2\pi\sigma^2 x)^{-1}\exp\{-\tfrac{1}{2}[(\ln x - \mu)/\sigma]^2\}$$

Where μ and σ are the expected value and variance, respectively, of Y, the transformed variable:

$$(V.24)\quad Y = \ln(x)$$

Y is normally distributed. The expected value and variance of X are given by Aitchison and Brown (1957):

$$(V.25)\quad \alpha = \exp(\mu + \sigma^2/2)$$

$$(V.26)\quad \beta^2 = \exp(2\mu + \sigma^2)[\exp(\sigma^2) - 1]$$

a. Estimation of μ and σ^2

Maximum likelihood estimation of μ and σ^2 for the uncensored case is equivalent to the normal case except that the observations are log transformed. The estimates and their variances are given by:

$$(V.27)\quad \hat{\mu} = \sum\ln(X_i)/n$$

$$(V.28)\quad \hat{\sigma}^2 = \sum[\ln(X_i-\hat{\mu})]^2/(n-1)$$

$$(V.29)\quad \mathrm{Var}(\hat{\mu}) = \sigma^2/n$$

$$(V.30)\quad \mathrm{Var}(\hat{\sigma}^2) = 2\sigma^4/n$$

Aitchison and Brown (1957) considered several methods for estimation from complete samples, and Gilbert and Kinnison (1981) described their use for estimation from censored samples. The method of quantiles (Aitchison and Brown, 1957) yields:

(V.31) $\hat{\mu} = \{v_{q2}\ln(x_{q1}) - v_{q1}\ln(x_{q2})\}/(v_{q2}-v_{q1})$

(V.32) $\hat{\sigma} = \{\ln(x_{q2}) - \ln(x_{q1})\}/(v_{q2} - v_{q1})$

where v_{q1} and v_{q2} are quantiles corresponding to percentiles q1 and q2 from the standard normal distribution, and x_{q1} and x_{q2} are the sample values corresponding to those quantiles. Maximum efficiency occurs when the quantiles are symmetrically placed, then $v_{1-q} = -v^{q}$, and $q = q_1 = 1 - q_2$.

(V.33) $\hat{\mu} = \frac{1}{2}\{\ln(x_{1-q}) + \ln(x_q)\}$

(V.34) $\hat{\sigma}^2 = \frac{1}{2}\{\ln(x_{1-q}) - \ln(x_q)\}/v_{1-q}$

The asymptotic efficiencies of (V.33) and (V.34) are reproduced from Aitchison and Brown (1957) as Figure V.3. Censoring may restrict the choice of q. Optimal choices are 0.27 and 0.07 for $\hat{\mu}$ and $\hat{\sigma}$ respectively. Censoring which exceeds those values force non-optimal choices of q and censoring exceeding 50% forces asymmetric choices of q_1 and q_2.

The graphical method (Aitchison and Brown, 1957) consists of plotting the observed points on log probability paper. Then,

(V.35) $\hat{\mu} = \ln(x_{.50})$

(V.36) $\hat{\sigma} = \ln\{\frac{1}{2}(x_{.50}/x_{.16} + x_{.84}/x_{.50})\}$

where the subscripts refer to percentile values. This method

is risky when more than 16% of the sample is censored (Gilbert and Kinnison, 1981).

b. Estimation of α and β^2

MLE, method of quantiles, method of moments, and graphical methods for estimation of α and β^2 for the lognormal distribution are described by Aitchison and Brown(1957). The MLE estimate due to Finney(1941) requires a computer, as a practical matter. Gilliom and Helsel(1984) use MLE to find $\hat{\mu}$ and $\hat{\sigma}$, and find $\hat{\alpha}$ and $\hat{\beta}$ from equations (V.25) and (V.26). Similarly, for the methods of quantiles and the graphical method for $\hat{\alpha}$ and $\hat{\beta}$, one can estimate $\hat{\mu}$ and $\hat{\sigma}$ and make the same transformation. The efficiency of the method of quantiles relative to MLE is shown in Figure V.4. Examples of MLE and probability plot estimation for LN data is provided in appendix A.1.

C. Goodness of fit tests

1. General concepts

Gilbert and Kinnison(1981) point out the importance of knowing the underlying distribution before using any distribution dependent methods.

Water quality random variables are believed by many to be lognormally or normally distributed. A goodness of fit test is needed for verification. Reasons for this belief are based on the distribution of elements in the earth's crust (Ahrens,1965), and the distribution of properties affecting

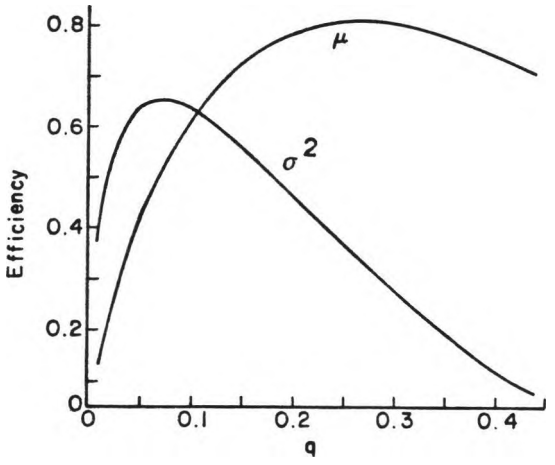

Figure V.3 Asymptotic efficiency relative to MLE of μ and σ^2 for a sample from a lognormal distribution by the method of quantiles (After Aitchison and Brown, 1957).

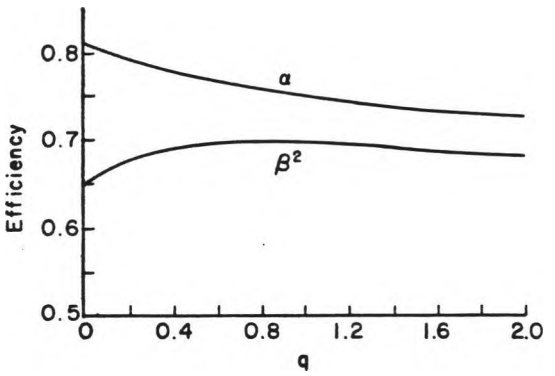

Figure V.4 Asymptotic efficiency relative to MLE of α and β^2 for a sample from a lognormal distribution by the method of quantiles (After Aitchison and Brown, 1957).

the movement of substances in the environment (Shaw,1961; Vistelius,1960). Gilliom and Helsel(1984) state that substances subject to similar transport processes will tend to have similarly shaped frequency distribution.

More formal studies of the distribution of water quality random variables include Ward and McBride (1984). They noted that dissolved oxygen and pH were normally distributed while flow, fecal coliform and turbidity were log normally distributed in the Waikato River at Mercer, New Zealand. Gilbert and Kinnison(1981) stated that radionuclide concentrations in the environment were log normally distributed. Gilliom, et.al.(1984) state that the lognormal distribution describes the distribution of many water quality random variables. Lettenmaier(1982) notes that fecal coliform in streams are often log normally distributed. Hirsch et.al.,(1982) found total phosphorus to be positively skewed and total dissolved solids to be nearly symmetric (skewness = 0.07 for 260 observations) for the Klamath river in Oregon from 1972 to 1979.

Gilliom and Helsel(1984) studied the distribution of trace elements in streams at 482 locations. The coefficient of variation ranged from 0.15 to 3.2. The coefficient of skewness ranged from -0.8 to 5.2 with a median value of 1.8. Six percent of the data sets were negatively skewed. They surmised that four parent distributions would adequately describe water quality trace elements: lognormal, contaminated lognormal, gamma (Γ), and the delta distribution.

There is also a strong argument for particular variables being left-skewed or symmetric. The concentration of trace metals in natural waters is limited on the high end by solubility. Such distributions are not uncommon (Skogerboe, 1986).

The "goodness of fit" problem refers to the testing of:

(V.37) H_o: $F(x) = F_o(x)$ vs. H_a: $F(x) \neq F_o(x)$

where $F(x)$ is unknown and $F_o(x)$ is a presumed distribution. If F_o is completely specified, as in "F_o is normally distributed with a mean of μ and variance σ^2, $[N(\mu,\sigma^2)]$", H_o is referred to as a simple hypothesis. When μ and σ must be estimated from the sample, H_o is a composite hypothesis.

Goodness of fit tests are selected on the basis of the type of hypothesis (simple or composite) to be tested, and the alternatives one expects to encounter. A desirable property of a goodness of fit test includes high power against particular alternatives. Often, no single goodness of fit test will be optimal against all possible alternatives, and it may be desirable to use several tests which have high power for different alternatives. A scheme for testing water quality data for goodness of fit might take the form:

1) Test whether the distribution is skewed or symmetrical

2) Test symmetric distributions for normal vs. other symmetric distributions

3) Test skewed distributions for log-normal vs. other skewed distributions

In addition, one might wish to test against F(x) being of

the same form as F_o but with different parameters. Several tests converging on the same conclusion may be taken as strong support for a given hypothesis (Ward and McBride, 1986).

Shapiro, et.al.(1968) conducted simulations using sample sizes from 10 to 50 to assess the power of several goodness of fit tests for normality against twelve families of alternate distributions. For uncensored data, tests suitable for case (1) above include Shapiro-Wilks and skewness tests against long tailed alternatives, and Shapiro-Wilks against short tailed alternatives. Gilbert and Kinnison (1981) and others recommend probability plotting. For case (2), Shapiro-Wilks, skewness, kurtosis, and the studentized range are most sensitive to long tailed alternatives, while Shapiro-Wilks, kurtosis, and studentized range are sensitive to short tailed alternatives (Shapiro, et.al.,1968). However, when alternatives become very close to normal, (such as T distribution with 10 or more degrees of freedom), none of these perform well. The kurtosis test for normality may be the best under that circumstance. Kolmolgorov-Smirnov(KS) and Cramer-von Mises(CM) perform poorly for both case 1 and 2. The chi-squared(CS) test performs erratically in that large changes in power occurred for relatively minor changes in the alternate distribution.

Case (3) may be the most difficult to assess as little is known about the power of commonly used tests when testing for log-normal vs. other skewed distributions. One might

proceed by testing log transformed data for normality against other log transformed skewed distributions. One could note the general form of such distributions, and use the guidelines provided for case 2.

Shapiro et.al.,(1968) also evaluated the effectiveness of tests intended for simple hypotheses when the parameters were misspecified. They found that KS, CM, and CS performed very poorly in that minor misspecification of μ and σ for a normal distribution resulted in a high rate of rejection of the null hypothesis. The rate of rejection was comparable to cases where the sampled distribution was extremely non-normal. This led Shapiro, et.al.(1968) to remark that this "throws into doubt their usefulness as practical statistical test procedures".

The Shapiro-Wilks test for normality is based on the regression of sample order statistics on their expected values. A correlation coefficient close to one indicates a good fit. This test has not been modified to accommodate censoring, though it is theoretically possible to do so.

2. Tests for censored data

A few goodness of fit tests have been modified for use with censored data, and for most of these, only large sample properties have been investigated.

a. Relative quartile range

Gilliom and Helsel(1984) develop means of estimating the form of the distribution using the relative quartile range (rqr) of the uncensored portion of the sample, and they assess the performance of a variety of estimators using small samples. Their index is:

(V.38) $rqr = (X_{q3}-X_{q1})/X_{LOD}$,

where X_{q3} and X_{q1} are the sample 75th and 25th percentiles, respectively based only on the uncensored portion of the sample. This statistic is not used to test a distribution related hypothesis, but is instead used to match sample properties with the most efficient estimators for population properties (mean, variance, median).

b. Chi-Squared(CS) test

The CS test is a distribution free test based on the statistic:

(V.39) $\chi^2 = \sum(n_i-nP_{oi})^2/nP_{oi}$

where n_i are the number of observations out of n falling in interval i and P_i is the probability of a given observation falling in that interval. The n_i are approximately multinomially distributed provided there are at least 5 observations in each interval (Kendall and Stuart, 1979), and asymptotically normal. The variable χ^2 is asymptotically distributed as chi-squared with k-1 degrees of freedom. The null hypothesis is rejected if $\chi^2 > \chi^2_{1-\alpha,k-1}$, where $1-\alpha$ is the confidence level of the test.

For the composite hypothesis, the P_i must be estimated from the data. Under this circumstance, χ^2 is chi-squared with between k-1 and k-s-1 degrees of freedom, where s is the number of parameters estimated (Kendall and Stuart, 1979). Also, the s parameters must be maximum likelihood estimates, or asymptotically both normal and efficient (Larson and Marx,1981). The null hypothesis should be rejected when χ^2 exceeds χ^2_{k-s-1}. Shapiro, et.al.(1968) recommend not using this test for composite hypotheses.

The power of the CS test is a rather complicated function of the choice of the k intervals. Equi-probable and equal size intervals have been examined. Kendall and Stuart (1979) recommend equal probability intervals with at least 5 observations per interval (to preserve the multinomial assumption). The choice of intervals must not be made with reference to the sample form. In other words, the null hypothesis must be stated prior to inferring the form of the distribution from the observations.

The choice of k can be made to maximize power against specified alternatives. However, censoring of the sample constrains this, and optimal choices may not be available. For equi-probable intervals, Kendall and Stuart (1979) recommend using:

$$(V.40) \qquad k = b\{\sqrt{2}(n-1)/(\tau_\alpha + F^{-1}(P_o))\}^{2/5}$$

P_o is the (asymptotic) desired power, F is the standard normal CDF, $\tau_\alpha = F^{-1}_{(1-\alpha)}$, and b is a constant between 2 and 4. For example, if $P_o = 0.80$, $\alpha = 0.05$, and b = 3, τ_α =

$F^{-1}(0.95) = 1.64$, $F^{-1}(0.80) = 0.84$, and $k = 3\{\sqrt{2(n-1)}/(1.64+0.84)\}^{2/5} = 2.4(n-1)^{2/5}$. With the constraint that there be 5 observations per class, $n \geq 63$ and $k = 12$. If more than about 10% of the sample were censored, this choice would not be possible.

Determining the asymptotic power of the test requires tables of the non-central CS distribution and specification of the alternative distribution (see Patnaik, 1949). Kendall and Stuart (1979) recommend testing the deviations (n_i-P_i) for independence. Regular patterns of deviations may result from particular types of violations of the null hypothesis. The CS test has been criticized because there is a loss of information when observations are grouped.

c. Komolgorov-Smirnov(KS) test

The KS test is based on the statistic:

(V.41) $$D_n = \text{Sup}\{|S_n(x_i)-F_o(x_i)|\}, \quad S_n(x_i) = i/n$$

Where "Sup" is the superior member, or largest of the elements in the brackets, S_n is the empirical sample CDF, and the x_i are the ordered observations with x_1 being the smallest. The KS test is distribution free as D_n depends only on differences between expected and observed CDF's. Because all observations are considered, more sample information is used than with the CS test. H_o is rejected when D_n exceeds $d(\alpha,n)$. Tables of $d(\alpha,n)$ can be found in CRC (1985). The KS test is consistent against any alternative and is more powerful than CS for continuous distributions, requiring

about $n^{0.80}$ as many samples to achieve the same power (asymptotically) as the CS test. For the composite case, where location and scale parameters must be estimated, D_n will depend on the form of F_o but not its parameters and the test is therefore not distribution free. Not much is known about the distribution of d_α for this case (Kendall and Stuart, 1979).

Barr and Davidson (1973) modified the KS statistic to accommodate right type I censored data. For a simple hypothesis,

$$(V.42) \qquad D_{N,T} = \sup\{|S_N(x_i)-F_o(x_i)|, \; |S_N(x_{i-1})-F_o(x_i)|, \; |S_N(T)-F_o(T)|\}$$

where T is the censoring level, and $F_o(T)$ is the fraction observed of sample size N under H_o. Here $S_N(x_i) = i/N$, $S_N(T) = (1+j)/N$, and x_j is the largest uncensored observation. The determination of $D_{N,T}$ is different from that of D_N for the uncensored case in that there are fewer comparisons. Tables of $D_{N,T}$ for $N \leq 25$ and $F_o(T) \leq 1$ prepared by Dufour and Maag (1978) are reproduced as Table V.1. Generally, full sample KS table values are greater than those for censored samples. Dufour and Maag (1978) provide approximate formulas for sample sizes greater than 25:

$$(V.43) \qquad D_{N,T,\alpha} \approx D^*_{T,\alpha}/(N)^{1/2} - 0.19/N$$

$$\text{for } 0.01 \leq \alpha \leq 0.20 \quad \text{and } F_o(T) \leq 0.25$$

where $D^*_{T;\alpha}$ is the asymptotic value for d which can be found in Table 1 of Koziol and Byar (1975), which is reproduced as

Table V.2. For this paper, truncation is equivalent to type I censoring.

For left censoring,

$$(V.44) \qquad D_{N,T} = \sup\{|S_N(x_i)-F_o(x_i)|, \; |S_N(x_{i+1})-F_o(x_i)|, \; |S_N(T)-F_o(T)|\}$$

where $S_N(T) = (j-1)/N$, and x_j is the smallest uncensored observation. When using tables, let $F_o(T) = 1 - F_o(T_r)$, where $F_o(T_r)$ is that calculated for right censored data. The power of this test can be found from a formula given by Steck (1971).

d. Cramer-von Mises(CM) test

The CM statistic is based on the same principle as KS, that is, deviations of the sample CDF from the hypothesized CDF. The basic form for uncensored data is (Durbin and Knott, 1972):

$$(V.45) \qquad W_n^2 = n\int_o^1 \{F_n(x) - F_o(x)\}^2 dF_o(x)$$

or alternatively,

$$(V.46) \qquad W_n^2 = \sum_{i=1}^{n} [(i-\tfrac{1}{2})/n - F(x_i)]^2 + 1/(12n)$$

where $F_n(x)$ is the empirical (sample) CDF. Pettitt and Stephens(1976) modified this to accommodate right type I censoring:

$$(V.47) \qquad W_n^2 = \Sigma_{i=1}^{R}[(2i-1)/2n - F_o(x_i]^2 - R(4R^2-1)/(12n^2) + nF_o(T)[R^2/n^2 - F_o(T)R/n + F_o(T)^2/3]$$

where the sum is from 1 to R, and R is the number of uncensored observations, $F_o(x_i) = i/n$, $F_o(T)$ is the value of the

CDF at the right censoring level under H_o, and x_i is the i^{th} ordered observation.

Pettitt and Stephens(1976) tabulated percentage points of ${}_pW^2$ for the simple asymptotic case where the fraction censored is 1-P. Pettitt(1976) tabulated percentage points of ${}_pW^2$ for the normal composite case for the asymptotic solution, and ${}_rW_n^2$ for sample sizes less than 100. These tables are reproduced as Table V.3.

For left censoring the sum is taken from i = r+1 to i = n where x_r is the smallest uncensored observation. Also, replace $F_o(T)$ with $1 - F_o(T)$ so that it will continue to reflect the fraction observed under H_o. Examples of the goodness of fit tests described in this section can be found in appendix A.2.

D. Two sample tests

The two sample test refers to comparisons between separate samples. Comparisons may be made on the basis of a distribution or a particular parameter. Goodness of fit tests may be modified to test whether two samples have come from the same distribution. For example, Hollander and Wolfe(1973) describe the use of the KS test for comparing two unknown distributions when data is left censored at the same fixed point, and the censoring level is less than the smallest uncensored observation. Censored data is considered tied for this application.

