DISSERTATION

WATER-QUALITY DATA ANALYSIS PROTOCOL DEVELOPMENT

Submitted by

Jonathan Brooks Harcum

Department of Agricultural and Chemical Engineering

In partial fulfillment of the requirements

for the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Summer, 1990

COLORADO STATE UNIVERSITY

June 20, 1990

WE HEREBY RECOMMEND THAT THE DISSERTATION PREPARED UNDER OUR SUPERVISION BY JONATHAN BROOKS HARCUM ENTITLED WATER-QUALITY DATA ANALYSIS PROTOCOL DEVELOPMENT BE ACCEPTED AS FULFILLING IN PART REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY.

<u>Committee on Graduate Work</u>

Robert M. Hirsch

Jones Sales

Jim C. Loftis

Coadviser

Robert C. Ward

Adviser

Vincent G. Murphy

Department Head

ABSTRACT

WATER-QUALITY DATA ANALYSIS PROTOCOL DEVELOPMENT

Several agencies have developed networks to routinely monitor water quantity and quality in an attempt to assess society's influence on the environment, including the impacts of modern agriculture. Data from these networks are often plagued with attributes that inhibit analysis and interpretation. As more and more emphasis and public pressure is placed upon demonstrating environmental results, it is increasingly necessary that a consistent protocol for analyzing data from water quality monitoring networks be developed.

Common data record attributes which inhibit data analysis include distribution applicability, variance heterogeneity, seasonality, serial correlation, extreme events, censoring, erroneous observations, small sample size, missing values, different sampling frequencies, multiple observations and measurement uncertainty. Each data record attribute is described in this study. In establishing a protocol to analyze water quality data, the handling of censored data and detection of trends in the presence of serial correlation and missing data are particularly difficult to quantify. This study focuses on these issues of protocol development.

Seventeen procedures are evaluated for estimating the mean, median, standard deviation and interquartile range from data sets with singly and multiply censored observations. The results from this evaluation support previous investigations. In addition, the "no censoring" rule was found superior to methods which used censored observations for estimation of the mean, median and standard deviation.

This study also compared the use of the Mann-Kendall tau test (and variations) for evaluating monotonic trends in water quality data. The Seasonal Kendall (Mann-Kendall) tau test should be used for data records with no serial correlation and five or less (ten or more) years of record. An ideal test for short data records which have serial correlation was not found in this study. The Seasonal Kendall tau test with serial correlation correction should be used for data sets of at least ten years of record and serial correlation. Furthermore, if monthly data sets have on the order of 40 to 50 percent missing values, monthly data should be collapsed to quarterly data by computing seasonal means or medians.

Jonathan Brooks Harcum
Agricultural and Chemical
Engineering Department
Colorado State University
Fort Collins, CO 80523
Summer, 1990

ACKNOWLEDGEMENTS

During the preparation of this dissertation, numerous individuals have either directly supported or have given invaluable personal time and recommendations to improve this study. Above all, I would like to thank my advisers, Robert Ward and Jim Loftis for their support, guidance and patience throughout my entire graduate program. My appreciation and gratitude is extended to Dr. Jose Salas and Dr. Robert Hirsch, who served as committee members and reviewed this work. Appreciation is extended to Dr. Peter Brockwell, who served as a committee member before leaving for Australia.

Funding for this study was provided in part by the United States Department of Agriculture National Needs Fellowship Program and a Colorado State Supercomputing Grant. Without this support, this project would not have been possible.

I am most indebted to the Systems Analysis Group of the United States Geological Survey in Reston, Virginia. Much of this study is based upon work initiated by their previous work or developed during my stay in Reston during summer of 1987. Special thanks also goes to Charlie Taylor for all of the statistical advice.

Finally, thanks should go to the Department of Agricultural and Chemical Engineering. In particular, I am grateful to Peter, Phil and Mathias--the coffee breaks helped a lot.

Thank you, Sarah.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Agricultural production has benefited the United States economy, and much effort has been devoted towards improving production. Arid regions generally unsuited for cash crops have been transformed into productive farm land, while improved crop varieties, tillage practices, pest management and weed control have increased agricultural production throughout the United States. Much less effort, however, has been placed on evaluating the environmental effects of modern agriculture.

Several agencies (U. S. Geological Survey, U. S. Environmental Protection Agency, U. S. Fish and Wildlife Service, etc.) have developed networks to routinely monitor water quantity and quality in an attempt to assess society's influence on the environment. In an ideal setting, monitoring networks would provide information identifying areas of environmental concern. It is apparent that information gathering has failed in areas such as the Chesapeake Bay in the Mid-Atlantic region and the Kesterson National Wildlife Refuge in California, where agriculture has been identified as a key nonpoint source polluter.

A. Evolution of Water Quality Monitoring

Considerable effort has been devoted over the years to developing water-quality sampling and laboratory analysis protocols in an attempt to improve knowledge about the behavior of water-quality variables (e.g., Everett, 1980; American Public Health Association, 1985). The vast majority of this effort has been focused on acquiring representative data, while the methods (protocols) for analyzing the data, particularly from a statistical viewpoint, have almost been ignored. Ward (1981) points out the problems that will arise if data analysis is not built into the water-quality monitoring program design and operation. Ward, et al. (1986) described what has happened in surface water-quality monitoring due to a lack of any well defined protocols for analyzing water-quality data--a "data rich but information poor" syndrome.

Many water-quality monitoring programs have reached the point where data are piling up and it is not clear to many monitoring program managers exactly how the data will be converted to information. In attempts to produce the required information, a monitoring network manager will often use statistical tests. Application of many statistical tests is not straightforward for water-quality records, a result of the natural variability of water-quality variables and the inherent difficulty associated with environmental monitoring. For example, to use certain statistical tests, the data are sometimes assumed to follow normal or lognormal probability distributions. Assumption adequacy is not checked in many instances. Depending on the statistical test used and the degree of assumption violation, false conclusions may result.

Problems associated with incomplete data records have been recognized for some time. GAO (1986) is one of the more recent publications citing this problem.

> Data on the performance of treatment plants are fairly abundant after completion of construction-grant upgrades but fairly sparse before that time. This makes before-and-after comparisons difficult. Much the same problem pertains, in reverse, to the evaluation of data from water quality monitoring stations. Consistent and complete sets of water quality monitoring data from before and after the construction of a construction-grant project are rare.

The above quote is in reference to analyzing data for improvements due to upgrades in treatment plants. While not a causal relationship, one may speculate on the difficulties in assessing water quality for nonpoint source pollution with incomplete data records.

B. Objectives

In order to effectively manage our natural resources, it is clearly necessary to have a consistent procedure for analyzing data from monitoring networks. At this point, no complete water-quality data analysis protocol is available. Several data analysis issues have been identified that must be addressed before it is possible to develop a data analysis protocol. These issues become the objectives of the study and are as follows:

1. Identify water-quality conditions and data record attributes common to many water-quality monitoring networks.
2. Discuss the assumptions and logic incorporated in the procedures that have been recommended for water-quality data analysis.
3. Recommend a framework which is appropriate for data analysis.
4. Recommend and evaluate procedures for alleviating the following shortcomings in past procedures:
    a) estimating average conditions in the presence of censoring,
    b) detecting changes in water quality in the presence of serial correlation and
    c) detecting changes in water quality in the presence of missing values.

Although a data analysis protocol is not produced in this study, this work is viewed as a further step towards developing a water-quality data analysis protocol.

C. Scope

There are many issues associated with environmental monitoring and data analysis. This study is necessarily limited in several respects.

1. This study is not intended for intensive monitoring strategies which are used to understand physical processes such as chemical dispersion, adsorption, etc.
2. This study is intended to be applicable to networks where one is interested in understanding water-quality conditions from a management standpoint.
4. The study is limited to protocols for evaluating average and changing conditions.
5. Water-quality indices and applications of biological monitoring will not be discussed.
6. The coupling of statistical data analysis and water-quality hydrologic models is not considered.
7. No new statistical techniques will be developed.

D. Organization

Chapter II is an introduction to water-quality conditions that are commonly encountered in water-quality monitoring networks. Data record attributes which inhibit standard data analysis are discussed. Chapter III contains a review and discussion of present data analysis protocols. A framework for this study is identified and developed in terms of information flow.

The next three chapters address specific water-quality data analysis issues that must be addressed more quantitatively before water-quality data analysis protocols can be developed. Chapter IV contains the details of a current study to determine effective procedures for evaluating average conditions from censored data. Both single and multiple censoring are considered. Chapter V presents details of another concurrent study addressing serial correlation and effective procedures for evaluating changing conditions. Chapter VI presents the results from a simulation study which addresses the computational problems introduced by missing values.

Chapter VII highlights key topics covered in this study and lists recommendations for further work.

CHAPTER II

ATTRIBUTES INHIBITING DATA ANALYSIS

Information goal assessment will typically involve the evaluation of the water-quality average, changing and extreme conditions. Ward and Loftis (1986) have found that many management decisions can be made by analyzing these conditions. This chapter serves as an introduction to water-quality conditions that are commonly encountered as well as data record attributes that are typically found in water-quality data records.

A. Water-Quality Conditions

What is the quality of water? To answer this question, the network manager is typically faced with describing the average water-quality conditions. The mean or median provides information about the central tendency and the standard deviation or interquartile range provides information about the variability (or spread) of a water-quality variable. Average conditions are useful in spatial displays (e.g., plume contours), load estimations and general reporting. Uncertainty should be indicated by reporting estimates with confidence limits or percentiles (Ward and Loftis, 1986).

Network managers are faced with detecting changing water-quality conditions. Has the quality of water gotten better or worse? Generally, we are interested in detecting changes over time or space. Step and monotonic trend tests are applicable for this data analysis aspect.

There have been a number of studies (e.g., U. S. Environmental Protection Agency, 1984a and 1984b; U. S. Environmental Protection Agency and U. S. Fish and Wildlife Service, 1984; U. S. Geological Survey, 1984; Association of State and Interstate Water Pollution Control Administrators, 1984 and Gilliom et al., 1985) which have focused on evaluating the average and changing water-quality conditions. For example, the before-and-after studies (U. S. Environmental Protection Agency, 1984a) evaluated construction grant program performance by comparing observations before and after the upgrade of treatment plants. The National Fisheries Survey (U. S. Environmental Protection Agency and U. S. Fish and Wildlife Service, 1984) examined the ability of the Nation's rivers to support fish. While the information goals are different for each study, the water-quality conditions evaluated were similar.

Extreme conditions are important as well, since most ecological systems have limited assimilative capacity. Management goals oriented towards compliance with regulatory requirements and environmental disaster prevention (e.g., fish kill) will typically involve extreme water-quality condition evaluation.

Much attention has been focused on what procedures are appropriate for analyzing average and changing water-quality conditions. As a result, some consensus regarding appropriate procedures (see section III.A for further discussion) has been reached. Unfortunately, the same is not true for extreme condition evaluation. Extreme conditions are usually connected to standard violations. In many situations, zero exceedences of a standard are acceptable and one or more exceedences are not acceptable. This type of regulation has been the subject of much debate. Many of the arguments against this type of evaluation have been summarized by McBride and Pridmore (1987) and Loftis and Ward (1981). Since water quality can be viewed as a random variable, it is possible that a violation will occur which is the result of natural variation--not the result of society. Designers must also "over-design" facilities so that accidental violation (e.g., due to possible digester breakdown) do not occur.

As an alternative approach, McBride and Pridmore (1987) suggested replacing an absolute standard with an upper percentile, where a specified number of exceedences are allowable over a specified time frame. They cited three primary reasons for supporting percentile standards in New Zealand.

1. Percentiles would aid treatment facility designers and reduce over-design.
2. Treatment plant performance would be more easily

evaluated. More efficient use of past data would result.

3. Percentile standards are more reliably estimated than absolute standards.

McBride and Pridmore (1987) recognized that a percentile standard does not protect against gross exceedences. They recommended that maximum allowable standards should be incorporated with percentile standards to produce a two-tier monitoring program.

Loftis and Ward (1981) used cumulative density functions to estimate the probability that a single grab sample would be in violation. A procedure was presented for calculating confidence limits around the cumulative density function for normal and nonparametric models. Their study describes a water-quality variables's past behavior based on collected data and compared the past behavior to a stream standard for changes in violation frequency.

More recently, debate has centered around biological monitoring applicability. Are fixed monitoring stations and physical properties appropriate monitoring strategies for evaluating extreme conditions? Should we place biological organisms below facilities to integrate water quality over time as well as different physical and chemical water-quality variables? As a result, it is difficult to judge what the most appropriate data analysis procedures are at this time. Due to the lack of consensus on general approaches to measuring extreme conditions, this study will not consider extreme condition evaluation.

B. Data Record Attributes

Water-quality condition evaluation is made more difficult by data record attributes. In order to understand the procedures which have been suggested to handle these attributes, it is first necessary to identify attributes most important in water-quality monitoring. Data record attributes can be divided into two groups: Statistical characteristics and data limitations. Both groups have an impact on the statistical methods which are used to meet information goals as well as method reliability.

Statistical characteristics are data record attributes resulting from the natural variation of water-quality variables. Five commonly found statistical characteristics are distribution applicability, variance homogeneity, seasonality, serial correlation and extreme events. While these attributes do not cause computational problems, they may violate typical statistical assumptions and result in false conclusions.

Data limitations are, for the most part, man-induced data record attributes. Data limitations often result in less reliable observations, less information for a given data set and greater analysis complexity. Common data limitations include: Missing values, different sampling frequencies, multiple observations, measurement uncertainty, censoring, small sample size and erroneous observations.

B.1 Statistical Characteristics

Water-quality data sets come from a large family of underlying probability density functions. Most applicable distributions are characterized by a coefficient of skew greater than or equal to zero. Montgomery et al. (1987) found 66 out of 172 ground water records which passed normality tests (the skewness coefficient was insignificantly different from zero and the null hypothesis of normality was not rejected with a chi-square goodness of fit test). The constituents for which most normal data sets (at least one-third of well records were normally distributed) were found to include: Chloride, total dissolved solids, pH, specific conductivity (microohms/cm), nitrate-nitrogen, sulfate, hardness (calcium carbonate) and temperature (degrees Fahrenheit). Iron, total organic carbon, total ammonia, boron, total Kjeldahl nitrogen, fluoride and sodium were the constituents for which less than one-third of the well records exhibited normality. Most well records were from shallow unconfined aquifers.

Hirsch and Slack (1984) have cited that temperature, pH and dissolved oxygen are the only commonly measured constituents that are typically normal or near normal. Although not stated, Hirsch and Slack's statement is most likely related to surface water-quality data records. Gilliom and Helsel (1986) examined 482 trace element data sets at U.S. Geological Survey river quality monitoring stations. They found skewness coefficients from -0.8 to 5.2 (six percent were negative) and coefficients of variation from 0.15 to 3.2.

Parametric statistical tests assume a particular distribution, usually normality. The degree to which this assumption is true will play a significant role in the power of the test used. Nonparametric statistical tests are less restrictive in that normality is not assumed. Nonparametric tests, however, do require that the samples come from distributions with the same shape (e.g., identically distributed) for the null hypothesis case. Assumption violation can cause significance levels to be incorrect.

Population variance may be different over time or between sampling locations. Constant variance is an often made assumption of commonly used two-sample and monotonic trend tests. Some nonparametric and modifications of parametric tests (Snedecor and Cochran, 1980) do not make this assumption. A new dam on a river is a classic example of changing variances caused by an intervention. The new dam acts as a buffer for suspended solids downstream. Not only does the mean concentration decrease, but the variance of the process will decrease as well. Incorrectly assuming constant variance (in this example) would result in an actual significance level higher than the nominal significance level for a monotonic trend test. That is, the probability of deciding that a trend occurs is higher than the stated significance under the assumption of no trend in the mean.

Seasonality has been defined as "the deterministic variation in ... concentration with time of year" (Montgomery et al., 1987). Seasonal variation in water quality is often associated with processes such as precipitation and stream flow (Harris et al., 1987). It is not uncommon to have seasonal cycles in both the data record mean and variance. Statistical tests are usually less powerful when seasonality is present (Hirsch et al., 1982). Data may be collapsed; however, information is lost when seasonal records are summarized over seasons. For example, yearly central tendency estimates may not be representative of constituent concentrations from seasonally variable streams.

When observations are spaced closely together, there may be information redundancy between observations. This redundancy is termed serial correlation. Most (parametric and non-parametric) statistical tests assume independence between observations. Serial correlation, in effect, reduces the sample size. Trend detection is more difficult when serial correlation is present in data (Hirsch et al., 1982 and Hirsch and Slack, 1984). For this study, serial correlation and autocorrelation are synonymous, referring to the stochastic dependence of successive observations after the removal of deterministic components (trend and seasonality).

Extreme events are observations which are "real" observations of the monitored random variable. Parametric tests tend to weigh these observations too heavily and may lead to biased results. Extreme events, in many cases, are the most important observations, as in the case of extreme condition evaluation, and should be included in data analyses. Unfortunately, it is sometimes difficult distinguish the difference between extreme events and erroneous observations (a data limitation which is discussed in the next section).

Much research has been devoted to detecting statistical characteristics as well as developing techniques for dealing with them. Procedures range from ignoring the problem, transforming the data to alleviate the problem (e.g., Brockwell and Davis, 1987; Salas et al., 1980 and Helsel, 1987), developing methods which explicitly handle or are robust against the statistical characteristic (e.g., Hirsch et al., 1982; Hirsch and Slack, 1984 and Gilliom and Helsel, 1986), subtracting deterministic components (e.g., Harris et al., 1987 and Harris, 1988), or reducing the effective sample size (e.g., Lettenmaier, 1976). The reader is referred to Harris et al. (1987), Harris (1988), Loftis et al. (1986) and Gilbert (1987) for a more complete discussion of detecting statistical characteristics described above.

B.2 Data Limitations

A missing value is a data record for which no measurement was recorded (e.g., no information is available for some time period). Missing values occur in a number of ways such as lost samples and equipment breakage. In addition to random missing values, systematic missing values can occur from lakes freezing over, wells or streams drying up, seasonal snow falls or budget cut-backs. Several problems result from missing values. Statistical tests which require regularly observed sequences are not readily applicable (e.g., standard time series analysis). Techniques which compare up-gradient and down-gradient observations have less comparisons available. Parameter reliability is decreased as a result of a smaller sample size. In general, summaries over time will be less accurate, since all time periods are not equally represented. The comparison of systematic versus random missing values has not been thoroughly evaluated in a water-quality context.

It is not uncommon to find data records with changing sampling frequencies. For example, an original monitoring design may call for monthly sampling. After a period of time, budget constraints force the monitoring scheme to quarterly sampling. Statistical techniques which require equally spaced data are not directly applicable. Summaries over time will be weighted more heavily to the period with higher sampling frequencies.

Similar problems exist when multiple observations are recorded for one time period. This occurs when quality assurance/quality control data are stored in the same record as the original observation. Summaries over time will be weighted more heavily for time periods with more observations. In addition, it is not straightforward how to use the "extra" observations in some statistical tests (van Belle and Hughes, 1984).

Water-quality data sets are actually observations of random variables and as such have observation error. Measurement uncertainty is the result of random analytical error and can vary with the magnitude of the signal and calibration of the analytical device. Trace organic measurements are one instance where the measurement error is relatively large. It is important to view these numbers as probability statements about the actual process, not as absolute numbers. The required information for this type of analyses is not often available and as a result, it is not uncommon to ignore measurement uncertainty altogether (Porter, 1986). Other types of error can exist as a result of poor quality assurance/quality control (Montgomery, 1987).

Some constituent concentrations exist close to the analytical detection limit. Trace metals and pesticides are common examples. When the signal from an analytical device is less than a predetermined detection limit, the data value is typically censored and reported as less-than the detection limit. There are two types of censoring (David, 1981). Type I censoring replaces all values below (above) a certain numerical level as less-than (greater-than). Type II censoring replaces a known percentage of observations with less-than (greater-than). Type I censoring is most common to water-quality variables and Type II censoring can be found in survival analysis.

Censoring may be complicated even further by multiple detection limits. Multiple censoring occurs in primarily three ways. First, different analytical techniques (and hence different detection limits) may be used as technological advances are achieved. Second, different laboratory protocols may suggest or require the use of different analytical techniques (e.g., different tests based on an initial specific conductance level). Finally, data sets from different sources may be combined for analysis.

Reliability decreases with sample size. Small sample size is not a data limitation per se; however, it does preclude the use of more sophisticated statistical techniques and reduces the power of statistical tests. Time series analysis, for example, is suggested when there are at least 50 observations available (Box and Jenkins, 1970 and van Latesteijn and Lambeck, 1986).

Erroneous observations should not be included in analyses. Although quality control/quality assurance programs are fashioned to reduce the frequency of erroneous observations, erroneous observations still occur in data sets. This largely results from the difficulty in detecting the difference between erroneous observations and extreme events. The protocols reviewed at the beginning of the next chapter support the use of techniques which are robust against erroneous observations and extreme events.

CHAPTER III

WATER-QUALITY DATA ANALYSIS PROTOCOLS

Water-quality data analysis has been the subject of much research in recent years. Much of the literature concerning data analysis is necessarily limited in scope. As one would anticipate, a complete protocol which incorporates all three information goals and accounts for all possible data record attributes is not available. There are, however, several protocols with a limited scope. The first section of this chapter is a water-quality data analysis protocol literature review. After reviewing these protocols, the assumptions and associated logic which have been incorporated into the protocols are presented. Finally, protocol framework and development for this study are presented.

A. Protocol Review

The first part of this review is centered around the work done by the U. S. Geological Survey. They have published a series of papers, when grouped together, can form a data analysis protocol base. The second part of this review focuses on two papers which actually package a data analysis protocol. The former is a review and synthesis of nonparametric trend tests coupled with a procedure for test selection and the later is a protocol directed at all three information goals.

A.1 U. S. Geological Survey

The U. S. Geological Survey's data analysis work has historically centered around surface water hydrology. Their emphasis has been towards describing the average and changing conditions of water quality. The U. S. Geological Survey has not been directly involved with evaluating extreme conditions since they are not a regulatory agency. Their work has lead to several key references and an in-house training program for their employees. The approaches identified here have been the result (at least in part) of the Systems Analysis Group located at the National Center in Reston, Virginia.

A.1.1 Average conditions

To describe the average conditions of water-quality variables, Helsel (1987) recommends estimating the central tendency and variability of the collected data. The median and interquartile range are used for estimates of the central tendency and variability, respectively. The interquartile range is defined as the difference between the observations corresponding to the upper and lower quartiles. These results may be displayed graphically with box plots. Box plots also contain information regarding symmetry and extreme values (Chambers et al., 1983). Box plots of constituent concentration may be made as a function of any explanatory variable such as season, year, or location.

A log-maximum likelihood technique is recommended (Gilliom and Helsel, 1986) for estimating the median and interquartile range when the data are singly censored. The mean and standard deviation can be estimated with a log-regression technique. This has been recently updated to account for multiple censoring. A log-regression technique for multiple censoring limits is recommended when the possibility of "severe non-lognormality" cannot be discounted (Helsel and Cohn, 1988). Improved estimation of the median and interquartile range can be made by using an adjusted maximum likelihood technique (Cohn, 1988) when "severe non-lognormality" is not anticipated. A technique for estimating confidence limits for single censoring is given as well (Gilliom and Helsel, 1986).

A.1.2 Changing conditions

Trend tests are primarily used to evaluate changing conditions. Tests are generally oriented towards detecting step or monotonic trends. In some cases, the information goal may infer which test type is appropriate; however, this choice is not always clear. The U. S. Geological Survey lists four rules in their lecture notes to follow:

1. Use a step trend test, if there is a distinct event in which the data may be viewed as "before" and "after".
2. Do not use a step trend test if there is no a priori explanation for a sudden step. The process of visually

selecting the data sets will bias the results.

3. Use a step trend test if there is a large gap in the record. This is probably most appropriate, if we view the earlier and later part of the record as snapshots and there is likely to be small differences within either part of the record. There is nothing wrong; however, with applying a monotonic test. A monotonic test would be more appropriate, if the trend is thought to be a gradual change.
4. For large multi-station studies it is "probably best" to use the same monotonic test for all data records. They note that this procedure is not optimal for all records, but will ease analysis and reporting. This rule is demonstrated by a national water-quality trend study of the Nation's rivers (Smith et al., 1987).

A.1.2.1 Flow adjustment

Variation due to stream flow should be removed in order to detect trends resulting from other causes (Harned et al., 1981 and Hirsch et al., 1982). For example, constituent concentration may be diluted by an increased stream flow during a wet year. If this variation due to stream flow were not removed, a trend test may indicate a declining trend, when in fact, constituent loading to the stream may be the same or even increasing. Flow adjustment has been demonstrated in surface water quality; however, a successful extension to ground water quality has not been found in literature. Other constituents may have other controlling variables such as solar radiation and temperature for biological measurements. It may be reasonable to adjust concentrations for other controlling variables.

Harned et al. (1981) demonstrates two discharge compensation techniques. These techniques have not gained wide acceptance in the U. S. Geological Survey. Hirsch et al. (1982) and Crawford et al. (1983) demonstrate a flow adjustment which has appeared in more recent literature (e.g., Smith et al., 1987). Alley (1988) demonstrates an improved procedure. To summarize, several general linear models (e.g., linear, quadratic, log-linear and log-log) are fitted to concentration-flow data. The most appropriate model is selected based on residual plots and r^2 values.

A flow adjusted concentration (FAC) is computed as the measured concentration minus the predicted concentration from the selected linear model. A trend test is then applied to the FAC data. These procedures are not valid for data which contain censored observations, although a sensitivity study could be made to handle a small number of censored observations.

A.1.2.2 Trend tests

The Seasonal Kendall tau (Hirsch et al., 1982) and the Seasonal Mann-Whitney Rank-Sum tests (e.g., Hirsch and Gilroy, 1985) are recommended for monotonic and step trend testing, respectively. Both tests are seasonal extensions of tests for grouped data. The magnitude of monotonic change may be estimated with the Sen slope estimator (Sen, 1968; Theil, 1950 and Hirsch et al., 1982). The seasonal Hodges-Lehmann estimator (Crawford et al., 1983 and Hirsch, 1988b) can be used for estimating step trend magnitudes.

An extension of the Seasonal Kendall tau test which is robust against moderate levels of correlation has been proposed (Hirsch and Slack, 1984). Serial correlation, in general, is ignored for monthly surface water-quality data (Hirsch, personal communication, 1987). As a result, the correlated-corrected Seasonal Kendall tau test, is not often applied to monthly surface water-quality observations. As an alternative, serially correlated data may be collapsed to a less frequent sampling scheme (e.g., use the median of four weekly observations to create monthly data).

In a recent water-quality trend study, Smith et al. (1987) demonstrated their analysis procedures using data from NASQAN stations. Records subject to laboratory method changes were not considered. Average concentrations were summarized with estimates of the median, lower quartile and upper quartile. All constituents were flow adjusted except for trace element data. The Seasonal Kendall tau test was then applied to the flow adjusted data. A cause and effect analysis compared the statistical association of trends with hydrologic and anthropogenic characteristics.

A.2 Nonparametric Trend Test Selection

Berryman et al. (1988) reviewed a number of nonparametric tests for detecting monotonic, step and multi-step trends. They have identified many of the trade-offs that are associated with competing trend tests. Results were summarized in table form (Table III.1). Recommendations

Table III.1 Proposed nonparametric tests for trend in water quality time series (after Berryman et al., 1988).

Data characteristics				Test	Key Reference	N Minimum*
S	1	N	N	Mann-Whitney	Conover (1980)	5**
S	1	M	N	Mann-Whitney	Lettenmaier (1976)	20
S	1	N	Y	Intrablock test	Berryman (1984)	12
S	1	Y	Y	Intrablock test	Berryman (1984)	?
M	1	N	N	Spearman	Conover (1980)	11
M	1	N	N	Kendall tau	Conover (1980)	9
M	1	M	N	Spearman	Lettenmaier (1976)	20
M	1	N	Y	Intrablock	Hirsch et al. (1982)	24 (2/12)
M	1	N	Y	Aligned tests	Farrell (1980)	20 (5/4)
M	1	Y	Y	Intrablock for persistent data	Hirsch and Slack (1984)	120 (10/12)
M	+	N	Y	Analysis of variance and chi-square test	van Belle and Hughes (1984)	240 (10/12 /2)

Data characteristics key:

trend type (column 1)	S - step trend M - monotonic trend
number of stations (column 2)	1 - one station + - one or more stations
correlation (column 3)	Y - correlated with unknown structure M - correlated with Markovian order 1 N - independent data
seasonality (column 4)	Y - seasonal data N - not seasonal data

* Minimum number of observations to apply method and the bracketed terms indicate the number of years of data, the number of observations per year and the number of stations that has been cited in literature.

**Ten observations are needed to conduct a test with a five percent significance level and five observations are needed for a ten percent significance level.

were based on the presence of seasonality, the presence and type of correlation structure and sample size. A flow chart was presented for selecting the most appropriate monotonic trend test.

In order to select the most appropriate monotonic trend test, the data are "pre-treated". First the data are detrended and then deseasonalized. They detrended an example data set by estimating and subtracting a linear regression line fitted to the original data. The detrended data are deseasonalized by differencing the data a year apart. Correlograms are calculated for data after detrending and after detrending and deseasonalizing the data. At this point, the data are checked for seasonality. It is suggested to plot the detrended data set as well as to compare the two computed correlograms. The data are seasonal if the detrended and deseasonalized correlogram shows less dependence than the correlogram after only detrending. It is necessary to check for seasonal trend homogeneity, if the data are deemed seasonal. According to Berryman et al. (1988), it is not necessary to use a formal test (such as van Belle and Hughes, 1984); however, a visual inspection is necessary.

Data are then checked for serial correlation. If the data are seasonal, the correlogram resulting from the detrended and deseasonalized data is used. If the data are not seasonal, the correlogram resulting from the detrended data is used. It is necessary to decide if the data are correlated, and if the correlation structure coincides with a first order Markov model. A first order Markov model is characterized by a significant lag-one correlation coefficient which decays exponentially to zero with increasing lag. The analyst may now use the information regarding seasonality, correlation and sample size to select the most appropriate monotonic trend detection test according to Table II.1. If the sample size is too small, it is not possible to do a trend test with that data set.

A.3 Ground Water Quality Data Analysis Protocol

Ward et al. (1988) have presented a ground water-quality data analysis protocol which incorporates the three water quality conditions important to many routine water quality monitoring networks. This protocol has been set up so that data analysis may be done with as little as two years of quarterly data. They assume from the beginning that the data record is derived from a monitoring program with an effective quality assurance/quality control program in place. The procedures in the following sections have been coded for use on an IBM-PC and are available from the authors.

A.3.1 Data preparation step

First, the raw data set is adjusted by averaging all laboratory duplicates and field replicates so that there is one observation per time interval. Collapsing data sets with different number of samples will result in different variances. Censored observations are replaced by one-half the detection limit and blanks are inserted at locations where no observations were taken. This procedure can result in trends due to changes in detection limits (e.g., multiple detection limits). No procedure is recommended for removing outliers from the raw data file. This file is referred to as the adjusted raw data file.

The adjusted raw data file is transformed into a data analysis input file which is of the proper form for subsequent statistical analysis. This transformation involves three steps. Values are first averaged until there is just one observation per quarter. Quarterly observations are then assigned sequence numbers for the year and season of the observation. A deseasonalized data record is created by subtracting seasonal means from the quarterly observations.

A.3.2 Preliminary graphical evaluation

After preparing a data analysis input file, a graphical evaluation is recommended to insure gross violations of basic assumptions are not made. Ward et al. (1988) recommend plotting time series plots and annual box plots using the adjusted raw data file. The analyst must then select (portions of) records for further analysis. Ward et al. (1988) point out that record segments should be selected based on information goals and the segments should also have homogeneous variance and trend.