Table V.1 Significance points for $D_{N,T}$
(After Dufour and Maag, 1978)

		Significance level α				
N	T	0.15	0.1	0.05	0.025	0.01
10	0.1	0.1358	0.1608	0.2135	0.2403	0.3070
	0.2	*	0.2234	0.2693	0.3192	0.3663
	0.3	0.2583	0.2935	0.3155	0.3564	0.4169
	0.4	0.2992	0.3114	0.3570	*	0.4388
	0.5	0.3090	0.3378	0.3855	0.4151	0.4631
	0.6	0.3260	0.3546	*	0.4309	0.4775
	0.7	0.3359	0.3636	0.4005	0.4392	0.4850
	0.8	0.3402	0.3673	0.4061	0.4439	0.4883
	0.9	0.3422	0.3686	0.4088	0.4455	0.4889
	1	0.3425	0.3687	0.4092	0.4456	0.4889
	0.2 α =	0.2534	d = 0.20^-	α =	0.1460	d = 0.20^+
	0.4	0.0276	0.40^-		0.0215	0.40^+
	0.6	0.0504	0.40^-		0.0433	0.40^+
15	0.1	0.1138	0.1300	0.1721	0.1936	0.2398
	0.2	0.1805	*	0.2198	0.2577	0.2959
	0.3	0.2164	*	0.2652	*	0.3345
	0.4	0.2399	*	0.2943	0.3319	0.3635
	0.5	0.2568	0.2814	0.3134	0.3472	0.3818
	0.6	0.2688	0.2921	0.3292	0.3573	0.3965
	0.7	0.2779	*	0.3336	0.3652	0.4003
	0.8	0.2805	0.3028	0.3358	0.3670	0.4027
	0.9	0.2823	0.3037	0.3375	0.3678	0.4042
	1	0.2823	0.3040	0.3376	0.3679	0.4042
	0.2 α =	0.1143	d = 0.200^-	α =	0.0791	d = 0.200^+
	0.3	0.0278	0.300^-		0.0230	0.300^+
	0.3	0.1182	0.233^-		0.0944	0.233^+
	0.4	0.1003	0.267^-		0.0857	0.267^+
	0.7	0.1019	0.300^-		0.0970	0.300^+

Table V.1 (continued)

20	0.1	0.1015	0.1152	0.1475	0.1672	0.2041
	0.2	0.1525	0.1695	*	0.2189	0.2575
	0.3	0.1888	0.2024	0.2336	0.2575	0.2947
	0.4	0.2075	0.2280	0.2563	0.2858	0.3168
	0.5	0.2244	0.2462	0.2744	0.3006	0.3353
	0.6	0.2359	0.2541	0.2857	0.3116	0.3457
	0.7	0.2424	0.2607	0.2914	0.3174	*
	0.8	0.2451	0.2636	0.2935	0.3198	0.3515
	0.9	0.2458	0.2646	0.2941	0.3206	0.3524
	1	0.2459	0.2647	0.2941	0.3206	0.3524
	0.2	α = 0.0549	d = 0.20^{-}	α = 0.0433	d = 0.20^{+}	
	0.7	0.0101	0.35^{-}	0.0096	0.35^{+}	
25	0.1	*	0.1048	0.1266	0.1497	0.1811
	0.2	0.1403	0.1591	0.1755	*	0.2268
	0.3	0.1692	0.1826	0.2107	0.2316	0.2617
	0.4	0.1885	0.2038	0.2330	0.2554	0.2846
	0.5	0.2018	*	0.2471	0.2707	*
	0.6	0.2110	0.2294	0.2560	*	0.3103
	0.7	0.2177	0.2343	0.2608	0.2856	0.3141
	0.8	0.2198	0.2371	0.2634	0.2872	0.3163
	0.9	0.2206	0.2376	0.2640	0.2879	0.3166
	1	0.2207	0.2377	0.2640	0.2879	0.3166
	0.1	α = 0.1913	d = 0.10^{-}	α = 0.1155	d = 0.10^{+}	
	0.2	0.0278	0.20^{-}	0.0240	0.20^{+}	
	0.5	0.1028	0.22^{-}	0.0965	0.22^{+}	
	0.5	0.0109	0.30^{-}	0.0099	0.30^{+}	
	0.6	0.0264	0.28^{-}	0.0249	0.28^{+}	

* Points of discontinuity; the exact probabilities (α) and critical points (d) are given at the end of the table for each N.

Table V.2 Percentage points for the Kolmogorov-Smirnov statistic for censored data. (After Koziol Byar, 1975)

TABLE 1—*Percentage points for the cumulative distribution* G_T *of the truncated Kolmogorov-Smirnov statistic* D_T *for truncation times* $T = .1(.1)1.0$

CUMULATIVE PROBABILITY	TRUNCATION TIME									
	.10	.20	.30	.40	.50	.60	.70	.80	.90	1.00
0.010	0.1587	0.2232	0.2717	0.3115	0.3454	0.3747	0.3999	0.4209	0.4362	0.4410
0.025	0.1761	0.2473	0.3006	0.3441	0.3810	0.4125	0.4394	0.4612	0.4764	0.4806
0.050	0.1938	0.2718	0.3299	0.3771	0.4168	0.4504	0.4786	0.5011	0.5160	0.5196
0.100	0.2182	0.3054	0.3700	0.4219	0.4652	0.5014	0.5311	0.5540	0.5683	0.5712
0.150	0.2376	0.3321	0.4015	0.4571	0.5029	0.5409	0.5716	0.5946	0.6082	0.6106
0.200	0.2550	0.3559	0.4297	0.4883	0.5363	0.5756	0.6069	0.6300	0.6428	0.6448
0.250	0.2716	0.3785	0.4562	0.5176	0.5675	0.6079	0.6398	0.6626	0.6748	0.6764
0.300	0.2878	0.4005	0.4820	0.5460	0.5976	0.6391	0.6713	0.6938	0.7054	0.7067
0.350	0.3041	0.4225	0.5078	0.5743	0.6275	0.6699	0.7023	0.7245	0.7353	0.7365
0.400	0.3207	0.4449	0.5339	0.6029	0.6576	0.7008	0.7333	0.7551	0.7652	0.7662
0.450	0.3379	0.4681	0.5608	0.6322	0.6885	0.7323	0.7649	0.7863	0.7956	0.7964
0.500	0.3559	0.4923	0.5889	0.6627	0.7204	0.7649	0.7975	0.8183	0.8270	0.8276
0.550	0.3750	0.5180	0.6185	0.6949	0.7541	0.7992	0.8316	0.8518	0.8597	0.8602
0.600	0.3956	0.5455	0.6503	0.7293	0.7899	0.8356	0.8678	0.8872	0.8944	0.8948
0.650	0.4181	0.5755	0.6849	0.7666	0.8287	0.8748	0.9068	0.9254	0.9318	0.9321
0.700	0.4431	0.6088	0.7231	0.8078	0.8715	0.9180	0.9496	0.9673	0.9729	0.9731
0.750	0.4714	0.6465	0.7663	0.8544	0.9195	0.9666	0.9976	1.0142	1.0190	1.0192
0.800	0.5045	0.6905	0.8168	0.9085	0.9756	1.0229	1.0533	1.0687	1.0727	1.0727
0.850	0.5449	0.7443	0.8734	0.9746	1.0438	1.0914	1.1208	1.1348	1.1379	1.1379
0.900	0.5935	0.8155	0.9597	1.0616	1.1334	1.1813	1.2094	1.2216	1.2238	1.2238
0.950	0.6825	0.9268	1.0868	1.1975	1.2731	1.3211	1.3471	1.3568	1.3581	1.3581
0.975	0.7589	1.0282	1.2024	1.3209	1.3997	1.4476	1.4717	1.4794	1.4802	1.4802
0.990	0.8512	1.1505	1.3419	1.4696	1.5520	1.5996	1.6214	1.6272	1.6276	1.6276
0.995	0.9157	1.2361	1.4394	1.5735	1.6583	1.7056	1.7258	1.7306	1.7308	1.7308
0.999	1.0523	1.4171	1.6456	1.7931	1.8828	1.9292	1.9464	1.9494	1.9495	1.9495

Table V.3 Percentage points for the Cramer-von Mises statistic for censored data. (After Pettit and Stephens, 1976; and Pettit, 1976).

Table 1. *Percentage points of* ${}_pW^2$

Percentage points	Value of p									
	0·50	0·55	0·60	0·65	0·70	0·75	0·80	0·85	0·90	0·95
1	0·0080	0·0095	0·0112	0·0130	0·0149	0·0167	0·0187	0·0206	0·0211	0·0240
2·5	0·0101	0·0120	0·0141	0·0163	0·0186	0·0210	0·0233	0·0256	0·0250	0·0294
5	0·0126	0·0150	0·0176	0·0202	0·0230	0·0258	0·0286	0·0312	0·0314	0·0356
10	0·0166	0·0197	0·0230	0·0264	0·0298	0·0333	0·0367	0·0399	0·0426	0·0450
50	0·0536	0·0619	0·0707	0·0793	0·0877	0·0957	0·1030	0·1093	0·1147	0·1178
90	0·1890	0·2153	0·2407	0·2645	0·2861	0·3048	0·3205	0·3327	0·3412	0·3462
95	0·2579	0·2931	0·3269	0·3581	0·3861	0·4102	0·4298	0·4446	0·4548	0·4599
97·5	0·3295	0·3742	0·4167	0·4558	0·4906	0·5201	0·5439	0·5616	0·5733	0·5791
99	0·4271	0·4847	0·5393	0·5891	0·6330	0·6701	0·6997	0·7212	0·7352	0·7419

Table 1. *Asymptotic percentage points of Cramér–von Mises statistics for testing normality with censored data*

α	$\omega^2 = {}_p\hat{W}^2$			$\omega^2 = {}_{1-p,p}\hat{W}^2$		
	$p = 0·5$	$p = 0·75$	$p = 0·9$	$p = 0·75$	$p = 0·85$	$p = 0·95$
0·50	0·017	0·035	0·045	0·019	0·033	0·046
0·85	0·032	0·063	0·082	0·037	0·061	0·083
0·90	0·037	0·073	0·094	0·044	0·070	0·095
0·95	0·046	0·089	0·114	0·055	0·087	0·116
0·975	0·054	0·105	0·135	0·066	0·104	0·138
0·99	0·065	0·127	0·163	0·081	0·120	0·166

Table 2. *Empirical percentage points of* ${}_r\hat{W}^2_n$ *based on 5000 samples for* $n < 60$ *and 2000 samples for* $n \geqslant 60$

(*a*) Percentage points of ${}_r\hat{W}^2_n$

	95% point						90% point					
	$r = 0·9n$		$r = 0·75n$		$r = 0·5n$		$r = 0·9n$		$r = 0·75n$		$r = 0·5n$	
n	ML	G	ML	G	ML	G	ML	G	ML	G	ML	G
20	0·092	0·100	0·071	0·078	0·043	0·048	0·076	0·081	0·058	0·062	0·034	0·034
40	0·095	0·102	0·073	0·082	0·044	0·048	0·078	0·083	0·060	0·065	0·035	0·035
60	0·095	0·101	0·073	0·081	0·044	0·048	0·079	0·083	0·060	0·065	0·036	0·036
80	0·093	0·100	0·071	0·083	0·048	0·048	0·078	0·084	0·060	0·058	0·038	0·036
100	0·095	0·100	0·073	0·082	0·047	0·047	0·082	0·085	0·062	0·065	0·038	0·037
∞	0·114		0·089		0·046		0·094		0·073		0·037	

Often, in water quality, one is concerned with testing whether a random variable from one location exceeds that from another. One way to accomplish this is to compare sample means and their confidence intervals using the methods of section V.B. Another more general method is to test non-parametrically whether a sample from one distribution tends to exceed that from another.

1. Halperin's extension of the Wilcoxon-Mann-Whitney test

Halperin(1960) proposed an extension of the Wilcoxon-Mann-Whitney test to samples censored at the same fixed point. The test considers the case where independent samples of size n and m are taken from populations with CDF's denoted by $F_X(x)$ and $G_Y(y)$ respectively. Observations greater than some point T (or less than, in the case of detection limit censoring) are not observed. The number of observations censored from F and G are denoted r_n and r_m, respectively. The null hypothesis is that the samples come from identical populations versus the alternative that F exceeds G. That is,

(V.48) $$H_o: \; F_X(z) = G_Y(z) \text{ vs.}$$
$$H_a: \; F_X(z) > G_Y(z)$$

The test is carried out as follows.

- Arrange all $(n-r_n)$ and $(m-r_m)$ uncensored samples in ascending order.
- Compute $U_c = [(m-r_m)/2]\cdot(2n + 1 + m - r_m) - S$

  S = sum of ranks of the uncensored Y's in the sequence of all uncensored observations.

- Reject H_o for $U_c \leq U_c(\alpha$,

where $U_c(\alpha)$ can be found for small sample sizes (n and m both less than 8) and test sizes 0.05 and 0.01 in tables from Halperin(1960). These tables, partially reproduced as Table V.4, provide upper tail areas when censoring is on the left. For large samples, U_c is approximately normal with mean μ_c and variance σ_c^2. The following formulas can be used to calculate μ_c and σ_c^2, and standard normal tables can then be used to conduct the test.

$$\text{(V.49)}\quad \mu_c = m\cdot n\cdot(m+n-r)\cdot(m+n+r-1)/[2\cdot(m+n)\cdot(m+n-1)]$$

$$\sigma_c^2 = m\cdot n\cdot(m+n-r)\cdot(A+B+C-D)/[4\cdot(m+n)]$$

$$A = [(m+n-r)^2-1]/[3(m+n-1)]$$

$$B = (m-1)(m+n-r-1)\{n + (2n+1)r/(m+n+2) + r(r-1)(n-1)/[(m+n-2)(m+n-3)]\}/(m+n-1)$$

$$C = n+(2n+1)r/(m+n-1) + r(r-1)(n-1)/[(m+n-1)(m+n-2)]$$

$$D = mn(m+n-r)(m+n+r-1)^2/[(m+n)(m+n-1)^2]$$

where $r = r_m + r_n$.

The statistic U_c is identical to the Wilcoxon-Mann-Whitney U statistic plus a correction for ties equal to the product of the number of uncensored Y's and censored X observations. The test was shown by Halperin(1960) to be consistent. That is, as m and n tend to infinity, the probability that U_c is in the critical region when H_a is true approaches 1. No assumptions are made as to the form of F and G. Other properties of the test have not been published.

Table V.4 Percentage points for Halperin's two sample statistic for censored data (After Halperin, 1960).

($n = 6$)

r/m	1	2	3	4	5	6	7	8
0	0(.143)	0(.036)	2(.048)	4(.057)	5(.041)	7(.047)	9(.051)	11(.054)
1	0(.143)	0(.036)	2(.048)	4(.057)	5(.041)	7(.047)	9(.051)	11(.054)
2	0(.286)	0(.036)	2(.048)	4(.062)	5(.043)	7(.049)	9(.052)	11(.056)
3	0(.429)	0(.107)	2(.060)	3(.048)	5(.052)	7(.055)	9(.058)	10(.045)
4	0(.571)	0(.214)	0(.048)	2(.043)	4(.043)	6(.048)	8(.050)	10(.052)
5	0(.714)	0(.357)	0(.119)	2(.071)	3(.033)	5(.037)	7(.041)	9(.045)
6	0(.857)	0(.536)	0(.238)	0(.071)	2(.039)	5(.046)	6(.042)	8(.042)
7		0(.750)	0(.417)	0(.167)	0(.045)	3(.053)	5(.051)	7(.043)
8			0(.667)	0(.333)	0(.121)	0(.030)	4(.054)	6(.048)
9				0(.600)	0(.273)	0(.091)	3(.070)	4(.039)
10					0(.545)	0(.227)	0(.070)	3(.055)
11						0(.500)	0(.192)	0(.055)
12							0(.462)	0(.165)
13								0(.429)

($n = 7$)

r/m	1	2	3	4	5	6	7	8
0	0(.125)	1(.056)	3(.058)	5(.055)	7(.053)	9(.051)	11(.049)	13(.047)
1	0(.125)	1(.056)	3(.058)	5(.055)	7(.053)	9(.051)	11(.049)	13(.047)
2	0(.250)	0(.028)	2(.042)	5(.061)	7(.057)	9(.056)	11(.051)	13(.048)
3	0(.375)	0(.083)	2(.058)	4(.052)	6(.047)	8(.044)	11(.057)	13(.053)
4	0(.500)	0(.167)	0(.033)	3(.038)	6(.059)	8(.052)	10(.051)	12(.047)
5	0(.625)	0(.278)	0(.083)	2(.045)	5(.052)	7(.050)	9(.045)	12(.056)
6	0(.750)	0(.417)	0(.167)	0(.045)	3(.045)	6(.049)	8(.047)	11(.055)
7	0(.875)	0(.583)	0(.292)	0(.106)	3(.071)	5(.053)	7(.045)	10(.054)
8		0(.778)	0(.467)	0(.212)	0(.071)	3(.055)	6(.051)	8(.044)
9			0(.700)	0(.382)	0(.159)	0(.049)	4(.059)	7(.050)
10				0(.636)	0(.318)	0(.122)	0(.035)	4(.044)
11					0(.583)	0(.269)	0(.096)	0(.026)
12						0(.538)	0(.230)	0(.077)
13							0(.500)	0(.200)
14								0(.467)

($n = 8$)

r/m	1	2	3	4	5	6	7	8
0	0(.111)	1(.044)	3(.042)	6(.055)	8(.047)	11(.054)	13(.047)	16(.052)
1	0(.111)	1(.044)	3(.042)	6(.055)	8(.047)	11(.054)	13(.047)	16(.052)
2	0(.222)	1(.066)	2(.053)	5(.040)	8(.050)	10(.043)	13(.049)	16(.055)
3	0(.333)	0(.067)	1(.042)	5(.047)	7(.041)	10(0.49)	13(.054)	15(.047)
4	0(.444)	0(.133)	0(.024)	4(.047)	7(.051)	9(.043)	12(.050)	15(.054)
5	0(.556)	0(.222)	0(.061)	3(.051)	6(.051)	8(.042)	11(.046)	14(.050)
6	0(.667)	0(.333)	0(.121)	0(.030)	5(.051)	8(.055)	10(.045)	13(.051)
7	0(.778)	0(.467)	0(.212)	0(.071)	3(.044)	7(.054)	9(.050)	12(.054)
8	0(.889)	0(.622)	0(.339)	0(.144)	0(.044)	4(.047)	8(.053)	10(.047)
9		0(.800)	0(.509)	0(.255)	0(.098)	4(.070)	5(.045)	9(.051)
10			0(.727)	0(.424)	0(.196)	0(.070)	4(.051)	8(.057)
11				0(.667)	0(.359)	0(.154)	0(.026)	4(.038)
12					0(.615)	0(.308)	0(.123)	0(.038)
13						0(.571)	0(.267)	0(.100)
14							0(.533)	0(.233)
15								0(.500)

2. Gehan's extension of the Wilcoxon-Mann-Whitney

Gehan(1965) extended the Wilcoxon test to situations which include type I, type II, and multiply censored samples when both samples have the same type of censoring. The test statistic is computed by comparing all possible pairs of samples drawn from the two population. For left censoring,

(V.50) $W = \Sigma U_{ij}$, $i \neq j$,

$U_{ij} = -1$ for X_i uncensored $< Y_j$ uncensored
or X_i censored $\leq Y_j$ uncensored
$= 0$ for X_i uncensored $= Y_j$ uncensored
or X_i and Y_j both censored
or X_i censored $> Y_j$ uncensored
or X_i uncensored $\leq Y_j$ censored
$= +1$ for X_i uncensored $> Y_j$ uncensored
or X_i uncensored $\geq Y_j$ censored

There are nm possible comparisons. The statistic W reduces to Halperin's(1960) U_c when both samples are censored at the same fixed point. The statistic W is asymptotically normal with mean zero and variance equal to:

$$\text{(V.51)}\quad \text{Var}(W) = nm\{\Sigma k_i K_{i-1}(K_{i-1}+1) + \Sigma l_i K_i(K_i+1) + \Sigma k_i(n+m-K_i-L_{i-1})(n+m-3K_{i-1}-k_i-L_{i-1}-1)\} / [(n+m)(n+m-1)]$$

The above sums are from 1 to the number of uncensored observations.

$$K_j = \Sigma k_i$$
$$L_j = \Sigma l_i$$

where the sums are from 1 to j.

k_i = number of uncensored observations at rank i in the ordering of distinct uncensored observations

l_i = number of left censored observations with values less than observations at rank i but greater than observations at rank i + 1.

For left censored data with no ties among the uncensored observations, all $k_i = 1$ (then $K_j = j$) and $l_1 = 0$. The normal approximation can be used for both n and $m \geq 5$, and at least four of the 10 observations uncensored (Gehan,1965). Tables for n and m up to 8 and $\alpha = 0.01$ and 0.05 are provided by Haseman and Hoel(1974) and partially reproduced as Table V.5. W is also consistent (Gehan, 1965). Examples of Halperin's and Gehan's two sample tests can be found in appendix A.3.

E. Tests for trend

1. General concepts

There are many methods available for detecting trends in censored data. Tests for independence in which one variable is time may also be used as a test for independence between two samples. The following section contains a description of the most well known methods. Many are recent developments and there has not been much work which compares them under conditions often encountered with water quality data. Also, few of these methods are computationally simple, and a computer is often required.

Table V.5 Percentage points for Gehan's two sample statistic for censored data (After Haseman and Hoel, 1974).

$n_2 = 7$

r/n_1	1		2		3		4		5		6		7		8	
0.1	0	0.125	0	0.028	0	0.008	1	0.006	3	0.009	4	0.007	6	0.009	7	0.007
			1	0.056	1	0.017	2	0.012	4	0.015	5	0.011	7	0.013	8	0.010
					2	0.033	4	0.036	6	0.037	8	0.037	11	0.049	13	0.047
					3	0.058	5	0.055	7	0.053	9	0.061	12	0.064	14	0.060
2	0.5	0.25	0	0.028	0	0.008	1	0.006	3	0.009	4	0.007	6.5	0.009	7.5	0.007
			1.5	0.083	1	0.017	2	0.012	4	0.014	5	0.010	7	0.013	8	0.010
					2.5	0.042	5	0.048	7	0.049	9	0.048	11	0.047	13.5	0.049
					3	0.05	5.5	0.061	7.5	0.057	9.5	0.054	11.5	0.051	14	0.058
3	1	0.375	1	0.083	0	0.008	2	0.009	3	0.008	4	0.006	6	0.009	7	0.007
					2	0.033	3	0.021	4	0.015	5	0.011	7	0.013	8	0.010
					3	0.058	4	0.033	6	0.035	8	0.036	11	0.048	13	0.047
							5	0.055	7	0.052	9	0.050	12	0.063	14	0.059
4	1.5	0.5	2	0.167	1.5	0.033	0	0.003	3	0.005	5	0.009	6.5	0.009	8	0.009
					4	0.083	2.5	0.015	3.5	0.010	5.5	0.012	7	0.011	8.5	0.011
							4.5	0.039	7	0.040	9	0.041	11	0.041	13.5	0.049
							5.5	0.052	7.5	0.056	9.5	0.053	11.5	0.052	14	0.054
5	2	0.625	3	0.278	3	0.083	2	0.015	3	0.008	5	0.009	6	0.008	8	0.009
							5	0.045	4	0.014	6	0.016	7	0.012	9	0.013
							6	0.076	7	0.045	9	0.045	11	0.046	13	0.044
									8	0.064	10	0.063	12	0.059	14	0.057
6	2.5	0.75	4	0.417	4.5	0.167	4	0.045	2.5	0.008	4.5	0.008	7	0.009	8	0.007
							7.5	0.106	6	0.027	5.5	0.011	7.5	0.013	8.5	0.010
									7	0.045	9.5	0.043	11.5	0.044	13.5	2.045
									8	0.064	10	0.060	12	0.053	14	0.050
7	3	0.875	5	0.583	6	0.292	6	0.106	5	0.027	3	0.004	7	0.008	8	0.008
									9	0.071	7	0.016	8	0.010	9	0.011
											9	0.041	11	0.039	13	0.038
											10	0.053	12	0.051	14	0.050
8			6	0.778	7.5	0.467	8	0.212	7.5	0.071	6	0.016	3.5	0.002	9.5	0.008
											10.5	0.049	8	0.010	10	0.012
											11.5	0.082	12	0.043	14	0.048
													13	0.051	15	0.061
9					9	0.7	10	0.382	10	0.159	9	0.049	27	0.010	9	0.007
											14	0.122	12	0.035	10	0.013
													13	0.059	15	0.048
															16	0.061
10							12	0.636	12.5	0.318	12	0.122	10.5	0.035	8	0.007
													16	0.096	13.5	0.026
															14.5	0.044
															15.5	0.063
11									15	0.583	15	0.269	14	0.096	12	0.026
															18	0.077
12											18	0.538	17.5	0.231	16	0.077
13													21	0.500	20	0.200
14															24	0.467

This section is arranged in order from the most general to the most specific types of tests. The most general tests are non-parametric and distribution free. These include rank tests for independence in time. More specific tests are parametric and distribution free. Tests for trend of this type estimate a slope parameter and consider the hypothesis that the slope is significantly different from zero. No assumption is made regarding the distribution of random fluctuations. More specific are tests for trend which assume a particular distribution for the error term. Examples of many of these may be found in appendix A.4.

2. Rank methods

Rank methods for trend have been derived from rank tests for independence, where one of the covariates is time. Rank methods rely on the rank of a result relative to the other observations rather than its actual value. Data censored at the same fixed point are considered to be "tied" in rank. Assigning ranks to multiply censored data presents a problem. The following is a description of several rank methods for censored data.

a. Mann-Kendall test

The Mann-Kendall test for trend considers the hypothesis:

(V.52) H_o: Data are randomly distributed in time

vs. H_a: Data are not randomly distributed in time

This is a non-parametric hypothesis test. The Mann-Kendall statistic is computed as follows:

1) Arrange all observations in order of occurrence
2) Count the number of times $X_k < X_l$, where k and l are the time indices, for all k < l

The total count, Mann's T, is asymptotically normal under H_o with mean n(n-1)/4 and variance n(n-1)(2n+5)/72. The asymptotic approximation is sufficient for sample sizes larger than 10. Tables for smaller sample sizes can be found in Mann(1945). The table values are found by computing T for all permutations of n distinct values, and calculating the mean and variance of the T values.

If data are censored at the same fixed point, and the censoring level is the smallest value of the data set, consider all censored results to be tied. Then compute:

(V.53) $$E(T) = n(n-1)/4 - r(r-1)/4$$

$$Var(T) = n(n-1)(2n+5)/72 - r(r-1)(2r+5)/72$$

where r is the number of tied observations. This test assumes that the samples are independent. It is consistent and unbiased for the situation where the probability of $X_i > X_j$ for any pair (i,j) is ½. When there are no ties, the asymptotic efficiency of T relative to a test of significance of the slope in a linear trend, given normally distributed errors, is 98%.(Kendall and Stuart, 1979).

The Mann-Kendall test is based on Kendall's K statistic. This statistic is computed as:

(V.54) $$K = \Sigma \, sgn(X_j - X_k) \text{ for all } j \neq k,$$

$$= +1 \text{ for } X_j > X_k$$

$$= 0 \text{ for } X_j = X_k$$

$$= -1 \text{ for } X_j < X_k$$

K is asymptotically normal with a mean of 0 and variance of $n(n-1)(2n+5)/18$. Tables for small samples ($n<40$) are provided by Hollander and Wolfe(1973). Both Mann's and Kendall's test are similar to a sign test in that the relative position of two observations, and not the magnitude of their difference is considered. These tests are similar to Spearman's ρ which takes into account the relative difference in ranks between two observations. The ρ statistic is equivalent to a linear correlation coefficient based on ranks.

b. Hoeffding's test

Hoeffding's test considers more alternatives to randomness than do Mann's, Kendall's, or Spearman's tests. It considers the hypothesis (Hollander and Wolfe, 1973):

(V.55) H_o: $P(X \leq x \text{ and } Y \leq y) = P(X \leq x)P(Y \leq y)$

for all x and y. The test statistic (D) is calculated as follows:

$$(V.56)\quad D = [Q - 2(n-2)R + (n-2)(n-3)S] \Big/ [n(n-1)(n-2)(n-3)(n-4)]$$

n is sample size, r_i = rank of X_i, s_i = rank of Y_i,

c_i = number of pairs (X_a, Y_a) where $(X_a < X_i)$ and $(Y_a < Y_i)$

$$= \sum_{a=1}^{n} \phi(X_a X_i) \cdot \phi(Y_a Y_i)$$

$\phi(u,v) = 1$, if $u < v$

$= 0$ otherwise

$$Q = \sum_{i=1}^{n} (r_i-1)(r_i-2)(s_i-1)(s_i-2)$$

$$R = \sum_{i=1}^{n} (r_i-2)(s_i-2)c_i$$

$$S = \sum_{i=1}^{n} c_i(c_i-1)$$

The statistic D is distributed as $d(\alpha,n)$, tables of which are provided by Hollander and Wolfe(1973). When there are ties, use average ranks (of the tied observations) and recalculate c_i:

$$(V.57)\quad c_i = \sum_{a=1}^{n} \phi(X_a X_i)\phi(Y_a Y_i), \text{ where}$$

$\phi(u,v) = 1$, if $u < v$

$= \frac{1}{2}$ if $u = v$

$= 0$ otherwise

For large samples ($n > 9$), use B instead of D where $B = D + 1/(36n)$ and is distributed as $b(\alpha)$. Tables for $b(\alpha)$ can also be found in Hollander and Wolfe(1973). Reject H_o if $\pi^4 nB/2 \geq b(\alpha)$.

c. Seasonal Kendall test

For $\underline{X} = \begin{bmatrix} X_{11} & \cdots & X_{1p} \\ \vdots & & \vdots \\ X_{n1} & \cdots & X_{np} \end{bmatrix}$

n = number of years of data, p = number of seasons per year

$\underline{R}$ = matrix of ranks, where ranks are assigned separately to each season. Ties are assigned midranks.

$$(V.58) \quad R_{jg} = [n + 1 + \sum_{i=1}^{n} \text{sgn}(X_{jg}-X_{ig})]/2$$

Compute a Kendall statistic for each season.

$$(V.59) \quad S_g = \sum_{i<j} \text{sgn}(X_{jg}-X_{ig}), \; g = 1,2, \ldots p$$

$$S' = \sum_{g=1}^{p} S_g$$

S' is asymptotically normal with mean 0 and

$$(V.60) \quad \text{variance} = \sum_g \sigma_g^2 + \sum_{g,h} \sigma_{g,h}, \; g \neq h$$

where $\sigma_g^2 = [n(n-1)(2n+5) - \sum_{j=1}^{m} t_j(t_j-1)(2t_j+5)]/18$

$$\hat{\sigma}_{gh} = [K_{gh} + 4\sum_{i=1}^{n} R_{ig}R_{ih} - n(n+1)^2]/3$$

$$K_{gh} = \sum_{i<j} \text{sgn}[(X_{jg}-X_{ig})(X_{jh}-X_{ih})]$$

This test is robust against serial correlation, non-normality, and censoring. However, it requires large samples (the author recommends 10 years of monthly data). An earlier test, (Hirsch et.al., 1982) is better for small samples (say two years of monthly data) but is not robust against serial correlation. The earlier test assumes that $\sigma_{gh} = 0$. A Fortran listing for this test can be found in Crawford, et.al., (1983). Power curves for the earlier test with censoring can be found in Hirsch et.al. (1982).

Rank tests are simple to apply. Small sample statistics can easily be calculated by hand. Large samples require no

special tables (except for Hoeffding's statistic). Even though information is said to be lost by the use of ranks, some are asymptotically highly efficient. A disadvantage of rank tests arises with multiply censored data or when there are observations smaller than censored observations. When this occurs, there is no simple way to assign ranks. e.g., a result of "< 5" is not necessarily larger than a result of "2" or one of "< 2".

3. Distribution free parametric methods for censored data

The following methods, developed for survival and failure time analysis, make no assumptions about censored data being tied. In addition, parameter estimates for a linear (or higher order) model are provided. (A "seasonal slope estimator" is described for the rank test developed by Hirsch et.al., 1982).

The problem in the linear case is to determine whether the slope estimate for a linear trend model is significant. We have:

$$(V.61)\quad \underline{Y} = \underline{X}\underline{B} + \underline{e}\ ,\ \underline{e} \approx IID(0,\sigma^2)$$

$$\underline{X} = [\underline{1}\ \underline{T}],\quad \underline{B}' = [B_o\ B_1]$$

$$Y_i = \{t_i,\ d_i\} \text{ are observed,}$$

$$d = 0 \text{ for censored data, and 1 otherwise}$$

$\underline{Y}$ is the n x 1 vector of observations coded by d_i, and $\underline{X}$ is the n x 2 design matrix. When testing for a linear trend in time, $\underline{T}$ is a vector of times of observation, and $\underline{e}$ is a vector of independently distributed random fluctuations.

a. Miller estimators

Miller estimators are described in Miller(1981). The slope and intercept are estimated using weighted, uncensored data. The weights depend on the "product limit" (PL) estimate, which is the empirical CDF for a censored sample. For left, type I censored data, the PL estimator is defined as:

$$(V.62) \quad F(i) = \prod_{j=i}^{n-1} [j/(j+1)]^{di},$$

where j is the rank of the observation, F(n) is defined as 1.0, and the smallest observation is always defined as uncensored (for purposes of calculating F(i)). Note that jumps in the PL estimator do not occur at censored observations. Miller(1981) proposed that

$$(V.63) \quad f(B_i) = \sum_{i=1}^{n} \hat{w}_i (y_i - B_o - B_1 x_i)^2$$

be minimized with respect to $\underline{B}$. Here, the w_i are the jumps in the PL estimator based on residuals of the previous estimate of $\underline{B}$. That is, the PL estimator is based on jumps in the empirical (PL) CDF of $\underline{Z}$, where

$$(V.64) \quad z_i = Y_i - b_o - b_1 X_i$$

and b_o and b_1 are estimates of B_o and B_1, respectively. Because the order of the z_i are not affected by b_o, one can use:

$$(V.65) \quad z_i = Y_i - b_1 X_i$$

The process of minimizing (V.63) "can be tedious" (Miller, 1981). A modified procedure is suggested. As an initial estimate of $\underline{B}$, use ordinary least squares (OLS) for the data which is not censored. Then subsequent estimates

are given by:

$$(V.66) \quad b_1^+ = \Sigma w_i y_i (x_i - \bar{X}_u) / \Sigma w_i (x_i - \bar{X}_u)^2$$

$$b_o^+ = \Sigma w_i (y_i - B_1 X_i)$$

where $\bar{X}_u = \Sigma_u w_i x_i$, and the subscript u refers to uncensored observations.

The ranks and therefore the w_i of the sample are functions of b_1. For censored observations, the PL estimate of F(i) does not change and the w_i are zero. One iterates until a stable solution is obtained. There may be situations where the iteration oscillates between two values. Miller recommends using their average when this occurs. The variance of b_1 is given by:

$$(V.67) \quad Var(b_1) = \Sigma w_i (y_i - b_o - x_i b_1) / \Sigma w_i (x_i - \bar{X}_u)^2$$

b. Buckley-James estimator

The Buckley-James estimator (Buckley and James, 1979) replaces censored observations with:

$$(V.68) \quad \hat{Y}_i(b_1) = b_1 X_i + \Sigma_k \hat{w}_k \hat{z}_k / F(\hat{z}_i)$$

where the sum is over all $z_k < z_i$, and $F(z_k)$ is as before (PL estimate for the residuals). Then, b_o, b_1, and $Var(b_1)$ are then estimated using ordinary least squares (OLS), where the data consists of uncensored Y values and $\hat{Y}$ values from equation (V.68).

4. Maximum likelihood for normally distributed data

The maximum likelihood solution to the linear model is analogous to that for estimating the mean and variance of a

distribution. From equation (V.2), the likelihood function for a censored sample is:

$$(V.69) \qquad L \propto \prod_{i=1}^{n-r} f(Y_i) \cdot \prod_{j=1}^{r} F(a_j)$$

$$\ln(L) \propto \sum_{i=1}^{n-r} \ln[f(Y_i)] + \sum_{j=1}^{r} \ln[F(a_j)]$$

where r is the number of censored observations, and a_j is the censoring point of the j^{th} censored observation. For the normal distribution,

$$(V.70) \qquad f(Y_i) = (2\pi\sigma^2)^{-\frac{1}{2}} \exp\{-\tfrac{1}{2}[(Y_i - B_o - B_1 X_i)/\sigma]^2\}$$

$$F(a_i) = \int f(y_i)\,dy,$$

where the integral is from $-\infty$ to a_i then,

$$(V.71) \qquad \partial \ln(L)/\partial B_o = \sum_u (Y_i - B_o - B_1 X_i)/\sigma^2 - \sum_c [f(\cdot)/F(\cdot)]/\sigma = 0$$

$$(V.72) \qquad \partial \ln(L)/\partial B_1 = \sum_u X_i (Y_i - B_o - B_1 X_i)/\sigma^2 - \sum_c [X_i f(\cdot)/[F(\cdot)]/\sigma = 0$$

$$(V.73) \qquad \partial \ln(L)/\partial \sigma = -(n-r)/\sigma + \sum_u (Y_i - B_o - B_1 X_i)^2/\sigma^3 - \sum_c [(\cdot) f(\cdot)/F(\cdot)\}]/\sigma = 0$$

where u refers to uncensored and c to censored observations, and $(\cdot) = (a_i - B_o - B_1 X_i)/\sigma$.

The covariance matrix of parameter estimates is given by the inverse of the matrix of negative second derivatives of the likelihood function with respect to each estimated parameter. From Wolynetz(1979b):

$$(V.74)\quad \mathrm{Var}[b_o \quad b_1 \quad \hat{\sigma}]' =$$

$$\begin{bmatrix} -\partial^2 \ln(L)/\partial B_o^2 & -\partial^2 \ln(L)/\partial B_o \partial B_1 & -\partial^2 \ln(L)/\partial B_o \partial \sigma \\ -\partial^2 \ln(L)/\partial B_1 \partial B_o & -\partial^2 \ln(L)/\partial B_1^2 & -\partial^2 \ln(L)/\partial B_1 \partial \sigma \\ -\partial^2 \ln(L)/\partial \sigma \partial B_o & -\partial^2 \ln(L)/\partial \sigma \partial B_1 & -\partial^2 \ln(L)/\partial \sigma^2 \end{bmatrix}^{-1}$$

$$(V.75)\quad \partial^2 \ln(L)/\partial B_o^2 = -(n-r)/\sigma^2 - \Sigma_c[(\cdot)f(\cdot) /F(\cdot)+f^2(\cdot)/F^2(\cdot)]/\sigma^2$$

$$(V.76)\quad \partial^2 \ln(L)/\partial B_1^2 = -\Sigma_u X_i^2/\sigma^2 - \Sigma_c\{X_i^2[(\cdot)f(\cdot)/F(\cdot)+f^2(\cdot)/F^2(\cdot)]\}/\sigma^2$$

$$(V.77)\quad \partial^2 \ln(L)/\partial \sigma^2 = -\Sigma_u(Y_i-B_o-B_1X_i)^2/\sigma^4-(n-r)/\sigma^2 - \Sigma_c\{[(\cdot)f(\cdot)/F(\cdot)]^2+(\cdot)^3 f(\cdot)/F(\cdot)\}/\sigma^2$$

$$(V.78)\quad \partial^2 \ln(L)/\partial B_o \partial \sigma = -\Sigma_u(Y_i-B_o-B_1X_i)/\sigma^3 - \Sigma_c\{(\cdot)^2 f(\cdot)/F(\cdot)+(\cdot)f^2(\cdot)/F^2(\cdot)\}/\sigma^2$$

$$(V.79)\quad \partial^2 \ln(L)/\partial B_1 \partial \sigma = -\Sigma_u X_i(Y_i-B_o-B_1X_i)/\sigma^3 - \Sigma_c\{X_i[(\cdot)^2 f(\cdot)/F(\cdot)+(\cdot)f^2(\cdot)/F^2(\cdot)]\}/\sigma^2$$

$$(V.80)\quad \partial^2 \ln(L)/\partial B_1 \partial B_o = -\Sigma_u X_i/\sigma^2 - \Sigma_c\{X_i[(\cdot)f(\cdot)/F(\cdot)+X_i f^2(\cdot)/F^2(\cdot)]\}/\sigma^2$$

Methods for solving this set of equations include a Newton-Raphson approach (Nelder and Wedderburn, 1972), a modification of this approach termed "iteratively re-weighted least squares" (IRLS) (Stirling, 1984), and expectation-maximization (EM) (Dempster et.al., 1977; Wolynetz, 1979b; Schmee and Hahn, 1979).

a. IRLS

The IRLS approach is based on:

$$(V.81)\quad \hat{\underline{\beta}}^+ = (\underline{X}'\hat{\underline{W}}\underline{X})^{-1}\underline{X}'\hat{\underline{W}}\hat{\underline{Z}}$$

$$(V.82)\quad \hat{\underline{Z}} = \hat{\underline{\eta}} - [\partial \ln L(\underline{Y})/\partial \eta]/(\partial^2 \ln L/\partial \eta^2)$$

$\hat{\underline{W}}$ is diagonal with elements $-\partial^2 \ln L(y_i)/\partial\eta^2$

$\hat{\underline{\eta}} = \underline{XB}$

For censored data,

$$(V.83) \quad z_i = \eta_i - \sigma[f(\cdot)/F(\cdot) + (\cdot)]^{-1}$$

$$(V.84) \quad w_i = \{[f(\cdot)/F(\cdot)]^2 + (\cdot)f(\cdot)/F(\cdot)\}/\sigma^2$$

For uncensored data,

$$(V.85) \quad z_i = y$$

$$(V.86) \quad w_i = 1/\sigma^2$$

The solution is iterative, as successive estimates of $\hat{\beta}$ result in new values for $\hat{Z}$. A usual initial estimate is that obtained from OLS using only uncensored observations.

b. Expectation maximization (EM)

The EM method is OLS with censored observations replaced by:

$$(V.87) \quad Y_i^* = \text{new Y for censored observation}$$

$$= E(Y_i/Y_i < a_i)$$

$$= X_o - \hat{\sigma}\Sigma[f(\cdot)/F(\cdot)]/(n-r),$$

$$= b_o + b_1 X_i - \hat{\sigma} f(\cdot)/F(\cdot)$$

This method is also iterative, as successive solutions to $\hat{\beta}_i$ from the normal equations result in new estimates for Y^*. Wolynetz(1979b) has provided a fortran program for its solution assuming a normal distribution. This method is analogous to the distribution free method of Buckley and James, with known f and F replacing that of the PL estimator (Miller, 1981).