Ward et al. (1988) recommend visual inspection of the time series plot and seasonal box plots and a Kruskal-Wallis test at the 90% confidence level to determine seasonality.

If any one of the three procedures indicate seasonality, the deseasonalized data are used for trend analysis.

A.3.3 Changing conditions

A.3.3.1 Monotonic trend tests

Ward et al. (1988) now consider each information goal separately. If the data are seasonal, apply a Kendall tau test at the 90% confidence level to the deseasonalized data. If the data are not seasonal, the Kendall tau test is applied to the original quarterly data. In addition, the Sen slope estimator is computed and plotted with the original quarterly observations.

A.3.3.2 Step trend tests

Ward et al. (1988) describe a medians comparison which is similar to step trend detection (described in this document). There are two primary decisions made in selecting the appropriate statistical test. Can the data be paired? If not, are the data seasonal? If the data can be paired, a Wilcoxon Signed Rank test is applied at the 90% confidence level (80% confidence level if there are only four data pairs).

If the data cannot be paired, a Mann-Whitney Rank Sum test (90% confidence level) is recommended. The test is applied to the deseasonalized data, if seasonality is present. Quarterly values are then plotted on the same graph with their respective median estimates.

A.3.3.3 Extreme conditions

The proportion of samples above the standard are computed and plotted with confidence limits. If the data are seasonal, the proportion is computed for each season. A proportion is calculated for the entire data set, if the data are not seasonal.

A.3.3.4 Reporting

Finally, Ward et al. (1988) propose a clear format to summarize the statistical results. The table they presented included well identification, data segment descriptions, quarterly medians, trend results, compliance results and medians comparison results.

B. Data Record Attribute Handling

The protocols reviewed above have incorporated techniques that address certain data record attributes. Other data record attributes have been essentially ignored. The purpose here is to review these areas as they pertain to the assumptions and logic incorporated into the above protocols.

B.1 Statistical Characteristics

B.1.1 Applicable distribution

In general, nonparametric approaches are more robust than their parametric counterparts for water-quality data. While there are special cases in which parametric approaches are statistically better, it is difficult in practice to determine when these special cases arise. Most parametric procedures make a normality assumption and many water-quality data variables are not normally distributed. From a practical standpoint, it is not straightforward to check for normality. Standard tests (e.g., chi-square goodness of fit test) are not robust against non-normality for small data sets. Stated in other words, large departures from normality are necessary to reject the normality hypothesis for small data sets. In addition, nonparametric tests perform well (high relative efficiency) even under normality conditions.

It should be remembered that there are assumptions associated with nonparametric approaches and violation of these assumptions can cause misleading results as well (Harris, 1988). All three protocols reviewed above have abandoned parametric tests.

B.1.2 Variance homogeneity

This issue is only addressed by the ground water protocol (Ward et al., 1988). While nonparametric tests do not assume normality, an assumption of identical distribution is often made. The identical distribution assumption implies that the sampled population is not changing over time or space. Variance heterogeneity is common in water quality. For example, increased stream flow from snow melt will typically increase the concentration of sediment carried constituents. The increased concentration will likely be accompanied by an increase in the variability of concentration as well. Ward et al. (1988) recommend visually examining time series and box and whisker plots and a

Kruskal-Wallis test to check for gross assumption violation and selecting homogeneous record parts for further analysis.

B.1.3 Seasonality

The U. S. Geological Survey and Ward et al. (1988) recommend different procedures for handling seasonality. The ground water protocol recommends deseasonalizing the data and applying an aligned test since typical statistical packages do not have seasonal (intra block) tests available. The U. S. Geological Survey recommends using seasonal (intra block) tests regardless of seasonality presence. Van Belle and Hughes (1984) demonstrate with asymptotic relative efficiencies that aligned monotonic tests are more powerful than intra block monotonic tests. This difference is more dramatic with shorter data sets (van Belle and Hughes, 1984).[1]

Recent work (Taylor and Loftis, 1989) indicates that a Mann-Kendall tau test (aligned) applied to short seasonal data records (after the data are deseasonalized) suffer from inflated significance levels. Stated in other words, if the Mann-Kendall tau test is carried out at the five percent nominal significance level, more than five in 100 data sets with no trend will be detected as having trend (e.g., false positives). If the nominal significance level is adjusted so that the Mann-Kendall tau test produces five percent

---

[1] Asymptotic relative efficiencies are only valid for large size. As a result, comparison of shorter data sets is not valid.

false positives, the power of the test is similar to that of a Seasonal Kendall tau test for trend (Taylor and Loftis, 1989). The process of deseasonalizing the data causes uncorrelated time series to have a negative serial correlation at lags of integer multiples of the number of seasons (Hirsch et al., 1982). As the sample size increases, the level of correlation is reduced (Hirsch et al., 1984). This work tends to support the use of intra block tests for short data records.

B.1.4 Serial correlation

Serial correlation clearly effects the evaluation of all three information goals. Most concern has been directed at the effects of serial correlation on trend detection. To properly analyze data for trends, it is necessary to discern the difference between trend and serial correlation. Two methods for dealing with serial correlation have been recommended in the protocols reviewed.

First, a test (Hirsch and Slack, 1984 and Lettenmaier, 1976) which explicitly handles serial correlation may be used. Second, observations may be collapsed by taking the mean (Ward et al., 1988) or median (Gilbert, 1987) of data groups until the serial correlation has been effectively eliminated. Quarterly or less frequent ground water quality observations and monthly or less frequent stream water quality observations are typically assumed to be independent. This finding should be used as a rule of thumb not as a fact. There are examples in which this rule of thumb is not true. Montgomery et al. (1987), in a survey of the quarterly ground water observations, found 20% of the data records were serially correlated. Close (1987) found monthly correlation as high as 0.87 for some ground water-quality variables in New Zealand. Serial correlation in lakes would depend on size and retention time.

While there is an appeal that the test recommended by Hirsch and Slack (1984) uses all the data available, serial correlation estimation is crucial to test performance. Applying a test which is corrected for serial correlation to independent data results in a less powerful test than an equivalent test which is not corrected for correlation. Additionally, Hirsch and Slack (1984) and Taylor and Loftis (1989) indicate that the significance level of the corrected Seasonal Kendall tau test is not preserved for short data records.

There have clearly been trade-offs made in the protocols reviewed above. Ward et al. (1988) have assumed that serial correlation is effectively removed by collapsing ground water quality observations to quarterly values by taking the mean of more frequently sampled variables. While not a fast rule, Hirsch (personal communication, 1987) has stated that most analysis by the U. S. Geological Survey of monthly stream water quality data is based on the independence assumption.

There are two controversial areas. First, the independence assumption is not always justified and it is not clear which of the two techniques described above most effectively handle this problem. Second, is it better to collapse the data by taking the mean or the median or should we subsample from the original data? These issues are addressed more completely in Chapter V.

B.2 Data Limitations

B.2.1 Irregular sampling

Missing values, multiple observations, and different sampling frequencies are not specifically addressed in the protocols put forth by the U. S. Geological Survey or Berryman et al. (1988). Ward et al. (1988) recommend collapsing all ground water observations taken more frequently than quarterly by taking the mean. Multiple observations (field and laboratory duplicates) are first averaged, then values within a given quarter are averaged to give one value per quarter. Their procedure eliminates associated problems with multiple observations and different sampling frequencies. The number of missing values has been reduced but not necessarily eliminated.

Statistically, averaging observations will reduce the process variance which is a good idea. The number of observations included in an average value will not, in general, be the same, resulting in different variances for different values. Many statistical techniques now employed do not account for changing variance. The reasons supporting this procedure are more practical in nature. Record segments, that have been sampled more frequently, will receive more weight and vice versa (e.g., sampling bias). Ward et al. (1988) elected to ignore the variance homogeneity problem and placed more emphasis on equal representation from all record segments.

B.2.2 Measurement uncertainty

Measurement error is ignored by all the protocols that have been reviewed. Clearly, this is not completely correct. The degree to which measurement error can be ignored is not well defined in the literature. Water-quality measurements are probability statements about a random variable and should ideally be reported as an estimate with some statement of error or confidence limits (Porter, 1986). However, the methodology for effectively handling water-quality measurement error is not well established, and information to properly handle measurement error is rarely available from typical monitoring data sets.

B.2.3 Censoring

With increased attention given to trace organics such as pesticides, more and more data records contain censored observations. Single censoring does not provide computational problems for many nonparametric trend tests. Comparisons between observations reported as less-than some detection limit are simply treated as ties. Replacement of censored observations with one-half the detection limit will result in the same conclusions from nonparametric trend tests which explicitly account for one censoring limit.

The extension to multiple detection limits has not been thoroughly researched for multiple censoring. Hirsch et al. (1982) recommended treating all values less than the highest detection limit as tied observations. The multiple detection limit scenario was not explicitly explored by Ward et al. (1988). By default, Ward et al. (1988) would recode all values as one-half of the corresponding detection limit. Both alternatives are clearly less than optimal. The latter technique (Ward et al., 1988) would result in artificial trends for multiple detection limits. Hughes and Millard (1988) and Millard and Deverel (1988) have made the most recent advances concerning trend detection with multiple censoring for monotonic and step trend detection, respectively. They indicated, however, practical limitations to their work.

Average condition estimation in the presence of censoring has received recent attention by the U. S. Geological Survey (Gilliom and Helsel, 1986; Helsel and Gilliom, 1986 and Helsel and Cohn, 1988) and concurrent research contained herein (see Chapter IV).

B.2.4 Sample size

Berryman et al. (1988) recognize the frequent existence or the importance of data sets that are too short for analysis in their flow chart procedure. Economically, the collection of ten years of monthly data may not be feasible. On the other hand, the application of trend tests to as little as two years of quarterly data (Ward et al., 1988) may be misleading. These viewpoints lead to a classic manager's decision: How much data is enough? The answer, unfortunately, cannot really be decided a priori of data collection. Water-quality variables which are not seasonal, not correlated, etc., will require less data to detect changes of a given magnitude than a water-quality variable which possess these statistical characteristics.

B.2.5 Outliers

Water-quality variables are highly variable. As a result, outliers that are erroneous values are difficult, if not impossible, to detect. Nonparametric tests are usually based on the ranks of data. As such, anomalous values should not dramatically effect the results of nonparametric tests. Conversely, nonparametric tests may be less sensitive to large changes in a short period of time.

Procedures for explicitly removing outliers are not recommended by Ward et al. (1988) but an active quality assurance/quality control program is recommended. A data screening program has been used by the U. S. Geological Survey (Peters, 1984). The U. S. Geological Survey's screening program is based on chemical balances such as cation/anion balances and electroconductivity/pH comparisons. Nonparametric procedures are preferred by all protocols reviewed.

C. Data Analysis Approaches

In addition to the techniques discussed above for handling statistical characteristics and data limitations, general data analysis approaches (or more appropriately philosophies) have evolved for water-quality variables. Trade-offs regarding flow adjustment, data characterization, and multivariate analysis are discussed below.

C.1 Flow Adjusted Concentration

Water-quality hydrology is often driven by exogenous factors such as precipitation, stream flow, temperature, or solar radiation. Since the goal of many monitoring networks is to assess man's influence on the environment, it may be desirable to remove the effects of these factors. This is primarily true when estimating water quality changes.

One of the key components of stream water quality is stream flow. Increased stream flow may act as either a diluting or a flushing agent. If there is a constant mass loading to the stream, an increased flow will cause a decrease in constituent concentration. On the other hand, the concentration of a sediment transported constituent (e.g., phosphorous) may increase from an increase flow due to a large runoff event which resulted from a high precipitation event. In either case, the variation in stream flow has hidden man's impact on the environment. The purpose of using flow adjusted concentrations is to remove variance from a signal, resulting in an increased ability to detect trend.

If the relationship between stream flow and concentration has not changed, it is appropriate to adjust for stream flow since long term changes in stream flow can mask important changes in random variable concentrations. The extension of flow adjustment to ground water or lakes has not appeared in literature; however, it may be reasonable to relate concentration to water level or change in water level with time in either lakes or wells.

C.2 Data Characterization

Appropriate statistical procedures should be selected before any data are collected. In addition to this "Catch 22" situation, Bell and DeLong (1988) note that a protocol (or test selection procedure) based on data characteristics can result in different procedures for different data segments or different data records in space. If different characteristics require different procedures, results over time or space would be difficult to assess. As a result, many researchers have abandoned procedures which require distribution assumptions or in some cases seasonality assessment.

Hirsch (1988a) does recommend using a priori information to anticipate statistical characteristics and data limitations. In particular, Hirsch (1988a) suggests that the investigator acquires knowledge concerning: Sampling frequency, distributional ranges, relationships with other variables, seasonality and censoring possibility.

C.3 Multi-Stations Versus Single Station

The public and/or their elected representatives typically want to know if the water quality is getting worse in their area. The answer to this question is unfortunately not a straightforward yes or no. Combining information from multiple sites is not a trivial problem. Collapsing information from multiple sites may result in lost information. Collapsing two trends, one increasing and one decreasing, may result in an overall no trend. On the other hand, evaluation of many trend tests can be rather burdensome.

The protocols reviewed above tend to support analyzing one station and one variable at a time (except for spatial step trend tests). There are exceptions to this philosophy such as van Belle and Hughes (1984). Since it is necessary to manage our natural resources on a regional basis, some procedure is necessary to delineate area wide information. The U. S. Geological Survey, for example, plots triangles indicating direction of trend and small circles indicating no trend on a map of the study area. Different maps are used for different water-quality variable. Trend maps are then related to explanatory variables in an attempt to relate trends with regional anthropogenic patterns (Smith et al., 1987).

D. Protocol Framework and Development

The previous sections of this chapter have identified and discussed techniques for evaluating average and changing water-quality conditions. Many procedures have become standard such as the Mann-Kendall tau test, or variations, for detecting monotonic trends. Unfortunately, data sets that are typically encountered do not always conform to "standard" techniques. As a result, it is necessary to adjust data sets or techniques for data record attributes. It is clear that a standard data analysis protocol is needed to deal with commonly found attributes. Similarly, it is necessary to use a framework in which a defensible protocol can be assembled. Key points of a data analysis protocol are discussed in terms of information flow for a total monitoring program.

D.1 Information Flow

The information flow of a monitoring program (Sanders et al., 1983 and Ward and McBride, 1986) is adopted as a basis for this study (Figure III.1). The flow chart generality is convenient since the same operational activities are used for any information goal. Many of the operational activities will only be discussed briefly as they are beyond the scope of this study. Yet, it is valuable to understand where a data analysis protocol fits into the total monitoring program.

Information goals provide the reasons for monitoring. Although network design is not a routine operation (Sanders et al. 1983), many of the decisions made at the network design stage have ramifications later. It is at the design phase where information goals are related to statistical

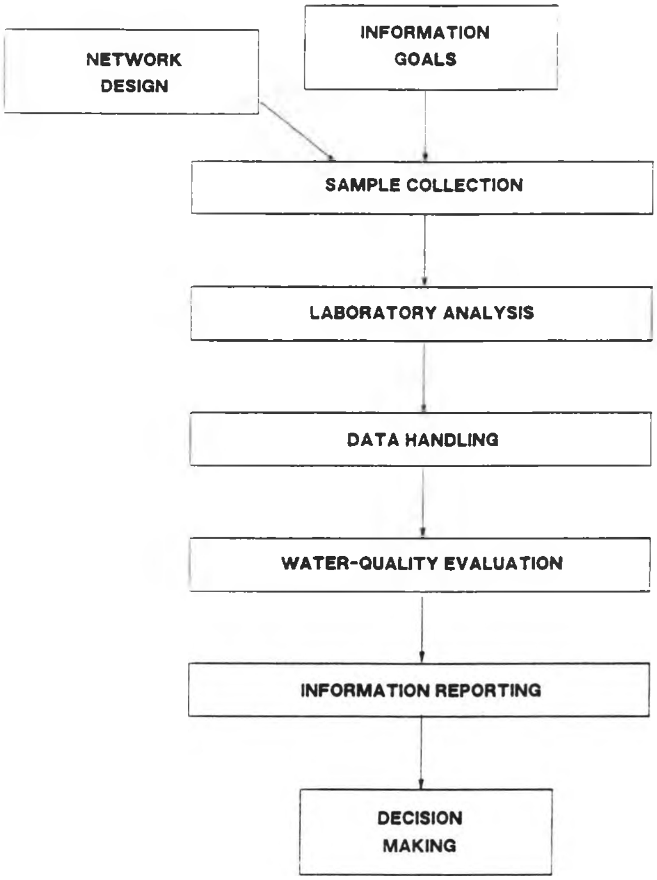

Figure III.1 Total monitoring program (adopted from Sanders et al., 1983).

tests and data collection strategies are developed. Data collection decisions include where to sample, what water-quality variables to sample, how often to sample, what analytical techniques should be used, etc. Some of these decisions have direct impacts on data analysis. For example, if the sampling frequency is high, serial correlation will be present which violates the independence assumption often made by trend tests. It also at the network design phase where data analysis procedures should be selected.

The next activities are sample collection and laboratory analysis. For this study, standard sample collection and laboratory analysis protocols are assumed. Proper quality assurance and quality control programs are assumed to be in place as well.

After results are received from the laboratory (data handling step), it is necessary to screen the results for erroneous values. Gilbert (1987, p186) recommends screening data for identification code errors and checking the data for consistency. He has suggested the use of control charts in some instances when checking the data for historical consistency. Peters (1984) describes a charge balance for major ions (Na^+, K^+, Mg^{2+}, Ca^{2+}, NH_4^+, H^+, Cl^-, SO_4^{2-}, and NO_3^-) and a comparison of electroconductivity and pH values as a check for consistency. If erroneous values are detected, duplicate samples should be analyzed or the station resampled. Well coordinated data storage and retrieval are vital to an effective monitoring system.

Water-quality evaluation is strongly linked to the previous activities. Poor performance at any of the previous activities may lead to poor results here. Information goal evaluation is performed by summary statistic estimation and hypothesis testing at this stage. Information reporting essentially transforms statistical results from the water-quality evaluation step into usable information which is acted upon at the management decision step. Reporting format has received little attention in the past; however, the issue of information transfer is gaining momentum.

D.2 Components of Data Analysis

The data analysis protocols reviewed earlier in this chapter suggest that there are five components to water-quality data analysis. These components are: Identification of information goals and transformation into water-quality conditions, data handling, identification of data record attributes, water-quality evaluation and information reporting. These components are not independent sequential steps. Rather each component is highly dependent on each other as well as the balance of the total monitoring program.

Data record attribute identification is not listed as a specific step in the total monitoring program (Figure III.1); however, data record attributes have an impact on which statistical test is appropriate for analysis. While some attributes may be anticipated before data collection as suggested by Hirsch (1988a), this is not always the case and preliminary data examination may be required.

D.3 Development

To develop a reasonable and defensible protocol, it is necessary to use accepted practices and consistent logic. Many procedures that have been developed have associated trade-offs. In general, the trade-offs are associated with robustness and practicality. Ideally, the data analyst should use the most robust technique available. This may result in the use of multiple procedures or tests. Increased analysis complexity is added if a single procedure must be selected among a suite of procedures. As an alternative, the analyst may use only one procedure which is less than optimal in certain instances but never performs poorly in all anticipated scenarios.

Water-quality data analysis protocols can be developed in several ways. Procedures that have gained wide acceptance in literature can be synthesized. There are key points agreed on by the protocols reviewed earlier and are adopted for this study.

1. Average conditions are often described with median and interquartile ranges (graphically displayed as box plots).
2. The Mann-Kendall tau test (or variations) are used for monotonic trend detection. The Sen slope estimator is used for monotonic trend magnitude estimation.
3. The Wilcoxon Signed Rank is used for step trend detection for paired data and the Wilcoxon Rank Sum (or variations) test is used for data which cannot be paired.
4. Nonparametric statistics are recommended to avoid problems of non-normality, potential outliers and the associated problems involved with data characterization.
5. Measurement error is ignored.
6. Analyze one water-quality variable at a time.

Protocols may also be developed via further simulation studies. The simulation studies that follow in the next three chapters are an attempt to unravel some of the controversy surrounding preliminary data handling. First, average condition estimation with censored data is addressed. This is a particular problem since more and more monitoring efforts will include constituents such as pesticides, trace metals and organics which are commonly found near or below the analytical detection limit. This issue is complicated with multiple detection limits. Several procedures have been recommended in previous protocols for handling serial correlation and subsequent trend detection. The second group of simulation experiments is a comparison of these recommended procedures. The third group of experiments addresses the issues of trend detection in the presence of missing values.

It is clear that not all possible data record attributes and subsequent water-quality conditions scenarios can be evaluated herein. The final technique for developing protocols is the extrapolation of logic and recommendations to those scenarios not specifically covered elsewhere. While not completely heuristic, this method is less than desirable. This method is not used in this study. It should be realized that the protocols developed here are the beginning and should evolve along with our knowledge of water quality, hydrology and statistics.

CHAPTER IV

AVERAGE CONDITIONS--ESTIMATION WITH CENSORING

It is often necessary to describe the average water-quality conditions at a particular monitoring station. Basic statistical summaries of the central tendency and variability (of a particular water-quality variable) form a basis for average water-quality condition description. Central tendency and variability estimates may be compared with regulatory standards or summary statistics from other nearby stations to evaluate, for example, the effect of remedial actions at hazardous waste facilities or best management practices used to control nonpoint source pollution from agriculture. Summary statistics may also be used for general reporting as required by law.

Central tendency and variability are most often described with estimates of the mean or median and the standard deviation or interquartile range, respectively. When the underlying distribution from which the observations are drawn is symmetric, the mean and median produce similar results. Water-quality variables are typically skewed to the right, and as a result, the mean is not a good estimate of the central tendency (Sanders et al., 1983 and Helsel, 1987). The previously reviewed protocols tend to support the use of the median and interquartile range. There are instances for which estimates of the mean and standard deviation are used. Common applications include deseasonalizing data, old reporting formats or control chart construction.

Basic summary statistic estimation is complicated with both single and multiple censoring. The following simulation is a comparison of different techniques which have been used to alleviate the censoring dilemma. Simulations for single and multiple censoring were performed separately. However, the simulations are presented together since the procedures and analyses are similar.

A. Simulation Methodology

Data sets with known population summary statistics (mean, standard deviation, median and interquartile range) are generated from a variety of parent distributions using procedures similar to those of Gilliom and Helsel (1986). Several modifications to their procedures are noted in Appendix A, along with a more detailed discussion of the data generation procedure. Artificial levels of censoring are then incorporated into the data sets. At this point, a variety of estimation techniques are applied to each data set. Performance is judged by evaluating the root-mean-squared-error (rmse) and bias of each estimation technique. The rmse and bias for the mean are calculated as

$$\text{rmse} = \left[N^{-1} \sum_{i=1}^{N} \left[\frac{\bar{x}_i - \mu}{\mu} \right]^2 \right]^{0.5} \qquad \text{IV.1}$$

and

$$\text{bias} = N^{-1} \sum_{i=1}^{N} \left[\frac{\bar{x}_i - \mu}{\mu} \right] \qquad \text{IV.2}$$

where $\bar{x}_i$, μ and N are the estimate of the mean for the ith data set, the true mean of the underlying distribution and the number of data sets simulated, respectively. Rmse and bias terms may be calculated for other summary statistics by substituting the estimated and true statistic for $\bar{x}_i$ and μ, respectively.

Next simulation results are summarized for the purpose of eliminating techniques which are not robust. The objective here is to determine which estimation techniques should not be used for estimating summary statistics from water-quality data records. Since the selection of the simulation parameters may be biased, it is necessary to evaluate technique performance over a wide range of scenarios--not just the scenarios selected for simulation.

To help determine which techniques are not robust, a general linear model is fitted to the rmse for each summary statistic. The independent variables used in the model are coefficient of variation, censoring, sample size, technique and distribution (e.g., all simulation parameters). Rmse will be summed for each technique and summary statistic using the raw simulation data and the general linear model. The rmse for the general linear model will be summed over all valid regions of the model (but not outside simulated regions), in essence a numerical integration of rmse and perfect estimation (e.g., rmse equal to zero). Techniques with large total rmse will not be considered further.

Techniques are then eliminated to reduce redundancy of suitable techniques. (e.g., If technique A is suitable for a portion of conditions and technique B is suitable for a larger range of conditions, including those for which technique A is suitable, it is reasonable to eliminate technique A.) A technique performing within an allowable rmse tolerance of the best technique for a given set of conditions will be considered a suitable technique for that set of conditions as well. If multiple techniques still exist, criteria (possibly heuristic in nature) for selecting one technique over the other will be established.

B. Analyses and Results

Five hundred data sets were generated for each set of simulation conditions. The parameters considered in this simulation were applicable distribution, coefficient of variation, sample size and censoring (Table IV.1 and IV.2). Each combination of parameters in Table IV.1 and IV.2 were considered in this simulation.

Table IV.1 Levels of distribution, coefficient of variation, and sample size considered in average condition estimation with censoring simulation.

Distribution	Coefficient of variation	Sample size
Normal	0.25	8*
Lognormal	0.50	12
Gamma	1.00	24
Mixed Lognormal	2.00	36
Delta		48

* A sample size of eight was not used for multiple censoring.

Table IV.2 Levels of censoring considered in average condition estimation with censoring simulation.

Single censoring scheme	Multiple censoring scheme					
all data censored at	one-third of data set is censored at each of the following levels					
0.2	0.15	0.20	0.25	0.40	0.50	0.60
0.4	0.35	0.40	0.45	0.20	0.50	0.80
0.6	0.55	0.60	0.65	0.20	0.85	0.80
0.8	0.75	0.80	0.85	0.80	0.60	0.70
	0.80	0.15	0.20	0.30	0.40	0.20

Numerous techniques for estimating the summary statistics were found in the literature. The techniques used in this simulation are summarized in Tables IV.3 and IV.4. In general the same technique was used to estimate all summary statistics with the exception of Techniques B and D. Based on analyses described later, it was found that Techniques B and D did not work well for estimating the mean of singly or multiply censored data. It was opted to estimate the

Table IV.3 Techniques chosen for estimating average conditions for single censoring.

Code	Ref	Description of technique
A	2	Do not censor observations
B*	1	Replace censored observations with zero
C	2,4	Replace censored observations with one-half of the detection limit
D*	1	Replace censored observations with the detection limit
E	1	Replace censored observations with numbers uniformly distributed from zero to the detection limit with x_i = dl (i-1)/(NC-1) where x_i, dl, and NC are the ith censored observation, the detection limit, and the number of censored observations; if NC is 1, let x_1 equal one-half of the detection limit
F	2	Replace censored observations with uniform random numbers from the interval [0.0, dl]
G	1	Fit a normal regression curve to the uncensored observations, extrapolate to the censored observations, and adjust all negative values to zero (for plotting position, see Gilliom and Helsel, 1986)
H	1	Fit a normal regression curve to the natural logs of the uncensored observations, extrapolate to the censored observations, and then back transform (for plotting position, see Gilliom and Helsel, 1986)
I	1,5	Normal maximum likelihood
J	1,3,5	Lognormal maximum likelihood
K	1,3	Delta distribution assumption
L	2,6	Linear estimator for normal distribution
M	2,6	Linear estimator for lognormal distribution
N		Replace censored observations with numbers uniformly distributed from zero to the detection limit with x_i = dl (i)/(NC+1)
O	7	Trimmed mean (the percentage of values trimmed on both sides of the ordered observations was equal to the percentage censored)
P	7	Winsorized mean and standard deviation
Q		Fit a normal regression curve to the natural logs of all observations using a maximum-likelihood technique (Wolynetz, 1979b) which incorporates censored observations, extrapolate to the censored observations, and then back transform

* The standard deviation for Techniques B and D were estimated using a mean estimate from Technique C.

Table IV.4 Techniques chosen for estimating average conditions for multiple censoring.

Code	Ref	Description of technique
A	2	Do not censor observations
B*	8	Replace all censored observations with zero
C	8	Replace censored observations with one-half their respective detection limit
D*	8	Replace censored observations with their respective detection limit
E		Replace censored observations with numbers uniformly distributed from zero to the detection limit with x_i = dl (i-1)/(NC-1) where x_i, dl, and NC are the ith censored observation, the detection limit, and the number of censored observations for each detection limit
F	8	Replace censored observations with numbers uniformly distributed from zero to the detection limit with x_i = dl (i)/(NC+1) for each detection limit
G	8	Fit a normal regression curve to the natural logs of the uncensored observations, extrapolate to the censored observations, and then back transform (for plotting position, see Helsel and Cohn, 1988)
H		Fit a normal regression curve to the natural logs of all observations using a maximum-likelihood technique (Wolynetz, 1979b) which incorporates censored observations, extrapolate to the censored observations, and then back transform
I	8,5	Normal maximum likelihood
J	8,5,3	Lognormal maximum likelihood

* The standard deviation for Techniques B and D were estimated using a mean estimate from Technique C.

Reference Key (applicable for Tables IV.3 and IV.4):

1 Gilliom and Helsel (1986)
2 Porter (1986)
3 Aitchison and Brown (1957)
4 Ward et al. (1988)
5 Wolynetz (1979b)
6 Persson and Rootzen (1977)
7 Gilbert (1987)
8 Helsel and Cohn (1988)

standard deviation (for Techniques B and D) using the estimated mean from Technique C (one-half censoring limit substitution).

Rmse and bias were computed for each simulation. A typical output is shown in Table IV.5. (The output file was edited for presentation purposes.)

B.1 Nonrobust Technique Elimination

Before fitting a linear model to the rmse simulation data, the simulation data were scanned to eliminate techniques which have extremely high rmses. By eliminating techniques which performed poorly, the linear model should have a better fit. Rmse was averaged for a given simulation parameter such as different censoring levels. The techniques were then ordered from lowest to highest average rmse. This process was repeated for different levels of sample size and coefficient of variation. These tables were visually reviewed to detect techniques which did not perform well.

As noted by Gilliom and Helsel (1986), technique performance did not change drastically relative to other techniques. In other words, a technique which performed well for a certain set of simulation conditions would perform well for many simulation conditions as compared to the performance of other techniques. The same pattern (as reported by Gilliom and Helsel (1986)) is noted in this study with a few exceptions. These exceptions are summarized below.

Table IV.5 Selected simulation results that are typical of the generated output.

INITIAL SIMULATION PARAMETERS

DISTRIBUTION ID............	3
GAMMA DISTRIBUTION	
DISTRIBUTION MEAN..........	1.00000
PERCENTILE CENSORED........	0.20000
CV.........................	1.00000
LIMIT OF DETECTION.........	0.22314
SAMPLE SIZE................	8

COMPUTATION OF BIAS AND RMSE

MEAN TECH	BIAS	RMSE	STDV TECH	BIAS	RMSE
A	-0.00563	0.34693	A	-0.09587	0.41639
B	-0.02697	0.35484	B	-0.07318	0.40858
C	-0.00426	0.34691	C	-0.09772	0.41639
D	0.01844	0.34101	D	-0.12066	0.42506
E	-0.00426	0.34691	E	-0.09639	0.41546
F	-0.00448	0.34674	F	-0.09694	0.41641
G	-0.01732	0.35109	G	-0.08514	0.41948
H	0.02175	0.34967	H	-0.12718	0.43233
I	-0.12777	0.39886	I	0.06952	0.51498
J	0.23229	0.58673	J	1.55648	4.06718
K	0.01823	0.38811	K	0.12670	0.65181
L	-0.21554	0.39937	L	0.13283	0.56508
M	0.12813	0.47197	M	1.08786	2.22979
N	-0.00426	0.34691	N	-0.09748	0.41621
O	-0.12986	0.38384	O	-99.99000	-99.99000
P	-0.11630	0.38292	P	-0.11879	0.58152

MEDIAN TECH	BIAS	RMSE	IQR TECH	BIAS	RMSE
A	0.10020	0.51939	A	0.83271	1.19815
B	0.09302	0.52915	B	0.91461	1.25818
C	0.09784	0.52241	C	0.84250	1.20401
D	0.10267	0.51653	D	0.77038	1.15601
E	0.10235	0.51696	E	0.79925	1.17217
F	0.09806	0.52203	F	0.84315	1.20408
G	0.09913	0.52387	G	0.87166	1.23353
H	0.10690	0.51565	H	0.73813	1.12605
I	0.25836	0.60323	I	1.08054	1.46727
J	-0.06488	0.38955	J	0.43064	0.70556
K	-0.04453	0.45437	K	0.25217	0.51492
L	0.07732	0.43536	L	1.20375	1.60489
M	-0.14123	0.40036	M	0.40150	0.68763
N	0.10061	0.51896	N	0.82751	1.19236
O	-99.99000	-99.99000	O	-99.99000	-99.99000
P	-99.99000	-99.99000	P	-99.99000	-99.99000

Techniques which typically performed poorly for a given summary statistic would perform well (relative to other techniques) for a coefficient of variation equal to 0.25. Techniques J and M, L and I and L performed well (relative to other techniques) for estimating the mean, median and interquartile range, respectively, for low to moderate levels of single censoring. Technique J performed well (relative to other techniques) for estimating the mean and standard deviation with low to moderate levels of multiple censoring. Technique D (relative to other techniques) performed well for estimating the median with moderate levels of multiple censoring. In all cases, there was an alternative technique which performed as well as these techniques.