5. Comparisons

Stirling (1984) found that the IRLS method converged faster than the EM method. Miller and Halperin (1982) compared several distribution free methods. They concluded that the Buckley and James method was the most reliable for the linear model. The method of Miller (1976) had methodological weaknesses related to the censoring pattern. Censoring patterns for water quality data tend to be heavily concentrated on the low end, while that for survival data tends to be more random. Though theoretical justification for the Buckley-James model is lacking, its usefulness has been justified by Monte-Carlo simulations.

McLeish(1983) stated that the Buckley-James method was a poor minimizer of the objective function (equation V.63). They suggest a modification of the Miller method.

Schmee and Hahn (1979) have developed a method of regression analysis for censored data in which data is grouped to form subsamples. For example, if there are several years of data, with several samples per year, one estimates the mean level for each year (using a method for censored data). Annual mean values are then regressed against time.

Zeger and Brookmeyer (1986) developed a method for regression analysis with serially correlated censored data which combines elements of MLE estimation for censored data with likelihood considerations for autoregressive models.

VI. SUMMARY AND CONCLUSIONS

A. Summary

Water quality management relies on information obtained from monitoring systems. For a variety of reasons, there is a great deal of concern with chemical substances which are present at levels too low to be measured precisely. Some of this concern is a consequence of poorly understood human health risks posed by many of these substances. It is widely believed that significant risk is incurred by any exposure. It has been suggested (Dowd, 1986) that drinking water standards be based on the ability of a measurement system to detect. A second aspect of this concern is a result of management's desire to obtain as much usable information as possible for a given amount of monitoring resources. This may mean spatially extending a monitoring network far from a point source. On a time scale, earliest possible detection is implied.

Obtaining information from a monitoring system which is operating at the fringes of analytical capabilities is not straightforward. This thesis is a discussion of the kinds of concerns one should have when statistically analyzing water quality data from such a system. Two general approaches are

discussed. The traditional approach is to regard all measurements as precise or imprecise. Precise results are simply numerical responses, for which statistical analysis may lead to valid and sound monitoring information. Imprecise results are reported as "ND", or not detected, with criteria for reporting based on categories of measurement precision. Data records which contain both precise and imprecise results may be analyzed using methods developed for censored data.

There are several problems with regarding data as ND or a single numerical result. One is that definitions of the limit of detection are not, nor indeed can be, standardized. What is deemed sufficient precision for one purpose may not be for a different purpose. Such a definition should be limited in purpose to its role as one of the fundamental performance characteristics of an analytical method. In this context, a standard definition, such as the IUPAC definition, makes sense in that it allows fair comparison between different analytical methods.

More generally, one should view all measurements as having some degree of imprecision rather than categorize data into distinct groups. Here, one attempts to consider all information produced by the measurement system. This includes, in addition to a numerical result, some estimate of data quality as well. This practice has been recommended by the American Chemical Society (1980).

A second problem with reporting ND is that data users may be mislead into thinking that measurements which are reported are made with high precision. This may or may not be the case. It was shown that data below the LOD does not resemble the population which was being measured. In this regard, censoring achieves its purpose. However, it was also shown that results between the LOD and LOQ may also not resemble the population from which they were taken. Reporting all results plus the observation error helps to eliminate misinterpretations of this variety.

Finally, a third problem with censoring is that it filters information from data. This is intuitively obvious, and is demonstrated for the example of estimating a mean from a set of measurements. It was demonstrated for the problem of trend detection by Gilliom et.al. (1984).

There are a variety of options for analyzing data for which observation error has been estimated. In this thesis, a noise model was used which is appropriate for measurements made near limits of detection. Models of this sort are useful for focusing error reduction efforts. In the case of measurement error, calibration design is one area where data precision can be improved at low cost.

It is conceivable that such a model could also be used for more than illustrative purposes. Data analysis would be improved if it were demonstrated that an error model accurately described actual data. For example, better use could be made of methods like SIDF in an effort to get at

population parameters as opposed to descriptions of the population of measurements. For real data, however, many factors which influence data precision are not constant, i.e., the measurement system cannot be assumed to be in a state of statistical control. Reasons for this abound, including data which are generated by more than one laboratory, and changes in measurement technology or procedures. There is an effort being made to deal with some of these problems. (For example, see Buonaccorsi, 1986b).

The influence of calibration design and some aspects of the statistical structure of measurements (as opposed to the population) are discussed. Measurement system - population combinations can be characterized as dominated by observation error, or relatively error free. It was shown that information can be obtained from systems which would be 100% censored with current practices. In addition, two methods for including observation error in an estimate of population variance are described. These are based on the POE estimate of variance and Satterthwaite's method as described by Gaylor and Hopper (1969).

Some additional uses can be made of this type of information. Monitoring systems can be classified according to their ability to detect changes in water quality. Sensitive systems are capable of detecting slight changes, while insensitive systems are dominated by observation error. Systems with low sensitivity produce results which are distributed differently from the underlying population.

For moderate degrees of insensitivity, subtracting components of variance due to observation error will produce good estimates of population variance, reducing the size of a confidence or tolerance interval. Subtracting components of variance is not useful for highly insensitive systems because of the high variance of the estimators used and the likelihood of negative variance estimates. Consideration of monitoring system sensitivity can lead to quantitative estimates of the magnitude of change they are capable of detecting.

To be useful, estimates of observation error must include all the sources of uncertainty in a monitoring system. It would be straightforward to develop a model for total error by including terms for sampling and sample preparation error.

Sometimes there is no choice but to use methods for censored data. When using these methods, it should be kept in mind that there can be significant observation error in the numbers which are reported. Also, many of these methods are sensitive to the underlying distribution. This refers to the distribution of measurements

While many methods are available, their properties for water quality applications have not been studied or compared to any great extent (with the exception of methods discussed by Gilliom and Helsel (1984) and Hirsch et.al.(1982)). Also, the point of censoring should depend on the eventual use to be made of the data.

B. Conclusions

The following conclusions are supported by this thesis.

1. Water quality data should not be censored

Censoring presumes that nothing is known about results below the censoring level (usually the LOD) and that all results above the LOD are reliable. Neither of these hold for water quality data.

2. Water quality results are probability statements

Information from water quality analyses are random variables. These include values which are descriptions of the measurement system, such as the LOD, LOQ, and relative error. It is usually not sound, from a data analysis perspective to regard them as constant.

3. Water quality measurements differ from the population

Observation error can significantly alter the statistical properties of a water quality random variable. Data with significant observation error will have statistical properties characteristic of the error.

4. Estimate monitoring system sensitivity

The ability of a monitoring system to achieve its objectives can be limited by system error. A quantitative assessment should be made of the impact of this error on statistical data analysis.

5. Methods for previously censored data are available

The literature contains methods for censored data and for many water quality statistical problems. These methods are useful when data has previously been censored and no information regarding such results is available. Some modification of published methods may be necessary due to differences between survival or failure time data and water quality data.

C. Recommendations

1. Report all results (do not censor)

The task of obtaining information from water quality would be simplified and improved if data were not censored. This requires that standard samples in the neighborhood of the LOD be analyzed.

The possibility of misinterpretation of an imprecise result can be reduced by careful presentation of monitoring information. The key is to keep in mind that analytical results are probability statements and should be presented as such.

2. Estimate and report observation error

Estimating the sources and magnitude of observation error serves several purposes. Identification of sources of uncertainty can properly focus efforts to reduce error. Secondly, when estimates of observation error are available, statistical methods which make use of this type of information can be used. This can lead to improved monitoring information.

Finally, information regarding observation error enables assessment of a monitoring system with regard to monitoring objectives. It is difficult to conceive of a monitoring system which is capable of meeting its objectives when this type of information has not been provided.

Estimates of observation error due to measurement can be extracted from the calibration information. Estimates of other sources of error may require additional effort and cost.

3. Apply above results to data which is consistent with the statistical characteristics of water quality data

Most of the discussion in this thesis is limited to independent samples from normal or lognormally distributed populations. In water quality, one must often deal with other distributions and seasonally or serially correlated data.

4. Study properties of methods for censored data

This thesis reviews a number of statistical methods which can be useful for water quality data with results of ND. However, the discussion is necessarily brief. Most of these have not been critically evaluated or compared. Exceptions include the methods evaluated by Gilliom and Helsel (1984) and by Hirsch et.al. (1982). Before other methods can be used reliably, their properties in situations representative of water quality problems should be the subject of further study. For example, it would be useful to compare the many tests for trend which are available.

D. Closing discussion

There is a need to statistically analyze water quality data in order to meet the evolving information needs of water quality management. Consequently, a discussion of observation error in monitoring systems in general, and of analytical error near LOD's in particular, is needed. Information users and producers should communicate more closely regarding water quality objectives. Information users should be concerned about data precision. Analysts should adhere to recommendations regarding the reporting of precision estimates, especially when imprecision is unavoidable.

Water quality managers have not created much demand for this type of information for a number of reasons. The cost

of an adequate quality assurance program can exceed the cost of analysis associated with monitoring alone. However, even a minimal quality assurance and quality control effort, combined with calibration results, would provide useful information. It should be noted that money spent on a monitoring system which fails to meet its objectives is money wasted. Data records which contain only ND's are not uncommon, and do not relate much information about the process being monitored. Data records which contain imprecision that has not been acknowledged may lead to false inferences, wasted expense, and embarrassment (Skogerboe and Koirtyohann, 1976).

In addition, there is a lack of statistical methodology which takes into account data precision. Information produced by an active quality assurance effort is not often utilized in a systematic fashion. It is not uncommon to see an analysis repeated because the result did not match the expectations of the data user, rather than for any objective criteria. As such, additional work is needed to evaluate a system error approach to reporting laboratory results and its ultimate usefulness in statistical analysis. The potential to improve water quality information generated by a monitoring program appears promising and worthy of additional research. Hopefully, this thesis will further enhance a dialogue concerning information needs and monitoring objectives between water quality managers and the analysts who generate the data.

VII. REFERENCES

Ahrens, L.H. 1965. Distribution of Elements in Our Planet. McGraw-Hill, New York, New York

Aitchison, J. and J.A.C. Brown. 1957. The Lognormal Distribution. Cambridge at the University Press

American Chemical Society Committee on Environmental Improvements. 1980. "Guidelines for Data Acquisition and Data Quality Evaluation in Environmental Chemistry". Analytical Chemistry, v.52, p.2242

Barr, D.R. and Davidson, T. 1973. "A Kolmogorov-Smirnov Test for Censored Samples". Technometrics, v.15, p.739

Bell, H. 1986. Senior chemist, Site Chemical/Environmental Services, IBM Corporation, East Fishkill, New York. Personal Communication

Beyer, W.H., ed. 1985. Handbook of Tables for Probability and Statistics. Chemical and Rubber Company Press, Boca Raton, Florida

Box, G.E.P., Hunter, W.G., Hunter, J.S. 1978. Statistics for Experimenters. John Wiley and Sons, New York.

Buckley, J. and James, I. 1979. "Linear Regression with Censored Data. Biometrika, v.3, p.429

Buonaccorsi, J.P. 1986a. "Design Considerations for Calibration". Technometrics, v.28, n.2, p.149

Buonaccorsi, J.P. 1986b. "Inferences Under Measuring Devices Having Random and Systematic Error". Submitted to Technometrics

Buonaccorsi, J.P. and Iyer, H.K. 1984. "A Comparison of Confidence Regions and Designs in Estimation of a Ratio." Communications in Statistical Simulation and Computation, v.13, p.723

Cardone, M.J. 1986a. "New Technique in Chemical Assay Calculations. 1. A Survey of Calculational Practices on a Model Problem". Analytical Chemistry, v.58, p.433

Cardone, M.J. 1986a. "New Technique in Chemical Assay Calculations. 2. Correct Solution of the Model Problem and Related Concepts". Analytical Chemistry, v.58, p.433

Cohen, A.C. 1957. "On the Solution of Estimating Equations for Truncated and Censored Samples from Normal Populations". Biometrika, v.44, p.225

Cohen, A.C. 1959. "Simplified Estimators for the Normal Distribution When Samples are Singly Censored or Truncated". Technometrics, v.1, n.3

Cohen, A.C. 1961a. "On the Solution of Estimating Equations for Truncated and Censored Samples from Normal Populations". Biometrika, v.44, p.225

Cohen, A.C. 1961b. "Tables for Maximum Likelihood Estimates: Singly Truncated and Singly Censored Samples". Technometrics, v.3, n.4, p.535

Cox, D.R. and Oakes, D. 1984. Analysis of Survival Data. Chapman and Hall, New York.

Council on Environment Quality. 1980. Final Report of the Interagency Task Force on Environmental Data and Monitoring. National Technical Information Service.

Crawford, C.G., Slack, J.R., and Hirsch, R.M. 1983. "Nonparametric Tests for Trends in Water-quality Data using the Statistical Analysis System". U.S. Geological Survey open-file report 83-550

Currie, L.A. 1968. "Limits for Qualitative Detection and Quantitative Determination". Analytical Chemistry, v. 40, p.586

Dempster, A.P. Laird, N.M. Rubin, D.B. 1977. "Maximum Likelihood from Incomplete Data via the EM Algorithm". Journal of the Royal Statistical Society, B., v.39, p.1

Dowd, R.M. 1985. "EPA Drinking Water Proposals: Round One". Environmental Science and Technology, v.19, p.1156

Draper, N., Smith, H. 1981. Applied Regression Analysis. John Wiley and Sons, New York.

Dufour, R. and Maag, U.R. 1978. "Distribution Results for Modified Kolmogorov-Smirnov Statistics for Truncated or Censored Samples". Technometrics, v.20, p.29

Durbin, J. and Knott, M. 1972. "Components of Cramér-von Mises Statistics.I". Journal of the Royal Statistical Society B., v.34, p.290

Elving, P.J.(ed.). 1984. Treatise on Analytical Chemistry. Part I. Theory and Practice. Interscience, Wiley and Sons, New York, N.Y.

Fieller, E.C. 1932. "The Distribution of the Index in a Normal Bivariate Population". Biometrika, v.24, p.428

Finney, D.J. 1941. "On the Distribution of a Variate Whose Logarithm is Normally Distributed". Journal of the Royal Statistical Society Supplement 7, p.151

Garden, J.S., Mitchell, D.G., and Mills, W.N. 1980. "Nonconstant Variance Regression Techniques for Calibration-Curve-Based Analysis". Analytical Chemistry, v.52, p.2310

Gaylor, D.W. and Hopper, F.N. 1969. "Estimating the Degrees of Freedom for Linear Combinations of Mean Squares by Satterthwaite's Formula". Technometrics, v.11, n.4, p.691

Gehan, E.A. 1965. "A Generalized Wilcoxon Test for Comparing Arbitrarily Singly-Censored Samples". Biometrika, v.52, p.203

General Accounting Office. 1981. Better Monitoring Techniques are Needed to Assess the Quality of Rivers and Streams. US General Accounting Office Report No. CED-81-30, Washington, D.C.

Gilbert, R.O. and R.R. Kinnison. 1981. "Statistical Methods for Estimating the Mean and Variance from Radionuclide Data Sets Containing Negative, Unreported or Less-Than Values". Health Physics, v.40,p.377

Gilliom, R.J. and D.R. Helsel. 1984. "Estimation of Distributional Parameters for Censored Trace-Level Water-Quality Data". United States Geological Survey open file report 84-729

Gilliom, R.J.; R.M. Hirsch; E.J. Gilroy. 1984. "Effect of Censoring Trace-Level Water-Quality Data on Trend-Detection Capability". Environmental Science and Technology, v.18, n.7, p.530

Glaser, J.A.;Forest, D.L.; McKee, G.D.; Quave, S.A; Buddle, W.L. 1981. Trace Analysis for Wastewaters. Environmental Science and Technology, v.15, n.12, p.1426

Graybill, F. 1976. Theory and Application of the Linear Model. Duxbury Press, North Scituate, Massachusetts.

Gupta, A.K. 1952. "Estimation of the Mean and Standard Deviation of a Normal Population from a Censored Sample". Biometrika, v.39, p.260

Hahn, G.J. 1982. "Removing Measurement Error in Assessing Conformance to Specifications". Journal of Quality Technology, v.14, n.3, p.117

Hald, A. 1949. "Maximum Likelihood Estimation of the Parameters of a Normal Distribution Which is Truncated at a Known Point". Skandinavisk Aktuarietidskrift, V.32, p.119

Halperin, M. 1952. "Estimation in the Truncated Normal Distribution". Journal of the American Statistical Association, v.47, p.457

Halperin, M. 1960. "Extension of the Wilcoxon-Mann-Whitney Test to Samples Censored at the Same Fixed Point". Journal of the American Statistical Society, v.55, p.125

Hanson, J.K. and Hoel, D.G. 1974. "Tables of Gehan's Generalized Wilcoxon Test with Fixed Point Censoring". Journal of Statistical Computations and Simulation, v.3. p.117

Hinkley, D.V. 1969. "On the Ratio of Two Correlated Normal Random Variables". Biometrika, v.56, p.635

Hirsch, R.M.; J.R. Slack; R.A. Smith. 1982. "Techniques of Trend Analysis for Monthly Water Quality Data". Water Resources Research, v.18, n.1, p.107

Hollander, M. and Wolfe, D.A. 1973. Nonparametric Statistical Methods. Wiley, New York

Hubaux, A. and Vos, G. 1970. "Decision and Detection Limits for Linear Calibration Curves". Analytical Chemistry, v.42, p.849

Hunter, W.G., Lamboy, W.F. 1981. "A Bayesian Analysis of the Linear Calibration Problem". Technometrics, v.23, n.4.

Iyer, H. 1986. Associate Professor, Colorado State University, Department of Statistics. Personal Communication.

Jaech, J.L. 1984. "Removing the Effects of Measurement Errors in Constructing Statistical Tolerance Intervals". Journal of Quality Technology, v.16, n.2, p.69

Kaiser, H. 1970a. "Quantitation in Elemental Analysis, Part I". Analytical Chemistry, v.42, p.24a

Kaiser, H. 1970b. "Quantitation in Elemental Analysis, Part II". Analytical Chemistry, v.42, p.26a

Kalbfleisch, J. and R.L. Prentice. 1980. The Statistical Analysis of Failure Time Data. Wiley, New York

Kateman, G. and Pijpers, F.W. 1981. Quality Control in Analytical Chemistry. John Wiley and Sons, New York

Kendall, M.G. and Stuart, A. 1979. The Advanced Theory of Statistics, Griffin, London, Volume 2, 4th edition

Koziol, J.A. and Byar, D.P. 1975. "Percentage Points of the Asymptotic Distributions of One and Two Sample K-S Statistics for Truncated or Censored Data". Technometrics, v.17, p.507

Larsen, R.J. and Marx, M.L. 1981. An Introduction to Mathematical Statistics and its Applications. Prentice-Hall, Englewood Cliffs, New Jersey

Lettenmaier, D.P.; L.L. Conquest; J.J. Hughes. 1982. Routine Streams and Rivers Water Quality Trend Monitoring Review. University of Washington, Department of Civil Engineering, Charles W. Harris Hydraulics Laboratory Technical Report No. 75

Long, G.L. and Winefordner, J.D. 1983. "Limit of Detection, A Closer Look at the IUPAC Definition". Analytical Chemistry, v.55, p.712a

Mann, H.B. 1945. "Nonparametric Tests Against Trend". Econometrica, v. 13, p.245

Marsaglia, G. 1965. "Ratios of Normal Variables and Ratios of Sums of Uniform Variables". Journal of the American Statistical Association, v.60, p.193

Massart, D.L., Dijkstra, A., Kaufman, L. 1978. Evaluation and Optimization of Laboratory Methods and Analytical Procedures. Techniques and Instrumentation in Analytical Chemistry, Volume 1, Elsevier, New York

McLeish, D.L. 1983. "A Note on Non-Parametric Censored Regression". Journal of Statistical Computation and Simulation, v.18, p.1

Mee, R.W. 1984. "Tolerance Limits and Bounds for Proportions Based on Data Subject to Measurement Error". Journal of Quality Technology, v.16, n.2, p.74

Mee, R.W., Owen, D.B., Shyu, J. 1986. "Confidence Bounds for Misclassification Probabilities Based on Data Subject to Measurement Error". Journal of Quality Technology, v.18, n.1, p.29

Menzies, A.C. and Kaiser, H. 1969. Two Papers on the Limit of Detection of a Complete Analytical Procedure. Hafner, New York

Miller, R.G. 1976. Least Squares Regression with Censored Data. Biometrika, v.63, p.449

Miller, R.G. 1981. Survival Analysis. Wiley and Sons, New York

Miller, R. and Halperin, J. 1982. "Regression with Censored Data". Biometrika, v.69, p.521

Mood, A.M., Graybill, F.A. and Boes, D.C. 1974. Introduction to the Theory of Statistics. McGraw-Hill, New York

Mulrow, J. 1986. M.S. Candidate, Colorado State University, Department of Statistics. Personal communication

Naszodi, L.J. 1978. "Elimination of the Bias in the Course of Calibration". Technometrics, v.20, p.201

National Academy of Sciences. 1977. Analytical studies for the US Environmental Protection Agency. Volume IV, Environmental Monitoring. National Academy of Sciences, Washington DC.