For presentation purposes, all simulation conditions were averaged together for a specific summary statistic (Table IV.6). The order of technique performance in Table IV.6 was similar to those of individual simulation parameters with the exceptions noted above.

Technique i+1 (and those with higher rmse) were eliminated from further consideration if the percentage increase in rmse from technique i to technique i+1 (in Table IV.6) was greater than a given critical percentage. The critical percentage was selected by plotting rmse as a function of technique. These plots were characterized by three distinct parts. Rmse was found to rise quickly for the first couple of techniques. This results from a comparison with the best

Table IV.6 Analysis of all raw data--average rmse for all simulations.

Single censoring

MEAN		STDV		MED		IQR	
A	.222	A	.367	A	.272	M	.821
P	.239	Q	.444	J	.298	J	.829
O	.242	H	.448	M	.319	D	.860
N	.268	D	.453	N	.351	H	1.044
C	.268	C	.456	E	.357	A	1.169
E	.268	N	.471	C	.378	C	1.511
F	.274	G	.486	F	.449	G	1.599
Q	.283	F	.487	H	.451	N	1.602
H	.295	E	.490	D	.507	E	1.644
G	.309	P	.583	G	.535	F	1.710
D	.368	B	.742	B	.677	K	2.242
B	.420	L	.899	K	.698	B	2.467
K	.443	K	1.017	L!	1.624	L!	3.621
L!	.847	I	1.310	I!	2.928	I!	3.820
M!	1.429	M!	2.999				
J!	1.903	J!	3.219				
I!	2.343						

Multiple censoring

MEAN		STDV		MED		IQR	
A	.195	A	.342	A	.232	D	.618
H	.229	H	.410	J	.246	J	.875
G	.234	G	.422	H	.308	A	1.053
C	.237	C	.431	G	.314	G	1.178
F	.237	F	.451	F	.320	H	1.224
E	.238	D	.464	E	.330	C	1.371
D!	.353	E	.505	C	.374	F	1.514
B!	.396	B	.662	D!	.649	E	1.740
J!	1.454	I!	1.074	B!	.675	B!	2.628
I!	1.513	J!	2.333	I!	1.838	I!	3.340

! These techniques were eliminated from further consideration due to high rmse.

overall technique (usually "no censoring"). The second part of these plots consisted of a gradual increase in rmse, followed by a sharp increase in rmse. A critical percentage, 40 percent, was chosen to eliminate techniques with sharply increasing rmse which occurred in the latter portion of these plots. Two to four techniques were eliminated for each simulation. In addition to eliminating poor performing techniques, one also notices from Table IV.6 that "no censoring" (technique A) is the best overall technique for estimating the mean, standard deviation and median.

The simulation data were fitted with a general linear model using SPSS. Typical input files are shown in Appendix B (see Figures B.1 and B.2). The linear model included all second order terms for coefficient of variation, sample size and censoring level and interaction terms for technique and distribution. In general, the model can be written as follows:

$$\text{rmse} = \underline{A}\ \underline{X} \qquad \text{IV.3}$$

where

$$\underline{A} = [\ \underline{a}_0,\ \underline{a}_1,\ \underline{a}_2,\ \underline{a}_3,\ a_4,\ a_5,\ a_6,\ a_7,\ a_8,\ a_9\ ],$$

$$\underline{X} = [\ 1.0,\ c_v,\ ss,\ p,\ c_v ss,\ c_v p,\ ssp,\ c_v^2,\ ss^2,\ p^2\ ]^T,$$

$$\underline{a}_0 = a_0 + a_0^i + a_0^j + a_0^{i,j},$$

$$\underline{a}_1 = a_1 + a_1^i + a_1^j,$$

$$\underline{a}_2 = a_2 + a_2^i + a_2^j \text{ and}$$

$$\underline{a}_3 = a_3 + a_3^i + a_3^j.$$

$\underline{X}$ is the covariate vector where c_v is the coefficient of variation, ss is the sample size and p is the censoring level. The censoring level for multiple censoring is defined as the weighted average of censoring levels. The first four terms of the coefficient vector, $\underline{A}$, incorporate distribution and technique factors. In general, $\underline{a}_0$ is composed of an overall constant, a_0; a distribution factor, a_0^i; a technique factor, a_0^j and a distribution-technique interaction factor, $a_0^{i,j}$. Similar definitions can be developed for $\underline{a}_1$, $\underline{a}_2$ and $\underline{a}_3$. The indices i and j refer to the distribution and technique used in the model, respectively. The choice of a second order model was partially based on the rmse levels reported by Gilliom and Helsel (1986). In their study, all rmse plots could be described with a second order model. It would be desirable in this study to use a third order model and to statistically demonstrate that none of the third order terms are significant. However, it was not possible to include higher order terms due to storage limitations of the mainframe computer. The storage limitation problem resulted from the large number of interaction terms.

A backwards step-wise process was used to eliminate insignificant terms (90% confidence level) until a parsimonious model was developed. The resulting coefficients are reported in Tables B.1a through B.1h. Coefficients reported as 0.0 correspond to covariates or factors that were not significant at the 90% confidence level.

The general linear model described above is not necessarily the most appropriate model to describe rmse for all summary statistics. For example, one would expect that the rmse for mean estimation would be inversely proportional to the square root of sample size. To check the adequacy of the selected general linear model, residuals were computed as the raw simulation rmse minus the general linear model rmse. Residuals were plotted as a function of coefficient of variation, sample size, censoring and distribution for each estimation technique. It was found that the residuals were zero mean (as expected by definition) and constant variance for various levels of simulation parameters. The residuals were also examined to consider the adequacy of the selected model. A curved residual plot would have indicated that the selected model was not satisfactory to describe the rmse surface, and a different model would be necessary. Based on visual inspection of scatter plots, there was no visable (to very little) curvature in the residual plots.

There were two exceptions to the generalization for variance homogeneity. First, there was occasional heterocedasity with respect to coefficient of variation. Some of the residual plots tended to have a larger variance for larger levels of coefficient of variation. Second, there were a few cases in which the residuals were heterocedastic with respect to technique. These techniques were characterized as techniques which had high rmse from Table IV.6, but had not been eliminated with the criterion developed for evaluating the raw simulation data. Both features were attributed to techniques that did not perform consistently over a wide range of conditions and contributed a great deal to the overall model variance. However, there was no consistent pattern that could be discerned. It was opted to continue with the model developed thus far recognizing some model fitting violations as opposed to creating a unique model for each summary statistic.

Response surfaces were made in the following way since it is difficult to display or visualize a model with more than two independent variables. For a given plot, distribution and sample size were held constant and censoring level and variability were allowed to vary over the simulated range. Rmse was computed at discrete points for each technique across the surface. The censoring level varied from zero to one and the coefficient of variation varied from zero to two, both in steps of 0.05. Multiple plots were made varying distribution and sample size. Sample size varied from eight to 48 in steps of four, and distribution varied across all five parent distributions listed in Table IV.1. Rather than contouring rmse, the technique code with minimum rmse was plotted. Figure IV.1 is a selected example displaying this technique. The "no censoring" technique is not included in these figures since retroactive uncensoring is not a feasible scenario.

The rmse can be evaluated with the general linear model. Since the linear model represents a more complete range of possibilities, it is reasonable to use the model to evaluate rmse in the same manner that the raw simulation results were analyzed in the preliminary scan.

```
           Dist: Lognormal  Sample size = 24
          |JJJJJJJJJCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
          |JJJJJJJJJJCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
          |JJJJJJJJJJCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
          |JJJJJJJJJJCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
      0.8—|JJJJJJJJJJCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
          |JJJJJJJJJJJCCCCCCCCCCCCCCCCCCCCCCCCCCCC
          |JJJJJJJJJJJCCCCCCCCCCCCCCCCCCCCCCCCCCCC
          |JJJJJJJJJJJCCCCCCCCCCCCCCCCCCCCCCCCCCCC
      0.6—|JJJJJJJJJJJCCCCCCCCCCCCCCCCCCCCCCCCCCCC
          |JJJJJJJJJJJJHCCCCCCCCCCCCCCCCCCCCCCCCCC
Censoring |JJJJJJJJJJJJHHHCCCCCCCCCCCCCCCCCCCCCBBB
level     |JJJJJJJJJJJJHHHHHHCCCCCCCCCCCCCCCCBBBBB
      0.4—|JJJJJJJJJJJJHHHHHHHHHCCCCCCCCCCBBBBBBBB
          |JJJJJJJJJJJJJDDDDHHHHHHCCCCBBBBBBBBBBBB
          |JJJJJJJJJJJJJDDDDDDDDDDDDDBBBBBBBBBBBBB
          |JJJJJJJJJJJJLLLDDDDDDDDDDDBBBBBBBBBBBBB
      0.2—|JJJJJJJJJJJIIIIIIIIIIIIIIIIIIIIIIIIIIII
          |JJJJJJJJIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
          |JJJJJJIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
          |JJJIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
          |JIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
          |________________________________________
              |    |         |                   |
            0.25  0.5       1.0                 2.0
                      Coefficient of variation
```

Figure IV.1 Selected rmse response surface displaying output from the analysis of a general linear model.

Rmse values in Table IV.7 are averaged over the range for which the technique was valid. Rmse values were computed for Table IV.7 in the following manner. A systematic grid which encompasses all simulation conditions was selected. The rmse was computed for each grid point was computed. All rmses were summed for all feasible grid points. (Techniques O and P are applicable only when less than one-half of the data are censored.) It was opted to divide the sum by the number of grid locations, producing an average rmse which would be more comparable to the results in Table IV.6. In Figure IV.1, each character in the

Table IV.7 Analysis of general linear model--average rmse for all valid regions.

Single censoring

MEAN		STDV		MED		IQR	
A	.228	A	.440	A	.301	M	.595
P	.240	H	.514	J	.322	J	.596
O	.250	D	.520	M	.347	D	.652
N	.277	C	.523	N	.383	H	.773
C	.277	Q	.526	E	.385	A	.903
E	.277	N	.537	C	.402	C!	1.224
F	.282	F	.547	F!	.475	N!	1.300
Q	.292	E	.551	H!	.479	G!	1.308
H	.304	G	.558	G!	.564	E!	1.355
G	.316	P!	.718	D!	.571	F!	1.409
D!	.383	B!	.783	B!	.697	K!	1.946
B!	.420	L!	.993	K!	.717	B!	2.127
K!	.441	K!	1.073				
		I!	1.609				

Multiple censoring

MEAN		STDV		MED		IQR	
A	.224	A	.323	A	.281	D	.590
H	.266	H	.385	J	.299	J	.801
G	.271	G	.398	H	.355	A	1.000
C	.271	C	.401	F	.367	G	1.077
F	.271	F	.413	G	.369	H	1.128
E	.272	D	.443	E	.383	C	1.268
		E	.465	C	.411	F	1.390
		B!	.604			E!	1.656

! These techniques were eliminated from consideration due to a high rmse from the general linear model analysis.

response surface represents a grid location. In all, 47,355 grid locations were selected for computing the average rmse for each summary statistic.

For actual application, it is necessary to know what the most appropriate technique is for a given set of conditions. Table IV.7 does not provide this information. As with any summary of this data, Table IV.7 may be biased to the particular choice of simulation conditions. However, presentation of all simulation data would result in an unmanageable table.

Techniques were now eliminated based on a critical percentage of 15 percent using the same reasoning that was developed for technique elimination earlier. One should notice that the order of technique performance in Table IV.6 and IV.7 is similar. In some cases, the absolute values of the rmse are different between the two tables. This is not surprizing since the average rmse computed in Table IV.6 is based on selected simulation conditions, perhaps introducing some bias. The purpose of the preliminary scan used in Table IV.6 was to eliminate techniques with the highest rmse (e.g., techniques that should always be avoided) so that the general linear model would fit better. If a more strict rule were used for technique elimination in Table IV.6, the same techniques would be eliminated as were eliminated in Table IV.7. This corraboration tends to support the use of either analysis procedure.

A number of techniques have been eliminated at this point. Many of the decisions have been based on corraborative evidence from analyzing the raw simulation data as well as the linear model. The objective thus far has been to remove techniques which should be avoided (techniques that are not robust). This philosophy follows the same premise behind the preference of nonparametric statistics. While there are certain instances where one of the eliminated techniques performs well, it would be difficult, in practice, to determine when these cases arise. On the other hand, incorrectly selecting techniques that are not robust may result in large errors. The remaining techniques do perform well over a wide range of conditions.

B.2 Robust Technique Elimination

I opted to refit the linear model to the raw simulation data. The raw simulation data developed for the eliminated techniques as well as the raw simulation data for the "no censoring" technique were not used since it was desired to develop a more sensitive model. In addition to the techniques eliminated above, techniques that are computationally equivalent were eliminated. Techniques E and N for mean estimation with single censoring and technique F for mean estimation with multiple censoring will not be considered further since these techniques provide the same mean estimate as technique C; however, either technique E or F would be a suitable replacement for technique C. The linear model (Equation IV.3) was fitted to the remaining simulation data using the backwards step-wise regression described previously. The resulting coefficients are reported in Tables B.2a through B.2h.

Average rmse values were recomputed using the refitted general linear model coefficients (Table IV.8). Results are similar to the values reported in Table IV.7 except for standard deviation and interquartile range estimation with single censoring. This is due to the techniques included in

Table IV.8 Analysis of refitted general linear model--average rmse for all valid regions.

Single censoring

MEAN		STDV		MED		IQR	
P	.241	Q	.410	J	.333	M	.796
O	.251	H	.415	M	.351	J	.797
C	.275	D	.419	N	.381	D	.833
F	.281	C	.422	E	.390	H	.998
Q	.291	N	.433	C	.401		
H	.302	F	.447				
G	.315	G	.448				
		E	.449				

Multiple censoring

MEAN		STDV		MED		IQR	
H	.265	H	.392	J	.311	D	.586
G	.270	G	.405	H	.371	J	.812
C	.271	C	.408	G	.376	G	1.082
E	.272	F	.421	F	.385	H	1.133
		D	.449	E	.400	C	1.271
		E	.472	C	.431	F	1.395

the previous model with excessively high rmse that were not eliminated in the preliminary scan. The residuals were rechecked for model adequacy. As expected, the overall residual variance for each summary statistic either decreased or remained approximately the same, which indicated a better fit to the raw simulation data for the remaining techniques. Variance heterocedasity was decreased in most instances although not completely removed.

B.2.1 Technique by technique comparison

Techniques are now compared, one to another, for each summary statistic. It was desired to remove techniques that were consistently worse than another technique. This does not imply that the techniques eliminated in this section performed poorly. Rather this was to reduce the number of competing techniques when performance was nearly the same. Response surfaces were developed to compare two techniques at a time. The technique with lower rmse was recorded at each grid point and the total number of grid points were summed. Rmse values from two different techniques within 5% were considered equal. The same comparison was made with the raw simulation data where the number of grid points summed is replaced by the number of simulations.

It was hoped that similar corraborative evidence between the evaluation of the raw simulation data and the general linear model would occur. While complete superiority between techniques did not occur often, superiority was detected usually with the raw simulation data. In some cases the general linear model evaluation did not support the evaluation of the raw data. I opted to use the results from the raw simulation data as the primary decision tool. Only results that lead to technique elimination are reported (Table IV.9).

There were two interesting features that were noticed when visually examining the response surfaces generated by the general linear model. First, maximum likelihood estimation techniques did not perform well when there was an extremely high level of censoring. This was clearly demonstrated when simple substitution techniques (such as technique C) worked better for high censoring levels. This was attributed to the algorithms used for maximum likelihood

Table IV.9 Rmse comparison for techniques which result in technique elimination.

Comparison description	Raw data comparison Linear model comparison			
	<	>	=	n/c
mean, single censoring, comparison of techniques C and F (F eliminated)	22.9	0.0	75.8	1.3
	56.1	40.0	0.0	
standard deviation, single censoring, comparison of techniques N and F (F eliminated)	28.2	0.8	69.7	1.3
	43.8	0.0	56.2	
interquartile range, single censoring, comparison of techniques J and H (H eliminated)	92.4	1.8	4.5	1.3
	77.9	8.1	14.0	
mean, multiple censoring, comparison of techniques C and E (E eliminated)	7.8	0.9	87.6	3.7
	8.2	0.0	91.8	
standard deviation, multiple censoring, comparison of techniques H and G (G eliminated)	15.5	0.7	80.1	3.7
	39.6	35.5	24.9	
interquartile range, multiple censoring, comparison of techniques D and G (G eliminated)	87.0	2.1	7.2	3.7
	64.8	18.1	17.1	
interquartile range, multiple censoring, comparison of techniques J and G (G eliminated)	90.3	4.3	1.7	3.7
	58.4	27.6	14.0	
interquartile range, multiple censoring, comparison of techniques D and H (H eliminated)	88.4	1.6	6.3	3.7
	68.1	15.3	16.6	

* Technique comparison is based on the percentage of time the rmse of the first technique listed is less than (column 1), greater than (column 2), or equal to (column 3) the rmse of the second technique listed in the description. Rmse was considered equal when the rmse from the two techniques is within five percent of each other. The last column is the percentage of times it was not possible to make a comparison of techniques.

which would not converge to realistic solutions. This suggests that one should not completely rely on maximum likelihood techniques.

Second, lognormal probability techniques (Q for single censoring and H for multiple censoring), which use a maximum likelihood technique for fitting a linear regression line to all data appear to work better than the techniques (H for single censoring and G for multiple censoring) which only use the uncensored observations to fit the linear regression line for higher censoring levels. Poor standard technique performance was attributed to extrapolating linear regression lines to censored observations based on only a few uncensored observations. The number of uncensored observations below each detection limit also appeared to give additional advantage to technique H for multiply censored data. The non-convergence problem associated with maximum likelihood techniques was a problem for technique H (single censoring) and technique G (multiple censoring) at high censoring levels as well.

B.2.2 Technique selection

So far the analysis has proceeded on the premise of eliminating techniques which are not robust or do not appear to perform any better than another available technique. Now an alternative approach is taken. How many and which techniques are necessary to adequately estimate the summary statistics? There are two extremes. First, only use the best technique that is listed in Table IV.6 and IV.7. Or second, use all of the techniques that have not been eliminated. The former is more easily implemented, but may result in poor performance for a specific set of conditions. The converse is true for the latter extreme.

There were several ways to approach this problem. After some trial and error, this analysis proceeded as follows. An average rmse was computed from the general linear model assuming that one has chosen to use the best technique available from those remaining (e.g., use the best technique available). This resulted in an optimal rmse. A ten percent rmse range was computed using the optimal rmse as the lower limit (Table IV.10). All technique combinations (up to three) which fall within the rmse range were considered as acceptable alternatives. After reviewing the list of acceptable combinations, it was noted that two techniques could meet the criterion established in Table

Table IV.10 Optimal rmse computations assuming the use of all techniques that have not been eliminated and the computed rmse range.

Censoring type	Summary statistic	Rmse range: Optimal rmse	Rmse range: Upper limit
Single censoring	Mean	.242	.266
	Standard Deviation	.371	.408
	Median	.281	.309
	Interquartile Range	.604	.664
Multiple censoring	Mean	.255	.281
	Standard Deviation	.342	.376
	Median	.253	.278
	Interquartile Range	.462	.508

IV.10 for the mean, standard deviation and median estimation. More than two techniques were needed to meet the interquartile range estimation. To maintain simplicity, only two techniques were selected for interquartile range estimation as they were less than 15% above the optimal rmse value.

B.2.3 Recommended techniques

In many instances, the criteria established above reduced the number of acceptable choices to two or three combinations. Final recommendations are summarized in Table IV.11. The best overall technique (Tables IV.6 and IV.7)

Table IV.11 Technique recommendations for summary statistic estimation.

Censoring type	Summary statistic	Technique recommendations
Single	Mean	CH or NH
	Standard Deviation	QN
	Median	JM
	Interquartile Range	MJ
Multiple censoring	Mean	HC or HF
	Standard Deviation	HF
	Median	JF
	Interquartile Range	DJ

was selected as one of the two techniques selected except for mean estimation with single censoring. In this case, Techniques C and H performed as well as Technique P and any other technique. Additionally, Technique P is only valid for censoring levels less than 50%. The second technique was selected based on which technique compliments the first technique (reduces rmse) the most. In instances were the second techniques were equivalent, the second technique was selected so as to reduce the overall number of different techniques used throughout the protocol.

There are several reasonable alternatives to those listed in Table IV.11. Technique N (F) may be substituted for technique C for mean estimation and single (multiple) censoring. Techniques which use a log-regression extrapolation procedure are similar to each other. In this study, the procedure which was modified after Wolynetz (1979b) appears to perform better than the original procedure described by Gilliom and Helsel (1986) for multiple censoring although the difference is small. A clear choice is not apparent for single censoring. The Wolynetz modified procedure requires more time to compute than the original procedure. This study did not develop a benchmark to compare computational time. However, it is reasonable to suspect that data sets with a high level of censored observations would reqire more time for the algorithms to converge than data sets with few censored observations. In some instances this difference in computation time may justify the use of the original procedure instead of the modified procedure.

Finally, the detection limit substitution technique is recommended for interquartile range and multiple censoring. As noted by Gilliom and Helsel (1986) and Helsel and Cohn (1988), this procedure produces a biased estimate of the interquartile range and was confirmed in this study (results not reported). However, based on the criteria developed in this chapter, this procedure does perform well (low rmse). In some instances, estimates of sample statistics do not provide good (unbiased) estimates of population statistics. The biased estimates reported by Gilliom and Helsel (1986) and in this study indicate that the sample interquartile range is a biased estimate of the population interquartile range. I am somewhat cautious about the use of the detection limit substitution recommendation and suggest careful review before applying the detection limit substitution technique to data sets since results may be biased.

C. Criteria for Selecting Appropriate Technique

Two techniques have been recommended for each type of summary statistic estimation problem identified in this study. A criteria is now developed to resolve which technique should be used for a given data set.

Gilliom and Helsel (1986) recommended using the relative quartile range for classification of singly censored data sets. The extrapolation of this technique to the multiply censored data sets may be reasonable; however, this procedure was not tested. In lieu, estimation of the coefficient of variation is recommended for deciding which technique is most appropriate.

For a given data set, the censoring level and sample size can be determined directly from the data set. It is recommended to make an initial estimate of the coefficient of variation (of the population) using the first technique listed in Table IV.11. Based on the initial coefficient of variation estiamte, Figure IV.2 and IV.3 may be consulted to determine which technique should be used for final summary statistic estimation. No attempt is made to incorporate distribution estimation into the technique selection procedure. Figures IV.2 and IV.3 are a summary of the response surfaces from the linear model, comparing the two recommended techniques. The results (in Figures IV.2 and IV.3) were compared to the raw simulation data and were found to be comparable. Several of the response surfaces in Figures IV.2 and IV.3 have two lines separating technique choices that are marked with a corresponding sample size (SS). In these instances, the choice between techniques is dependent on sample size in addition to censoring and variability.

The techniques listed in Table IV.11 have been found to perform relatively well in comparison to other techniques for a wide range of conditions typically found in water quality data. It is not recommended; however, to extrapolate beyond the simulation study scope.

D. Summary

Network managers are often faced with describing the average water-quality conditions of a particular location. Standard practices include estimating the central tendency and variability of a single water-quality random variable.

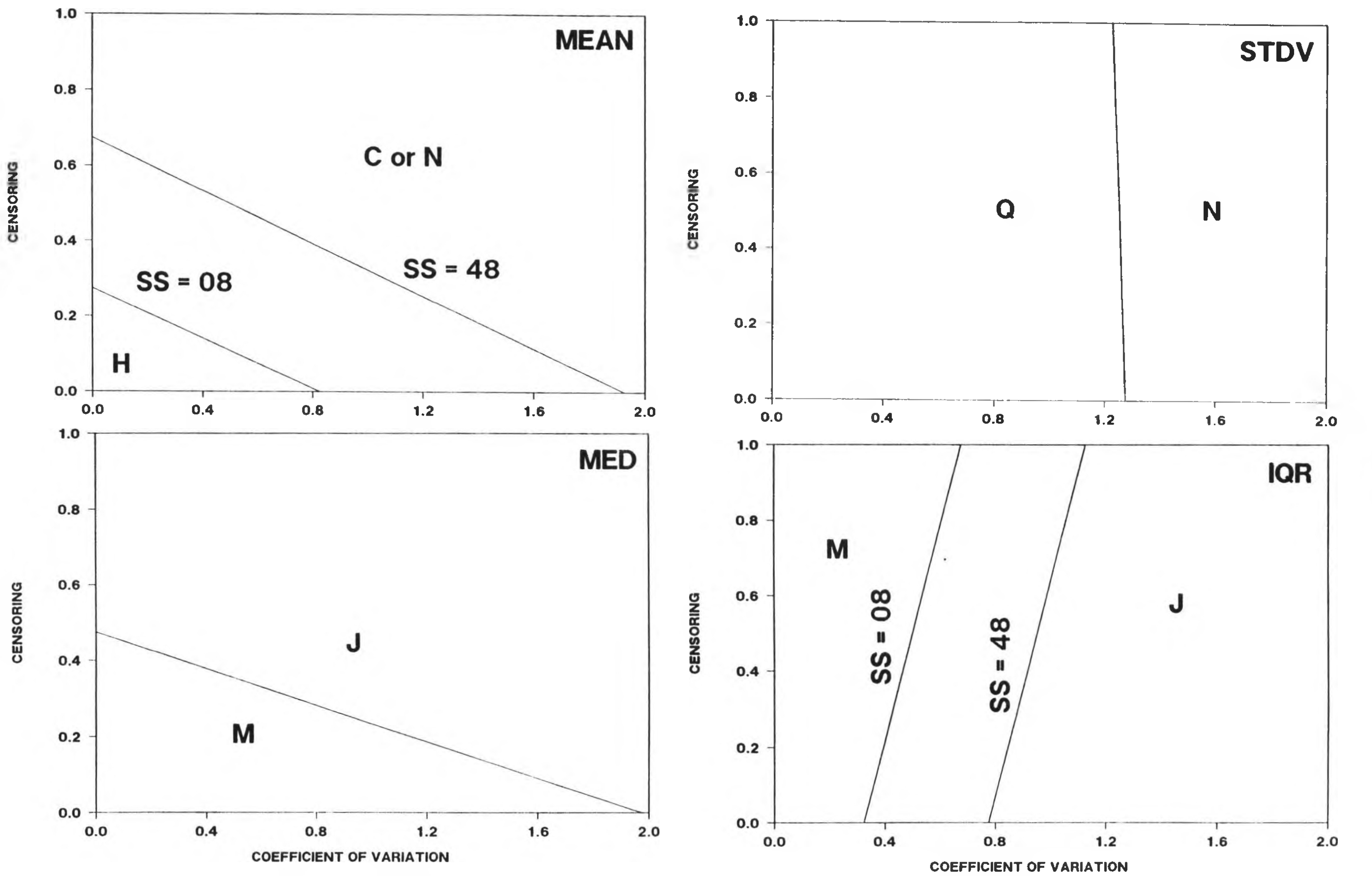

Figure IV.2 Response surface summary for single censoring.

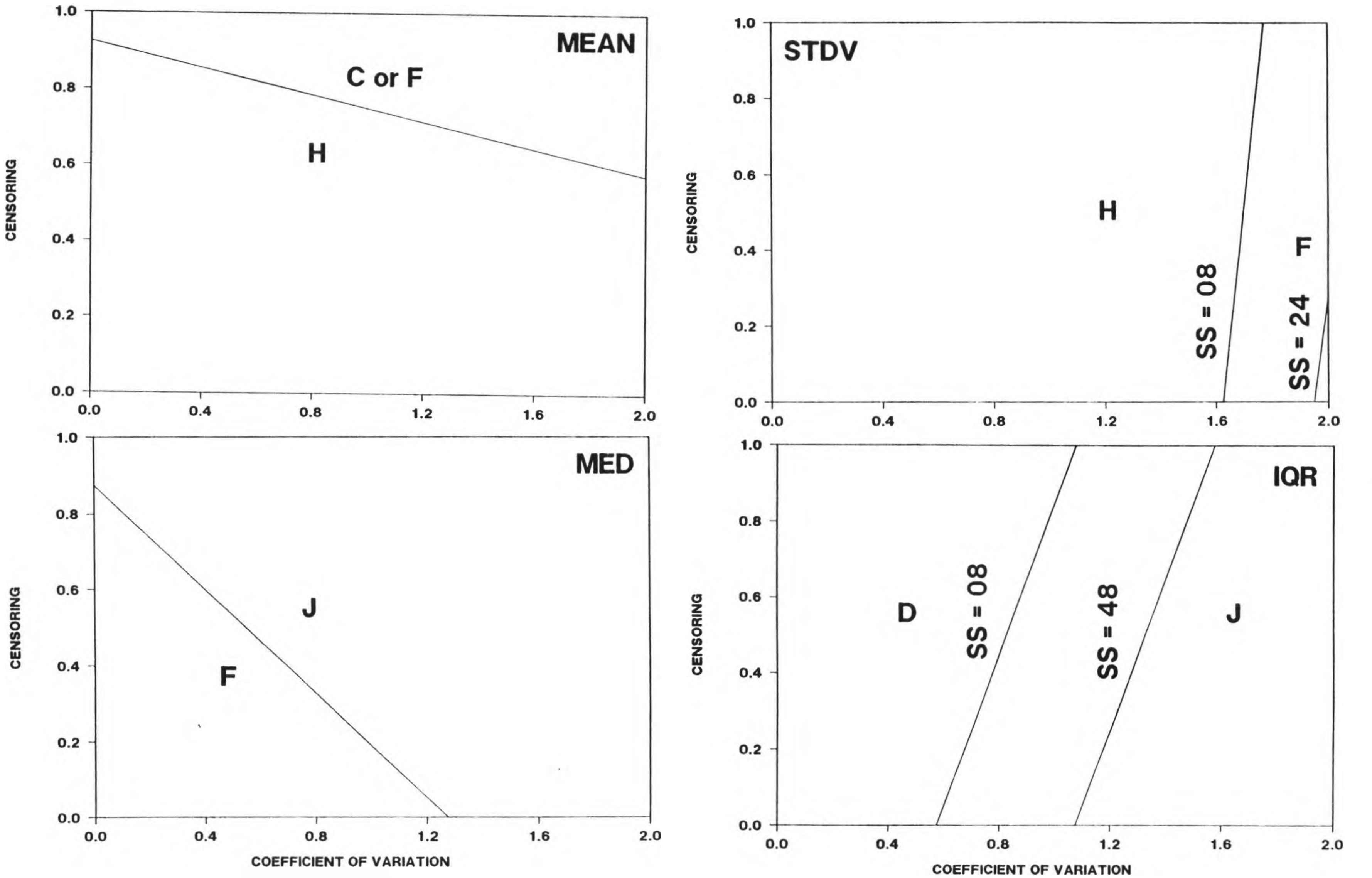

Figure IV.3 Response surface summary for multiple censoring.