Nelder, J.A. and Wedderburn, R.W.M. 1972. "Generalized Linear Models". Journal of the Royal Statistical Society A., v.135, p.370

Ott, R.L. and Myers, R.H. 1968. "Optimal Experimental Designs for Estimating the Independent Variable in Regression". Technometrics, v.10, n.4, p.811

Patnaik, P.B. 1949. "The Non-central x^2 and F-distributions and Their Applications". Biometrika, v.36, p.202

Persson, T. and Rootzen, H. 1977. "Simple and Highly Efficient Estimators for a Type I Censored Normal Sample. Biometrika, v.64, n.1, p. 123

Pettitt, A.N. 1976. "Cramr-von Mises Statistics for Testing Normality with Censored Samples". Biometrika, v.63, p.475

Pettitt, A.N. and Stephens, M.A. 1976. "Modified Cramr-von Mises Statistics for Censored Data". Biometrika, v.63, p.291

Pomeroy, R.D. and Orlob, G.T. 1967. "Problems of Setting Standards and of Surveillance for Water Quality Control". California State Water Quality Control Commission, publication number 36, Sacramento, California

Prudnikov, E.D. 1981. "Theoretical Calculation of the Standard Deviation in Atomic Emission Spectroscopy". Spectrochimica Acta, v.36B, p. 385

Prudnikov, E.D. and Shapkina, Y.S. 1984. "Random Errors in Analytical Methods". Analyst, v.109, p. 305

Ross, J.W. and Fraser, M.D. 1977. "Analytical Clinical Chemistry Precision, State of the Art for Fourteen Analytes". American Journal of Clinical Pathology, v.68, p.130

Satterthwaite, F.E. 1946. "An Approximate Distribution of Estimates of Variance Components". Biometrics Bulletin, v.2, p.110.

Saw, J.G. 1961. "Estimation of the Normal Population Parameters Given a Type I Censored Sample", Biometrika, v.3, p.367

Schmee, J. and G.J. Hahn. 1979. "A Simple Method for Regression Analysis with Censored Data". Technometrics, v.21, p.417

Shapiro, S.S.; Wilk, M.B.; Chen, H.J. 1968. "A Comparative Study of Various Tests for Normality". Journal of the American Statistical Association, v.39, p.1343

Shaw, D.M. 1961. "Element Distribution Laws in Geochemistry". Geochemica Cosmochimica Acta, v.23, p.116

Skogerboe, R.K. 1982. "The Analytical Blank: Sources and Effects on Lead Analyses". Journal of the Association of Analytical Chemists, v.65, p.957

Skogerboe, R.K. 1986. Professor of Chemistry, Colorado State University. Personal communication

Skoog, D.A. 1985. Principles of Instrumental Analysis. Saunders College Publishing, New York

Skougstad, M.W., Fishman, M.J., Friedman, L.C., Erdmann, D.E., Duncan, S.S., eds. 1979. Techniques of Water-Resources Investigations of the United States Geological Survey, Book 5.

Steck, G.P. 1971. "Rectangle Probabilities for Uniform Order Statistics and the Probability that the Empirical Distribution Function Lies Between Two Distribution Functions". The Annals of Mathematical Statistics, v.42, p.1

Stirling, W.D. 1984. "Iteratively Reweighted Least Squares for Models with a Linear Part". Applied Statistics, v.33, p.7

Taylor J.K., Stanley, T.W. 1985. Quality Assurance for Environmental Measurements. ASTM Special Technical Publication 867. ASTM, Philadelphia, Pa.

Taylor, C., Iyer, H. 1986. "The Calibration Problem". Colorado State University Statistics Department, Technical Report.

Thompson, M. and Howarth, R.J. 1976. "Duplicate Analysis in Geochemical Practice". Analyst, v.101, p. 690

Thompson, M. and Howarth, R.J. 1980. "The Frequency Distribution of Analytical Error". Analyst, v.105, p.1188

U.S. House of Representatives, Ninety-Eighth Congress, First Session. Hearings before the Subcommittee on Natural Resources, Agriculture Research and Environment of the Committee on Science and Technology; National Environmental Monitoring; May and June 1983. U.S. Government Printing Office, Washington, D.C.

Vistelius, A. 1960. "The Skew Frequency Distributions and the Fundamental Law of the Geochemical Process". Journal of Geology, v.68, p.1

Ward, R.C. and McBride, G.B. 1984. Design of Water Quality Monitoring Systems in New Zealand. Water Quality Centre, Ministry of Works and Development, draft copy

Ward, R.C. and McBride, G.B. 1986. Design of Water Quality Monitoring Systems in New Zealand. Water Quality Centre, Ministry of Works and Development, Publication No. 8

Welch, B.L. 1947. "The Generalisation of 'Student's' Problem When Several Different Population Variances are Involved". Biometrika, v.34, p.28

West Publishing. 1983. Selected Environmental Law Statutes. St. Paul, Minnesota

Williford, W.O., Carter, M.C., and Field, J.E. 1979. "A Further Look at the Bayesian Approach to Calibration". Journal of Statistical Computation and Simulation, v.9, p.47

Wilson, A.L. 1970a. "The Performance Characteristics of Analytical Methods - I". Talanta, v.17, p.21

Wilson, A.L. 1970b. "The Performance Characteristics of Analytical Methods - II". Talanta, v.17, p.31

Wilson, A.L. 1970c. "The Performance Characteristics of Analytical Methods - III". Talanta, v.20, p.725

Winefordner, J.D. and Ward, J.L. 1980. "The Reliability of Detection Limits in Analytical Chemistry". Analytical Letters, v.13, p. 1293

Wolynetz, M.S. 1979a. "Maximum Likelihood Estimation from Confined and Censored Normal Data", Applied Statistics, v.28, p.185

Wolynetz, M.S. 1979b. "Maximum Likelihood Estimation in a Linear Model from Confined and Censored Normal Data", Applied Statistics, v.28, p.195

Yao, Y. 1984. Associate Professor, Colorado State University, Department of Statistics. Personal communication

Appendix A. Numerical examples for censored data methods

A.1 Estimation of the mean and variance from censored samples

.1 Assign zero or the detection limit
.2 Assign uniform random numbers
.3 Assign normal random numbers
.4 Assign lognormal random numbers
.5 Maximum likelihood for normally distributed data
.6 Linear estimator for normally distributed data
.7 Maximum likelihood for lognormally distributed data
.8 Probability plot for lognormally distributed data

A.2 Goodness of fit tests for censored samples

.1 Chi-squared
.2 Kolmogorov-Smirnov
.3 Cramer-Von Mises

A.3 Two sample tests for censored samples

.1 Halperin's extension of Wilcoxon-Mann-Whitney
.2 Gehan's extension of Wilcoxon-Mann-Whitney

A.4 Tests for trend

.1 Mann and Kendall's rank tests
.2 Hoeffding
.3 Miller estimators
.4 Buckley-James
.5 Iteratively reweighted least squares
.6 Expectation maximization

A.1 Estimation of the mean and variance from censored samples

Appendix A.1 contains numerical examples of methods which can be used to estimate the mean and variance of censored samples. The purpose of the examples is to illustrate the use of the methods described in Chapter V. The data used are shown in Table A.1. Set 1 was generated from a normal distribution with a mean and variance of 5, and was used in examples A.1.1, A.1.2, A.1.3, A.1.4, A.1.5, and A.1.6. Set 2 is simply the natural log of set 1 and was used in examples A.1.7, and A.1.8.

Table A.1

Set 1	Set 2
6.27	1.84
7.18	1.97
<5	<1.61
<5	<1.61
5.14	1.64
<5	<1.61
6.03	1.80
<5	<1.61
<5	<1.61
<5	<1.61
<5	<1.61
<5	<1.61
7.65	2.03
<5	<1.61
<5	<1.61
8.33	2.12
10.36	2.34
<5	<1.61
5.19	1.65
5.75	1.75

A.1.1 Assign zero or the detection limit

Data set 1 in Table A.1.
Method introduced on page 94 of text

Result	Assign 0	Assign 5
6.27	6.27	6.27
7.18	7.18	7.18
<5	0.0	5.0
<5	0.0	5.0
5.14	5.14	5.14
<5	0.0	5.0
6.03	6.03	6.03
<5	0.0	5.0
<5	0.0	5.0
<5	0.0	5.0
<5	0.0	5.0
<5	0.0	5.0
7.65	7.65	7.65
<5	0.0	5.0
<5	0.0	5.0
8.33	8.33	8.33
10.36	10.36	10.36
<5	0.0	5.0
5.19	5.19	5.19
5.75	5.75	5.75

estimate of sample mean:

61.9/20 = 3.10 116.9/20 = 5.85

estimate of sample variance:

= 13.54 = 2.14

Note that these methods are biased in that zero is too small for all of the censored points, and five is too large. The properties of this method when used for multiply censored data have not been studied.

A.1.2 Assign uniform random numbers

Data set 1 in Table A.1.
Method introduced on page 95 of text

Result	Uniform(0,5)Random Number	New Data Set
6.27		6.27
7.18		7.18
<5	2.78	2.78
<5	0.74	0.74
5.14		5.14
<5	4.92	4.92
6.03		6.03
<5	2.82	2.82
<5	1.26	1.26
<5	2.44	2.44
<5	0.71	0.71
<5	2.04	2.04
7.65		7.65
<5	4.78	4.78
<5	2.32	2.32
8.33		8.33
10.36		10.36
<5	4.81	4.81
5.19		5.19
5.75		5.75

estimate of sample mean = 91.52/20

= 4.58

estimate of sample variance = 7.06

The properties of this method when used for multiply censored data have not been studied.

A.1.3 Assign normal random numbers

Data set 1 in Table A.1.

This method is introduced on page 95 of text. Figure A.1 is a normal probability plot of the data. Normal scores were assigned using equation V.1 on page 95 of the text. "New results" are the ordinate corresponding to a normal score of Figure A.1.

Result	Ordered Results	Normal Score	New Result	Assign
6.27	<5	-1.665	-1.539	0.0
7.18	<5	-1.311	-0.325	0.0
<5	<5	-1.067	0.51	0.51
<5	<5	-0.878	1.16	1.16
5.14	<5	-0.713	1.72	1.72
<5	<5	-0.565	2.23	2.23
6.03	<5	-0.432	2.69	2.69
<5	<5	-0.303	3.13	3.13
<5	<5	-0.179	3.56	3.56
<5	<5	-0.060	3.96	3.96
<5	<5	0.060	4.38	4.38
<5	5.14	0.179		
7.65	5.19	0.303		
<5	5.75	0.432		
<5	6.03	0.565		
8.33	6.27	0.713		
10.36	7.18	0.878		
<5	7.65	1.067		
5.19	8.33	1.311		
5.75	10.36	1.665		

slope of probability plot = 3.428

intercept of probability plot = 4.169

estimate of sample mean = 85.24/20 = 4.26

estimate of sample variance = 8.41

The properties of this method when used for multiply censored data have not been studied. Multiple censoring may present a problem with ranking. Also note that "new results" of less than zero are reassigned zero.

Wolynetz (1979a) provides a computer program which computes $f(\cdot)/[1-F(\cdot)]$ (the reciprocal of the Mills' ratio), where $f(\cdot)$ and $F(\cdot)$ are values of the PDF and CDF, respectively, for a standard normal distribution, and $(\cdot)$ is the normal score. This program can be used to find normal scores for this method. One can choose $(\cdot)$, compute $f(\cdot)/[1-F(\cdot)]$

with the program, solve for F(·) and compare the result with r/(n+1), incrementing (·) until program values agree with the chosen (·).

A.1.4 Assign lognormal random numbers

Data set 1 in Table A.1.

This method is introduced on page 95 of text. Figure A.2 is a normal probability plot of the log transformed data. Normal scores were assigned using equation V.1 on page 95 of the text. "New results" are the ordinates corresponding to normal scores of Figure A.2.

Natural Log of Result	Ordered Results	Normal Score	New Result
1.84	<1.61	-1.665	0.75
1.97	<1.61	-1.311	0.91
<1.61	<1.61	-1.067	1.03
<1.61	<1.61	-0.878	1.12
1.64	<1.61	-0.713	1.20
<1.61	<1.61	-0.565	1.27
1.80	<1.61	-0.432	1.33
<1.61	<1.61	-0.303	1.39
<1.61	<1.61	-0.179	1.45
<1.61	<1.61	-0.060	1.50
<1.61	<1.61	0.060	1.56
<1.61	1.64	0.179	
2.03	1.65	0.303	
<1.61	1.75	0.432	
<1.61	1.80	0.565	
2.12	1.84	0.713	
2.34	1.97	0.878	
<1.61	2.03	1.067	
1.65	2.12	1.311	
1.75	2.34	1.665	

slope of probability plot = 0.471
intercept of probability plot = 1.532

new sample mean (of logs) = 1.533
new sample variance (of logs) = 0.176

transformation to population mean and variance:

$$\hat{\alpha} = \exp(1.533 + 0.176/2) = 5.06$$

$$\hat{\beta}^2 = \exp(2 \cdot 1.533 + 0.176) \cdot [\exp(0.176) - 1] = 4.92$$

Multiple censoring may present a problem with ranking. Also note that "new results" of less than zero cannot occur.

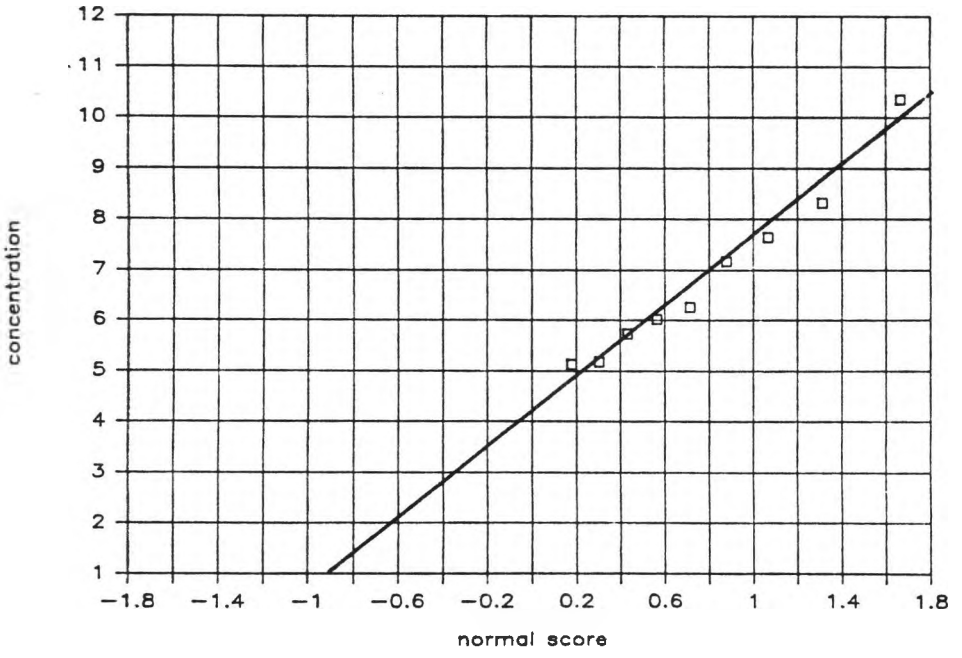

Figure A.1 Normal probability plot

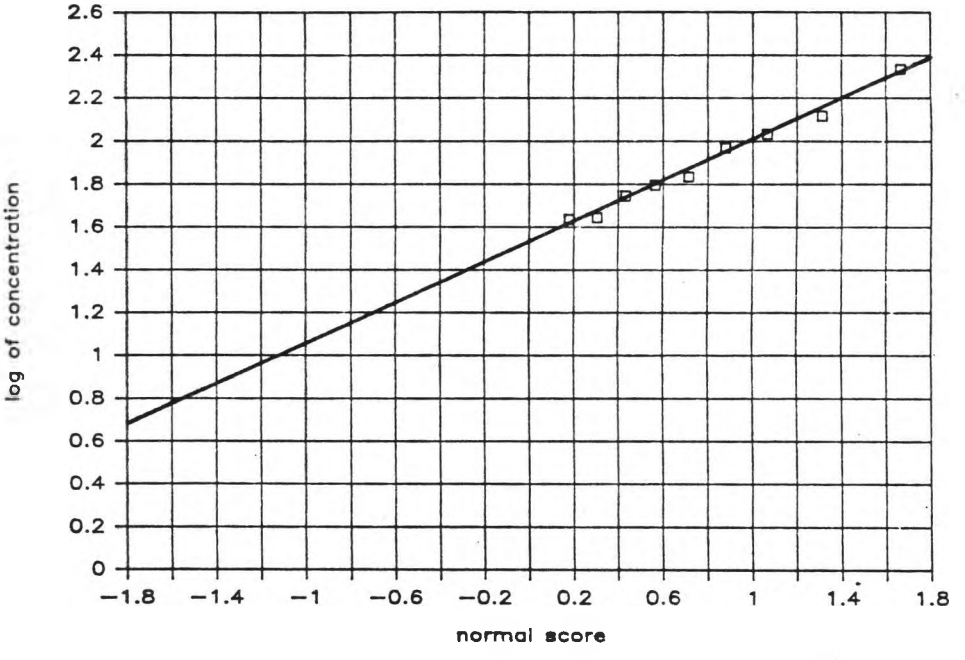

Figure A.2 Lognormal probability plot

A.1.5 Maximum likelihood for normally distributed data

Data set 1 in Table A.1, n = 20, r = 11, n-r = 9
This method is based on equations V.8 - V.11 (page 97).

iteration

1) X_o $= 61.89/9$
$= 6.88$

S_o^2 $= 23.17/9$
$= 2.57$

$(a-\hat{\mu})/\hat{\sigma}$ $=(5.00-6.88)/0.219 = -1.175$

$f[(a-\hat{\mu})/\hat{\sigma}] = 0.200$ $F[(a-\hat{\mu})/\hat{\sigma}] = 0.120$

2) $\hat{\mu}$ $= X_o-(\hat{\sigma}r/(n-r))f((a-\hat{\mu})/\hat{\sigma})/F((a-\hat{\mu})/\hat{\sigma})$

$= 6.88-(1.60)(11/9)(0.200/0.120) = 3.62$

$(a-\hat{\mu})/\hat{\sigma}$ $=(5.00-3.62)/1.60 = 0.861$

$f[(a-\hat{\mu})/\hat{\sigma}] = 0.275$ $F[(a-\hat{\mu})/\hat{\sigma}] = 0.805$

$\hat{\sigma}^2$ $= [S_o^2+(X_o-\hat{\mu})^2]\{1+[r/(n-r)][(a-\hat{\mu})/\hat{\sigma}]$
$\cdot f[(a-\hat{\mu})/\hat{\sigma}]/F[(a-\hat{\mu})/\hat{\sigma}]\}^{-1}$

$= [2.57 + (6.88-3.62)^2]$
$\cdot[1 + (11/9)(0.861)(0.275/0.805)]^{-1}$

$= 9.704$

$(a-\hat{\mu})/\hat{\sigma}$ $= (5.00-3.62)/3.115 = 0.142$

$f[(a-\hat{\mu})/\hat{\sigma}] = 0.395$ $F[(a-\hat{\mu})/\hat{\sigma}] = 0.556$

3) $\hat{\mu}$ $= 6.88-(3.115)(11/9)(0.395/0.556) = 4.18$

$(a-\hat{\mu})/\hat{\sigma}$ $=(5.00-4.18)/3.115 = 0.265$

$f[(a-\hat{\mu})/\hat{\sigma}] = 0.385$ $F[(a-\hat{\mu})/\hat{\sigma}] = 0.605$

$\hat{\sigma}^2$ $= [2.57 + (6.88-4.18)^2]$
$\cdot[1 + (11/9)(0.265)(0.385/0.605)]^{-1}$

$= 8.17$

$(a-\hat{\mu})/\hat{\sigma}$ $=(5.00-4.18)/2.858 = 0.287$

$f[(a-\hat{\mu})/\hat{\sigma}] = 0.383$ $F[(a-\hat{\mu})/\hat{\sigma}] = 0.613$

4) $\hat{\mu} = 6.88-(2.858)(11/9)(0.383/0.613) = 4.70$

$(a-\hat{\mu})/\hat{\sigma} = (5.00-4.70)/2.858 = 0.105$

$f[(a-\hat{\mu})/\hat{\sigma}] = 0.397 \quad F[(a-\hat{\mu})/\hat{\sigma}] = 0.542$

$\hat{\sigma}^2 = [2.57 + (6.88-4.70)^2]$

$\cdot[1 + (11/9)(0.105)(0.397/0.542)]^{-1}$

$= 6.69$

$(a-\hat{\mu})/\hat{\sigma} = (5.00-4.70)/2.587 = 0.116$

$f[(a-\hat{\mu})/\hat{\sigma}] = 0.396 \quad F[(a-\hat{\mu})/\hat{\sigma}] = 0.546$

5) $\hat{\mu} = 6.88-(2.587)(11/9)(0.396/0.546) = 4.59$

$(a-\hat{\mu})/\hat{\sigma} = (5.00-4.59)/2.587 = 0.160$

$f[(a-\hat{\mu})/\hat{\sigma}] = 0.394 \quad F[(a-\hat{\mu})/\hat{\sigma}] = 0.564$

$\hat{\sigma}^2 = [2.57 + (6.88-4.59)^2]$

$\cdot[1 + (11/9)(0.160)(0.394/0.564)]^{-1}$

$= 6.84$

$(a-\hat{\mu})/\hat{\sigma} = (5.00-4.59)/2.615 = 0.157$

$f[(a-\hat{\mu})/\hat{\sigma}] = 0.394 \quad F[(a-\hat{\mu})/\hat{\sigma}] = 0.562$

6) $\hat{\mu} = 6.88-(2.615)(11/9)(0.394/0.562) = 4.64$

$(a-\hat{\mu})/\hat{\sigma} = (5.00-4.64)/2.615 = 0.138$

$f[(a-\hat{\mu})/\hat{\sigma}] = 0.395 \quad F[(a-\hat{\mu})/\hat{\sigma}] = 0.555$

$\hat{\sigma}^2 = [2.57 + (6.88-4.64)^2]$

$\cdot[1 + (11/9)(0.138)(0.395/0.555)]^{-1}$

$= 6.62$

Assume $\hat{\mu} = 4.64$ and $\hat{\sigma}^2 = 6.62$

Six iterations were used for this example. The number required depends on convergent criteria and sample size. Small samples may take longer. The computer program developed by Wolynetz (1979a) is based on this method. This method can be used with multiply censored data.

A.1.6 Linear estimator for normally distributed data

Data set 1 in Table A.1

This method is based on equations V.19 through V.22 on page 101 of the text.

r=11, r/n=0.55, n-r = 9, $f(0.55)=0.343$, $F^{-1}(0.55)=0.126$

observed mean = 61.92/9 = 6.88

$$\Sigma(X_i - a)^2 = 54.87$$

$$\sigma_{rml} = \tfrac{1}{2}[F^{-1}(r/n)(X_o-a) + \{[F^{-1}(r/n)(X_o-a)]^2 + [4/(n-r)]\Sigma(x_i-a)^2\}^{\frac{1}{2}}\}$$

$$= 0.5\{\cdot[(0.126)(6.88 - 5.00)] + \{[(0.126)(6.88 - 5.00)]^2 + (4/9)(54.87)\}^{0.5}\} = 2.61$$

$$\hat{\sigma}^2 = S_o^2 - \{[n/(n-r)]f(r/n)F^{-1}(r/n) - [n/(n-r)]^2[f(r/n)]^2\}\sigma_{rml}^2$$

$$= 2.57 - \{(20/9)(0.343)(0.126) - (20/9)^2(0.343)^2\}(2.61)^2$$

$$= 5.87$$

$$\hat{\mu} = X_o - [n/(n-r)]f(r/n)\sigma_{rml}$$

$$= 6.88 - (20/9)(0.343)(2.61)$$

$$= 4.89$$

The advantage of this method over MLE$_2$ is that no iterations are required. This is because μ and σ^2 are functions of the data only, and not of the estimates.

A.1.7 Maximum likelihood for lognormally distributed data

Data set 2 in Table A.1.

This method is identical to MLE for normally distributed data except for the transformations given by equations V.25 and V.26 in the text.

n = 20, r = 11, n-r = 9

<u>iteration</u>

1) $X_o = 17.14/9 = 1.904$

$S_o^2 = 0.430/9 = 0.0478$

$(a-\hat{\mu})/\hat{\sigma} = (1.610-1.904)/0.219 = -1.342$

$f[(a-\hat{\mu})/\hat{\sigma}] = 0.162 \quad F[(a-\hat{\mu})/\hat{\sigma}] = 0.0898$

2) $\hat{\mu} = X_o-[\hat{\sigma}r/(n-r)]f[(a-\hat{\mu})/\hat{\sigma}]/F[(a-\hat{\mu})/\hat{\sigma}]$

$= 1.904-(0.219)(11/9)(0.162/0.0898)$

$= 1.422$

$(a-\hat{\mu})/\hat{\sigma} = (1.610-1.422)/0.219 = 0.858$

$f[(a-\hat{\mu})/\hat{\sigma}] = 0.276 \quad F[(a-\hat{\mu})/\hat{\sigma}] = 0.805$

$\hat{\sigma}^2 = [S_o^2+(X_o-\hat{\mu})^2]\{1+[r/(n-r)][(a-\hat{\mu})/\hat{\sigma}]$

$\cdot f[(a-\hat{\mu})/\hat{\sigma}]/F[(a-\hat{\mu})/\hat{\sigma}]\}^{-1}$

$= [0.0478 + (1.904-1.422)^2]$

$\cdot[1 + (11/9)(0.858)(0.276/0.805)]^{-1}$

$= 0.206$

$(a-\hat{\mu})/\hat{\sigma} = (1.610-1.422)/0.4542 = 0.414$

$f[(a-\hat{\mu})/\hat{\sigma}] = 0.366 \quad F[(a-\hat{\mu})/\hat{\sigma}] = 0.660$

3) $\hat{\mu} = 1.904-(0.454)(11/9)(0.366/0.660) = 1.592$

$(a-\hat{\mu})/\hat{\sigma} = (1.610-1.592)/0.454 = 0.0397$

$f[(a-\hat{\mu})/\hat{\sigma}] = 0.399 \quad F[(a-\hat{\mu})/\hat{\sigma}] = 0.516$

$\hat{\sigma}^2 = [0.0478 + (1.904-1.592)^2]\cdot$

$$[1 + (11/9)(0.0397)(0.399/0.516)]^{-1}$$

$= 0.140$

$(a-\hat{\mu})/\hat{\sigma} = (1.610-1.592)/0.374$

$= 0.0481$

$f[(a-\hat{\mu})/\hat{\sigma}] = 0.399 \quad F[(a-\hat{\mu})/\hat{\sigma}] = 0.519$

4) $\hat{\mu} = 1.904-(0.374)(11/9)(0.399/0.519) = 1.553$

$(a-\hat{\mu})/\hat{\sigma} = (1.610-1.553)/0.374 = 0.152$

$f[(a-\hat{\mu})/\hat{\sigma}] = 0.394 \quad F[(a-\hat{\mu})/\hat{\sigma}] = 0.560$

$\hat{\sigma}^2 = [0.0478 + (1.904-1.553)^2]\cdot$

$$[1 + (11/9)(0.152)(0.394/0.560)]^{-1}$$

$= 0.151$

$(a-\hat{\mu})/\hat{\sigma} = (1.610-1.553)/0.389 = 0.147$

$f[(a-\hat{\mu})/\hat{\sigma}] = 0.395 \quad F[(a-\hat{\mu})/\hat{\sigma}] = 0.558$

5) $\hat{\mu} = 1.904-(0.389)(11/9)(0.395/0.558) = 1.567$

$(a-\hat{\mu})/\hat{\sigma} = (1.610-1.567)/0.389 = 0.111$

$f[(a-\hat{\mu})/\hat{\sigma}] = 0.397 \quad F[(a-\hat{\mu})/\hat{\sigma}] = 0.544$

$\hat{\sigma}^2 = [0.0478 + (1.904-1.567)^2]\cdot$

$$[1 + (11/9)(0.111)(0.398/0.544)]^{-1}$$

$= 0.147$

$(a-\hat{\mu})/\hat{\sigma} = (1.610-1.567)/0.383 = 0.112$

$f[(a-\hat{\mu})/\hat{\sigma}] = 0.396 \quad F[(a-\hat{\mu})/\hat{\sigma}] = 0.545$

6) $\hat{\mu} = 1.904-(0.3783(11/9)(0.396/0.545) = 1.564$

$(a-\hat{\mu})/\hat{\sigma} = (1.610-1.564)/0.383 = 0.120$

$f[(a-\hat{\mu})/\hat{\sigma}] = 0.396 \quad F[(a-\hat{\mu})/\hat{\sigma}] = 0.548$

$\hat{\sigma}^2 = [0.0478 + (1.904-1.546)^2]\cdot$

$$[1 + (11/9)(0.120)(0.396/0.548)]^{-1}$$

$= 0.148$

Assume $\hat{\mu} = 1.56$, and $\hat{\sigma}^2 = 0.148$

Estimate of the population mean:

$$\hat{\alpha} = \exp(1.56 + 0.148/2)$$

$$= 5.12$$

Estimate of the population variance:

$$\hat{\beta}^2 = \exp(2 \cdot 1.56 + 0.148) \cdot [\exp(0.148) - 1]$$

$$= 4.19$$

By adding the transformation noted, the computer program developed by Wolynetz (1979a) can be used for this method.

A.1.8 Probability plot for lognormally distributed data

Data set 2 in Table A.1.

$$\hat{\mu} = \ln(X_{0.50})$$

$$= 1.53$$

$$\hat{\sigma} = \ln[0.50 \cdot (X_{0.50}/X_{0.16} + X_{0.84}/X_{0.50})]$$

$X_{0.50} = \exp(1.53)$ $X_{0.16} = \exp(1.05)$ $X_{0.84} = \exp(2.00)$

$$\hat{\sigma} = 0.475$$

transformation to population mean:

$$\hat{\alpha} = \exp(1.53 + 0.475/2)$$

$$= 5.86$$

transformation to population variance:

$$\hat{\beta}^2 = \exp[2 \cdot 1.53 + (0.475)^2] \cdot [\exp(0.475)^2 - 1]$$

$$= 6.76$$

This method is based on equations V.35 and V.36 in the text. Figure A.2 is the probability plot used for the example. This method is similar to A.1.4.

A.2 Goodness of fit tests for censored samples

Appendix A.2 contains examples of goodness of fit tests for censored samples. The data is shown with each example. Methods include the chi-squared, the Kolmolgorov-Smirnov, and Cramer-Von Mises tests. Descriptions of these are given in section V.C of the text. Each test is of the simple hypothesis:

H_o: F(x) is normally distributed with mean 1.0 and variance 1.0

vs. H_a: F(x) is not distributed as described

A.2.1 Chi-Squared

X	ranked X	interval	numb.	numb.exp.	$(n_i-n_e)^2/(nP_i)$
1.35	<0.5	<0.5	6	6.17	0.0046
1.81	<0.5				
<0.5	<0.5	0.5 to 1	6	3.83	1.2294
0.68	<0.5				
0.88	<0.5	1 to 1.5	3	3.83	0.1798
<0.5	<0.5				
1.24	0.60	1.5 to 2	3	3.83	0.1798
1.53	0.61				
<0.5	0.68	>2.19	2	2.34	0.0494
<0.5	0.72				
<0.5	0.88				
0.60	0.89				
2.07	1.12				
0.61	1.24				
<0.5	1.35				
2.49	1.53				
4.00	1.81				
0.72	2.07				
0.89	2.49				
1.12	4.00				

sum = 1.643

$\chi^2_{0.95,4}$ = 9.488, therefore accept H_o

Note that n_i < 5 in some cases, therefore the test is invalid. This method is based on equation V.39 in the text. Although this example is a test for normality, it can be used for any distribution. Assumptions inherent in the test require that there be at least five observations per interval. Power considerations suggest that there be as many intervals as possible. This can be difficult for small samples, especially when they are censored.

A.2.2 Kolmogorov-Smirnov

X	ordered	$S_n(X_i)$	$S_n(X_{i-1})$	$F_o(X_i)$	D1	D2
1.35	<0.5					
1.81	<0.5					
<0.5	<0.5					
0.68	<0.5					
0.88	<0.5		$F_o(T) = 0.3085$			
<0.5	<0.5					
1.24	0.6	0.35	0.25	0.344	0.0054	-0.0946
1.53	0.61	0.4	0.3	0.348	0.0517	-0.0483
<0.5	0.68	0.45	0.35	0.374	0.0755	-0.0245
<0.5	0.72	0.5	0.4	0.389	0.1103	0.0103
<0.5	0.88	0.55	0.45	0.452	0.0978	-0.0022
0.6	0.89	0.6	0.5	0.456	0.1438	0.0438
2.07	1.12	0.65	0.55	0.547	0.1022	0.0022
0.61	1.24	0.7	0.6	0.594	0.1052	0.0052
<0.5	1.35	0.75	0.65	0.636	0.1132	0.0132
2.49	1.53	0.8	0.7	0.701	0.0981	-0.0019
4.00	1.81	0.85	0.75	0.791	0.059	-0.041
0.72	2.07	0.9	0.8	0.857	0.0423	-0.0577
0.89	2.49	0.95	0.85	0.931	0.0181	-0.0819
1.12	4.00	1.00	0.9	0.998	0.0013	-0.0987

maximum = 0.1438

$d_{0.05,20} = 0.2336$ for $F_o(T) = 0.30$

Therefore Accept H_o

True significance level > 0.15

This method is based on equation V.44 and Table V.1 of chapter V.

A.2.3 Cramer-Von Mises

X	Rank	$F_o(x_i)$	$[F_o(x_i)-(2i-1)/2n]^2$
1.35	15	0.6368	0.007779
1.81	17	0.7910	0.001156
<0.5	1-6		
0.68	9	0.3745	0.0101
0.88	11	0.4522	0.005299
<0.5	1-6		
1.24	14	0.5948	0.006432
1.53	16	0.7019	0.005344
<0.5	1-6		
<0.5	1-6		
<0.5	1-6		
0.60	7	0.3446	0.000384
2.07	18	0.8577	0.000299
0.61	8	0.3483	0.000713
<0.5	1-6		
2.49	19	0.9319	0.000047
4.00	20	0.9987	0.000562
0.72	10	0.3897	0.00728
0.89	12	0.5478	0.0141
1.12	13		0.00596

sum = 0.0655

S = 7, R = 13 $F_o(T) = p = 1-F_o(0.50) = 1-0.3815 = 0.6915$

Above Sum - $R(4R^2-1)/12n^2 + np(R^2/n^2-pR/n^2+p^3/3)$

$= 0.0655 - 13\cdot(4\cdot169-1)/4800 +$

$+ 20\cdot(0.6915)(169/400-(0.6915)\cdot13/20+(0.6915)$

$= 0.06868$

Table value for $_{0.7}W^2 = 0.386$, for $\alpha = 0.05$

Therefore, accept H_o

True significance level approximately 0.50

This test is based on equation V.47 and Table V.3 of chapter V.

A.3 Two sample tests for censored samples

Appendix A.3 contains examples for Halperin's (1960) and Gehan's (1965) two sample tests. Each considers the hypothesis:

$$H_o: \quad F_x(z) = G_y(z)$$

$$\text{vs.} \quad H_o: \quad F_x(z) > G_y(z)$$

These are single tail hypotheses. The data used for the examples are shown in Table A.2. Set 1 was used for the Halperin test and set 2 for the Gehan test. Note that the only difference between the two is that the Halperin samples are censored at the same fixed point, while the Gehan sample Y is multiply censored.

Table A.2 Data for Two Sample Tests

Set 1		Set 2	
Sample X	Sample Y	Sample X	Sample Y
6.27	<4	6.27	<4
7.18	13.95	7.18	13.95
<5	<4	<5	<5
<5	10.11	<5	10.11
5.14	6.78	5.14	6.78
<5	<4	<5	<6
6.03	8.83	6.03	8.83
<5	<4	<5	<5
<5	<4	<5	<4
<5	5.23	<5	5.23
<5	<4	<5	<5
<5	10.88	<5	10.88
7.65	4.75	7.65	4.75
<5	<4	<5	<3
<5	4.64	<5	4.64
8.33	<4	8.33	<6
10.36	7.91	10.36	7.91
<5	<4	<5	<5
5.19	7.52	5.19	7.52
5.74	<4	5.74	<5

A.3.1 Halperin extension of Wilcoxon-Mann-Whitney

S = Sum of ranks of uncensored Y's in the sequence of all uncensored observations.

$= 1 + 2 + 5 + 9 + 11 + 13 + 15 + 16 + 18 + 19 = 109$

$U_c = [(m-r_m)/2](2n+1+m-r_m) - S$

$= [(20-10)/2][(2)(20)+1+20-10) - 109 = 146$

$\mu_c = mn(m+n-r)(m+n+r-1)/[2(m+n)(m+n-1)]$

$= (20)(20)(40-21)(61-1)/[2(40)(39)] = 146.15$

$\sigma_c^2 = mn(m+n-r)(A+B+C-D)/[4(m+n)]$

$= (20)(20)(40-21)(3.077+404.9+47.5-449.7)/[4(40)]$

$= 274.4$

where $A = [(m+n-r)^2-1]/[3(m+n-1)]$

$= [(40-21)^2-1]/[3(40-1)] = 3.077$

$B = (m-1)(m+n-r-1)\{n + (2n+1)r/(m+n+2) + r(r-1)(n-1)/[(m+n-2)(m+n-3)]\}/(m+n-1)$

$= (19)(40-21-1)\{20+(41)(21)/(40+2)+(21)(21-1)(20-1)/[(40-2)(40-3)]\}/(40-1) = 404.9$

$C = n+(2n+1)r/(m+n-1) + r(r-1)(n-1)/[(m+n-1)(m+n-2)]$

$= 20 + (40+1)21/(40-1) + 21(21-1)(19)/[(40-1)(40-2)]$

$= 47.462$

$D = mn(m+n-r)(m+n+r-1)^2/[(m+n)(m+n-1)^2]$

$= (20)(20)(40-21)(40+21-1)^2/[40(40-1)^2)$

$= 449.7$

At the 99% confidence level for a single tail test, reject H_o if $Z > 2.33$. The Z value is given by:

$Z = (U_c - \mu_c)/\sigma_c = (146.00 - 146.15)/16.6$

$= -0.00904$

Therefore H_o cannot be rejected. This test is based on equation V.49 in the text.

A.3.2 Gehan's extension of Wilcoxon-Mann-Whitney

Test: H_o: $F_X(z) = G_Y(z)$ vs.

H_a: $F_X(z) > G_Y(z)$ vs.