The mean and median are most often used as an estimate of the central tendency and the standard deviation and interquartile range are used as estimates of the variability. Summary statistic estimation is complicated with both single and multiple censoring. Numerous techniques have been proposed to estimate summary statistics, when part of the data have been censored. The problem facing the network manager is choosing the most appropriate technique.

A simulation study was used to compare the performance of different techniques. A methodology was developed to reduce the number of techniques a network manager would have to consider for estimating summary statistics. It should be noted that there were a number of techniques which perform well for many possible water-quality data records.

This chapter confirms the results of previous researchers (Helsel and Gilliom, 1986; Gilliom and Helsel, 1986; and Helsel and Cohn, 1988). The techniques which performed well for their study performed well in this study. Additional support is also given to the "no censoring" rule recommended by Porter (1986) and Porter et al. (1988).

CHAPTER V

CHANGING CONDITIONS--DETECTION WITH SERIAL CORRELATION

There are several motivations for detecting trends in water-quality variables. Primarily, the network manager is interested in deciding whether the water quality, over time, has: (1) not changed; (2) changed for the better or (3) changed for the worse. Trend tests are a statistical way to evaluate the significance of apparent changes in a water-quality variable and to estimate the magnitude of change.

Gradual and abrupt changes are two trend types that are of most concern the network managers. Monotonic and step trend tests are typically used to detect gradual and abrupt changes, respectively. Step trend tests are also used to detect changes over space (e.g., from one monitoring station to another). In many instances, water-quality data do not meet the normality assumption underlying parametric tests. Typical data records may also contain outliers or censored observations. As a result, most water-quality trend detection investigations use nonparametric methods (Smith et al., 1987 and Ward et al., 1988).

There are still assumptions affiliated with nonparametric methods. Usual assumptions relating to water quality include identically distributed errors and no serial correlation. Assumption violations can cause misleading results. The former assumption suggests variance homogeneity and no seasonality. Some nonparametric tests have been modified to implicitly deal with seasonality in the mean (Hirsch et al., 1982; Hirsch and Slack, 1984 and Gilbert, 1987). Most literature recommends, at least, visually looking at the data (or box plots) to check for variance homogeneity (e.g., Ward et al., 1988 and Berryman et al., 1988).

The independence assumption implies no stochastic relationship between successive observations in time (Harris, 1988). Some trend tests have also been modified to accommodate serial correlation. Tests which have been corrected for serial correlation are less powerful than uncorrected tests when no serial correlation is present. Corrected tests also require more data, and correlation coefficients are difficult to estimate with small data sets. (Tasker (1983a) gives approximate probability density functions for estimated correlation coefficients and small sample sizes which could be used for hypothesis testing.)

On the other hand, uncorrected test significance levels maybe greatly inflated when serial correlation is present. Unfortunately, standard water-quality sampling frequencies are usually in the range where serial correlation may or may not be important. This places the network manager at an impasse--what technique should be used? This chapter investigates several techniques found in the literature for monotonic trend detection when serial correlation exists. Techniques studied are: (1) ignoring serial correlation; (2) using tests which explicitly account for serial correlation or (3) creating new times series by collapsing the former time series to less frequent values and then applying a standard test.

A. Simulation Methodology

A.1 Data Generation

Data sets with known trends are generated from a variety of parent distributions similar to the procedures described by Taylor and Loftis (1989). Data were generated using an International Mathematical and Statistical Libraries generating routine for the normal distribution.

Monthly data were created by first generating 500 normal warm-up values with zero mean and unit variance (e.g., N(0,1)). These values were transformed to account for serial correlation by:

$$w_i = (1-\rho^2)^{0.5}\, z_i + \rho\, z_{i-1} \qquad \text{V.1}$$

where

w_i = ith generated serially correlated value,

z_i = ith generated N(0,1) random number and

ρ = selected monthly serial correlation.

The last value, z_{500}, is used as z_0 during the next generation of random numbers.

For each simulation, a new set of N(0,1) data is generated. (e.g., Sixty new N(0,1) random numbers would be created for five years of monthly data, that is $z_1, \ldots, z_{60}$.) If the simulation conditions specify a lognormally distributed error term, generated N(0,1) data are transformed by:

$$v_i = \exp(0.833 z_i - 0.347) - 1.0 \qquad \text{V.2}$$

creating lognormal (0,1) data, otherwise v_i is equal to z_i. A serially correlated series is then obtained from Equation V.1 where v_i is substituted for z_i.

Seasonality in the variance is added by:

$$x_i = w_i \sigma_r \qquad \text{V.3}$$

where σ_r is equal to the ratio of standard deviations in high variability months to low variability months. The seasonality was formulated so that the high variance months were the second and fourth quarters of the year. This results in a low--high--low--high quarterly variance pattern. Taylor and Loftis (1989) demonstrate that the trend tests considered in their study handle mean seasonality equally well. As a result, this study (which uses a subset of the tests used by Taylor and Loftis, 1989) does not explicitly consider mean seasonality.

Finally, trend is added to the simulated data by:

$$y_i = x_i + b\,i + c \qquad \text{V.4}$$

where b is the incremental slope and c is a small constant so that all simulated data are positive.

A.2 Evaluation

Monotonic trend detection techniques are then applied to each data set, recording the number of simulations for which trend was detected. All tests are run at a nominal five percent significance level. Two criteria (empirical significance level and power) are used for evaluating test performance. First, the nominal significance level is compared with the simulated significance level. A test which preserves the nominal significance level is a test that results in an actual significance level in the range $0.032 \leq \alpha \leq 0.068$. The range was selected as the 95% confidence interval for α when 500 simulations are used (Snedecor and Cochran, 1980). A larger number of simulations would result in a smaller confidence interval. Tests are termed conservative and liberal when the actual significance level is below and above the range, respectively.

According to Taylor and Loftis (1989), second order linear models were too crude for technique comparison. Therefore, a technique comparison similar to that presented in Chapter IV is not attempted. Consequently, the second criterion is based on direct comparison of power curves.

Power curves can be graphically interpreted, or the actual power levels can be compared statistically. The null hypothesis is that the proportion of detects for test A is the same as the proportion of detects for test B, or equivalently, the difference in detect proportions for test A and B is zero. Snedecor and Cochran (1980) show that the normal deviate, z, for comparing two proportions with equal sample sizes is given by:

$$z = (p_A - p_B)/(pq\ (2/N))^{0.5} \qquad \text{V.5}$$

where

p_A = proportion of detects for test A,

p_B = proportion of detects for test B,

$p = (p_A+p_B)/2$,

$q = 1-p$ and

N = the number of samples for either test.

The null hypothesis is rejected when the absolute value of z is greater than 1.96 (for a two-sided test at the 95% confidence level). Test A is said to be more <u>powerful</u> than test B when p_A is significantly greater than p_B for simulations in which there is a real trend in the data.

It is unfair to apply the test described above when the comparison is made with one trend test which has an inflated nominal significance level. A trend test which starts with an inflated actual significance level will likely have more power than a test which is accurate. To protect against this bias, comparisons will be limited to those test pairs for which most nominal significance levels were insignificantly different.

B. Analyses

B.1 Simulation Conditions

Five hundred sets of monthly data were generated for each set of simulation conditions. The parameters considered in this simulation were applicable distribution, variance seasonality, serial correlation and sample size (Table V.1). A total of 60 combinations were selected for simulation purposes.

Table V.1 Levels of distribution, variance seasonality, serial correlation and sample size considered in changing condition detection with serial correlation simulation.

Distribution	Normal Lognormal		Ratio of seasonal standard deviations	1.0 5.0
Monthly lag-1 correlation coefficient	0.0 0.2 0.4	0.6 0.8	Years of data	5 10 15

Mean seasonality was not considered since Taylor and Loftis (1989) demonstrated that the ratio of the seasonal means, a measure of seasonality, was not significant in the second order linear model they fitted to their simulation results. In other words, all the tests they considered effectively deal with mean seasonality. Incremental slopes (b of Equation V.4) were selected as 0.000, 0.002, 0.005, 0.020, 0.050, 0.200 and 0.500. Preliminary simulations indicated that these seven incremental slopes adequately cover the range of slopes that are interesting.

B.2 Trend Detection Procedures

Trend detection techniques studied were the Mann-Kendall tau test on deseasonalized data (MKD), the Seasonal Kendall tau test (SK) and the Seasonal Kendall tau test with correction for serial correlation (SKC). Data were deseasonalized by the seasonal mean and variance. Taylor and Loftis (1989) deseasonalized for the seasonal mean, whereas, Hirsch et al. (1982) deseasonalized for the mean and variance before applying a seasonal regression technique. Each test was applied to each data set of simulated monthly observations. The data sets were then collapsed to quarterly values in an attempt to reduce serial correlation. Seasonal quarters were defined to coincide with the seasonal variance pattern. (The first quarter contains the first three months of the year, etc.) Quarterly values were determined by computing the mean, median and middle observation for each quarter. The three trend tests were then reapplied to each of the three collapsed data sets. These three data sets are referred to as quarterly collapsed data. Quarterly averaged data refer to the first two procedures of collapsing data. Two yearly time series were constructed by computing the yearly mean and median from the corresponding quarterly averaged data (e.g., the mean of the quarterly means). The Mann-Kendall tau test (MK) was applied to both yearly averaged series. The fourteen procedures are summarized in Table V.2.

Table V.2 Summary of statistical procedures used for trend detection study.

Code	Technique
Tests applied to monthly data:	
A	Mann-Kendall tau test is applied to data which have been deseasonalized for mean and variance
B	Seasonal Kendall tau test
C	Seasonal Kendall tau test with serial correlation correction
Tests applied to quarterly averaged data (monthly data were collapsed by computing quarterly means):	
D	Mann-Kendall tau test is applied to data which have been deseasonalized for mean and variance
E	Seasonal Kendall tau test
F	Seasonal Kendall tau test with serial correlation correction
Tests applied to quarterly averaged data (monthly data were collapsed by computing quarterly medians):	
G	Mann-Kendall tau test is applied to data which have been deseasonalized for mean and variance
H	Seasonal Kendall tau test
I	Seasonal Kendall tau test with serial correlation correction
Tests applied to quarterly collapsed data (monthly data were collapsed by selecting the middle observation of each quarter):	
J	Mann-Kendall tau test is applied to data which have been deseasonalized for mean and variance
K	Seasonal Kendall tau test
L	Seasonal Kendall tau test with serial correlation correction
Tests applied to yearly averaged data:	
M	Mann-Kendall tau test is applied to yearly values. The yearly values were computed as the mean of the quarterly means.
N	Mann-Kendall tau test is applied to yearly values. The yearly values were computed as the median of the quarterly medians.

B.3 Correlation Levels in Collapsed Data

Data set collapsing is an attempt to reduce the serial correlation in a time series. It is interesting to consider a typical situation for which monthly data have been collected and it is suspected that the data may contain some serial correlation. If a new time series is constructed by calculating quarterly means, what serial correlation level is in the new time series?

Suppose a monthly time series, X_t, is defined as a normal autoregressive process with order one. The variance (σ_x^2) and lag-one serial correlation coefficient (ρ) of X_t are constant. A new quarterly time series, Y_t, is defined such that the first value of Y_t is the mean of the first three values of X_t and so on. Specifically, Y_t is:

$$Y_t = (X_s + X_{s+1} + X_{s+2}) / 3 \qquad \text{V.6}$$

where s is equal to 3(t-1)+1. The variance of Y_t is computed as:

$$\mathrm{Var}(Y_t) = \frac{1}{9}\left[\sum_{i=1}^{3} \mathrm{Var}(X_i) + 2 \sum_{j<k} \mathrm{Cov}[x_j, x_k]\right] \qquad \text{V.7}$$

$$\text{or} \quad \mathrm{Var}(Y_t) = \frac{1}{9}\{\, 3\,\sigma_x^2 + 2\,\mathrm{Cov}[X_s, X_{s+1}] + 2\,\mathrm{Cov}[X_{s+1}, X_{s+2}] + 2\,\mathrm{Cov}[X_s, X_{s+2}]\} \qquad \text{V.8}$$

$$\text{or} \quad \mathrm{Var}(Y_t) = \frac{(3 + 4\rho + 2\rho^2)}{9}\,\sigma_x^2 \,. \qquad \text{V.9}$$

The covariance is:

$$\mathrm{Cov}[Y_t, Y_{t+1}] = \mathrm{Cov}[(X_s+X_{s+1}+X_{s+2})/3 \ , \ (X_{s+3}+X_{s+4}+X_{s+5})/3]. \qquad \text{V.10}$$

After expanding the right hand side of Equation V.10 and combining like terms, the covariance may be expressed as:

$$\mathrm{Cov}[Y_t, Y_{t+1}] = \frac{(\rho + 2\rho^2 + 3\rho^3 + 2\rho^4 + \rho^5)}{9} \sigma_x^2. \qquad \text{V.11}$$

The lag-one serial correlation of Y_t is the ratio of Equation V.11 to V.9 or

$$\mathrm{Corr}[Y_t, Y_{t+1}] = \frac{(\rho + 2\rho^2 + 3\rho^3 + 2\rho^4 + \rho^5)}{(3 + 4\rho + 2\rho^2)}. \qquad \text{V.12}$$

The lag-one serial correlation for any new time series may be computed by generalizing the above relationship such that:

$$\rho^* = \frac{N\rho^N + \sum_{i=1}^{N-1} i(\rho^{2N-i} + \rho^i)}{N + \sum_{i=1}^{N-1} 2(N-i)\rho^i} \qquad \text{V.13}$$

where ρ^* and ρ are the new and old time series lag-one serial correlation coefficients, respectively, and N is the number of values used to compute the mean of the new time series. For example, the serial correlation for the quarterly averaged time series given by Equation V.12 may be computed from Equation V.13 with N equal to three. The above development does not account for variance seasonality. While this restricts practical use, it does provide the basis for a qualitative discussion later. Table V.3 provides the lag-one serial correlation values of mean collapsed time series.

Table V.3 Lag-one serial correlation levels for generated and collapsed data sets assuming no variance seasonality.

Generated monthly serial correlation	Serial correlation for mean collapsed data		Serial correlation for middle collapsed data	
	Quarterly	Yearly	Quarterly	Yearly
0.0	0.000	0.000	0.000	0.000
0.2	0.079	0.018	0.008	0.000039
0.4	0.198	0.043	0.064	0.002
0.6	0.377	0.092	0.216	0.020
0.8	0.637	0.245	0.512	0.165

Recall that the yearly mean is computed from the quarterly means. To compute the serial correlation level in the yearly mean averaged time series, Equation V.13 may be applied with N equal to 12 and ρ equal to the monthly lag-one serial correlation coefficient. It is not appropriate to apply Equation V.13 successively since the quarterly averaged time series is an ARMA(1,1) process (Obeysekera and Salas, 1986). (Equation V.13 was derived assuming that the original time series was an AR(1) process.) For comparison the serial correlation of the middle collapsed time series (ρ^N) is provided as well. The process variance of the time series created by selecting the middle value is the same as that of the original time series. The process variance of the time series created by computing quarterly means has an upper bound of the original time series process variance and a lower bound of one-third of the original process variance (for $0 \leq \rho \leq 1$).

C. Results

Power tables resulting from the above simulation are presented in Appendix C. Techniques are now compared based on their actual significance level and power.

C.1 Actual Significance Level Comparison

The number of detected trends should equal five percent when there is no trend in the data. Actual significance levels can be read from the tables in Appendix C. Significance levels were indexed and summarized in Tables V.4 through V.6. A negative index indicates a "conservative" test (actual significance levels were significantly less than nominal significance levels) and increasing positive indices indicate increasingly "liberal" tests (actual significance levels were significantly greater than nominal significance levels). A zero index indicates that the nominal significance level was preserved at the 95% confidence level.

Significance levels performed as expected. The SKC was consistently conservative for short data records ("short" being defined as five years of record) regardless of serial

Table V.4 Nominal significance index for trend detection (years of data=05).

SIMULATION DESCRIPTION			MONTHLY			QUARTERLY									YEARLY	
						MEAN			MEDIAN			MIDDLE			MEAN	MED
Dist.	σ_r	ρ	MKD	SK	SKC	MKD	SK	SKC	MKD	SK	SKC	MKD	SK	SKC	MK	MK
Normal	1	0.0	1	0	-1	1	0	-1	1	0	-1	0	-1	-1	-1	-1
	5	0.0	1	0	-1	1	0	-1	1	0	-1	0	0	-1	-1	-1
Lognormal	1	0.0	1	0	-1	0	0	-1	0	-1	-1	0	0	-1	-1	-1
	5	0.0	1	0	-1	0	0	-1	1	0	-1	0	0	-1	-1	-1
Normal	1	0.2	2	1	-1	1	0	-1	1	0	-1	0	0	-1	-1	-1
	5	0.2	2	1	-1	1	0	-1	1	0	-1	1	0	-1	-1	-1
Lognormal	1	0.2	2	1	-1	0	-1	-1	1	0	-1	0	-1	-1	-1	-1
	5	0.2	2	2	-1	0	0	-1	1	0	-1	0	0	-1	-1	-1
Normal	1	0.4	3	2	-1	2	1	-1	1	0	-1	1	0	-1	-1	-1
	5	0.4	3	2	-1	2	0	-1	1	0	-1	1	0	-1	-1	-1
Lognormal	1	0.4	3	2	-1	2	1	-1	1	0	-1	1	0	-1	-1	-1
	5	0.4	3	2	-1	2	1	-1	2	0	-1	1	0	-1	-1	-1
Normal	1	0.6	4	3	-1	2	2	-1	2	2	-1	2	1	-1	-1	-1
	5	0.6	4	3	-1	2	2	-1	2	1	-1	2	1	-1	-1	-1
Lognormal	1	0.6	5	4	-1	3	2	-1	2	2	-1	2	1	-1	-1	-1
	5	0.6	4	3	-1	2	1	-1	2	1	-1	2	0	-1	-1	-1
Normal	1	0.8	6	5	-1	3	3	-1	3	3	-1	3	2	-1	0	-1
	5	0.8	6	5	-1	3	3	-1	3	2	-1	3	2	-1	-1	-1
Lognormal	1	0.8	6	5	0	4	3	0	3	3	-1	3	3	-1	-1	0
	5	0.8	5	5	-1	3	2	-1	3	2	-1	3	2	-1	-1	-1

Index code					
-1	$0.000 \leq \alpha < 0.032$	2	$0.100 < \alpha \leq 0.200$	5	$0.400 < \alpha \leq 0.500$
0	$0.032 \leq \alpha \leq 0.068$	3	$0.200 < \alpha \leq 0.300$	6	$0.500 < \alpha$
1	$0.068 < \alpha \leq 0.100$	4	$0.300 < \alpha \leq 0.400$		

Table V.5 Nominal significance index for trend detection (years of data=10).

SIMULATION DESCRIPTION			MONTHLY			QUARTERLY									YEARLY	
						MEAN			MEDIAN			MIDDLE			MEAN	MED
Dist.	σ_r	ρ	MKD	SK	SKC	MKD	SK	SKC	MKD	SK	SKC	MKD	SK	SKC	MK	MK
Normal	1	0.0	0	0	-1	0	0	-1	0	0	0	0	0	-1	0	0
	5	0.0	0	0	0	0	0	-1	0	0	0	0	-1	0	0	0
Lognormal	1	0.0	1	1	0	1	0	0	1	0	0	0	0	0	0	0
	5	0.0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Normal	1	0.2	1	1	0	0	0	0	0	0	-1	0	0	0	0	0
	5	0.2	1	0	-1	0	0	-1	0	0	0	0	0	0	0	0
Lognormal	1	0.2	2	2	0	0	0	0	0	0	0	0	0	-1	0	0
	5	0.2	2	2	0	0	0	0	1	0	0	0	-1	-1	0	0
Normal	1	0.4	3	2	0	2	1	0	2	1	0	0	0	0	0	0
	5	0.4	3	2	-1	2	1	-1	1	1	-1	1	0	0	0	0
Lognormal	1	0.4	3	3	0	2	1	0	2	1	0	0	0	0	0	0
	5	0.4	3	2	0	1	1	0	1	0	0	1	1	0	0	0
Normal	1	0.6	4	4	0	2	2	0	2	2	0	2	1	0	0	0
	5	0.6	4	3	0	2	2	0	2	2	0	2	1	0	0	0
Lognormal	1	0.6	4	3	0	2	2	0	2	2	0	2	1	0	0	0
	5	0.6	4	4	1	2	2	0	2	2	1	2	1	0	0	0
Normal	1	0.8	5	5	1	3	2	1	3	3	0	3	2	0	1	0
	5	0.8	6	5	2	3	3	2	3	3	1	3	3	1	1	1
Lognormal	1	0.8	6	5	2	4	3	1	4	3	1	3	3	1	2	1
	5	0.8	6	5	1	3	3	1	3	3	1	3	3	1	1	1

Index code					
-1	$0.000 \leq \alpha < 0.032$	2	$0.100 < \alpha \leq 0.200$	5	$0.400 < \alpha \leq 0.500$
0	$0.032 \leq \alpha \leq 0.068$	3	$0.200 < \alpha \leq 0.300$	6	$0.500 < \alpha$
1	$0.068 < \alpha \leq 0.100$	4	$0.300 < \alpha \leq 0.400$		

Table V.6 Nominal significance index for trend detection (years of data=15).

SIMULATION DESCRIPTION			MONTHLY			QUARTERLY									YEARLY	
						MEAN			MEDIAN			MIDDLE			MEAN	MED
Dist.	σ_r	ρ	MKD	SK	SKC	MKD	SK	SKC	MKD	SK	SKC	MKD	SK	SKC	MK	MK
Normal	1	0.0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
	5	0.0	0	0	0	1	0	0	1	0	0	0	0	0	0	0
Lognormal	1	0.0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	5	0.0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Normal	1	0.2	2	2	0	0	0	0	1	0	0	0	0	0	0	0
	5	0.2	2	1	-1	1	1	0	0	0	0	0	0	0	0	0
Lognormal	1	0.2	2	2	0	1	0	0	0	0	0	0	0	0	0	0
	5	0.2	2	2	0	1	0	0	0	0	0	0	0	0	0	0
Normal	1	0.4	2	2	0	1	1	0	1	0	0	0	0	0	0	0
	5	0.4	3	2	0	1	1	0	2	1	0	1	0	0	0	0
Lognormal	1	0.4	3	3	0	2	1	0	1	1	0	1	1	0	0	0
	5	0.4	2	2	0	1	0	0	1	1	0	0	0	0	0	0
Normal	1	0.6	4	4	0	2	2	0	2	2	0	2	1	0	0	0
	5	0.6	4	3	0	2	2	0	2	2	0	1	1	0	0	0
Lognormal	1	0.6	4	4	0	2	2	0	2	2	0	2	1	0	1	0
	5	0.6	4	4	0	2	2	0	2	2	0	2	2	0	0	0
Normal	1	0.8	5	5	1	4	3	1	3	3	1	3	2	1	1	1
	5	0.8	6	5	1	4	3	1	4	3	1	3	3	1	1	1
Lognormal	1	0.8	6	5	2	3	3	1	3	3	2	3	3	2	2	1
	5	0.8	5	5	1	3	3	1	3	3	1	3	3	1	0	1

Index code					
-1	$0.000 \leq \alpha < 0.032$	2	$0.100 < \alpha \leq 0.200$	5	$0.400 < \alpha \leq 0.500$
0	$0.032 \leq \alpha \leq 0.068$	3	$0.200 < \alpha \leq 0.300$	6	$0.500 < \alpha$
1	$0.068 < \alpha \leq 0.100$	4	$0.300 < \alpha \leq 0.400$		

correlation level. As more data became available (ten and fifteen years of record), the nominal significance level was preserved in most cases. The exception is for high serial correlation levels (ρ = 0.8). These results are similar to those of Hirsch and Slack (1984) and Taylor and Loftis (1989).

Different standard deviation ratios (σ_r) were not expected to effect the significance levels. If changes were apparent, they would have been attributed to the violation of the identically distributed test assumption. Based on inspection of Tables V.4, V.5 and V.6, no significant effect was detected. Standard deviation ratio does have a dramatic effect on power and will be discussed in the next section of this chapter.

It is interesting to note that collapsing data did not improve the nominal significance level for the SKC when the serial correlation was equal to 0.8. This result is attributed to the serial correlation in the data spaced one year apart. The SKC formulation does not account for serial correlation in data across years, only serial correlation between seasons of the same year. From Table V.3, the lag-twelve (data one year apart) serial correlation can be approximated as 0.165 when the monthly lag-one serial correlation is 0.8. The liberal performance of the SKC for ten and fifteen years of record is attributed to the serial correlation between years.

In Table V.4, the SK was occasionally conservative (four times) for quarterly collapsed data sets. The actual significance levels for these cases were just outside the range considered acceptable for significance level. These four cases were not significant at the 99% confidence level. A portion of the simulation was rerun with a new random number seed and the significance levels were preserved at the 95% confidence level for these cases. Based on these results, the slightly conservative results for the SK did not warrant further investigation.

The conservative results for the yearly collapsed data were based on the MK with just five observations (Table V.4). Due to the truncation associated with the MK test statistic and a sample size of five, it was necessary for all observations to be in rank order (e.g., all increasing or all decreasing) to detect a significant trend. In this case, the critical values for α equal to 0.05 and 0.02 were identical for the MK to detect trend. In essence, the test could have been performed at a smaller significance level than was actually specified and still yield identical results. Therefore it is expected that the actual significance levels for the MK to be conservative for a sample size of five and a 95% confidence level. The MK nominal significance level was preserved for ten and fifteen years of record, similar to the SKC.

The MKD was liberal for five years of monthly record and independent data, corresponding to previous work (Taylor and Loftis, 1989). This is attributed to the negative correlation added into the time series by deseasonalizing the data (Hirsch et al., 1982). The level of added correlation decreases with increasing sample size. The SK test preserved the nominal significance level when the data were independent regardless of sample size. When serial correlation was introduced into the data at levels as low as 0.2, both the MKD and SK were liberal for monthly data, with the MKD more liberal than the SK. The serial correlation added to the data is inversely proportional to the sample size, resulting in the observed pattern.

Collapsing the data was moderately successful. For low levels of serial correlation, actual significance was nearly returned to nominal values. A decrease in actual significance was also present for high levels of correlation, but collapsing alleviated only some of the problems. For the extreme case of collapsing from monthly to yearly, the MK was slightly liberal for the highest serial correlation levels and ten or more years of data.

Collapsing data by the middle value was more effective than the median or mean data collapse at returning the actual significance level back to the nominal significance level. This was an anticipated result since the serial correlation level in the middle collapsed data is smaller than the quarterly averaged data. The median collapse was more effective than the mean collapse as well. None of the techniques which collapse data from monthly observations to quarterly values work well for high correlation levels.

Although this study did not attempt to estimate the serial correlation level which significantly effects the performance of uncorrected tests, some inference is possible. Similar observations can be made with shorter records (Tables V.4 and V.5); however, only the longest record (Table V.6) is considered for simplicity.

1. The SK applied to monthly data is liberal for serial correlation equal to 0.2.
2. If the monthly serial correlation is 0.2 and 0.4, the serial correlation level of quarterly mean collapsed data is 0.079 and 0.198 (Table V.3), respectively. From Table V.6, the SK applied to quarterly mean data was liberal for the latter case and not the former.
3. If the monthly serial correlation is 0.4 and 0.6, the serial correlation of quarterly middle collapsed data is 0.064 and 0.216 (Table V.3), respectively. From Table V.6, the SK applied to quarterly middle collapsed data was liberal for the latter case and not the former.
4. The MK actual significance level for yearly mean collapsed data was significantly effected for yearly a serial correlation equal to 0.245 and not significantly effected for yearly a serial correlation equal to 0.092.

It appears reasonable to conclude that serial correlation levels greater than 0.2 significantly effect the actual significance levels of uncorrected trend tests and serial correlation levels below 0.1 do not significantly effect the actual significance levels. No inference is made about serial correlation in the range from 0.1 to 0.2.

C.2 Power Curve Comparison

All trend procedures are compared (two at a time) using the test described in section A.2 for power curve comparison. A comparison is made for all incremental slopes except for 0.5 since all tests were able to detect this slope with a power of one. Power curves were not compared when the detect proportions for both tests were unity.

The results for all incremental slope comparisons and common simulation conditions are placed in a single column from no slope to highest slope. In the following tables, a period indicates no statistical difference in power between the two tests. An underscore indicates that the power was unity for both techniques. A letter indicates that the power or significance level was significantly higher for the technique which corresponds to that letter.

Power comparisons (in the following tables) are useful only if the comparisons are made between tests that have actual significance levels which are insignificantly different from each other (e.g., a "period" is reported in the first column). The actual significance level comparisons made earlier in this chapter indexed tests based on ranges of actual significance levels. Only the index "0", was statistically based for the nominal significance level equal to 0.05. Therefore two actual significance levels, say

0.100 and 0.101, would be indexed differently; yet, there is no practical difference between the two values.

In the following tables, it will be common to compare two tests which were indexed differently earlier in this chapter. The comparison will be valid if p_A is insignificantly different from p_B when there was no trend in the data (e.g., no significant difference in actual significance levels). Comparisons for tests which have statistically different significance levels will not be reported. Results for the SKC and MK applied to five years of record will not be presented either since it was concluded in the previous section (Section C.1) that neither test should be used for short data records.

The first comparison is a contrast between the different methods of collapsing data. If it were necessary to collapse the data from monthly to quarterly, which of the three methods is better? The power curve results from the median and middle collapsing of monthly observations to quarterly observations are summarized in Table V.7. A comparison of the mean and middle quarterly values are similar to those shown in Table V.7 for the median and middle comparison.

In general, collapsing data by computing the quarterly median is more powerful than selecting the quarterly middle value. The same result is true when comparing quarterly mean and middle collapsed data. This result can be attributed to a reduced variance level in the median or mean

Table V.7 Power comparison for median and middle collapsing of monthly observations to quarterly observations.

SIMULATION DESCRIPTION			MKD (median) - G MKD (middle) - J			SK (median) - H SK (middle) - K			SKC (median) - I SKC (middle) - L		
			years of record			years of record			years of record		
Dist.	σ_r	ρ	5	10	15	5	10	15		10	15
Normal	1	0.0	..GGG_	..GG__	.GG___	..HHH_	..HH__	.HH___		..II__	.II___
	5	0.0	...GGG	..GG._	..GG__	...HHH	...H__	..HH__	n/c	...I._	..II__
Lognormal	1	0.0	...GG_	.GG___	.GG___	...HH_	.HH___	.HH___		.II.__	.II___
	5	0.0	...GG.	..GG__	.GG___	...HH.	..HH__	.HH___		..II__	.II___
Normal	1	0.2	...GG_	..GG__	.GG___	...HH_	..HH__			..II__	.II___
	5	0.2	...GG.	..GG._	..GG__	....HH	...H._	..HH__	n/c	.I.I._	..II__
Lognormal	1	0.2	..GGG_	.GG.__	.GG___	..HHH_	.HH.__	.HH___		..I.__	.II___
	5	0.2	...GG.		.GG.__	...HH.		.HH.__		..II._	..I___
Normal	1	0.4	.G.GG_		.GG___	...HH_		.HH___		..II__	.II___
	5	0.4	...GG.	...G._	.GG.__	...HH.	...H._	..HH__	n/c	...I._	...I__
Lognormal	1	0.4	...GG_		.GG___	..HHH_		..H___		....__	..I___
	5	0.4	...GG.	.GGG._	.GG.__	...HH_	.HHH._	.HH.__		.I.I._	....__
Normal	1	0.6	.G.G._		..G___	....H_				....__	..I___
	5	0.6	....G.	..GG._	....__	..H.HH	..HH._	..H.__	n/c	....._	....__
Lognormal	1	0.6		....__	..G___	....H_	....__			....__	...___
	5	0.6	......	...G__	....__		..HH._	....__		....._	....__
Normal	1	0.8	....._	....__	...___	....._	....__	...___		....__	...___
	5	0.8	......	....._	....__		....._	....__	n/c	....._	....__
Lognormal	1	0.8	....._	....__	...___	....._	....__	...___		....__	....__
	5	0.8	......	....._	....__	......	....._	....__		....._	....__

collapsed data sets. Trends are more easily detected in data sets with smaller variances. This comparison is somewhat weighted in favor of the median/mean collapsed data sets. Typically, the tests applied to median/mean collapsed data will have a higher actual significance level than tests applied to middle collapsed data although most differences are insignificant at the 95% confidence level. This bias is attributed to the higher correlation levels in the median/mean collapsed data than in the middle collapsed data sets for data that were originally serially correlated at the monthly level.