Uncensored Values		Ranks	
4.64	(Y)	1	$K_j = j$
4.75	(Y)	2	
5.14	(X)	3	All $k_i = 1$
5.19	(X)	4	
5.23	(Y)	5	$l_1 = 0$; $L_i = 0$
5.75	(X)	6	
6.03	(X)	7	$l_2 = 16$; $L_2 = 16$
6.27	(X)	8	
6.78	(Y)	9	$l_6 = 2$; $L_6 = 18$
7.18	(X)	10	
7.52	(Y)	11	all other $l_i = 0$
7.65	(X)	12	
7.91	(Y)	13	n = m = 20
8.33	(X)	14	
8.83	(Y)	15	
10.11	(Y)	16	
10.36	(X)	17	
10.88	(Y)	18	
13.95	(Y)	19	

X_i uncensored < Y_j uncensored		54·-1 =	-54
or X_i censored $\leq^j Y_j$ uncensored		8	- 8
X_i uncensored = Y_j uncensored		0· 0 =	0
or X_i and Y_j both censored		110	0
or X_i censored > Y_j uncensored		22	0
or X_i uncensored $\leq^j Y_j$ censored		6	0
X_i uncensored > Y_j uncensored		36·+1 =	36
or X_i uncensored $\geq$ Y_j censored		90	90
	sum 20·20 =	400	64

$$nm/[(n+m)(n+m-1)] = 20^2/(40^2-40)$$

$$= 0.2564$$

$$\Sigma k_i K_{i-1}(K_{i-1}+1) = 0+2+6+12+20+30+42+(7)(8)+(8)(9)+(9)(10)$$
$$+(10)(11)+(11)(12)+(12)(13)+(13)(14)$$
$$+(14)(15)+(15)(16)+(16)(17)+(17)(18)+(18)(19)$$
$$= 2280$$

$$\Sigma l_i K_i(K_i+1) = (16)(2))(3) + (2)(6)(7) = 180$$

$$\Sigma k_i(n+m-K_i-L_{i-1})(n+m-3K_{i-1}-k_i-L_{i-1}-1)$$
$$= (39)(38)+(38)(35)+(21)(16)+(20)(13)+(19)(10)$$
$$+(18)(7)+(15)(2)+(14)(-1)+(13)(-4)+(12)(-7)$$
$$+(11)(-10)+(10)(-13)+(9)(-16)+(8)(-19)+$$
$$+(7)(-22)+(6)(-25)+(5)(-28)+(4)(-31)$$
$$= 2500$$

$$Var(W) = nm\{\Sigma k_i K_{i-1}(K_{i-1}+1) + \Sigma l_i K_i(K_i+1)$$
$$+ \Sigma k_i(n+m-K_i-L_{i-1})(n+m-3K_{i-1}-k_i-L_{i-1}-1)\}$$
$$/[(n+m)(n+m-1)]$$
$$= 0.2564(2280 + 180 + 2500) = 1271.7$$

Standard Deviation = 35.7

Assuming normality, (64 - 0)/35.7 = 1.793 For a one sided test, the probability that the test statistic will exceed 1.793 under H_o is about 0.96. Therefore accept H_o. The significance level of the test is about (1-0.96) = 0.04. We could reject H_o at the 95% confidence level. This example is based on equations V.50 and V.51 in the text.

A.4 Tests for trend

Appendix A.4 provides examples of tests for trend which can be used with censored data. The data used is shown in Table A.3. Table A.3 also contains the OLS estimates of B_o and B_1 based on the uncensored observations which are used as initial estimates for the iterative methods.

Table A.3 Data for trend detection examples

X	Y	d	$d(X-\bar{X})^2$	$dY(X-\bar{X})$
1	1.35	1	90.25	-12.825
2	1.81	1	72.25	-15.385
3	<2	0	0	0
4	<2	0	0	0
5	<2	0	0	0
6	<2	0	0	0
7	1.24	1	12.25	-4.34
8	1.53	1	6.25	-3.825
9	<1	0	0	0
10	<1	0	0	0
11	<1	0	0	0
12	<1	0	0	0
13	2.07	1	6.25	5.175
14	<0.5	0	0	0
15	<0.5	0	0	0
16	2.49	1	30.25	13.695
17	4	1	42.25	26
18	<0.2	0	0	0
19	<0.2	0	0	0
20	1.12	1	90.25	10.64
sums	84	15.61	8	350	19.15

B_1-start = 19.15/350 = 0.054671

B_o-start = $\bar{Y} - B_1 \bar{X}$ = 15.61/8 - (0.0547)(84/8) = 1.377

var-start = 0.868558

The data for rank tests is singly censored at 0.5, whereas the data for other examples is censored as above.

A.4.1 Mann and Kendall's rank tests

time	Y	rank of Y	Mann's T	Kendall's K
1	1.35	15	5	-9
2	1.81	17	3	-12
3	<0.5	<7	12	12
4	0.68	9	9	2
5	0.88	11	7	-1
6	<0.5	<7	10	10
7	1.24	14	4	-5
8	1.53	16	3	-6
9	<0.5	<7	8	8
10	<0.5	<7	8	8
11	<0.5	<7	8	8
12	0.6	7	7	6
13	2.07	18	2	-3
14	0.61	8	5	4
15	<0.5	<7	5	5
16	2.49	19	1	-2
17	4	20	0	-3
18	0.72	10	2	2
19	0.89	12	1	1
20	1.12	13	0	0
			100	25

E(Mann Score) = (20)(19)/4 - (6)(5)/4 = 87.5
E(Kendall's K) = 0

Var(Mann's T) = [(20)(19)(40+5)/72-(6)(5)(12+5)]/72
= 230.4166

Var(Kendall's K) = [(20)(19)(40+5)/18-(6)(5)(12+5)]/18
= 921.6666

Mann's Z value: $(100 - 87.5)/(230.7)^{1/2} = 0.822974$

Kendall's Z value: $(25)/(921.667)^{1/2} = 0.823479$

Level of Test = 0.21 for a Mann single tail test
0.42 for a Kendall two tailed test

Mann's and Kendall's tests for independence in time are based on equations V.53 and V.54 in the text. These tests may not be applicable to multiply censored samples (see discussion in the text). The examples make use of the large sample approximation to normality.

A.4.2 Hoeffding

time	Y	rank of Y	c	Q	R	S
1	1.35	15	0	0	0	0
2	1.81	17	1	0	0	0
3	<0.5	3.5	0	7.5	0	0
4	0.68	9	1	336	14	0
5	0.88	11	2	1080	54	2
6	<0.5	3.5	0.5	75	3	-0.25
7	1.24	14	4	4680	240	12
8	1.53	16	6	8820	504	30
9	<0.5	3.5	1	210	10.5	0
10	<0.5	3.5	1.5	270	18	0.75
11	<0.5	3.5	2	337.5	27	2
12	0.6	7	5	3300	250	20
13	2.07	18	12	35904	2112	132
14	0.61	8	6	6552	432	30
15	<0.5	3.5	2.5	682.5	48.75	3.75
16	2.49	19	15	64260	3570	210
17	4	20	16	82080	4320	240
18	0.72	10	9	19584	1152	72
19	0.89	12	11	33660	1870	110
20	1.12	13	12	45144	2376	132
				306982.5	17001.25	996.25

$$D = [Q-2(n-2)R+(n-2)(n-3)S]/[n(n-1)(n-2)(n-3)(n-4)]$$

$$= [306982.5 - 2(18)(17001.25) + 18(17)(996.25)]$$

$$/[(20)(19)(18)(17)(16)]$$

$$= -0.00011$$

$$B \text{ statistic} = (20\pi^4/2)(D+1/(36n))$$

$$= 1.242954$$

$b(\alpha=0.05)$ from table A.26 in Hollander and Wolfe(1973)

$$= 2.85$$

Therefore, accept Ho at the 95% level

Actual level approximately 0.40

This example is based on equations V.55 through V.57 in the text. As with other rank tests, there may be some problem when samples are multiply censored.

A.4.3 Miller estimators

First Iteration: Initial Slope Estimate = 0.054671

ranked z	X	Y	d	F(PL)	wi
3.070593	17	4	1	1	0.05882
1.835987	3	2	0	1	0.06274
1.781316	4	2	0	1	0.06274
1.726645	5	2	0	1	0.06274
1.700658	2	1.81	1	0.941176	0.06274
1.671974	6	2	0	0.941176	0.06274
1.615264	16	2.49	1	0.878431	0.10457
1.359277	13	2.07	1	0.815686	0.26144
1.295329	1	1.35	1	0.752941	0.26144
1.092632	8	1.53	1	0.690196	
0.857303	7	1.24	1	0.627450	
0.507961	9	1	0	0.627450	
0.45329	10	1	0	0.627450	
0.398619	11	1	0	0.627450	
0.343948	12	1	0	0.627450	
0.02658	20	1.12	1	0.522875	
-0.26539	14	0.5	0	0.522875	
-0.32006	15	0.5	0	0.522875	
-0.78407	18	0.2	1	0.522875	
-0.83874	19	0.2	0	0.261438	

X_u	Y_u	$w(x-\bar{X}_u)^2$	$wy(x-\bar{X}_u$	$w(y-b_1x)$	$w(y-\hat{\beta}_o-x\hat{\beta}_1)$
17	4	0.386136	0.602845	0.1806	36.0419
2	1.81	9.706765	-1.41255	0.1067	4.1841
16	2.49	0.153106	0.244053	0.1013	14.5209
13	2.07	0.129730	-0.18675	0.0852	9.2689
1	1.35	11.33034	-1.13826	0.0812	1.7891
8	1.53	2.600575	-0.61803	0.0685	3.9755
7	1.24	5.785357	-0.96449	0.0896	2.3297
20	1.12	8.088068	1.628638	0.0069	4.4214
18	0.2	5.441223	0.238540	-0.219	0.4717
	Sums:	43.62130	-1.60604	0.5011	77.00

$\bar{X}_u$ = 14.438

New slope estimate = -1.606/43.62 = -0.03681

New intercept estimate = 0.5011

Estimate of the variance of B_1 = 77.004/43.62 = 1.765

Miller Estimator: Fourth Iteration

Slope Estimate: -0.02777

ranked z	X	Y	d	F(PL)	wi
4.47209	17	4	1	1	0.05
2.93432	16	2.49	1	0.95	0.05
2.43101	13	2.07	1	0.9	0.06429
2.16662	6	2	0	0.9	0.06429
2.13885	5	2	0	0.9	0.06429
2.11108	4	2	0	0.9	0.06429
2.08331	3	2	0	0.9	0.06429
1.86554	2	1.81	1	0.835714	0.28929
1.75216	8	1.53	1	0.771428	0.28929
1.6754	20	1.12	1	0.707142	
1.43439	7	1.24	1	0.642857	
1.37777	1	1.35	1	0.578571	
1.33324	12	1	0	0.578571	
1.30547	11	1	0	0.578571	
1.2777	10	1	0	0.578571	
1.24993	9	1	0	0.578571	
0.91655	15	0.5	0	0.578571	
0.88878	14	0.5	0	0.578571	
0.72763	19	0.2	0	0.578571	
0.69986	18	0.2	1	0.289285	

X_u	Y_u	$w(x-\bar{X}_u)^2$	$wy(x-\bar{X}_u$	$w(y-b_1x)$	$w(y-\hat{\beta}_o-x\hat{\beta}_1)$
17	4	2.204005	1.327857	0.2236	16.959
16	2.49	1.590077	0.702091	0.1467	3.7770
13	2.07	0.447803	0.351213	0.1562	1.5170
2	1.81	4.493670	-0.97282	0.1199	0.1865
8	1.53	0.358262	-0.23219	0.1126	0.0737
20	1.12	5.973160	0.694028	0.1077	0.0266
7	1.24	0.726068	-0.26789	0.0922	0.0316
1	1.35	25.34807	-3.65569	0.3985	0.0664
18	0.2	16.88233	0.441987	0.2024	1.4799
	sums:	58.02345	-1.61143	1.560109	24.11

$\bar{X}_u$ = 10.361

New slope estimate = -1.61143/58.02345 = -0.02777

New intercept estimate = 1.560109

Estimate of the variance of B_1 = 24.11/58.02345 = 0.4157

This example is based on equations V.62 through V.67 in the text.

A.4.4 Buckley-James

Initial Iteration: $B_1 = 0.054671$

X	Y	d	z	F(z)	w	wz/F(z)	sums(z_k<z_i)	new Y
17	4	1	3.07	1.0	0.0588	0.1806		4
3	2	0	1.83	1.0			0.280	0.444
4	2	0	1.78	1.0				0.499
5	2	0	1.72	1.0				0.553
2	1.81	1	1.70	0.941	0.0627	0.1067		1.81
6	2	0	1.67	0.941			0.184	0.512
16	2.49	1	1.61	0.878	0.0627	0.1013		2.49
13	2.07	1	1.35	0.815	0.0627	0.0852		2.07
1	1.35	1	1.29	0.752	0.0627	0.0812		1.35
8	1.53	1	1.09	0.690	0.0627	0.0685		1.53
7	1.24	1	0.85	0.627	0.1045	0.0537		1.24
9	1	0	0.50	0.627			-0.345	0.146
10	1	0	0.4	0.627				0.201
11	1	0	0.39	0.627				0.256
12	1	0	0.34	0.627				0.311
20	1.12	1	0.0	0.522	0.2614	0.0027		1.12
14	0.5	0	-0.2	0.522			-0.419	0.346
15	0.5	0	-0.3	0.522				0.400
18	0.2	0	-0.7	0.522				0.564
19	0.2	1	-0.8	0.261	0.2614	-0.219		0.2

New $B_1 = 0.024701$

Final Iteration: $B_1 = 0.006225$

X	Y	d	z	F(z)	w	wz/F(z)	sums(z_k<z_i)	new Y
17	4	1	3.89	1	0.05	0.1947		4
16	2.49	1	2.39	0.95	0.05	0.1195		2.49
13	2.07	1	1.98	0.9	0.05	0.1278		2.07
3	2	0	1.98	0.9			0.76181	0.2562
4	2	0	1.97	0.9				0.3109
5	2	0	1.96	0.9				0.3655
6	2	0	1.96	0.9				0.4202
2	1.81	1	1.79	0.835	0.064	0.1155		1.81
8	1.53	1	1.48	0.771	0.064	0.0951		1.53
1	1.35	1	1.34	0.707	0.064	0.0863		1.35
7	1.24	1	1.19	0.642	0.064	0.0769		1.24
20	1.12	1	0.99	0.578	0.064	0.2879		1.12
9	1	0	0.94	0.578			0.04086	0.0726
10	1	0	0.93	0.578				0.1273
11	1	0	0.93	0.578				0.1820
12	1	0	0.92	0.578				0.2366
14	0.5	0	0.41	0.578				0.3460
15	0.5	0	0.40	0.578				0.4006
18	0.2	0	0.08	0.578				0.5647
19	0.2	1	0.08	0.289	0.289	0.0236		

New B_1 = 0.006225, New B_o = 0.925096,

Variance of the slope = 0.001646

This test for trend is based on equation V.68 in the text, and OLS estimation using the new set of data. This method can be used for multiply censored samples. It is the distribution free equivalent of the EM method. Only two iterations are shown (the first and the last), whereas several were needed. This method requires a computer as a practical matter.

A.4.5 Iteratively reweighted least squares

Initial Iteration: $B_1 = 0.0547$, $B_o = 1.377$, $\sigma^2 = 0.8686$

$(\cdot)$	$\underline{F}(\cdot)$	$\underline{f}(\cdot)$	$\underline{\eta}_i$	$\underline{z}_i$	$\underline{w}_i$
NA	NA	NA	NA	1.35	1.151
NA	NA	NA	NA	1.81	1.151
0.492	0.684	0.353	1.541	0.617	0.600
0.433	0.67	0.363	1.596	0.640	0.608
0.374	0.64	0.371	1.650	0.675	0.639
0.316	0.63	0.379	1.705	0.690	0.637
NA	NA	NA	NA	1.24	1.151
NA	NA	NA	NA	1.53	1.151
-0.932	0.176	0.258	1.869	0.125	0.903
-0.991	0.16	0.244	1.924	0.178	0.937
-1.049	0.15	0.229	1.978	0.047	0.851
-1.108	0.13	0.215	2.033	0.342	1.053
NA	NA	NA	NA	2.07	1.151
-1.762	0.039	0.084	2.142	-0.176	1.001
-1.821	0.034	0.075	2.197	-0.054	1.064
NA	NA	NA	NA	2.49	1.151
NA	NA	NA	NA	4	1.151
-2.318	0.01	0.027	2.361	-0.015	1.223
-2.377	0.0087	0.023	2.416	-0.345	1.055
-1.449	0.24	0.139	2.470	1.12	1.151

Result: $B_1 = -0.00472$, $B_o = 1.046$, $\sigma^2 = 9.82$

13th Iteration: $B_1 = -0.00564$, $B_o = 0.615$, $\sigma^2 = 2.29$

$(\cdot)$	$\underline{F}(\cdot)$	$\underline{f}(\cdot)$	$\underline{\eta}_i$	$\underline{z}_i$	$\underline{w}_i$
NA	NA	NA	NA	1.35	0.421
NA	NA	NA	NA	1.81	0.421
0.894	0.81	0.267	0.598	-0.634	0.170
0.898	0.82	0.266	0.592	-0.641	0.167
0.902	0.82	0.265	0.587	-0.644	0.167
0.906	0.82	0.264	0.581	-0.647	0.167
NA	NA	NA	NA	1.24	0.421
NA	NA	NA	NA	1.53	0.421
0.268	0.61	0.384	0.564	-1.125	0.239
0.272	0.61	0.384	0.559	-1.124	0.239
0.276	0.61	0.384	0.553	-1.124	0.240
0.279	0.61	0.383	0.547	-1.124	0.241
NA	NA	NA	NA	2.07	0.421
-0.037	0.48	0.398	0.536	-1.383	0.278
-0.033	0.49	0.398	0.530	-1.421	0.268
NA	NA	NA	NA	2.49	0.421
NA	NA	NA	NA	4	0.421
-0.216	0.41	0.389	0.514	-1.563	0.294
-0.212	0.42	0.390	0.508	-1.622	0.280
NA	NA	NA	NA	1.12	0.421

Result: $B_1 = -0.00564$, $B_o = 0.615$, Variance = 2.29

Slope Variance = 0.00445 (same as EM method)

This method is based on equations V.81 through V.86 in the text. It is not practical without a computer. It can be used with multiply censored data.

A.4.6 Expectation maximization

Initial Iteration: B_o = 0.05467, B_1 = 1.377, Var. = 0.8686

Standardized Y	F(std.Y)	f(std.Y)	new Y
NA	NA	NA	1.35
NA	NA	NA	1.81
0.492	0.688	0.353	1.062
0.433	0.666	0.363	1.087
0.374	0.644	0.371	1.112
0.316	0.626	0.379	1.140
NA	NA	NA	1.24
NA	NA	NA	1.53
-0.93	0.176	0.258	0.501
-0.99	0.161	0.244	0.511
-1.05	0.147	0.229	0.521
-1.10	0.134	0.215	0.532
NA	NA	NA	2.07
-1.76	0.039	0.084	0.135
-1.82	0.034	0.075	0.138
NA	NA	NA	2.49
NA	NA	NA	4
-2.31	0.01	0.027	-0.16
-2.37	0.008	0.023	-0.11
NA	NA	NA	1.12

Result: B_1 = -0.01234, B_o = 1.233, Var. = 2.091

13th Iteration: B_1 = -0.00566, B_o = 0.6297, Var. = 2.302

Standardized Y	F(std.Y)	f(std.Y)	new Y
NA	NA	NA	1.35
NA	NA	NA	1.81
0.914	0.820	0.263	0.127
0.918	0.821	0.262	0.123
0.922	0.822	0.261	0.120
0.925	0.823	0.260	0.116
NA	NA	NA	1.24
NA	NA	NA	1.53
0.278	0.609	0.384	-0.38
0.281	0.611	0.383	-0.38
0.285	0.612	0.383	-0.38
0.289	0.613	0.382	-0.38
NA	NA	NA	2.07
-0.03	0.487	0.398	-0.69
-0.03	0.488	0.398	-0.69
NA	NA	NA	2.49
NA	NA	NA	4
-0.22	0.415	0.389	-0.90
-0.21	0.416	0.389	-0.90
NA	NA	NA	1.12

Result: B_1 = -0.00565, B_o = 0.629, Variance = 2.3049
Slope variance = 0.00445
This method is based on equations V.87 and V.88 in the text.