Table V.8 is a comparison of the mean and median collapsed data. The difference between the mean and median collapsed data is very small. There is one interesting point. While granted the number of differences is small, the technique which used the median collapsed data performed better than the mean collapsed data when the error terms were lognormally distributed. The opposite was true when the error terms were normally distributed. This implies that if one were to believe that water-quality data were normally distributed and data were to be collapsed, it is probably best to use the mean. Whereas, if one were to believe that water-quality data were lognormally distributed and data were to be collapsed, it is then best to use the median. It does not appear as though the error associated with incorrectly assuming (log)normality is large based on the simulations in this study.

Table V.8 Power comparison for mean and median collapsing of monthly observations to quarterly observations.

SIMULATION DESCRIPTION			MKD (mean) - D MKD (median) - G			SK (mean) - E SK (median) - H			SKC (mean) - F SKC (median) - I		
			years of record			years of record			years of record		
Dist.	σ_r	ρ	5	10	15	5	10	15	5	10	15
Normal	1	0.0	...D__	..D___	..D___	...E._	..E___	.EE___		..F___	..F___
	5	0.0	....D_	...D__	..D___	....E_	...E__	..E.__	n/c	...F__	..F.__
Lognormal	1	0.0	...G__	..G___	.GG___	...H__	.HH___	.HH___		..I___	.II___
	5	0.0	....G_	..G.__	..G___	....H_	..H.__	..H___		....__	..I___
Normal	1	0.2	....._	..D___	...___	...E._	...___	..E___		....__	...___
	5	0.2	......	...D__	...___	...EE_	...E__	...___	n/c	...F__	...___
Lognormal	1	0.2	...G._	..G___	.G.___	...H._	..H___	.H.___		...___	.I.___
	5	0.2	....._	...G__	..G___	....H_	...H__	..H___		...I__	..I___
Normal	1	0.4	....._	....__	...___	....E_	....__	..E___		....__	...___
	5	0.4	......	....__	....__	....E.	....__	....__	n/c	....._	....__
Lognormal	1	0.4	....._	...___	..G___	....._	...___	...___		....__	...___
	5	0.4	....._	....__	...___	......	....__	...___		....__	....__
Normal	1	0.6	....._	....__	...___	..EE._	....__	...___		....__	...___
	5	0.6	......	....._	....__	......	....._	....__	n/c	....._	....__
Lognormal	1	0.6	....._	....__	...___	....._	....__	...___		....__	...___
	5	0.6	.D....	....._	....__	......	....._	....__		....._	....__
Normal	1	0.8	....._	....__	...___	....._	....__	...___		....__	...___
	5	0.8	......	....._	....__	......	....._	....__	n/c	....._	....__
Lognormal	1	0.8	....._	....__	...___	....._	....__	...___		....__	....__
	5	0.8	......	....._	....__	......	....._	....__		....._	....__

This result is interesting in light of the nominal significance level discussion and the level of correlation in collapsed data sets. Tables V.4 through V.6 show that the tests which use middle collapsed data tend to preserve the nominal significance levels better than tests which use the median/mean collapsed data. This is anticipated since the serial correlation level is lower for middle collapsed data than it is for mean collapsed data (Table V.3). Statistically testing (with Equation V.5) the difference between significance levels demonstrates that there is little gain (within the simulation constraints) in using the middle collapsed data.

The next comparison contrasts the performance of the SKC under conditions of collapsing data (Table V.9). Middle quarterly collapsed values were not included, since it was determined from the first comparison that this technique did not perform well in comparison to the other two methods for collapsing data. From inspection of Table V.9, it is clear that data should not be collapsed when using the SKC for trend detection.

Yearly averaged time series trend detection is compared in Table V.10. Should one ever attempt to detect trends by collapsing data from either monthly to yearly or quarterly to yearly? In the previous comparison, it was shown that data should not be collapsed when using the SKC.

This comparison (Table V.10) demonstrates that the SKC for quarterly collapsed data is more powerful than the MK

Table V.9 Power comparison for mean and median collapsing of monthly observations to quarterly observations.

SIMULATION DESCRIPTION			SKC (monthly) - C SKC (mean) - F			SKC (monthly) - C SKC (median) - I		
			years of record			years of record		
Dist.	σ_r	ρ	5	10	15	5	10	15
Normal	1	0.0		...___	...___		..C___	..C___
	5	0.0	n/c	....__	...___	n/c	...C__	..C.__
Lognormal	1	0.0		.CC___	.CC___		.CC___	.C.___
	5	0.0		.CCC__	.CC___		..CC__	.CC___
Normal	1	0.2		...___	...___		....__	..C___
	5	0.2	n/c	....__	...___	n/c	...C__	..C___
Lognormal	1	0.2		..C___	.CC___		..C___	.CC___
	5	0.2		..CC__	.CC___		...C__	.CC___
Normal	1	0.4		....__	...___		...C--	..C___
	5	0.4	n/c	....._	....__	n/c	...C.-	...C__
Lognormal	1	0.4		..C.__	..C___		....__	..C___
	5	0.4		...C__	..C___		...C__	....__
Normal	1	0.6		....__	...___		...C__	...___
	5	0.6	n/c	....C_	....__	n/c	....C_	....__
Lognormal	1	0.6		...C__	..C___		....__	...___
	5	0.6		...C._	..CC__		...C._	...C__
Normal	1	0.8		....__	...___		....__	...___
	5	0.8	n/c	....._	....__	n/c	....._	....__
Lognormal	1	0.8		....._	....__		....._	....__
	5	0.8		....._	....__		....._	....__

for yearly collapsed data. By induction, this comparison also demonstrates that the SKC for monthly data would be more powerful than the MK for yearly collapsed data as well. Similarly, the SKC is the least powerful option when compared to the MKD or SK. Since the SKC applied to quarterly collapsed data is more powerful than the MK applied to yearly collapsed data, the MKD or SK (for monthly or

Table V.10 Power comparison for mean and median collapsing of quarterly observations (SKC) to yearly observations (MK).

SIMULATION DESCRIPTION			SKC (mean) - F MK (mean) - M			SKC (median) - I MK (median) - N		
			years of record			years of record		
Dist.	σ_r	ρ	5	10	15	5	10	15
Normal	1	0.0		...___	...___		...I__	...___
	5	0.0	n/c	..FFF_	.FFF__	n/c	...I._	..II__
Lognormal	1	0.0		..F___	.FF___		..I___	.II___
	5	0.0		..FFF_	.FFF__		..II__	.II___
Normal	1	0.2		..M.__	...___		.N..__	...___
	5	0.2	n/c	...FF_	.FFF__	n/c	...II_	..II__
Lognormal	1	0.2		..FF__	..F___		....__	.II___
	5	0.2		..FFF_	.FFF__		..II._	.II.__
Normal	1	0.4		....__	...___		....__	...___
	5	0.4	n/c	...FF_	..FF__	n/c	...II_	..II__
Lognormal	1	0.4		...F__	.FF___		....__	..I___
	5	0.4		..FFF_	.FFF__		...II_	.III__
Normal	1	0.6		....__	...___		....__	..I___
	5	0.6	n/c	...FF_	.FFF__	n/c	...II_	...I__
Lognormal	1	0.6		...F__	..F___		....__	..I___
	5	0.6		...FF_	..FF._		...II_	..II__
Normal	1	0.8		....__	....__		...I._	....__
	5	0.8	n/c	...FF_	..FFF_	n/c	...II_	...I._
Lognormal	1	0.8		...F__	....__		...I__	....__
	5	0.8		...FF_	..FFF_		...II_	..II__

quarterly collapsed data) are more powerful than the MK applied to yearly data as well.

It is clear that for data records with at least ten years of record, yearly values should not be used when either quarterly or monthly observations are available. Data that are suspected to be serially correlated can be analyzed more powerfully with the SKC than by collapsing data to yearly values. It is not powerful to collapse data that are not serially correlated to yearly values either. Both of the tests (SKC applied to quarterly collapsed data and MK applied to yearly averaged data) compared in Table V.10 are conservative (low actual significance levels) for five years of record and should be avoided. The significance level of the MK applied to five observations can be determined exactly. It is not possible to perform a 95% two-sided trend test that will preserve the nominal significance level with only five observations. This is due the truncation of the MK test statistic discussed earlier.

The final comparison requires some level of judgement. The key trade-off is between power and preserving the significance level. When there is no serial correlation in the data record, tests which do not correct for serial correlation are more powerful as expected. Uncorrected tests do not preserve the nominal significance level in the presence of serial correlation. Figures V.1 through V.6 are a comparison of the MKD, SK and SKC applied to monthly data and the MKD and SK applied to quarterly median collapsed data. Although it has been demonstrated that the SKC should not be applied to data sets with five years of record, it was opted to maintain the power curves in Figures V.1, V.3 and V.5 for comparison purposes. Distribution and variance seasonality were held constant throughout all of the figures. In general, the same relative test performance existed for other distributions and variance patterns. Fifteen years of

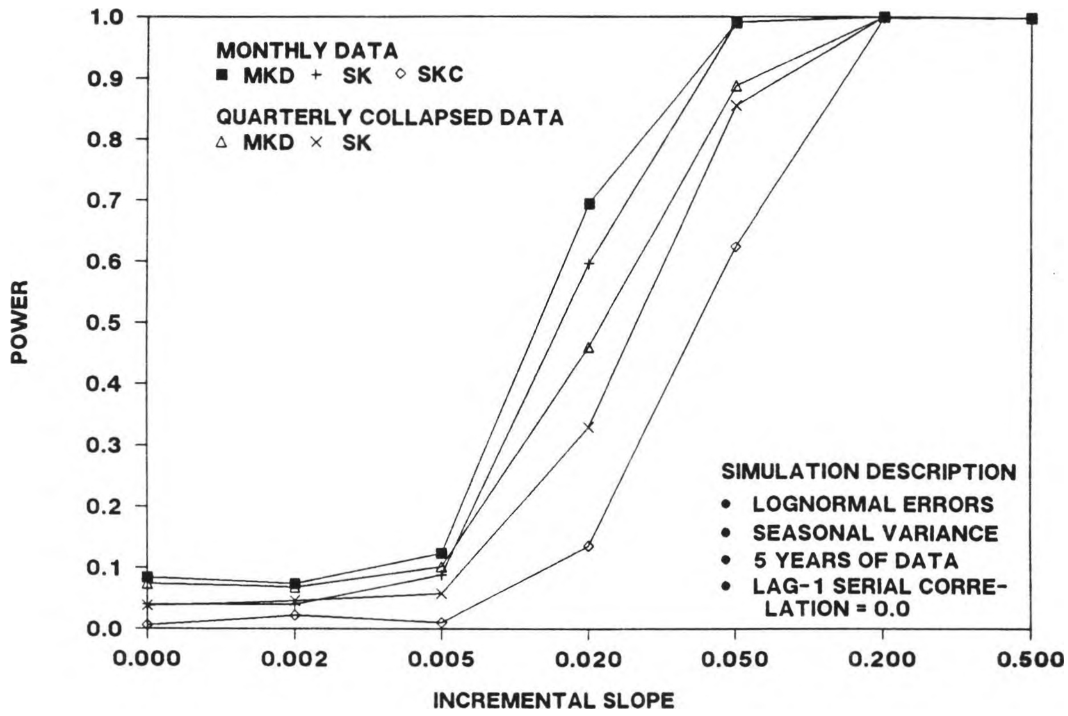

Figure V.1 Power curve comparison for monthly and quarterly (median) averaged data for five years of data and lag-one serial correlation equal to 0.0.

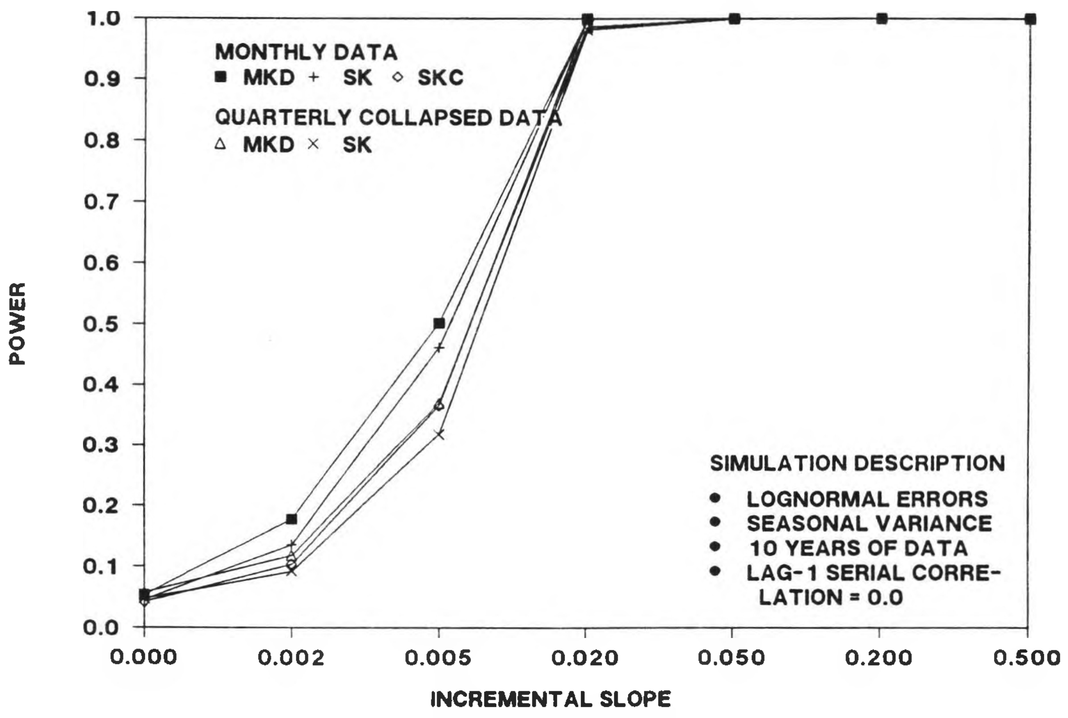

Figure V.2 Power curve comparison for monthly and quarterly (median) averaged data for ten years of data and lag-one serial correlation equal to 0.0.

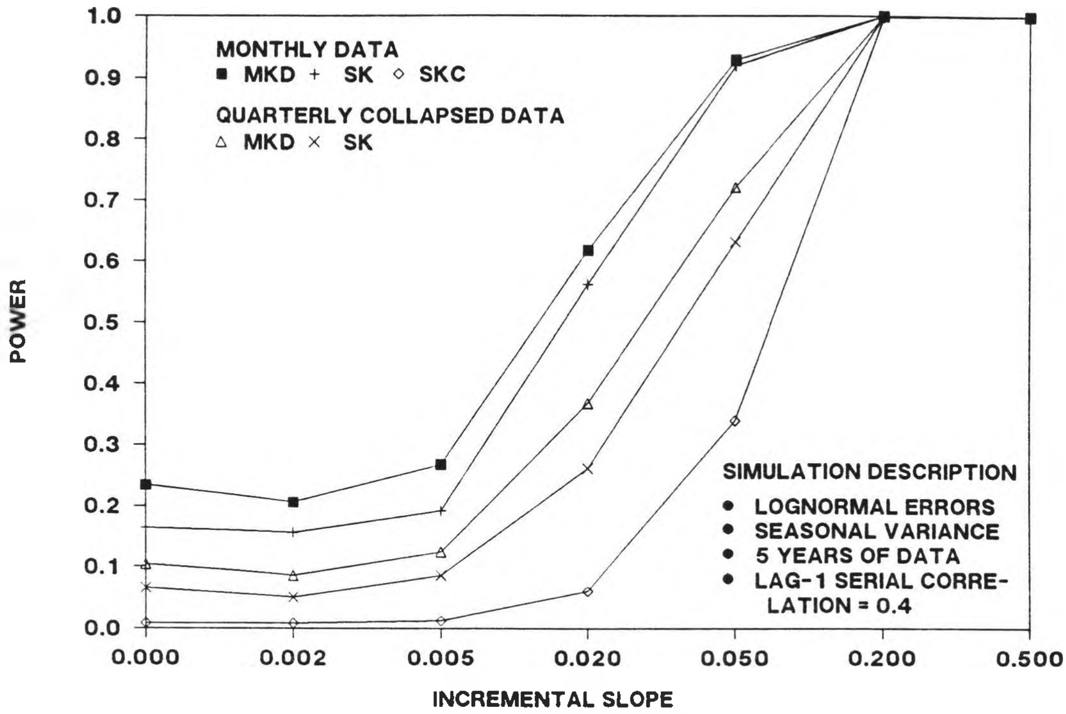

Figure V.3 Power curve comparison for monthly and quarterly (median) averaged data for five years of data and lag-one serial correlation equal to 0.4.

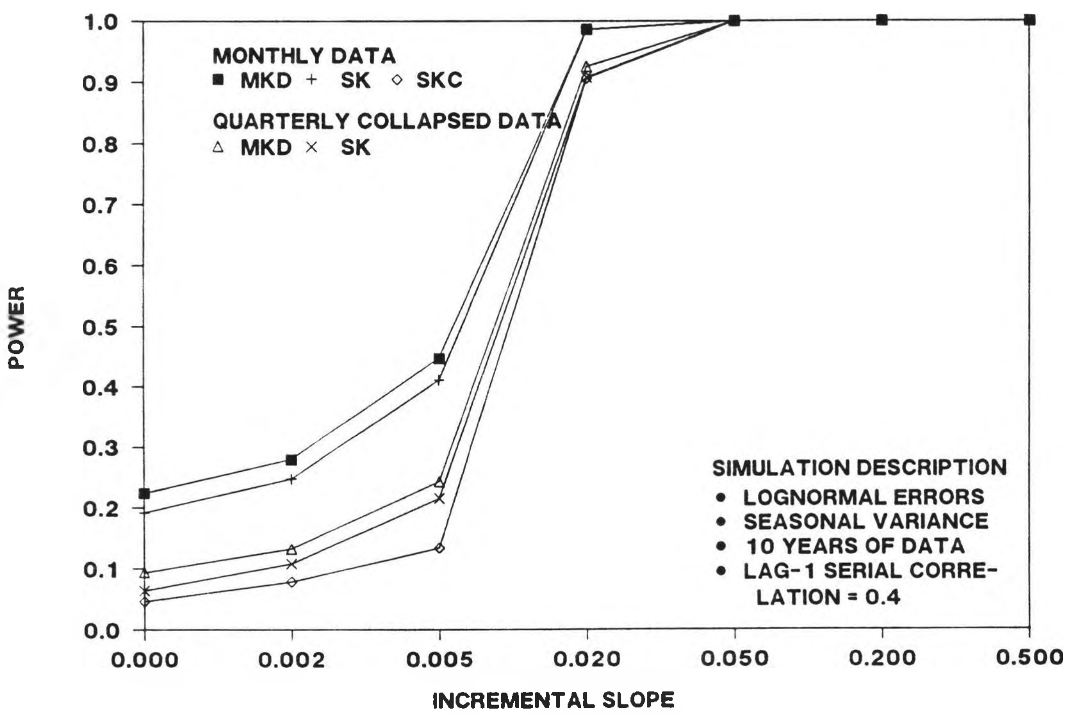

Figure V.4 Power curve comparison for monthly and quarterly (median) averaged data for ten years of data and lag-one serial correlation equal to 0.4.

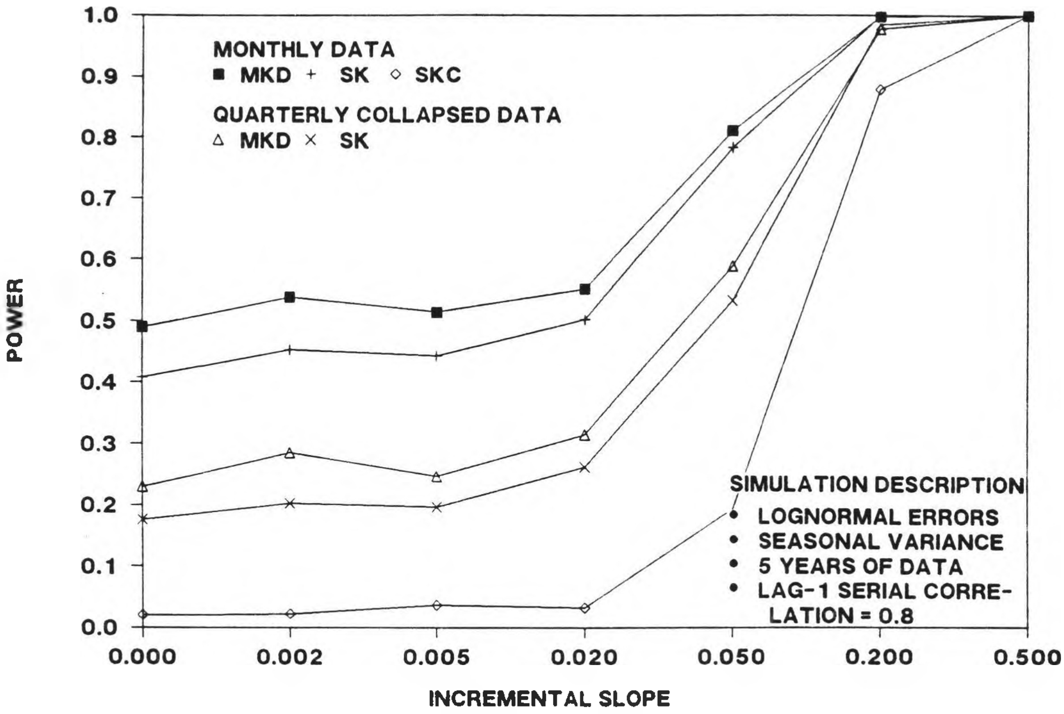

Figure V.5 Power curve comparison for monthly and quarterly (median) averaged data for five years of data and lag-one serial correlation equal to 0.8.

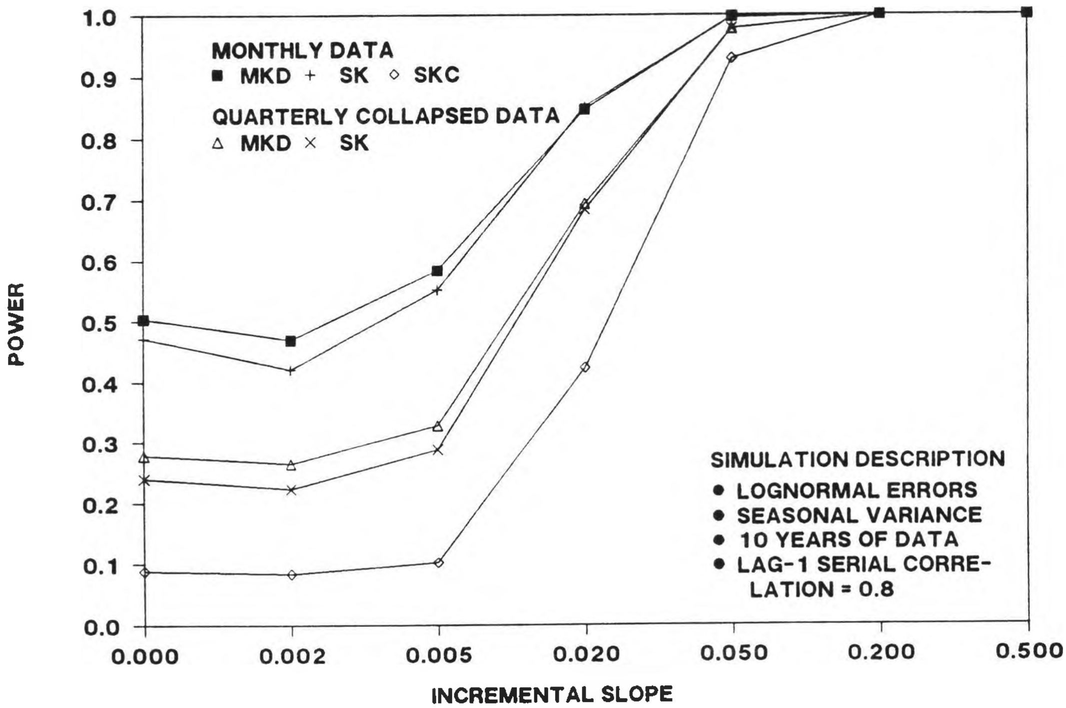

Figure V.6 Power curve comparison for monthly and quarterly (median) averaged data for ten years of data and lag-one serial correlation equal to 0.8.

record are not shown since the conclusions for ten years of record apply equally to fifteen years of record.

When no serial correlation is present, the SK and MKD are the most appropriate tests for short (Figure V.1) and long (Figure V.2) data records, respectively. The SK is more appropriate than the MKD for five years of record since the MKD is liberal for short records. The liberal behavior of the MKD is reduced for ten years of record and the MKD is more powerful than the SK. As the record length increases, the performance difference between the least powerful test and the most powerful test decreases.

For non-zero serial correlation values (Figures V.3-V.6), the MKD and SK do not preserve the significance level. Procedures which collapse data sets from monthly to quarterly values are marginally successful at returning the nominal significance level (Figures V.3 and V.4). The SKC is too conservative and does not have a reasonable power level for use with short data records (Figure V.3). This problem is eliminated for longer records. Figures V.5 and V.6 are the extreme example of the highest serial correlation simulated in this study. In this case, none of the tests perform well.

D. Summary and Conclusions

One of the primary objectives of a monitoring network is to detect changing conditions. The two types of trends that a network manager is most concerned with are:

Monotonic and step trends. Trend tests are a statistical means for detecting trend. Most water-quality literature has recommended the use of nonparametric tests over parametric tests due to the non-normality and potential for outliers encountered. Nonparametric tests, however, also have associated assumptions. The key assumptions include independence and homogeneous variance. Violation of both assumptions have been noted in the literature.

A simulation study was used to compare different non-parametric procedures for detecting monotonic trends in simulated data records. Procedures were evaluated based on their robustness and power. All conclusions summarized below are based on the simulations in this chapter and extrapolation to conditions not covered herein is not suggested.

D.1 Significance Levels

1. The SKC requires more than five years of data in order to preserve the nominal significance level, otherwise, the actual significance level is lower than the nominal significance level (e.g., conservative). For longer records (10 and 15 years), the nominal significance level is preserved for all but the highest serial correlation levels (ρ = 0.8).
2. The actual significance level is higher than the nominal significance level (e.g., liberal) for the MKD with five years of record and no serial correlation. For longer records (10 and 15 years) and no serial correlation, the

nominal significance level was preserved. This high significance level is the result of the negative correlation added to the data by the deseasonalizing process. For similar simulation conditions, the SK will generally have more exact significance levels than the MKD.

3. Serial correlation levels larger than 0.2 can significantly effect the actual significance level of uncorrected tests while serial correlation levels below 0.10 do not significantly effect actual significance levels. No conclusions regarding levels of correlation between 0.1 and 0.2 were possible with the simulations in this chapter.

D.2 Power

1. If it is necessary to collapse data, collapsing the data by computing the median or mean is better than using the middle value.
2. If it is necessary to collapse data, collapsing the data by computing the mean or median are appropriate for normal and lognormally distributed errors, respectively. Based on the simulations in this study the error associated with incorrectly assuming (log)normally distributed errors is small.
3. Do not collapse the data if using the SKC.
4. Do not collapse the data to yearly values if it is possible to analyze more frequently sampled observations.

5. The MKD is slightly more powerful than the SK for 10 and 15 years of record; however, the MKD has significantly higher actual significance levels than the SK for five years of record and a power comparison is not appropriate.
6. The SKC applied to monthly data is more appropriate than the MKD or SK applied to quarterly collapsed data if serial correlation is a concern and there are enough years of record (e.g., at least ten years). When the data are truly independent, the MKD or SK are more powerful than the SKC. However, the difference in power appears to decrease as the record length increases.

E. Application

The conclusions in the previous section give some guidance to which statistical procedure is more appropriate in certain instances. There remains, however, a fuzzy boundary between significance and power. The simulations performed in this chapter are somewhat crude in the sense that only three lengths of record were used. As a result, it is difficult to pin down the exact point where one should select another test procedure based on record length. In addition, serial correlation levels are difficult to estimate accurately with small data sets. For example, if the serial correlation were estimated as 0.2 from a data set with 87 observations, the null hypothesis of no serial correlation would be accepted at the 95% confidence level (Harris, 1988).

From a data analysis standpoint, most analysts set the nominal significance level (Type I error) and hope the test they use produces a small Type II error. However, due to the nature of water-quality data, one cannot assume that the Type I error (or α) is preserved. The manager must, as a result, decide in these instances which is more important. If early trend detection is a must, the network manager may consider using a test which may be liberal in terms of significance but more powerful than a slightly conservative test. On the other hand, declaration that clean-up operations are effective when, in fact, the detected trends are actually the result of serial correlation can be misleading. Since serial correlation can result in "trends" (periods of increasing or decreasing consistently), it may be important to detect significant trends regardless of the culprit.

In light of the this discussion, the following recommendations are made based on the findings of this chapter assuming one wishes to preserve the nominal significance level and still use a powerful test. (No recommendations can be made about records between six and nine years of record since the simulations in this chapter considered only five, ten and fifteen years of record. It is suggested that further simulations be done to further refine the following recommendations.)

1. Use the SK test on the original data (e.g., do not collapse the data) when there are five years of record and no serial correlation.
2. Use the SK test on mean or median collapsed data when there are five (or less) years of record and serial correlation. This is not an ideal solution. In fact, when monotonic trend is detected in data sets with only five years of record, it should be noted that the detected trend can be due to serial correlation. The mean and median collapsed data sets are more appropriate for normally and lognormally distributed data, respectively. Based on the simulations in this study, the reduction in power for incorrectly choosing (log)normality is small.
3. Use the MKD test on the original data when there are ten or more years of record and no serial correlation. Based on the simulation studies in this chapter, the MKD is more powerful than the SK. This result is confirmed by van Belle and Hughes (1984) in a comparison of aligned and intrablock tests using asymptotic relative efficiencies. However, the MKD is liberal and should not be used for shorter data records.
4. Use the SKC test on the original data when there are ten or more years of record and serial correlation. The MKD and the SK are liberal for modest serial correlation levels. Techniques designed to reduce serial correlation (quarterly averaging) were not entirely successful

and as a result are not recommended when there are enough data to use the SKC.

This chapter does not address the issue of serial correlation detection. A study is currently planned to consider the sequential testing for serial correlation and trend detection using the above rules as a decision base.

CHAPTER VI

CHANGING CONDITIONS--DETECTION WITH MISSING DATA

Data may be collapsed from frequent to less frequent observations (e.g., monthly observations to quarterly values) to avoid problems associated with serial correlation and computational problems with missing values. The former problem is addressed in Chapter V. The latter issue is addressed here. How many missing values are needed in a data set before it is necessary to collapse the data?

A. Simulation Methodology

A.1 Data Generation

Data are generated for the missing value simulation using the same procedure described in Section V.A.1. After the water-quality time series is generated, a set of uniform (0,1) random numbers are generated corresponding to each "observation" of the water-quality time series. If the uniform random number is below the selected missing level percentile, the corresponding value in the water-quality time series is recoded as a missing value.

A.2 Evaluation

The same evaluation procedure described in Section V.A.2 is used here.

B. Analyses

B.1 Simulation Conditions

Five hundred sets of monthly data were generated for each set of simulation conditions. The parameters considered in these simulations are summarized in Table VI.1.

Table VI.1 Levels of distribution, variance seasonality, serial correlation, sample size, and levels of missing values or replicated data.

Distribution	Normal, Lognormal
Ratio of seasonal standard deviations	1.0, 5.0
Monthly lag-1 serial correlation	0.0, 0.4
Years of data	5, 10, 15
Percentage of missing data	0.0, 0.1, 0.3, 0.5 0.6, 0.7
Percentage of replicated data	0.0, 0.1, 0.2

B.2 Trend Detection Procedures

For the simulations in this chapter, monthly data were tested for trend with the MKD, SK and SKC. The same three tests were applied to time series which were constructed from quarterly medians. Quarterly means were not analyzed since the results from Chapter V indicate that there are only slight differences between the two procedures. Quarterly middle collapsed data or yearly summaries were not applied either since the results from Chapter V indicate that there is always a better approach. In this chapter, quarterly averaged data refers exclusively to data that were created by computing the quarterly median. The procedures used here are summarized in Table VI.2. With increasing

Table VI.2 Summary of statistical procedures used for trend detection study.

Code	Technique
Tests applied to monthly data:	
A	Mann-Kendall tau test is applied to data which have been deseasonalized for the mean and variance
B	Seasonal Kendall tau test
C	Seasonal Kendall tau test with serial correlation correction
Tests applied to quarterly values (monthly data were collapsed by computing quarterly medians):	
G	Mann-Kendall tau test is applied to data which have been deseasonalized for the mean and variance
H	Seasonal Kendall tau test
I	Seasonal Kendall tau test with serial correlation correction

levels of missing values, it was possible to estimate a negative or very small test statistic variance for the SKC (Equation 5 in Hirsch and Slack, 1984). When the test statistic variance estimation is negative, no trend was assumed. For the simulations in this chapter this condition was rarely achieved and did not warrant further investigation.

C. Results--Missing Data

Power tables resulting from the above simulation are presented in Appendix D. Techniques are now compared based on their actual significance level and power.

C.1 Actual Significance Level Comparison

Actual significance levels were indexed and tabulated similar to results in Chapter V (Tables VI.3 through VI.5). The significance levels for five years of data were not presented since the significance levels were conservative for all SKC simulations (Table VI.3). The performance of the MKD and SK and no missing values was the same as the performance in Chapter V. Recall that both the MKD and SK were liberal in the presence of serial correlation and the MKD was liberal for five years of record with uncorrelated data.

As the percentage of missing values increased from zero to 70%, the actual significance level decreased for the MKD and SK by 0.10 to 0.15 for monthly correlated data. Monthly uncorrelated data yielded decreases of the significance level up to 0.026. Decreases in actual significance levels for the MKD and SK with quarterly collapsed data were less than 0.02. This result was expected since the quarterly collapsed data sets have relatively few missing values. For short data records (five years), the SK was conservative when 60% and 70% of the data were missing for uncorrelated and correlated data, respectively. The nominal significance level of the MKD for monthly data was preserved for 70%

Table VI.3 Nominal significance index for trend detection (years of data=05).

SIMULATION DESC.				MONTHLY			QUART.				MONTHLY			QUART.		
Dist.	σ_r	p	ρ	MKD	SK	SKC	MKD	SK	SKC	ρ	MKD	SK	SKC	MKD	SK	SKC
Nor.	1	0.0	0.0	1	0		1	0		0.4	2	2		1	0	
	5			1	0		1	0			3	2		1	0	
Log.	1			1	0		0	-1			3	2		1	0	
	5			1	0		1	0			3	2		1	0	
Nor.	1	0.1	0.0	1	0		1	0		0.4	2	2		1	0	
	5			1	0		0	0			3	2		1	0	
Log.	1			1	0		0	0			3	2		1	0	
	5			0	0		0	-1			3	2		1	0	
Nor.	1	0.3	0.0	0	0		0	0		0.4	2	1		1	0	
	5			1	0		0	0			2	2		1	0	
Log.	1			1	0		0	0			2	1		1	-1	
	5			0	-1		0	-1			2	2		2	0	
Nor.	1	0.5	0.0	1	0		0	0		0.4	2	0		1	0	
	5			1	0		1	-1			2	1		1	0	
Log.	1			1	-1		1	-1			2	0		1	0	
	5			0	0		0	0			2	0		1	0	
Nor.	1	0.6	0.0	1	-1		0	0		0.4	2	0		1	0	
	5			1	0		0	-1			2	0		1	0	
Log.	1			1	-1		1	0			2	0		1	0	
	5			0	-1		0	-1			2	0		1	0	
Nor.	1	0.7	0.0	0	-1		1	-1		0.4	1	-1		1	-1	
	5			1	-1		0	-1			2	0		1	0	
Log.	1			0	-1		0	-1			1	-1		0	-1	
	5			0	-1		0	-1			1	-1		1	-1	

Index code			
-1	$0.000 \le \alpha < 0.032$	3	$0.200 < \alpha \le 0.300$
0	$0.032 \le \alpha \le 0.068$	4	$0.300 < \alpha \le 0.400$
1	$0.068 < \alpha \le 0.100$	5	$0.400 < \alpha \le 0.500$
2	$0.100 < \alpha \le 0.200$	6	$0.500 < \alpha$

missing values. The MKD was not conservative for any simulation conditions. Since the SK test went from preserved to conservative, it is suspected that it is coincidental with the simulation parameters that some of

Table VI.4 Nominal significance index for trend detection (years of data=10).

SIMULATION DESC.				MONTHLY			QUART.				MONTHLY			QUART.		
Dist.	σ_r	p	ρ	MKD	SK	SKC	MKD	SK	SKC	ρ	MKD	SK	SKC	MKD	SK	SKC
Nor.	1	0.0	0.0	0	0	-1	0	0	0	0.4	2	2	0	0	0	-1
	5			0	0	0	0	0	0		2	2	0	1	0	0
Log.	1			1	1	0	1	0	0		3	2	0	1	0	0
	5			0	0	0	0	0	0		3	2	0	1	0	0
Nor.	1	0.1	0.0	0	0	-1	0	-1	-1	0.4	2	2	0	0	0	-1
	5			0	0	0	0	0	0		2	2	0	0	0	0
Log.	1			0	0	0	1	0	0		3	2	0	1	1	0
	5			0	0	0	0	0	0		3	2	0	1	1	0
Nor.	1	0.3	0.0	0	0	0	0	0	-1	0.4	2	1	0	0	0	-1
	5			0	0	0	0	0	0		2	1	0	1	0	-1
Log.	1			1	0	0	1	0	0		2	2	0	1	0	0
	5			1	0	0	0	0	0		2	2	0	2	1	0
Nor.	1	0.5	0.0	0	0	0	0	-1	-1	0.4	2	0	0	0	0	-1
	5			0	0	-1	0	0	-1		2	1	-1	1	0	-1
Log.	1			1	0	0	0	0	-1		2	2	0	1	1	0
	5			1	0	0	0	0	0		2	2	0	2	1	0
Nor.	1	0.6	0.0	0	0	-1	1	0	0	0.4	1	0	-1	0	0	-1
	5			0	0	0	0	0	-1		2	0	0	1	0	0
Log.	1			1	0	-1	0	0	-1		2	1	0	1	0	-1
	5			0	0	0	0	0	0		2	1	0	1	0	-1
Nor.	1	0.7	0.0	0	0	-1	0	0	0	0.4	0	0	-1	0	-1	-1
	5			0	-1	-1	0	0	0		2	0	0	1	0	0
Log.	1			0	0	-1	0	0	-1		2	0	0	1	0	0
	5			0	0	-1	0	0	-1		1	0	-1	1	0	0

Index code	-1	$0.000 \leq \alpha < 0.032$	3	$0.200 < \alpha \leq 0.300$
	0	$0.032 \leq \alpha \leq 0.068$	4	$0.300 < \alpha \leq 0.400$
	1	$0.068 < \alpha \leq 0.100$	5	$0.400 < \alpha \leq 0.500$
	2	$0.100 < \alpha \leq 0.200$	6	$0.500 < \alpha$

these tests preserved the nominal significance level for certain missing value levels, and the test did not necessarily reach an asymptotic significance level equal to the nominal significance level.

Table VI.5 Nominal significance index for trend detection (years of data=15).

SIMULATION DESC.				MONTHLY			QUART.				MONTHLY			QUART.		
Dist.	σ_r	p	ρ	MKD	SK	SKC	MKD	SK	SKC	ρ	MKD	SK	SKC	MKD	SK	SKC
Nor.	1	0.0	0.0	0	0	0	0	0	0	0.4	3	2	0	1	1	0
	5			0	0	0	1	0	0		2	2	0	1	1	0
Log.	1			0	0	0	0	0	0		3	3	0	1	1	0
	5			0	0	0	0	0	0		3	3	0	1	1	0
Nor.	1	0.1	0.0	0	0	0	0	0	-1	0.4	3	2	0	2	1	0
	5			0	0	0	0	0	0		2	2	0	1	1	0
Log.	1			0	0	0	0	0	0		3	2	0	2	1	0
	5			0	0	-1	0	0	-1		3	2	0	1	0	0
Nor.	1	0.3	0.0	0	0	0	0	0	0	0.4	2	2	0	1	1	0
	5			0	0	0	0	0	0		2	2	0	1	1	0
Log.	1			0	0	0	0	0	0		2	2	0	1	1	0
	5			0	0	0	0	0	0		2	2	0	1	1	0
Nor.	1	0.5	0.0	0	0	0	0	0	-1	0.4	2	2	0	1	0	0
	5			0	0	0	0	0	0		2	1	0	1	1	0
Log.	1			0	0	0	0	0	0		2	1	0	1	1	0
	5			1	-1	-1	0	0	0		2	1	0	1	0	0
Nor.	1	0.6	0.0	1	0	0	0	0	0	0.4	2	1	0	1	1	0
	5			0	0	0	0	0	-1		2	1	0	1	1	0
Log.	1			0	0	0	0	0	0		2	1	0	1	1	0
	5			0	0	0	0	-1	-1		1	1	0	0	0	0
Nor.	1	0.7	0.0	0	0	0	0	0	0	0.4	2	1	0	1	1	0
	5			0	0	0	0	0	0		1	0	0	0	0	0
Log.	1			0	0	0	0	0	0		1	0	0	1	0	0
	5			0	0	0	0	0	0		1	1	0	0	0	-1

Index code			
-1	$0.000 \leq \alpha < 0.032$	3	$0.200 < \alpha \leq 0.300$
0	$0.032 \leq \alpha \leq 0.068$	4	$0.300 < \alpha \leq 0.400$
1	$0.068 < \alpha \leq 0.100$	5	$0.400 < \alpha \leq 0.500$
2	$0.100 < \alpha \leq 0.200$	6	$0.500 < \alpha$

At ten years of record, the SKC was conservative for larger levels of missing values although the pattern was somewhat erratic. The SKC preserved the nominal significance level at 15 years of record and all levels of missing values. There were no appreciable decreases in actual significance levels for the SKC with increasing missing values levels. The actual significance levels were similar to those reported by Hirsch and Slack (1984).

C.2 Power Curve Comparison

The power curves developed for this simulation are similar to those developed in Chapter V. As expected, power decreased with increasing missing value levels.

In Table VI.6, the MKD and SK for monthly data are more powerful than the MKD and SK for quarterly collapsed data up to 50% missing values and uncorrelated data. The converse is true for greater than 50% missing values. A comparison of the serially correlated data is not meaningful since the actual significance levels of the monthly tests are significantly higher than the actual significance levels of the quarterly tests.

The SKC is more affected by the level of missing values than the MKD and SK. In Table VI.7, the SKC for quarterly collapsed data is more powerful than the SKC for monthly data once 50% percent of the data are missing. The two tests compared in Table VI.7, are equivalent for serially correlated data at the 30% missing data level; although the monthly test is more powerful for uncorrelated data. It would be reasonable to average monthly data to quarterly observations if 40% of the data are missing.

The MKD is more powerful than the SK for all levels of missing values and monthly data (Table VI.8) or quarterly

Table VI.6 Power comparison for the MKD and SK test for monthly and quarterly collapsed data.

SIMULATION DESCRIPTION			A MKD (monthly) G MKD (quarterly)			B SK (monthly) H SK (quarterly)		
			(ρ = 0.0)			(ρ = 0.0)		
			years of record			years of record		
Dist.	σ_r	p	05	10	15	05	10	15
Nor.	1	0.0	...A_	..A___	.AA___	...B._	..B___	.BB___
	5		...AA_	..AA__	..A___	....B_	..BB__	..B.__
Log.	1		..AA__	.AA___	.A____		...B__	.B____
	5		...AA_	.AAA__	.AA___	...BB_	.BBB__	.BB___
Nor.	1	0.1	...A__	..A___	..A___	...B._	..B___	..B___
	5		...AA_	...A__	..A.__	....B_	...B__	..B___
Log.	1		.AAA__	.AA___	.A.___	...B__	..B___	.B.___
	5		..AAA_	.AAA__	.AA___	...BB_	.BBB__	.BB___
Nor.	1	0.3	....._	....__	..A___	....._	....__	...___
	5			...A__	..A___		...B__	....__
Log.	1		..AA._	.AA___	.AA___	...B._	..B___	.BB___
	5		...AA_	..AA__	.AA___	...BB_	..BB__	.BB___
Nor.	1	0.5	....._	..A.__	...___	....H_	....__	...___
	5		......	...A._	...A__	.....H	....._	....__
Log.	1		....._	..A___	.AA___	....._	..B___	..B___
	5		......	..AA._	.AA___	......	....._	..B.__
Nor.	1	0.6	...GG.	..A.__	...___	...HHH	...H__	..H___
	5		.....G	....G_	....__	....HH	....H_	....__
Log.	1		..A.GG	....__	.AA___	....HH	....__	...___
	5		....GG	....._	..A.__	....HH	....._	..B.__
Nor.	1	0.7	...GGG	...G._	...___	.H.HHH	.H.H._	..H.__
	5		....GG	....G_	....._	....HH	...HH_	...HH_
Log.	1		..AGGG	...G__	...___		...H._	...___
	5		...GGG	...GG_	...G__	...HHH	...HH.	...H._

data (Table VI.9). Five years of data are not compared since the actual significance levels of the MKD is significantly greater than the actual significance level of the SK. The MKD should be used when there is no serial

Table VI.7 Power comparison for the SKC test for monthly and quarterly collapsed data.

SIMULATION DESCRIPTION			C - SKC (monthly data) I - SKC (quarterly median data)					
			(ρ = 0.0)			(ρ = 0.4)		
			years of record			years of record		
Dist.	σ_r	p	05	10	15	05	10	15
Nor.	1	0.0	....._	..C___	..C___	....._	...C__	...___
	5			...C__	..C.__	C	...C._	....__
Log.	1		...CC_	.CC___	.C.___		..C.__	..C___
	5		...CC.	..CC__	.CC___	...CCC	...C__	...___
Nor.	1	0.1	....I_	....__	..C___	....._		...___
	5		CC	...C__	...___	.C...C	....._	....__
Log.	1		...C._	.CC___	.CC___	.C..C_	..C.__	..C___
	5		...CC.	..CC__	.CC___	...C.C	...C__	...___
Nor.	1	0.3	....I.	....__	...___	....I_	....__	...___
	5		I	....__	....__	......	....._	....__
Log.	1			..C___	.CC___	....._	....__	..C___
	5		.C.C..	...C__	.CC___	......	....__	...___
Nor.	1	0.5		...I__	...___	II	....__	...___
	5			....._	....__	I	...II_	....__
Log.	1		II	.C..__	...___	II	....__	...___
	5		II	....._	..C.__	I	...I._	....__
Nor.	1	0.6	II	...I__	..I.__	II	..II__	....__
	5		II	...II_	...I._	.C...I	...II_	...I._
Log.	1		II	...I__	...___	...III	...I__	...___
	5		II	...II_	...I._	II	...II_	....__
Nor.	1	0.7	.I..II	...II_	..I.__	II	...II_	...I__
	5		I	...III	...II_	I	....I.	...II_
Log.	1		II	...I._	..I___	...III	..II._	..I.__
	5		II	..III.	...I._	.I..II	...II.	..II._

correlation and at least ten years of record since it is the most powerful and preserves the nominal significance level. For five years of record, the MKD is liberal in comparison

Table VI.8 Power comparison for MKD and SK with monthly observations.

SIMULATION DESCRIPTION			A - MKD (monthly data) B - SK (monthly data)					
			(ρ = 0.0)			(ρ = 0.4)		
			years of record			years of record		
Dist.	σ_r	p	05	10	15	05	10	15
Nor.	1	0.0		..A___	...___			...___
	5	0.0	n/c	....__	..A___	n/c	....__	...___
Log.	1	0.0		...___	..____		...___	...___
	5	0.0		...___	...___		....__	...___
Nor.	1	0.1		..A___	..A___			...___
	5	0.1	n/c	..A.__	..A___	n/c	.AA.__	...___
Log.	1	0.1		.AA___	..____			...___
	5	0.1		...___	...___		....__	...___
Nor.	1	0.3		..A.__	..A___			...___
	5	0.3	n/c	...A__	...___	n/c		..A.__
Log.	1	0.3		..A___	...___		..A___	...___
	5	0.3		..A.__	..A___			...___
Nor.	1	0.5		..AA__	..A___			.AA___
	5	0.5	n/c	.A.AA_	.AAA__	n/c		
Log.	1	0.5		..A___	.A.___			...___
	5	0.5		..AA._	..A___			...___
Nor.	1	0.6		.AAA__	.AA___		..AA__	..A___
	5	0.6	n/c		...A__	n/c		
Log.	1	0.6			.AA___			..A___
	5	0.6		..AA._	.A..__			..A.__
Nor.	1	0.7		.AAA._	.AA.__		..AA._	..A.__
	5	0.7	n/c	...AA_	...A._	n/c		..AA._
Log.	1	0.7			.AA___			.AA___
	5	0.7		.AAA..	....._			.AAA._

to the SK (Table VI.3) and should be avoided. The SK would be more appropriate for shorter records.

As in Chapter V, the decision between tests is not clear cut when the possibility of serially correlated data

Table VI.9 Power comparison for MKD and SK median collapsing of monthly observations to quarterly values.

SIMULATION DESCRIPTION			G - MKD (quarterly median data) H - SK (quarterly median data)					
			(ρ = 0.0)			(ρ = 0.4)		
			years of record			years of record		
Dist.	σ_r	p	05	10	15	05	10	15
Nor.	1	0.0		..G___	.G.___		..G.__	...___
	5	0.0	n/c	..GG__	..G.__	n/c	....__	....__
Log.	1	0.0		..G___	..____		...___	...___
	5	0.0		....__	...___		....__	...___
Nor.	1	0.1		..G___	...___		..G.__	..G___
	5	0.1	n/c	....__	..G.__	n/c	....__	....__
Log.	1	0.1		...___	...___		....__	...___
	5	0.1		....__	...___		....__	...___
Nor.	1	0.3		....__	...___		..G.__	...___
	5	0.3	n/c	...G__	....__	n/c	..GG._	....__
Log.	1	0.3		..G___	...___		....__	...___
	5	0.3		....__	...___		...G__	...___
Nor.	1	0.5		G...__	..G___		....__	...___
	5	0.5	n/c	...G._	..G.__	n/c		....__
Log.	1	0.5		..G___	...___		.G..__	...___
	5	0.5		....._	....__		....._	....__
Nor.	1	0.6		G.GG__	..G___		....__	..G___
	5	0.6	n/c	...GG_	....__	n/c	...G._	...G__
Log.	1	0.6		..G.__	...___		....__	...___
	5	0.6		....._	....__		....._	....__
Nor.	1	0.7		..GG__	..G___		G.GG__	..G.__
	5	0.7	n/c	...GG_	...G__	n/c	GG.G._	...G._
Log.	1	0.7		..GG__	.GG___		..G.__	...___
	5	0.7		..GG._	....__		...G._	...G__

is present. The SKC is not appropriate for five years of record (conservative significance levels), and quarterly averaging does not remove all correlation in records so that the MKD and SK can be readily applied. Figures VI.1 through

VI.6 demonstrate some of the trade-offs involved with percentage of missing values, years of record and serial correlation.

The SKC is appropriate for longer records (ten or more years) when data are serially correlated. The SKC does have a tendency to be slightly conservative at ten years of record but does preserve the nominal significance level for 15 years of record (Tables VI.4 and VI.5). The SKC is considered too conservative for five years of record. As an alternative, the SK is recommended with quarterly averaged data, five years of record and serial correlation. For large missing value levels (70%), the SK is conservative (Table VI.3).

D. Summary and Conclusions

This chapter investigated the effect of missing values on trend detection. Procedures for handling this data record attribute were suggested and evaluated. Key conclusions are summarized below.

1. The SK actual significance level is usually closer to the nominal significance than the MKD actual significance level for short data records (five years of record). When the SK and MKD have comparable actual significance levels (10 and 15 years of record), the MKD is more powerful than the SK regardless of missing value level.
2. The actual significance level of the MKD and SK with

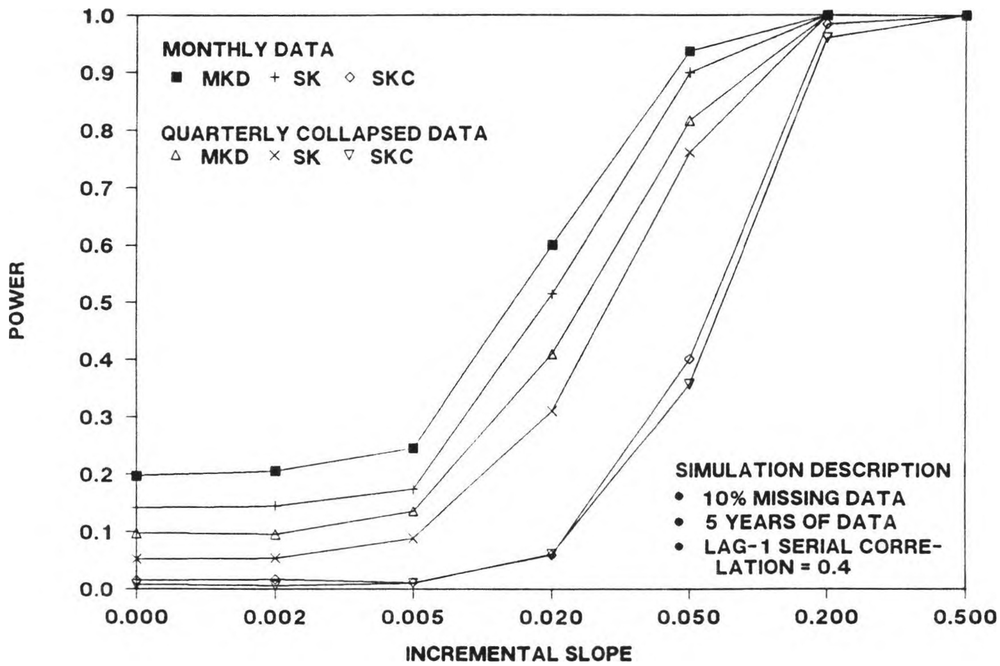

Figure VI.1 Comparison of power curves for monthly and quarterly (median) averaged data for five years of data, lag-one autocorrelation equal to 0.4, and missing values percentage equal to 0.1.

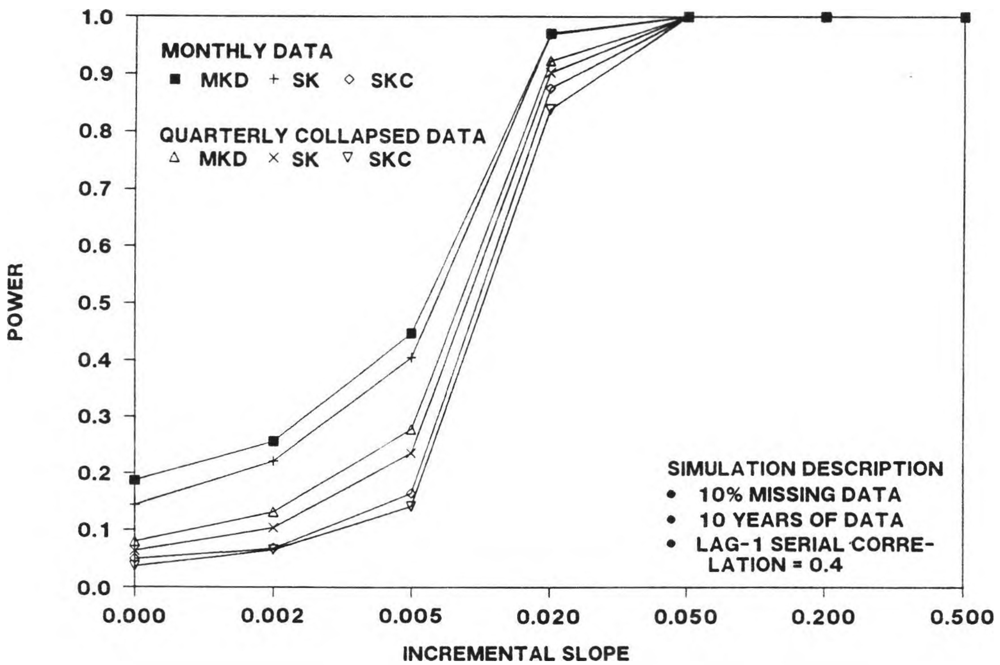

Figure VI.2 Comparison of power curves for monthly and quarterly (median) averaged data for ten years of data, lag-one autocorrelation equal to 0.4, and missing values percentage equal to 0.1.

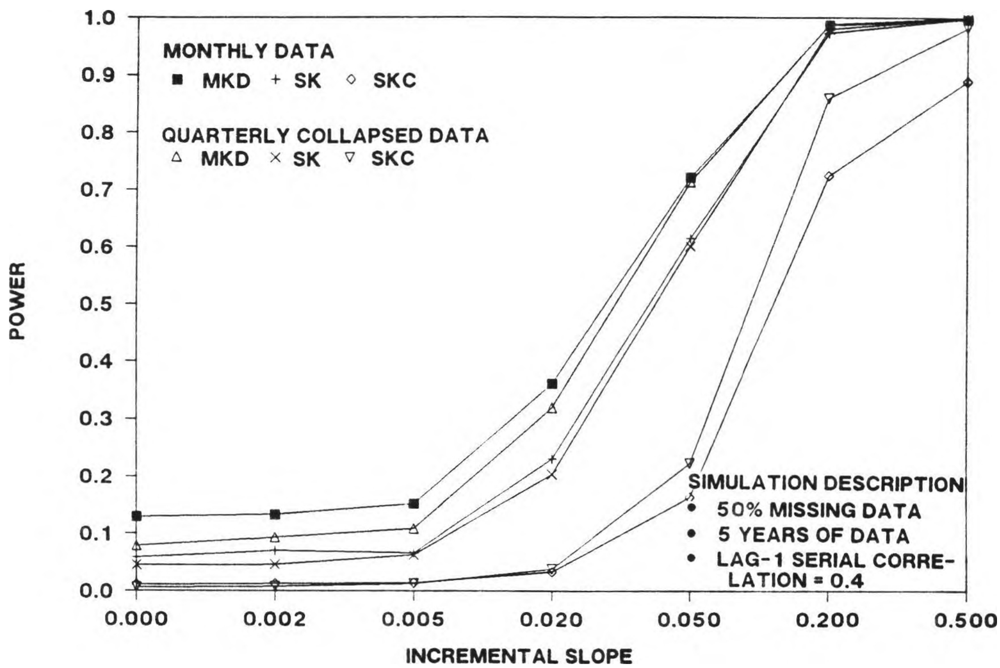

Figure VI.3 Comparison of power curves for monthly and quarterly (median) averaged data for five years of data, lag-one autocorrelation equal to 0.4, and missing values percentage equal to 0.5.

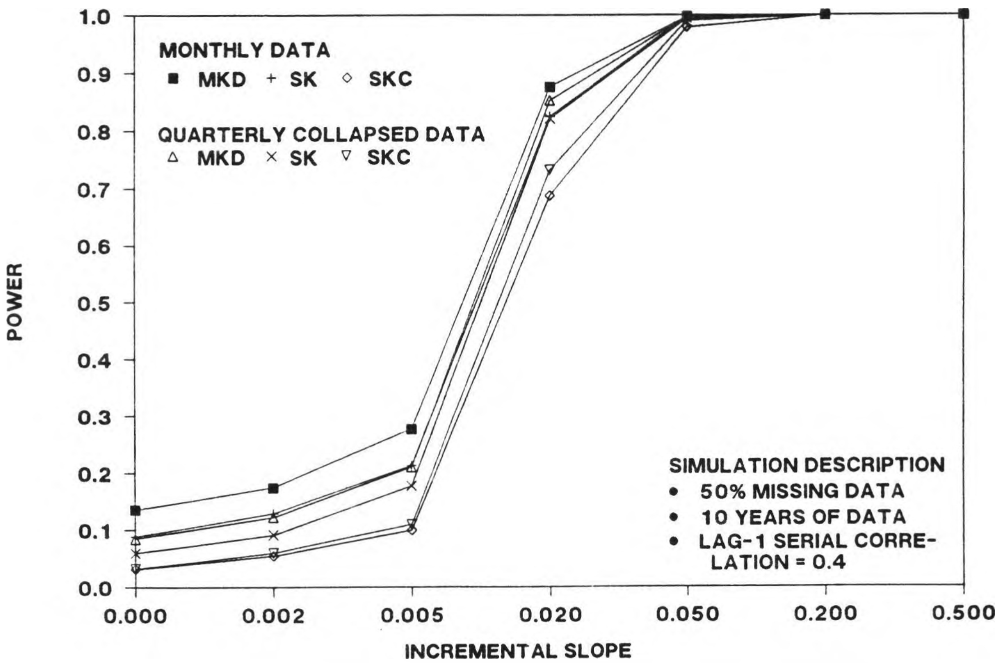

Figure VI.4 Comparison of power curves for monthly and quarterly (median) averaged data for ten years of data, lag-one autocorrelation equal to 0.4, and missing values percentage equal to 0.5.

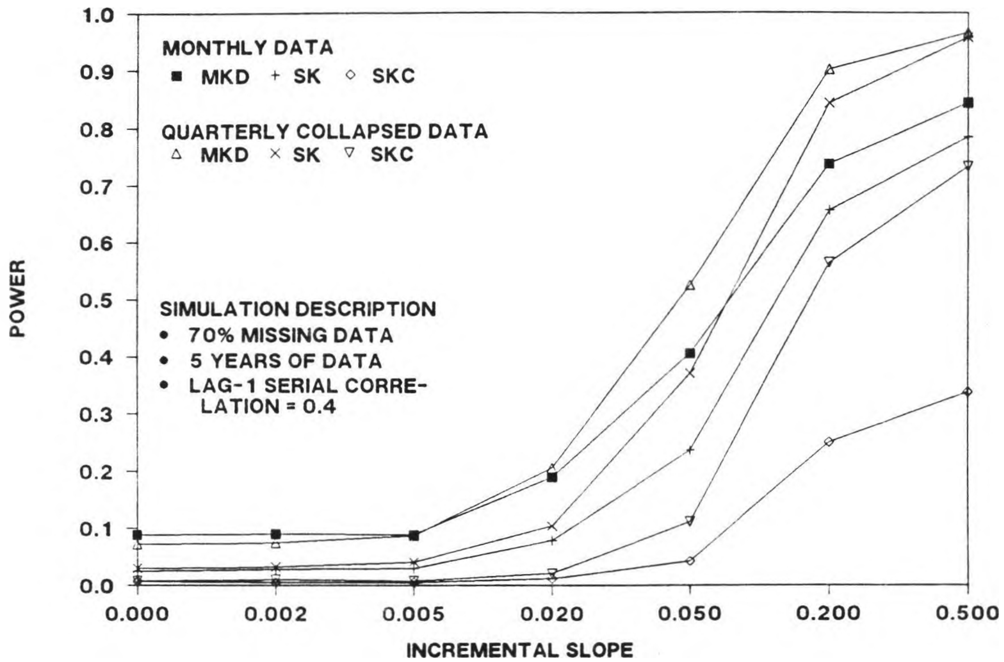

Figure VI.5 Comparison of power curves for monthly and quarterly (median) averaged data for five years of data, lag-one autocorrelation equal to 0.4, and missing values percentage equal to 0.7.