Appendix B. Example of subtracting components of variance

An example is provided of the use of Satterthwaite's improved degrees of freedom for calculating a confidence interval for a population mean. The sample is assumed to come from a normal population and has homogenous measurement error. Here, X_p is normally distributed with mean μ_p and variance σ_p^2 and $\hat{X}$ is normally distributed with mean μ_p and variance $\sigma_p^2 + \sigma^2(\varepsilon)$ The sample consists of 4 replicate measurements each of four samples. The sample of measurements is:

$\hat{X}$ =	$\hat{X}_1$ =	1.5	1.1	1.8	1.0	$X_{i.}$ = 1.350	$X_{..}$ = 1.988
	$\hat{X}_2$ =	2.5	2.4	1.7	2.9	2.375	
	$\hat{X}_3$ =	2.0	1.7	2.9	2.2	2.200	
	$\hat{X}_4$ =	2.4	2.0	1.6	2.1	2.025	

$$\hat{\sigma}^2(\varepsilon) = \Sigma\Sigma(X-X_{i.})^2/[n(n-1)]$$
$$= (0.410 + 0.747 + 0.780 + 0.327)/12$$
$$= 0.189$$
$$\hat{\sigma}_m^2 = \Sigma(X_{i.}-X_{..})^2/(n-1)$$
$$= 0.201$$
$$Q = 0.201/(0.189/4) = 4.261$$
$$\hat{f} = (Q - 1)^2/[Q^2/f_1 + 1/f_2]$$
$$= (4.261-1)^2/(4.261^2/3 + 1/12)$$
$$= 1.733$$

A confidence interval is given by:

$$CI = \bar{X} \pm t(1-\alpha/2, \hat{f})\hat{\sigma}_p S/(n)^{\frac{1}{2}}$$

$$= 1.988 \pm (5.00)(0.201 - 0.189/4)^{\frac{1}{2}}/2$$

for $\alpha = 0.05$

$$= [1.008, 2.968]$$

Appendix C. PDF's used in the simulation studies

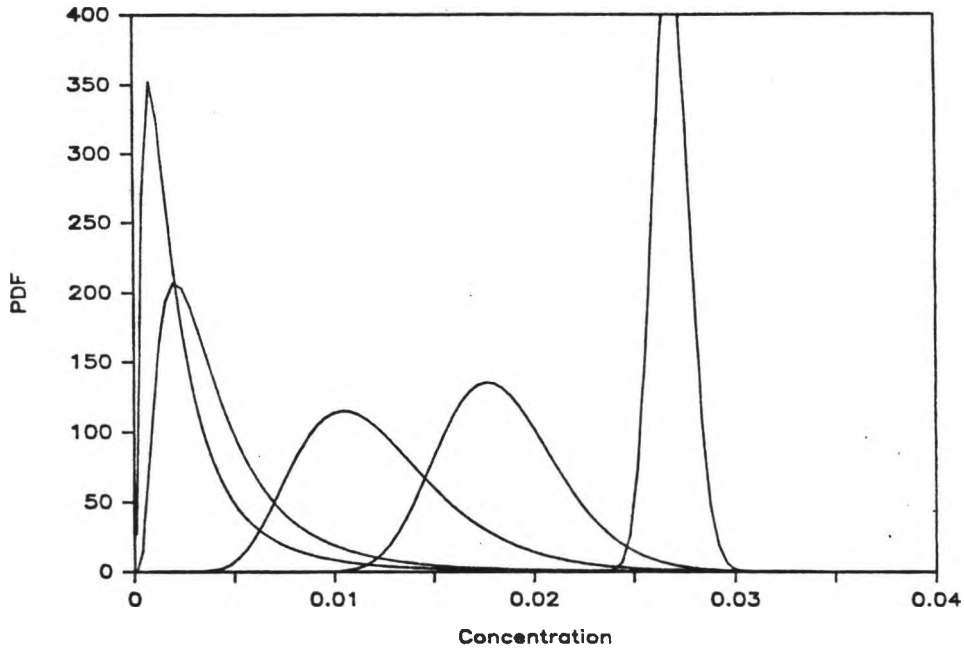

Figure C.1 PDF used in simulation studies with R = 1.0 and FC:FBQ = 1,1.

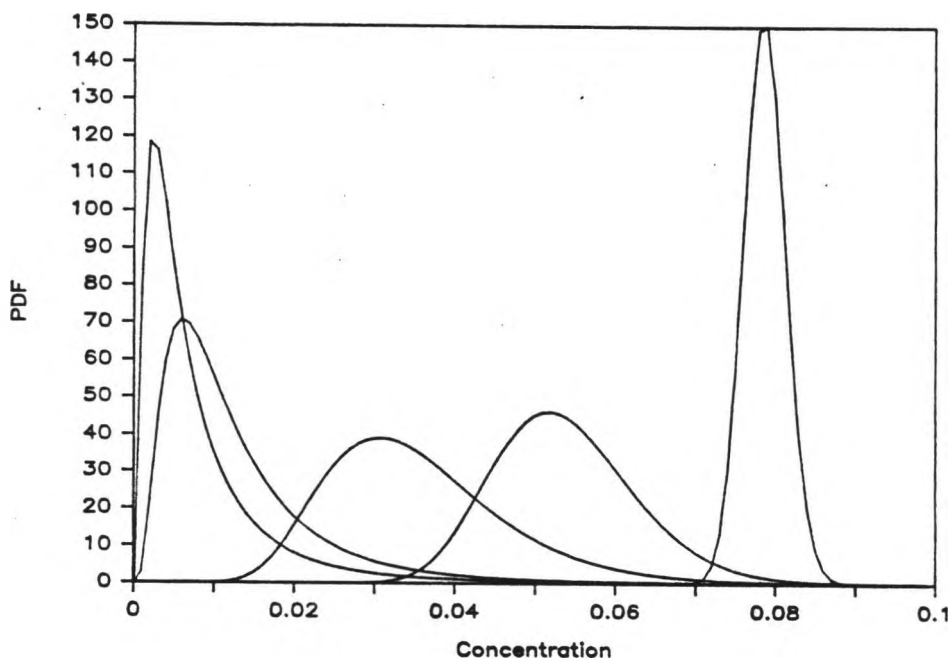

Figure C.2 PDF used in simulation studies with R = 0.1 and FC:FBQ = 1,1.

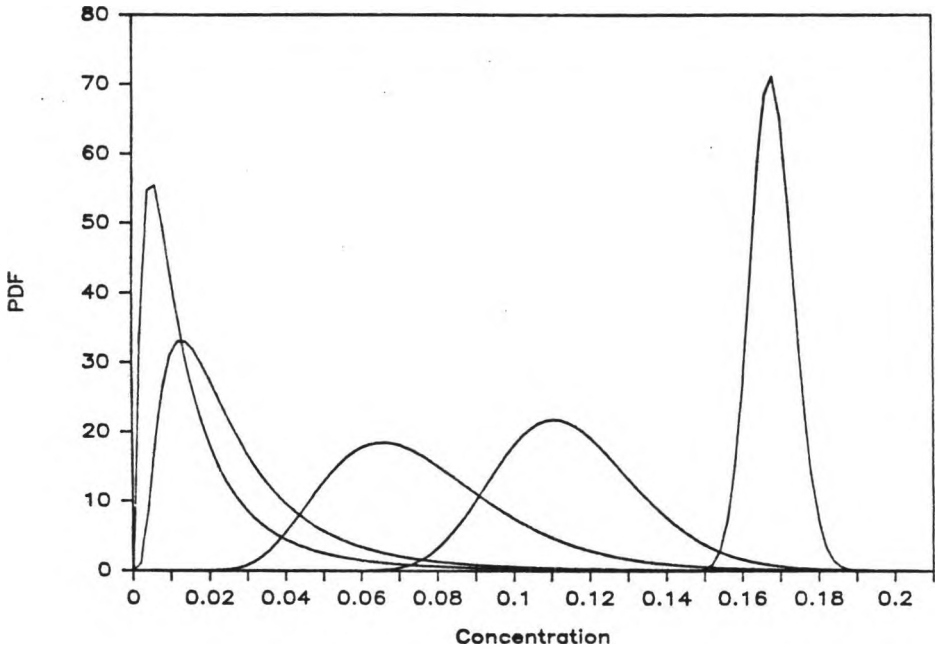

Figure C.3 PDF used in simulation studies with R = 0.01 and FC:FBQ = 1,1.

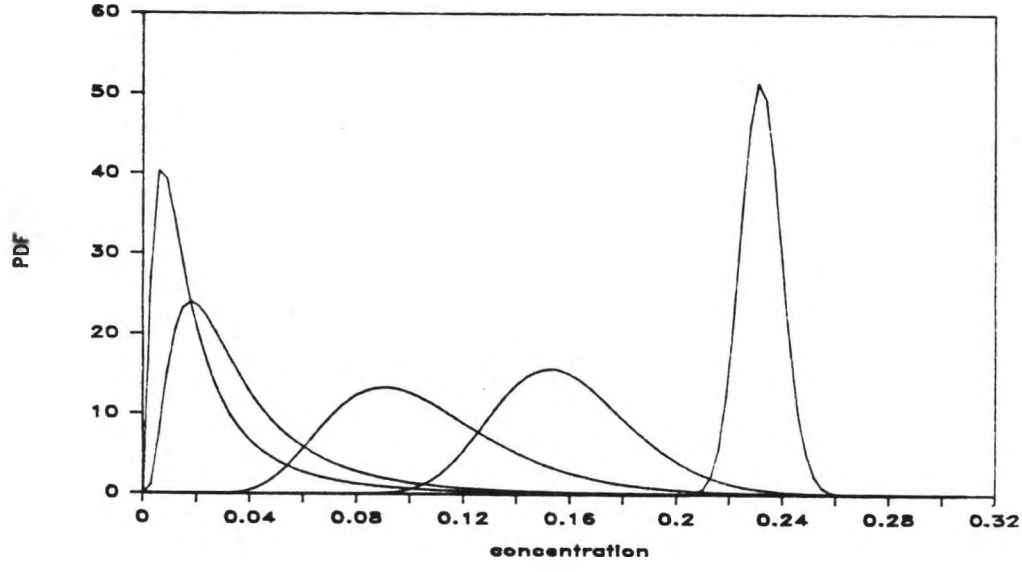

Figure C.4 PDF used in simulation studies with R = 0.001 and FC:FBQ = 1.0,1.0

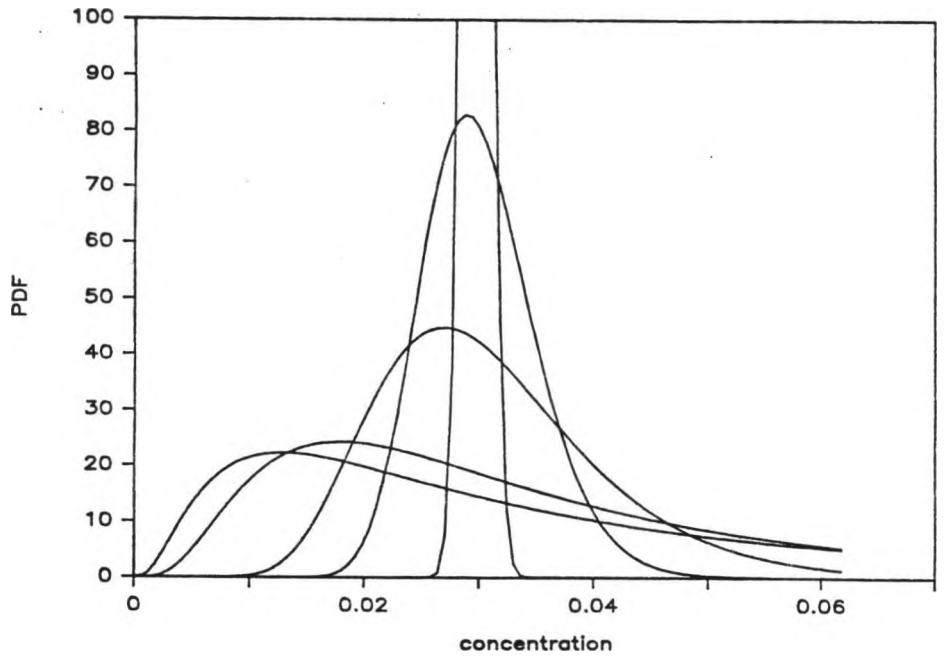

Figure C.5 PDF used in simulation studies with R = 1.0 and FC:FBQ = 0.5,1.

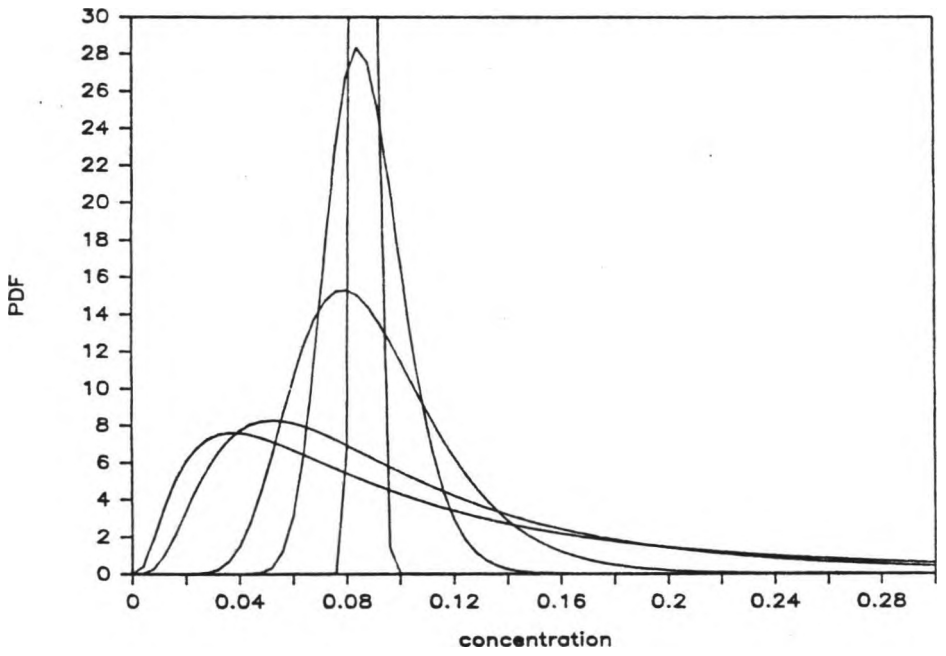

Figure C.6 PDF used in simulation studies with R = 0.1 and FC:FBQ = 0.5,1.

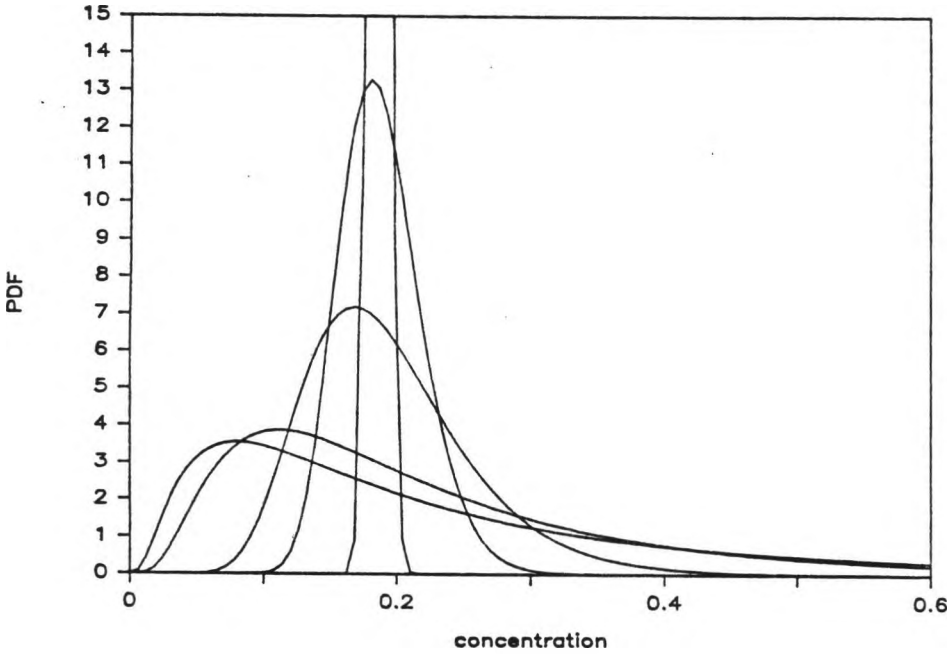

Figure C.7 PDF used in simulation studies with R = 0.01 FC:FBQ = 0.5, 1.0

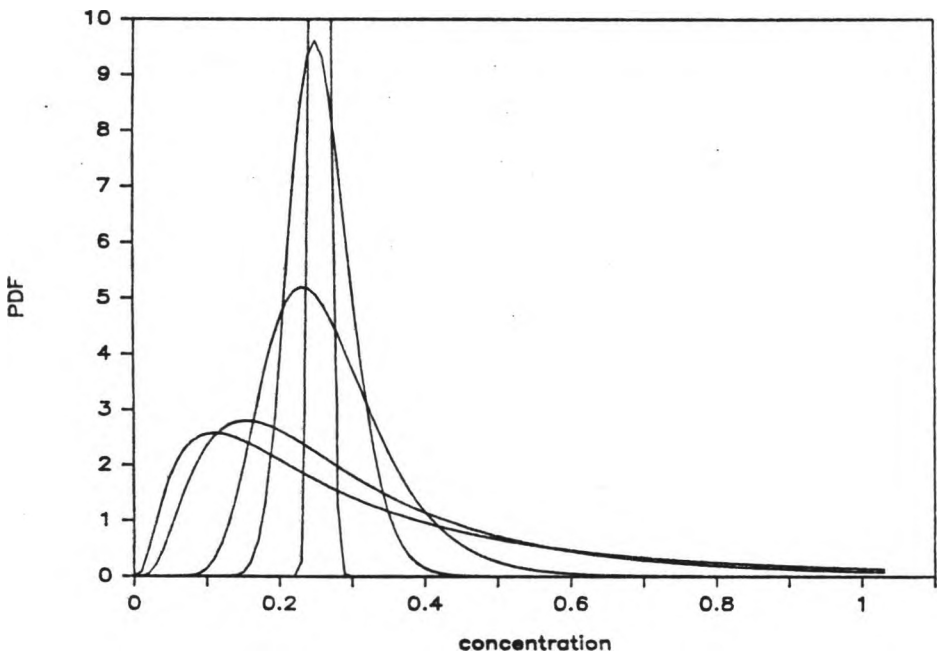

Figure C.8 PDF used in simulation studies with R = 0.001 and FC:FBQ = 0.5,1.

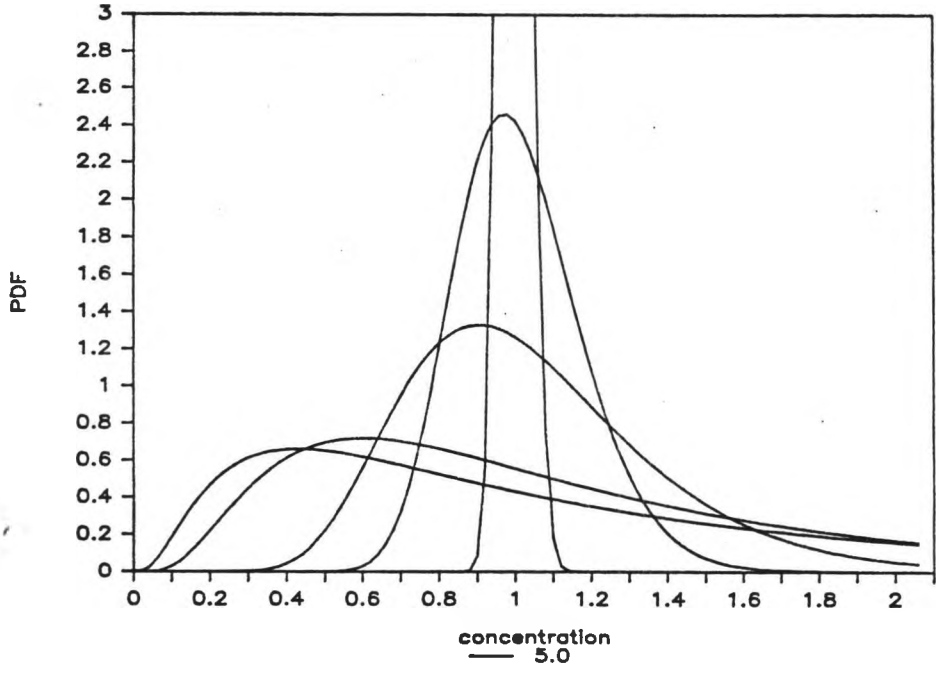

Figure C.9 PDF used in simulation studies with FC:FBQ = 0, 0.5.