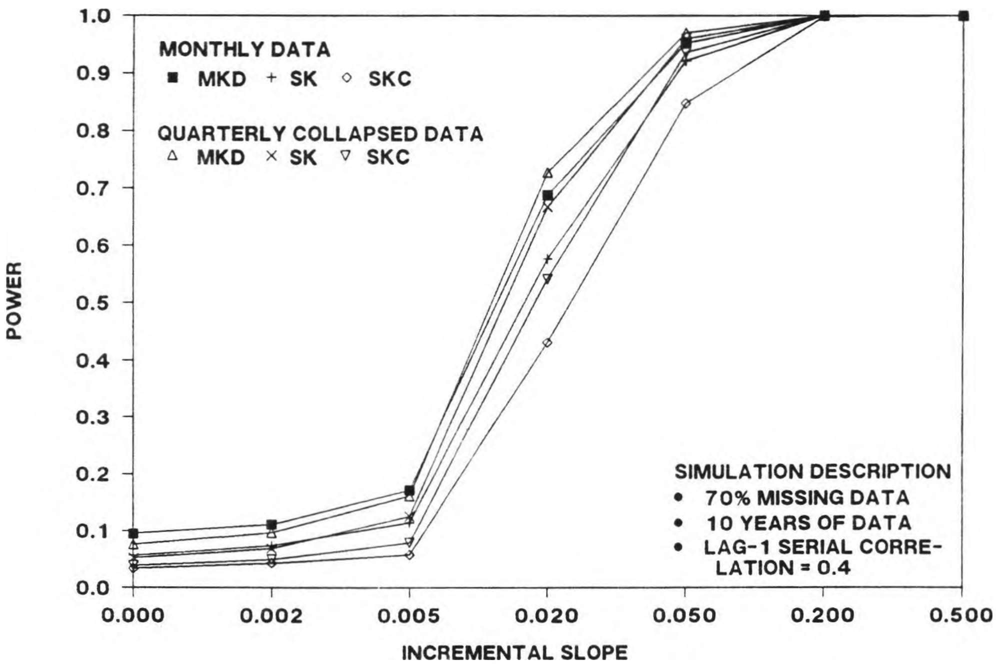

Figure VI.6 Comparison of power curves for monthly and quarterly (median) averaged data for ten years of data, lag-one autocorrelation equal to 0.4, and missing values percentage equal to 0.7.

correlated monthly data decreased as the percentage of missing values increased. The actual significance level of the SKC was not strongly affected by missing values.

3. All tests are affected by missing values in terms of power. Monthly data should be collapsed to quarterly values when more than 50% of the monthly data are missing and the SK or MKD are to be used. Monthly data should be collapsed to quarterly values when more than 40% of the monthly data are missing and the SKC is to be used.
4. There is not a good alternative when there are five years of record and more than 50% missing values regardless of correlation level. The SKC is too conservative for five years of record. The SK is too conservative for five years of record and large levels of missing values and short records.

E. Application

In general, the same discussion regarding the trade-offs between significance and power that were made in Section V.E is applicable here. The recommendations made in Chapter V are now amended to account for missing values.

1. Collapse monthly data to quarterly values if more than 50% of the monthly data are missing when applying the MKD or SK.
2. Collapse monthly data to quarterly values if more than 40% of the monthly data are missing when applying the

SKC.

3. When there are only five years of record and more than 50% missing values, there is not a good alternative.

CHAPTER VII

SUMMARY AND CONCLUSIONS

A. Review and Synthesis

Four data analysis tasks were identified that must be accomplished before a water-quality data analysis protocol can be identified. These issues were: Identify common water-quality conditions and data record attributes, review and discuss assumptions and logic incorporated in past protocols, recommend a framework appropriate to data analysis and recommend and evaluate procedures for alleviating some shortcomings of past protocols.

Many management decisions can be made by evaluating three types of water-quality conditions: average, changing and extreme conditions. This study was limited to the first two conditions. Data record attributes which inhibit statistical analysis were classified into two groups. Statistical characteristics are attributes resulting from the natural variability of water-quality variables. These attributes do not cause computational problems per se; however, they may violate typical statistical assumptions which may lead to incorrect conclusions. Data limitations are, for the most part, man-induced attributes which result in less reliable observations, less information for a given data set and increase analysis complexity. Common data record attributes in this study are summarized in Table VII.1.

Table VII.1 Data record attributes common to water-quality data sets.

Statistical characteristics	Data limitations
distribution applicability	censoring
variance homogeneity	erroneous observations
seasonality	small sample size
serial correlation	missing values
extreme events	different sampling frequencies
	multiple observations
	measurement uncertainty

Based on the review of past protocols, several "standard" procedures have gained wide acceptance for water-quality data analysis.

1. Average conditions are described with median and interquartile range estimates, often displayed as box plots.
2. The Mann-Kendall tau and the Wilcoxon Rank Sum tests (and variations) are used for monotonic and step trend analysis, respectively. The Wilcoxon Signed Rank test is used for step trends when the data can be paired. The Sen slope estimator is used for monotonic trend estimation.
3. Nonparametric procedures are recommended to avoid problems of non-normally distributed data and are outliers.

4. Measurement error is included in overall process variance and is not explicitly included in statistical analyses.
5. One water-quality random variable is analyzed at a time.

In terms of information flow, five components were identified that are integral to data analysis. These five components are: Identification of information goals and transformation of goals into water-quality conditions, data handling, identification of data limitations and statistical characteristics, water-quality evaluation and information reporting. In a well-conceived network design, many of the specific details are determined at the initial network design phase.

Several protocol shortcomings were identified for simulation. First, the average condition estimation work initiated by Gilliom and Helsel (1986) and Helsel and Gilliom (1986) was extended. Second, several monotonic trend detection procedures were evaluated to compare their effectiveness against serial correlation. The performance of these techniques was then evaluated under increasing levels of missing values.

B. Concluding Remarks

Specific conclusions for each task are identified at the end of the respective chapters and are not repeated here. There are some remarks which relate to water-quality data analysis protocols in general.

1. Much research effort has been spent developing techniques for water-quality data analysis. The greatest task facing network managers is not one of simply finding a methodology for evaluating information goals. There are several choices available. Rather the most difficult task is selecting the most appropriate method and coupling the method with appropriate data preparation procedures that account for specific data record attributes.
2. Protocols have been defined for data collection and laboratory analysis with the purpose of reducing associated errors and biases. A protocol for data analysis should provide a means for consistent and defensible water-quality evaluation as well.
3. This study is a further attempt to address water-quality data analysis issues. This study and similar studies should evolve along with our knowledge of water quality, hydrology and statistics.

C. Recommendations for Future Work

The following is a brief discussion of topics which are water-quality data analysis weaknesses. These issues should be addressed with additional resources.

1. Numerous techniques have been recommended to detect erroneous observations (e.g., Gilbert, 1987). Little justification is available for the use of one procedure over other procedures for detecting erroneous

observations. Although the recommended procedures are robust against outliers, this does not mean that the procedures are immune to the effects of poor data. Effort should be placed on determining which random variables (and their respective sampling and laboratory analysis procedures) are most susceptible to error.

2. Better physical understanding of the monitored region will assist in subsequent data analysis. Limited physical knowledge is incorporated into data analysis. Additional research should be directed towards determining what physical information is most useful for subsequent water-quality evaluation.
3. This study did not address extreme condition evaluation. Key challenges in this area would include determining both appropriate statistics and simulation procedures. The trade-offs between absolute standards and percentile standards are known; however, the appropriateness of either is not well defined. McBride and Pridmore (1987) suggest a two-tier program combining the two approaches. A second concern involves simulation experiments. The data generation procedures used in Chapter IV were less than ideal (Appendix A). It was difficult to transform data from normal to lognormal or mixed lognormal probability distributions. For some cases, the sample variance was drastically different from the simulated variance. This was primarily due to a few very large numbers. While this issue is probably not critical for

average and changing condition evaluation, the tails of distributions would be critical for extreme condition evaluation.

4. This study did not address information reporting. In terms of information flow, this component of data analysis is probably the most critical. Report frequency and format and included information may vary depending on the management structure and efforts.
5. How much data is enough? This study did not attempt to answer this question. Although it is possible to run some statistical tests with a minimum of data, the amount of misinformation with short data records is not known. Some standards should be set for minimum data requirements before meaningful analysis is possible. These standards should be tempered with timely data analysis. The amount of data necessary to check distributional assumptions is not known. Although this study emphasizes nonparametric procedures, improved analysis is possible with knowledge of the distribution.
6. One of the key recommendations of past work (e.g., Porter et al., 1988) is to not censor data at the laboratory. As a result, negative values of concentration are likely. Some of the recommended procedures rely on log-transformations. This "new" data limitation has not been dealt with in this study or in the literature.
7. Multiple censoring has been dealt with quite extensively

in recent literature. The issue of systematic versus progressive censoring has been dealt with to some extent for trend detection. The same issues have not been dealt with for average condition estimation.

8. Some of the implications made in this study are sequential processes. These sequential recommendations are based on individual simulations. For example, serial correlation detection is a significant part of the decision making process for selecting a trend test. However, the issues regarding the significance level and power of sequential tests should be examined with a simulation procedure which incorporates the entire process.
9. Several key issues regarding average condition estimation are discussed in the protocol. However, one issue relating to the simulation in Chapter IV is left unresolved. The "no censoring rule" was the overall best estimation procedure for the mean, standard deviation, and median. "No censoring" was not the best procedure for the interquartile range. The latter result is similar to the results by Helsel and Cohn (1988). This result is counter-intuitive and should be investigated further. It has been suggested that the sample interquartile range may not be the best estimate of the population interquartile range available.
10. There are several ancillary issues arising from the trend detection simulations. First does the Mann-

Whitney test behave similarly to the Mann-Kendall tau test for correlation, collapsed data sets, and missing values? Second, the recommendations for monotonic trend tests were based on five, ten and fifteen years of simulated data. The results tend to indicate that the trade-off between the Seasonal Kendall and Mann-Kendall tau tests occurs between five and ten years. This result should be confirmed with more refined simulations. Finally, it appeared that the difference in power of the Seasonal Kendall tau test with correlation correction (SKC) and the Mann-Kendall tau test on deseasonalized data (MKD) was decreasing for longer periods of records and no correlation. At what point (years of data) do the SKC and MKD perform equally well on data with no correlation? At this point, it would make sense to use the SKC test exclusively.

11. Lettenmaier (1976) adapted the Mann-Whitney test to account for Markovian correlation structure (AR(1)). The Mann-Whitney test should be modified to account for arbitrary correlation structure.
12. Little research has focused on trend magnitude estimation in the presence of single or multiple censoring limits. Gilbert (1987) and Ward et al. (1988) recommended one-half of the detection limit; however, little justification is provided.

In summary, there is considerable additional research needed if a scientifically defensible water-quality data analysis protocol is to be developed. Hopefully, this study has helped organize and extend some of the current thinking on the subject.

BIBLIOGRAPHY

Aitchison, John. 1955. On the distribution of a positive random variable having a discrete probability mass at the origin. Journal of the American Statistical Association, 50(271):901-908.

Aitchison, John and J. A. C. Brown. 1957. The Lognormal Distribution. The Syndics of the Cambridge University Press, Cambridge, 176 p.

Alley, William M. 1988. Using exogenous variables in testing for monotonic trends in hydrologic time series. Water Resources Research, 24(11):1955-1961.

American Public Health Association. 1985. Standard Methods for the Examination of Water and Wastewater. 16th ed., American Public Health Association, 1015 Eighteenth Street, N.W., Washington, D. C. 20036.

Association of State and Interstate Water Pollution Control Administrators. 1984. America's Clean Water: The States' Evaluation of Progress 1972-1982. Washington, D. C.

Averett, R. C. 1978. Selecting Water-Quality Parameters for the Specified Program or Study Goals. Central Region, Water Resources Division, U. S. Geological Survey - Handout. 17 p.

Azencott, Robert and Didier Dacunha-Castelle. 1986. Series of Irregular Observations, Forecasting and Model Building. Springer-Verlag, New York, 236 p.

Baracos, Paul C., Keith W. Hipel, and A. Ian McLeod. 1981. Modeling hydrologic time series from the Artic. Water Resources Bulletin, 17(3):414-422.

Barnett, V. and T. Lewis. 1978. Outliers in Statistical Data. Wiley, New York, 365 p.

Bell, Harry F. and Harry P. DeLong. 1988. Data characteristics: Groundwater monitoring's "Catch 22". Presented before the Division of Environmental Chemistry, American Chemical Society, Los Angeles, California, September.

Berryman, David. 1984. La détection des tendances dans les séries temporelles de paramétres de la qualité de l'eau des tests non-paramétriques. M.S. Thesis, Institut national de la recherche scientifique (INRS-Eau), Saint-Foy, Québec.

Berryman, David, Bernard Bobée, Daniel Cluis, and John Haemmerli. 1988. Nonparametric tests for trend detection in water quality time series. Water Resources Bulletin, 24(3):545-556.

Beyer, William H. (ed.). 1978. CRC Handbook of Mathematical Sciences. 5th ed., CRC Press, Boca Raton, Florida, 982 p.

Bilonick, R. A. 1984. Further discussion. Atmospheric Environment, 18:232-234.

Bilonick, R. A. 1984. The statistical confirmation of trends in autocorrelated data. Atmospheric Environment, 18:229-230.

Box, G. E. P. and Gwilym M. Jenkins. 1970. Time Series Analysis, Forecasting and Control. Holden-Day, San Francisco, 553 p.

Box, G. E. P. and G. C. Tiao. 1975. Intervention analysis with applications to economic and environmental problems. Journal of the American Statistical Association, 70(349):70-79.

Bradley, James V. 1968. Distribution-Free Statistical Tests. Prentice-Hall, Englewood Cliffs, New Jersey, 388 p.

Brockwell, Peter J. and Richard A. Davis. 1987. Time Series, Theory and Methods. Springer-Verlag, New York, 519 p.

Buckley, J. and I. James. 1979. Linear regression with censored data. Biometrika, 66(3):429-436.

Chambers, John M., William S. Cleveland, Beat Kleiner, and Paul A. Tukey, 1983. Graphical Methods for Data Analysis. Duxbury Press, Boston, 395 p.

Close, M. A. 1987. Effects of irrigation on water quality of a shallow unconfined aquifer. Water Resources Bulletin, 23(5):793-802.

Cochran, W. G. 1977. Sampling Techniques, 3rd ed., Wiley, New York. 428 p.

Cohen, A. Clifford, Jr. 1959. Simplified estimators for the normal distribution when samples are singly censored or truncated. Technometrics, 1(3):217-237.

Cohn, Timothy A. 1988. Adjusted maximum likelihood estimation of the moments of lognormal populations from Type I censored samples. U. S. Geological Open File Report 88-350.

Condie, Robert and Keith A. Lee. 1982. Flood frequency analysis with historic information. Journal of Hydrology, 58:47-61.

Conover, W. J. 1980. Practical Nonparametric Statistics, 2nd ed. Wiley, New York, 493 p.

Crawford, Charles G., James R. Slack, and Robert M. Hirsch. 1983 Nonparametric tests for trends in water-quality data using the statistical analysis system. U. S. Geological Open File Report 83-550, 102 p.

David, H. A. 1981. Order Statistics, 2nd ed., John Wiley, New York, 360 p.

Deininger, R. A. 1979. Calculating geometric mean. Deeds and Data, published by the Water Pollution Control Federation, 16(2):13-14.

Dempster, A. P., N. M. Laird, and D. B. Rubin. 1977. Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society, B., 39(1):1-22.

Devroye, Luc. 1986. Non-Uniform Random Variate Generation. Springer-Verlag, New York, 843 p.

Dwass, Meyer. 1970. Probability and Statistics, An Undergraduate Course. W. A. Benjamin, Menlo Park, California, 635 p.

Egar, David L., A. L. Wilson, and R. K. Aylesworth. 1979. DISCUSSION on "Design considerations for ambient stream quality monitoring" by Dennis P. Lettenmaier. Water Resources Bulletin, 15(6):1781-1784.

El-Shaarawi, A. H. and Eivind Damsleth. 1988. Parametric and nonparametric tests for dependent data. Water Resources Bulletin 24(3):513-519.

Everett. L. G. 1980. Groundwater Monitoring. General Electric Company, Schenectady, New York.

Farrel, R. 1980. Methods for classifying changes in environmental conditions. Technical Report VRF-EPA7.4-FR80-1, Vector Research Inc., Ann Arbor, Michigan.

Finney, D. J. 1941. On the distribution of a variate whose logarithm is normally distributed. Journal of the Royal Statistical Society, B., 7(2):155-161.

General Accounting Office. 1986. The Nation's Water: Key Unanswered Questions About the Quality of Rivers and Streams. Report no. GAO/PEMD-86-6, US General Accounting Office, Washington, D. C., 20548.

Gianessi Leonard P., Henry M. Peskin and G. K. Young. 1981. Analysis of national water pollution control policies, 1, A national network model. Water Resources Research, 17(4):796-802.

Gianessi, Leonard P. and Henry M. Peskin. 1981. Analysis of national water pollution control policies, 2, Agricultural sediment control. Water Resources Research, 17(4):803-821.

Gilbert, Richard O. 1987. Statistical Methods for Environmental Pollution Monitoring. Van Nostrand Reinhold Company, New York, 320 p.

Gilliom Robert J., Robert M. Hirsch, and Edward J. Gilroy. 1984. Effect of censoring trace-level water-quality data on trend detection capability. Environmental Science and Technology, 18(7):530-535.

Gilliom, Robert J., Richard B. Alexander, and Richard A. Smith. 1985. Pesticides in the Nation's Rivers, 1975-1980, and Implications for Future Monitoring, U. S. Geological Survey Water-Supply Paper 2271, 26 p.

Gilliom, Robert J. and Dennis R. Helsel. 1986. Estimation of distributional parameters for censored trace level water quality data, 1, Estimation techniques. Water Resources Research, 22(2):135-146.

Graybill, Franklin A. 1976. Theory and Application of the Linear Model. Wadsworth Publishing Company, Belmont, California, 704 p.

Harned, D. A., C. C. Daniel III, and J. K. Crawford. 1981. Methods of discharge compensation as an aid to the evaluation of water quality trends. Water Resources Research, 17(5):1389-1400.

Harris, Jane, Jim C. Loftis, and Robert H. Montgomery. 1987. Statistical methods for characterizing ground-water quality. Ground Water, 25(2):185-193.

Harris, Jane. 1988. Statistical analysis of ground water quality, interaction of deterministic and stochastic components. Ph.D. Dissertation. Department of

Agricultural and Chemical Engineering, Colorado State University, Fort Collins, Colorado, 145 p.

Harter, H. Leon and Albert H. Moore. 1968. Maximum likelihood estimation, from doubly censored samples, of the parameters of the first asymptotic distribution of extreme values. Journal of the American Statistical Association, 63(323):889-901.

Healy, M. J. R. 1968. Multivariate normal plotting. Applied Statistics, 17(2):157-161.

Helsel, Dennis R. 1983. Mine drainage and rock type influences on eastern Ohio stream water quality. Water Resources Bulletin, 19(6):881-887.

Helsel, Dennis R. 1986. Estimation of distributional parameters for censored water quality data. Developments in Water Science, Statistical Aspects of Water Quality Monitoring, 27:137-157.

Helsel, Dennis R. and Robert J. Gilliom. 1986. Estimation of distributional parameters for censored trace level water quality data, 2, Verification and applications. Water Resources Research, 22(2):147-155.

Helsel, Dennis R. 1987. Advantages of nonparametric procedures for analysis water quality data. Hydrological Sciences Journal, 32(2):179-190.

Helsel, Dennis R. and Timothy A. Cohn. 1988. Estimation of descriptive statistics for multiply-censored water-quality data. Water Resources Research 24(12):1997-2004.

Hipel, Keith William, W. C. Lennox, T. E. Unny, and Angus Ian McLeod. 1975. Intervention analysis in water resources. Water Resources Research, 11(6):855-861.

Hipel, Keith William, Angus Ian McLeod, and William C. Lennox. 1977. Advances in Box-Jenkins modeling, 1, Model construction. Water Resources Research, 13(3):567-575.

Hipel, Keith William, Angus Ian McLeod, and R. R. Weiler. 1988. Data analysis of water quality time series in Lake Erie. Water Resources Bulletin 24(3):533-544.

Hirsch, Robert M. 1979. Synthetic hydrology and water supply reliability. Water Resources Research, 15(6):1603-1615.

Hirsch, Robert M. 1982. A comparison of four streamflow record extension techniques. Water Resources Research, 18(4):1081-1088.

Hirsch, Robert M., James R. Slack, and Richard A. Smith. 1982. Techniques of trend analysis for monthly water quality data. Water Resources Research, 18(1):107-121.

Hirsch, Robert M. and Edward J. Gilroy. 1984. Methods of fitting a straight line to data, examples in water resources. Water Resources Bulletin, 20(5):705-711.

Hirsch, Robert M. and Norman E. Peters. 1984. Author's Reply. Atmospheric Environment, 18:230-232.

Hirsch, Robert M. and James R. Slack. 1984. A nonparametric trend test for seasonal data with serial dependence. Water Resources Research, 20(6):727-732.

Hirsch, Robert M. and Edward J. Gilroy. 1985. Detectability of step trends in the rate of atmospheric deposition of sulfate. Water Resources Bulletin, 21(5):773-783.

Hirsch, Robert M. 1987. Acting Assistant Chief Hydrologist for Research and External Coordination, Water Resources Division, U. S. Geological Survey, personal communication.

Hirsch, Robert M. and Jery R. Stedinger. 1987. Plotting positions for historical floods and their precision. Water Resources Research, 23(4):715-727.

Hirsch, Robert M. 1988a. Selection of trend analysis techniques for water quality data. Presented before the Division of Environmental Chemistry, American Chemical Society, Los Angeles, California, September.

Hirsch, Robert M. 1988b. Statistical methods and sampling design for estimating step trends in surface-water quality. Water Resources Bulletin, 24(3):493-503.

Hughes, James P. and Steven P. Millard. 1988. A tau-like test for trend in the presence of multiple censoring points. Water Resources Bulletin, 24(3):521-531.

Hollander, Myles and Douglas A. Wolfe. 1973. <u>Nonparametric Statistical Methods</u>. Wiley, New York, 503 p.

Iman, R. L., D. Quade, D. A. Alexander. 1975. Exact probability levels for the Kruskal-Wallis test. Selected Tables in Mathematical Statistics 3:329-384.

Kirby, W. 1974. Straight line fitting of an observation path by least normal squares. U. S. Geological Survey, Open-File Report No.74-197, 11 p.

Kitagawa, Genshiro. 1980. Outlap, an outlier analysis program. Computer Science Monographs No. 15, Institute of Statistical Mathematics, Tokyo, Japan, 44 p.

Koch, George S., Jr. and Richard F. Link 1980. Statistical Analyses of Geological Data, Volume 1. Wiley, New York, 375 p.

Koch, George S., Jr. and Richard F. Link 1980. Statistical Analyses of Geological Data, Volume 2. Wiley, New York, 438 p.

Krause, D. J. and D. Graupe. 1973. Identification of predictor and filter parameters by ARMA methods. International Journal of Control, 17(5):1021-1027.

Kuczera, George. 1982. Combining site-specific and regional information: An empirical Bayes approach. Water Resources Research, 18(2):306-314.

Kuczera, George. 1982. Robust flood frequency models. Water Resources Research, 18(2):315-324.

Kuczera, George. 1983. Effect of sampling uncertainty and spatial correlation on an empirical Bayes procedure for combining site and regional information. Journal of Hydrology, 65(4):373-398.

Larsen, Richard J. and Morris L. Marx. 1981. An Introduction to Mathematical Statistics and Its Applications. Prentice-Hall, Englewood Cliffs, New Jersey, 536 p.

Leese, Morven N. 1973. Use of censored data in the estimation of Gumbel distribution parameters for annual maximum flood series. Water Resources Research, 9(6):1534-1542.

Lettenmaier, Dennis P. 1976. Detection of trends in water quality data form records with dependent observations. Water Resources Research, 12(5):1037-1046.

Lettenmaier, D. P. 1978. Design considerations for ambient stream quality monitoring. Water Resources Bulletin, 14(4):884-902.

Lettenmaier, Dennis P. 1979. REPLY to DISCUSSION on "Design considerations for ambient stream quality monitoring" by Dennis P. Lettenmaier. Water Resources Bulletin, 15(6):1785-1786. Lettenmaier, Dennis P. 1980. Intervention analysis with missing data. Water Resources Research, 16(1):159-171.

Lettenmaier, Dennis P. 1988. Multivariate nonparametric tests for trend in water quality. Water Resources Research, 24(3):505-512.

Liebetrau, A. M. 1979. Water quality sampling, some statistical considerations. Water Resources Research, 15(6):1717-1725.

Loftis, Jim C. and Robert C. Ward. 1980. Sampling frequency selection for regulatory water quality monitoring. Water Resources Bulletin, 16(3):501-507.

Loftis, Jim C. and Robert C. Ward. 1980. Water quality monitoring, some practical sampling frequency considerations. Environmental Management, 4(6):521-526.

Loftis, Jim C. and Robert C. Ward. 1981. Evaluating stream standard violations using a water quality data base. Water Resources Bulletin, 17(6):1071-1078.

Loftis, Jim C., Robert C. Ward, and Gary M. Smillie. 1983. Statistical model for water quality regulation. Journal of the Water Pollution Control Federation, 55(8):1098-1104.

Loftis, Jim C., Robert H. Montgomery, Jane Harris, David Nettles, P. Steven Porter, Robert C. Ward, and Thomas G. Sanders. 1986. Final Report USGS Grant 14-08-0001-G-1060. Colorado Water Resources Research Institute, Fort Collins, Colorado. 170 p.

Loftis, Jim C., Jane Harris, and Robert H. Montgomery. 1987a. Detecting changes in ground water quality at regulated facilities. Ground Water Monitoring Review, 7(1):72-76.

Loftis, J. C., P. S. Porter, and G. Settembre. 1987b. Statistical analysis of industrial wastewater monitoring data. Journal Water Pollution Control, 59(3):145-151.

Mann, H. B. 1945. Nonparametric tests against trend. Econometrica, 13:245-259.

Mann, Nancy R. 1968. Point and interval estimation procedures for the two-parameter Weibull and extreme-value distributions. Technometrics, 10(2):231-256.

Matalas, N. C. 1967. Mathematical assessment of synthetic hydrology. Water Resources Research, 3(4):937-945.

McBean, Edward A. and Frank A. Rovers. 1985. Analysis of variances as determined from replicates vs. successive sampling. Ground Water Monitoring Review, 5(3):61-64.

McBride, G. B. and R. D. Pridmore. 1987. A statistically defensible procedure for expressing percentile effluent standards. Presented at New Zealand Water Supply and Disposal Association Annual Conference, September 2-4, Paper No. 15, Rotorua, New Zealand.

McLeod, A. I. 1977. Improved Box-Jenkins estimators. Biometrika, 64(3):531-534.

McLeod, Angus Ian, Keith William Hipel, and William C. Lennox. 1977. Advances in Box-Jenkins modeling, 2, Applications. Water Resources Research, 13(3):577-586.

McLeod, A. Ian, Keith W. Hipel, and Fernando Comancho. 1983. Trend assessment of water quality time series. Water Resources Bulletin, 19(4):537-547.

Meijers, E. M. J. 1986. Defining confusions--confusing definitions. Environmental Monitoring and Assessment, 7(2):157-159.

Millard, Steven P. and Steven J. Deverel. 1988. Nonparametric statistical methods for comparing two sites based on data with multiple nondetect limits. Water Resources Research 24(12):2087-2098.

Miller, R. G. 1981. Survival Analysis. Wiley and Sons, New York. 238 p.

Montgomery, Robert H. and Kenneth H. Reckhow. 1984. Techniques for detecting trends in lake water quality. Water Resources Bulletin, 20(1):43-52.

Montgomery, Robert. H. 1987. Uncertainty in ground water quality data. Ph.D. Dissertation, Department of Civil Engineering, Colorado State University, Fort Collins, Colorado, 354 p.

Montgomery, Robert H. and Jim C. Loftis. 1987. Applicability of the t-test for detecting trends in water quality variables. Water Resources Bulletin 23(4):653-662.

Montgomery, Robert H., Jim C. Loftis, and Jane Harris. 1987. Statistical characteristics of ground-water quality variables. Ground Water, 25(2):176-184.

Mosteller, F. and J. W. Tukey. 1977. Data Analysis and Regression, A Second Course in Statistics. Addison-Wesley, Reading, Mass., 588 p.

Nelder, J. A. and R. W. M. Wedderburn. 1972. Generalized linear models. Journal of the Royal Statistical Society A., 135(3):370-384.

Obeysekera, J. T. B. and J. D. Salas. 1986. Modeling of aggregated hydrologic time series. Journal of Hydrology. 86:197-219.

Owen, W. J. and T. A. DeRouen. 1980. Estimation of the mean for lognormal data containing zeroes and left-censored values, with applications to the measurement of worker exposure to air contaminants. Biometrics, 36:707-719.

Parzen, Emanuel (ed.). 1984. Time Series Analysis of Irregularly Observed Data. Proceedings of a Symposium held at Texas A & M University, College Station, Texas, February 10-13, 1983. Springer-Verlag, New York, 363 p.

Persson, T. and H. Rootzen. 1977. Simple and highly efficient estimators for a type I censored normal sample. Biometrika, 64(1):123.

Peters, N. E. 1984. Quality analysis of U. S. Geological Survey precipitation chemistry data for New York. Atmospheric Environment. 18:1041-1042.

Porter, Paul Steven, 1986. Statistical analysis of water quality data affected by limits of detection. Ph.D. Dissertation. Department of Agricultural and Chemical Engineering, Colorado State University, Fort Collins, Colorado, 197 p.

Porter, Paul Steven, Robert C. Ward, and Harry F. Bell. 1988. The detection limit. Water quality monitoring data are plagued with levels of chemicals that are too low to be measured precisely. Environmental Science and Technology 22:856-861.

Rajagopal, R. 1986. The effect of sampling frequency on ground water quality characterization. Ground Water Monitoring Review, 6(4):65-73.

Reckhow, K. H. and S. C. Chapra. 1983. Engineering Approaches for Lake Management, Volume 1, Data Analysis and Empirical Modeling. Butterworth, Boston, 340 p.

Reckhow, K. H. and S. C. Chapra. 1983. Engineering Approaches for Lake Management, Volume 2, Mechanistic Modeling. Butterworth, Boston, 492 p.

Rosner, B., 1983. Percentage points for a generalized ESD many-outlier procedure. Technometrics 25(2):165-172.

Salas, J. D., J. W. Delleur, V. Yevjevich, and W. L. Lane. 1980. Applied Modeling of Hydrologic Time Series. Water Resources Publications, Fort Collins, Colorado, 484 p.

Salas, J. D. and J. T. B. Obeysekera. 1982. ARMA model identification of hydrologic time series. Water Resources Research, 18(4):1011-1021.

Sanders, Thomas G., Robert C. Ward, Jim C. Loftis, Timothy D. Steele, Donald D. Adrian, and Vujica Yevjevich. 1983. Design of Networks for Monitoring Water Quality. Water Resources Publications, Littleton, Colorado, 328 p.

Schmee, J. and G. J. Hahn. 1979. A simple method for regression analysis with censored data. Technometrics, 21(4):417-432.

Sen, P. K. 1968. Estimates of the regression coefficients based on Kendall's tau, Journal of the American Statistical Association. 63:1379-1389.

Showalter, Patricia. 1985. Developing objectives for the ground water quality monitoring network of the Salinas River Drainage Basin. Ground Water Monitoring Review, 5(2):37-45.

Sichel, H. S. 1952. New methods in the statistical evaluation of mine sampling data. Transactions Institute of Mining and Metallurgy, 61:261-288.

Sichel, H. S. 1966. The estimation of means and associated confidence limits for small samples from lognormal populations, in Proceedings of the Symposium on Mathematical Statistics and Computer Applications in Ore Valuation. South African Institute of Mining and Metallurgy, Johannesburg, pp. 106-123.

Smith, Richard A., Richard B. Alexander, and M. Gordon Wolman. 1987. Water-quality trends in the nation's rivers. Science, 235:1607-1615.

Snedecor, George W. and William G. Cochran. 1980. Statistical Methods, 7th ed. The Iowa State University Press, Ames Iowa, 507 p.

Stedinger, Jery R., Kenneth Potter, David F. Kibler, and Gary Tasker. 1983. Comment on "Regional flood frequency estimation and network design" by N. P. Greis and E. F. Wood. Water Resources Research, 19(5):1343-1345.

Stedinger, Jery R. and Gary D. Tasker. 1985. Regional hydrologic analysis, 1, Ordinary, weighted, and generalized least squares compared. Water Resources Research, 21(9):1421-1432.

Stedinger, Jery R. and Gary D. Tasker. 1986. Correction to "Regional hydrologic analysis, 1, Ordinary, weighted, and generalized least squares compared" by J. R. Stedinger and G. D. Tasker. Water Resources Research, 22(5):844.

Stedinger, Jery R. and Gary D. Tasker. 1986. Regional hydrologic analysis, 2, Model-error estimators, estimation of Sigma and Log-Pearson Type 3 distributions. Water Resources Research, 22(10):1487-1499.

Stirling, W. D. 1984. Iteratively reweighted least squares for models with a linear part. Applied Statistics, 33(1):7-17.

Tasker, Gary D. 1980. Hydrologic regression with weighted least squares. Water Resources Research, 16(6):1107-1113.

Tasker, Gary D. and Edward J. Gilroy. 1982a. Comparison of estimators of standard deviation for hydrologic time series. Water Resources Research, 18(5):1503-1508.

Tasker, Gary D. 1982b. Simplified testing of hydrologic regression regions. Journal of the Hydraulics Division, 108(HY10):1218-1222.

Tasker, Gary D. 1983a. Approximate sampling distribution of the serial correlation coefficient for small samples. Water Resources Research, 19(2):579-582.

Tasker, Gary D. 1983b. Effective record length for the t-year event. Journal of Hydrology, 64:39-47.

Taylor, Charles H. and Jim C. Loftis. 1989. Testing for trend in lake and ground water quality time series. Water Resources Bulletin 25(4):715-726.

Theil, H. 1950. A rank-invariant method of linear and polynomial regression analysis, 1, 2, and 3. Ned. Akad. Wetensch Proc., 53:386-392; 53:521-525; and 53:1397-1412.

U. S. Environmental Protection Agency. 1984a. Before-and-After Case Studies: Comparisons of Water Quality Following Municipal Treatment Plant Improvements. Washington, D. C.

U. S. Environmental Protection Agency. 1984b. National Water Quality Inventory: 1982 Report to Congress. Washington, D. C.

U. S. Environmental Protection Agency and U. S. Fish and Wildlife Service. 1984. 1982 National Fisheries Survey. Washington, D. C.

U. S. Geological Survey. 1984. National Water Summary 1983--Hydrologic Events and Issues. Washington, D. C.

van Belle, Gerald and James P. Hughes. 1984. Nonparametric tests for trend in water quality. Water Resources Research 20(1):127-136.

van Latesteijn, H. C. and R. H. D. Lambeck. 1986. The analysis of monitoring data with the aid of time-series analysis. Environmental Monitoring and Assessment, 7:287-297.

Wallis, J. R., N. C. Matalas, and J. R. Slack. 1974. Just a moment! Water Resources Research, 10(2):211-219.

Ward, Robert C., Jim C. Loftis, Knud S. Nielsen, and R. Dennis Anderson. 1979. Statistical evaluation of sampling frequencies in monitoring networks. Journal of the Water Pollution Control Federation, 51(9):2292-2300.

Ward, Robert C. 1981. Ground-water quality monitoring, what information is to be obtained? Ground Water, 19(2):March-April.

Ward, Robert C. and Jim C. Loftis. 1986. Establishing statistical design criteria for water quality monitoring systems, review and synthesis. Water Resources Bulletin, 22(5):759-767.

Ward, Robert C. and Graham B. McBride. 1986. Design of Water Quality Monitoring Systems in New Zealand. Water Quality Centre, Ministry of Works and Development for the National Water and Soil Conservation Authority, Water Quality Centre Publication, 0112-689x; no. 8, 65 p.

Ward, Robert C., Jim C. Loftis, and G. B. McBride. 1986. The "data-rich but information-poor syndrome in water quality monitoring". Environmental Management, 10(3):291-297.

Ward, Robert C., Jim C. Loftis, Harry P. DeLong, and Harry F. Bell. 1988. Groundwater quality: A data analysis protocol. Journal Water Pollution Control Federation, 60(11):1938-1945.

Wolman, M. Gordon. 1971. The nation's rivers. Science, 174:905-918.

Wolynetz, M. S. 1979a. Maximum likelihood estimation from confined and censored normal data. Applied Statistics 28(2):185-195.

Wolynetz, M. S. 1979b. Maximum likelihood estimation in a linear model from confined and censored normal data. Applied Statistics 28(2):195-206.

Yevjevich, V. 1972. Probability and Statistics in Hydrology. Water Resources Publications, Fort Collins, Colorado. 302 p.

APPENDIX A

DATA GENERATION

A. Average Conditions

Data (for average condition estimation, Chapter IV) were generated using International Mathematical and Statistical Libraries generating routines for the normal, gamma, and uniform distributions. Lognormal, mixed lognormal, and delta distributions were produced with transformations of normal and uniform distributions described by Gilliom and Helsel (1986) and Aitchison (1957) with one exception. The proportion of zeroes for the delta distribution was increased from 5 to 15 percent. Five hundred data points were generated and discarded as a "warm-up" of the random number generator.

Large data sets (10000 numbers) were generated for each distribution and different coefficient of variation level to check for the adequacy of the generator. Probability density functions were computed from the this simulation (Figure A.1). Data sets were checked for cycling, goodness of fit, autocorrelation, and parameter reproducibility. During the simulation, the generator did not cycle (reproduce the same number) within 75000 numbers. A typical chi-square goodness of fit test is shown in Figure A.2 for a

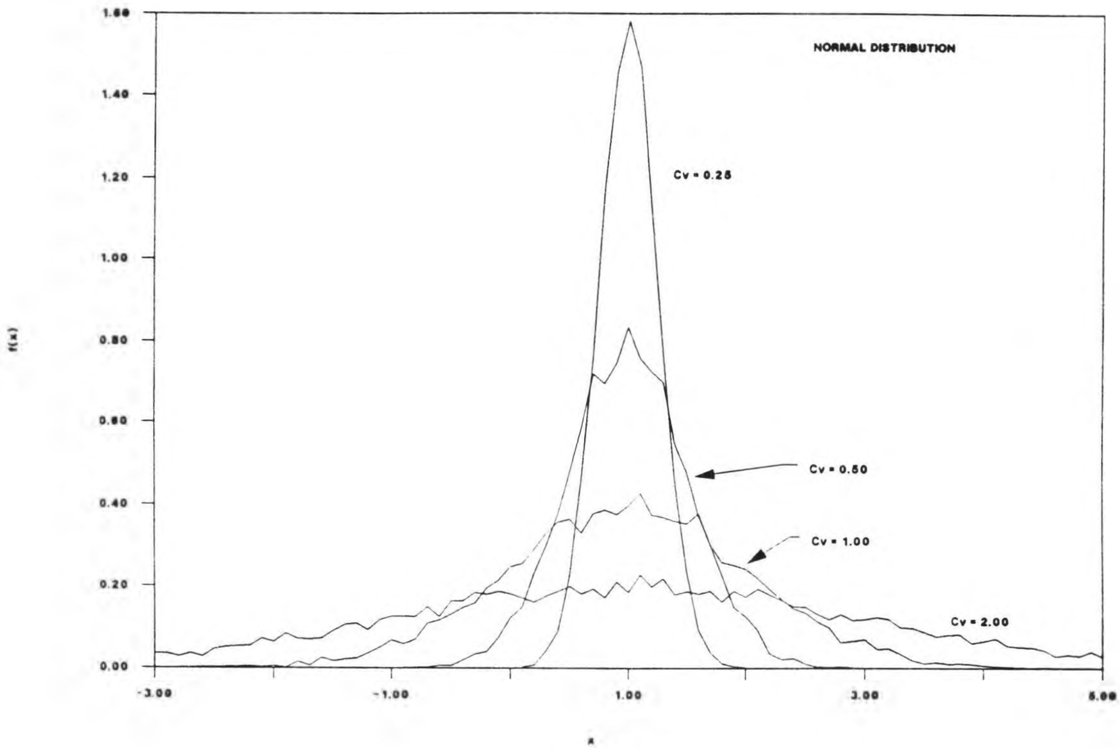

Figure A.1 Simulated probability density function for parent distributions used in simulation (a) normal distribution.

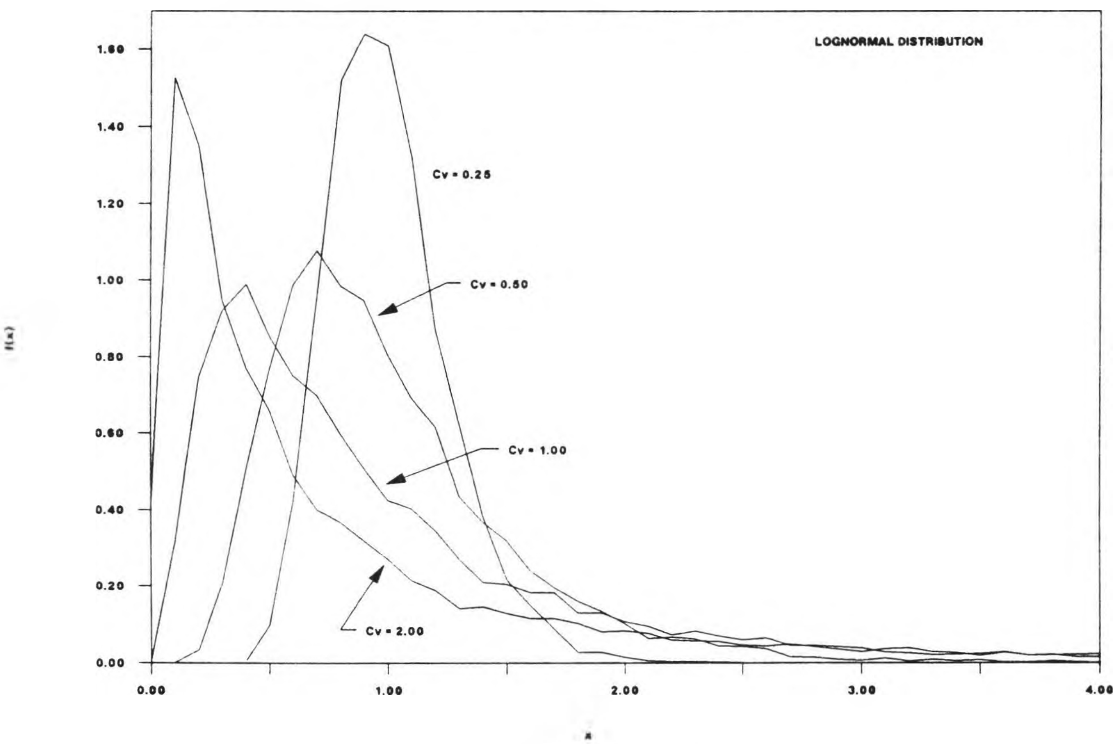

Figure A.1 Simulated probability density function for parent distributions used in simulation (b) lognormal distribution.

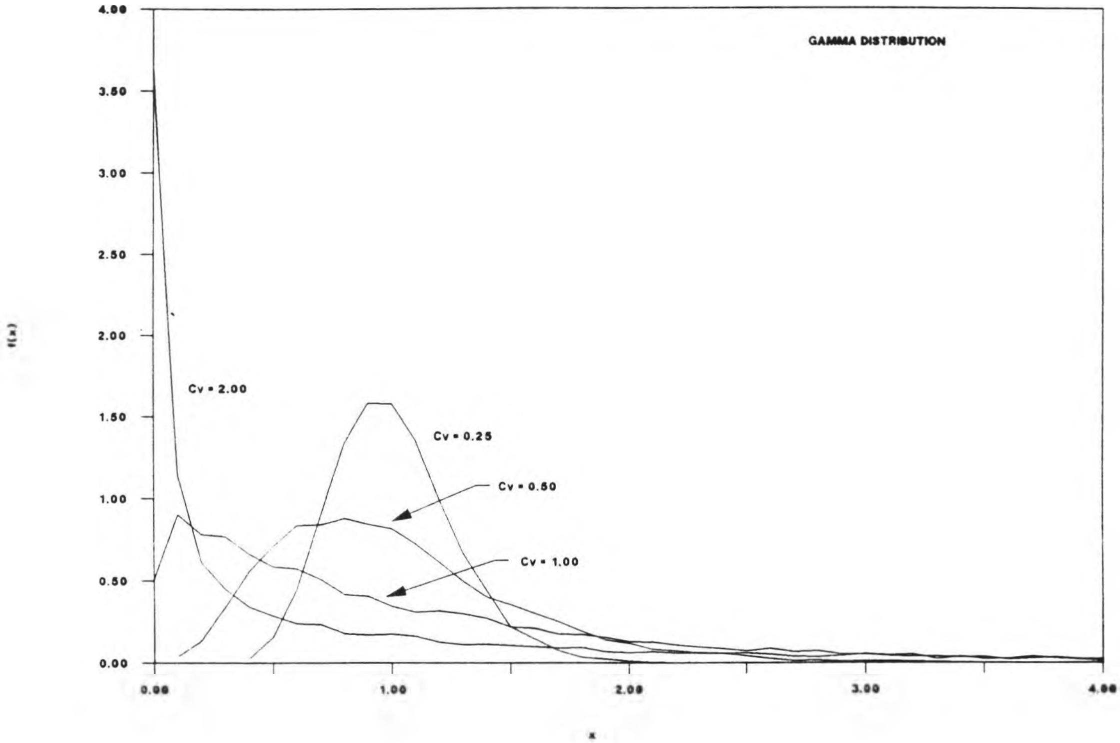

Figure A.1 Simulated probability density function for parent distributions used in simulation (c) gamma distribution.

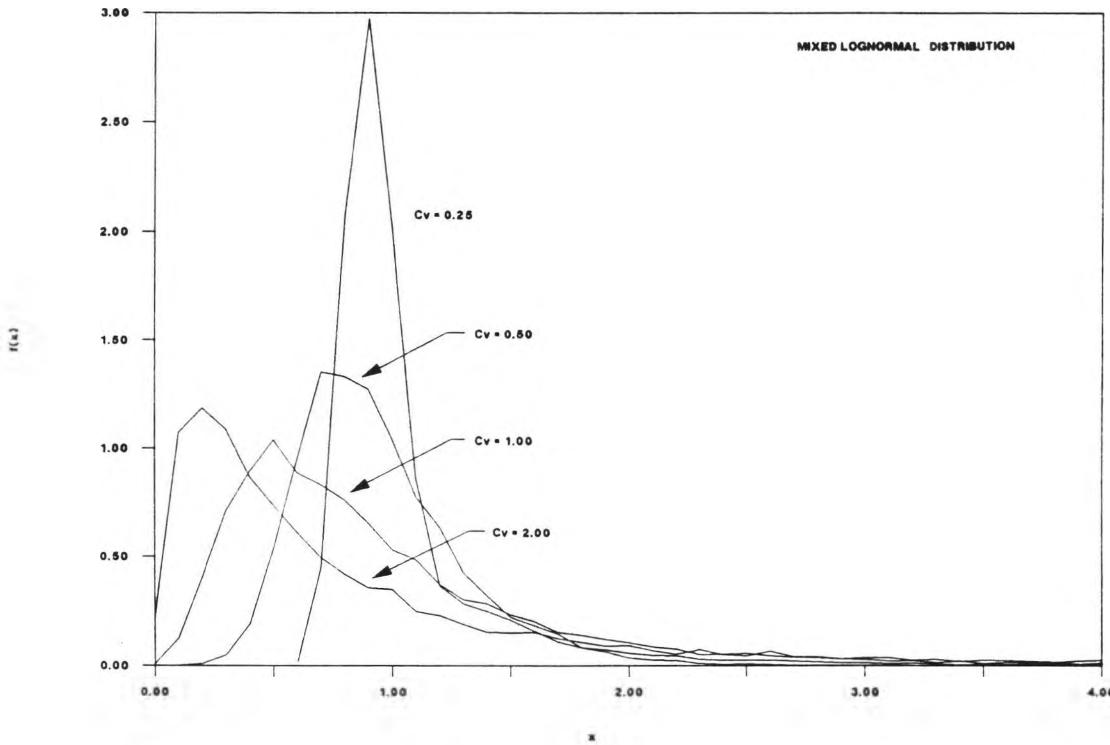

Figure A.1 Simulated probability density function for parent distributions used in simulation (d) mixed lognormal distribution.

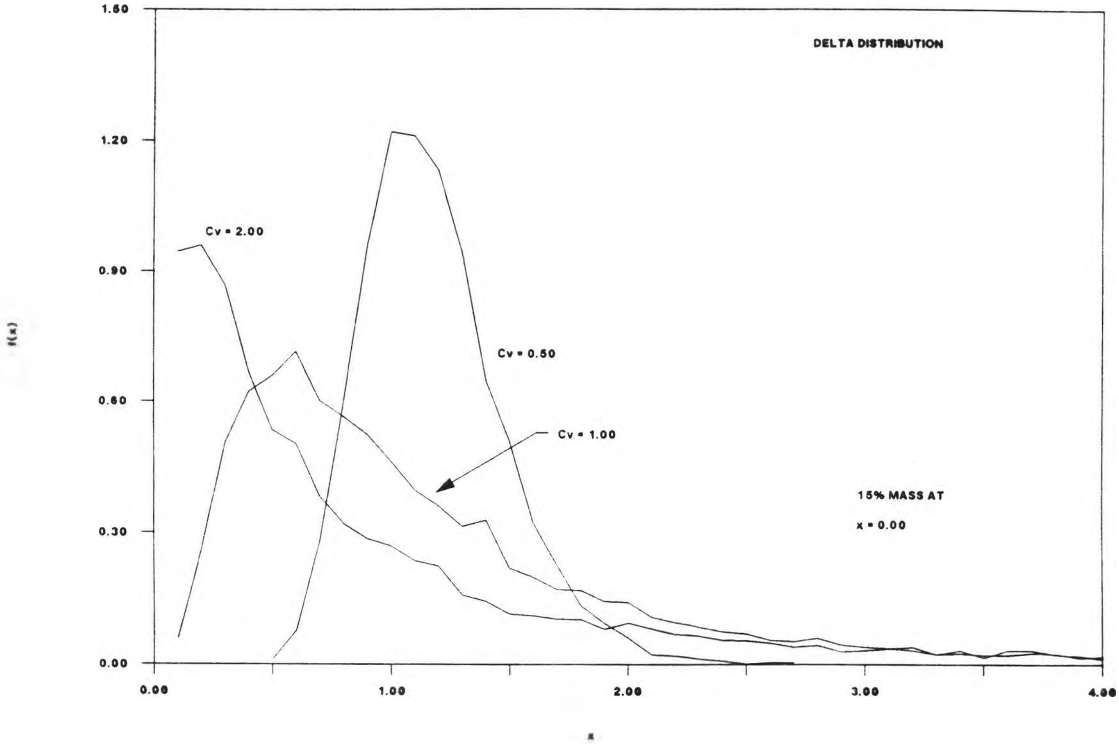

Figure A.1 Simulated probability density function for parent distributions used in simulation (e) delta distribution.

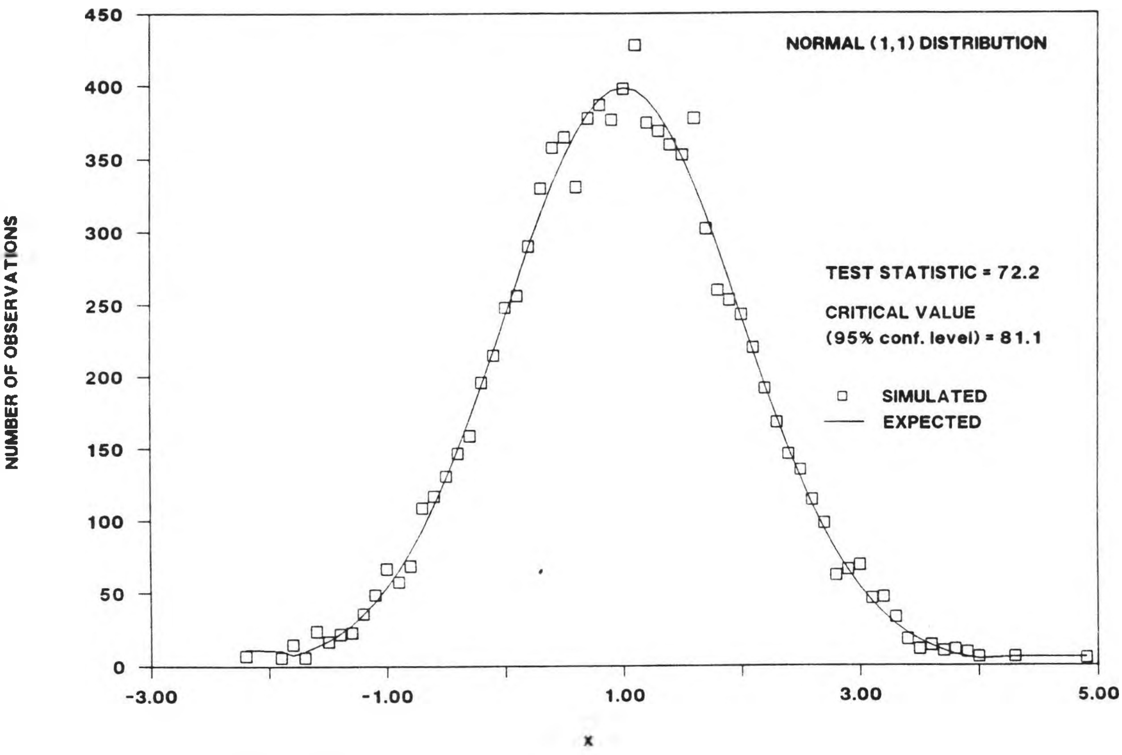

Figure A.2 Chi-square goodness of fit test for normal (mean=1.0, variance=1.0) distribution used in simulation.

normal (mean = 1.00, coefficient of variation = 1.00) distribution. Although the probability density functions were more jagged than expected, distributions did pass chi-square goodness of fit tests.

The data were checked for independence before transformation (Table A.1). The 95% confidence interval for autocorrelation for sample size equal to 10000 is (-0.0197, 0.0195). From inspection of Table A.1 there is a small negative correlation at lag-1 but the other lags were not significant.

Table A.1 Summary of autocorrelation coefficients for simulated data before transformation.

lag-k	r_k	lag-k	r_k
1	-0.0250*	6	-0.0018
2	-0.0116	7	-0.0126
3	0.0127	8	0.0110
4	-0.0059	9	0.0027
5	0.0039	10	-0.0022

* not independent at the 95% significance level

For each distribution, the mean and coefficient of variation was specified. The mean was always set equal to one. The coefficient of variation was set equal to 0.25, 0.50, 1.00, and 2.00 (which corresponds to a variance of 0.0625, 0.25, 1.00, and 4.00). It is not theoretically possible to simulate a delta distribution with 15% zeroes at the 0.25 coefficient of variation level. The mean and variance were computed for each simulation and were compared to the specified conditions (Table A.2 and A.3).

Table A.2 Summary of computed means for simulated data.

	Coefficient of Variation			
	0.25	0.50	1.00	2.00
	Variance			
Distribution	0.0625	0.25	1.00	4.00
Normal	0.9998	0.9995	0.9990	0.9981
Lognormal	1.0002	1.0014	1.0055	1.0140
Gamma	1.0041	1.0082	1.0043	1.0253
Mixed Lognormal	1.0004	1.0026	1.0077	1.0174
Delta	n/a	0.9988	0.9989	1.0009

Table A.3 Summary of computed variances for simulated data.

	Coefficient of Variation			
	0.25	0.50	1.00	2.00
	Variance			
Distribution	0.0625	0.25	1.00	4.00
Normal	0.0634	0.2535	1.0139	4.0554
Lognormal	0.0637	0.2564	1.0350	4.0226
Gamma	0.0640	0.2559	1.0058	4.3050*
Mixed Lognormal	0.0641	0.2632*	1.1039*	4.7316*
Delta	n/a	0.2507	1.0053	3.7983*

* significantly different from the specified variance at the 95 confidence level

All means fell within the 95% confidence level of the specified mean. Five of the variances did not fall within the 95% confidence level of the specified variance. The simulations which did not pass the variance test were from the mixed lognormal distribution or from the highest coefficient of variation level.

Upon closer scrutiny of the data, it was noticed that the higher than specified variances appeared to be the result of one extremely large number in the data set. For example, the largest simulated value for the mixed lognormal distribution at a coefficient of variation of 2.00 was 132.25. The variance due to this observation is $(132.25-1.0174)^2/10000$ or 1.72, 36% of the total variance.

The tests performed in this section indicate that even commercially available generators are not ideal. Developing a better generator, however, is beyond the scope of this study. Therefore, the simulation study will be run with the number generator described above.

APPENDIX B

GENERAL LINEAR MODEL DEVELOPMENT

SPSS was used to fit raw simulation data with a general linear model described in Chapter IV. Example submit files and computed coefficients are contained in this appendix. The first group of tables (B.1a through B.1h) contain the computed coefficients from the initial fit for rmse. The second group of tables (B.2a through B.2h) contain the computed coefficients from the refitted model for rmse. Table headings correspond to Equations IV.3 and IV.4. Technique code order refers to the coefficients indexed with j. The index i refers to the parent distributions used in simulation. In order, the parent distributions are: Normal, lognormal, gamma, mixed lognormal, and delta. A further discription of simulation procedures is given in Chapter IV and Appendix A.

Figure B.1 SPSS submit file used to create general linear model for average condition estimation with single censoring.

```
/JOB
JH61,T504.
/USER
/NOSEQ
ROUTE,OUTPUT,DEF,DC=WT,ID=0.
ATTACH,SPSSX/UN=LIBRARY.
SPSSX.
SKIP,DAYF.
   EXIT.
   DAYFILE,BLOWUP.
   REPLACE,BLOWUP.
   EXIT.
ENDIF,DAYF.
/EOR
TITLE RMSE -- MEAN -- SINGLE
DATA LIST FREE / ID CV P TECH SS BIAS RMSE
MISSING VALUES BIAS TO RMSE (-99.99)
BEGIN DATA
1  .25 .2  1  8   -.04782    .12509
1  .25 .2  2  8   -.05001    .13315

 ... all simulation data ...

5 2.00 .8  6 48 -99.99000 -99.99000
5 2.00 .8  7 48   -.21332    .35472
END DATA
COMPUTE CVSS = CV * SS
COMPUTE CVP  = CV * P
COMPUTE SSP  = SS * P
COMPUTE CVCV = CV * CV
COMPUTE SSSS = SS * SS
COMPUTE PP   = P  * P
MANOVA RMSE,CV,SS,P,CVSS,CVP,SSP,CVCV,SSSS,PP BY
TECH(1,7),ID(1,5)/
 METHOD = SSTYPE(UNIQUE)/
 ANALYSIS = RMSE/
 DESIGN =   ID,          TECH,          ID BY TECH,
            CV,          SS,            P,
            CVSS,        CVP,           SSP,
            CVCV,        SSSS,          PP,
            ID BY CV,    ID BY SS,      ID BY P,
            TECH BY CV,  TECH BY SS,    TECH BY P/
FINISH
/EOR
```

Figure B.2 SPSS submit file used to create general linear model for average condition estimation with multiple censoring.

```
/JOB
JH65,T504.
/USER
/NOSEQ
ROUTE,OUTPUT,DEF,DC=WT,ID=0.
ATTACH,SPSSX/UN=LIBRARY.
SPSSX.
SKIP,DAYF.
   EXIT.
   DAYFILE,BLOWUP.
   REPLACE,BLOWUP.
   EXIT.
ENDIF,DAYF.
/EOR
TITLE RMSE -- MEAN -- MULTIPLE
DATA LIST FREE / ID CV P1 P2 P3 TECH SS BIAS RMSE
MISSING VALUES BIAS TO RMSE (-99.99)
BEGIN DATA
1  .25 .15 .20 .25 12  1   -.0481     .1006
1  .25 .15 .20 .25 12  2   -.0899     .1321

 ... all simulation data ...

5 2.00 .30 .40 .20 48  7    .0215     .3200
5 2.00 .30 .40 .20 48  8    .0191     .3200
END DATA
COMPUTE P    = (P1 + P2 + P3) / 3.0
COMPUTE CVSS = CV * SS
COMPUTE CVP  = CV * P
COMPUTE SSP  = SS * P
COMPUTE CVCV = CV * CV
COMPUTE SSSS = SS * SS
COMPUTE PP   = P  * P
MANOVA RMSE,CV,SS,P,CVSS,CVP,SSP,CVCV,SSSS,PP BY
TECH(1,4),ID(1,5)/
 METHOD = SSTYPE(UNIQUE)/
 ANALYSIS = RMSE/
 DESIGN =    ID,          TECH,        ID BY TECH,
             CV,          SS,          P,
             CVSS,        CVP,         SSP,
             CVCV,        SSSS,        PP,
             ID BY CV,    ID BY SS,    ID BY P,
             TECH BY CV, TECH BY SS, TECH BY P/
FINISH
/EOR
```

Table B.1a General linear model coefficients used for rmse estimation (mean, single censoring).

Technique code order ABCDEFGHKNOPQ

	a_0	a_1	a_2	a_3	a_4
	.1820141	.1747887	-.0069182	-.2287104	-.0035658

	a_5	a_6	a_7	a_8	a_9
	-.0740795	.0009461	.0685827	.0001153	.5333083

i	a_0^i	a_1^i	a_2^i	a_3^i
1	-.1083777	.0853836	-.0000813	.1393306
2	.0240089	-.0335978	.0001369	-.0275813
3	.0440856	.0173827	-.0009580	-.0574932
4	.0132322	-.0412996	.0009951	-.0252809
5	.0270509	-.0278689	-.0000926	-.0289751

j	a_0^j	a_1^j	a_2^j	a_3^j
1	.0425729	.0197042	-.0005248	-.2582133
2	.1010015	-.1720475	.0001866	.3532876
3	.0221156	.0065147	.0001999	-.1330108
4	-.2737796	.0976228	.0011853	.4239972
5	.0221156	.0065147	.0001999	-.1330108
6	.0257048	.0028907	.0000162	-.1110686
7	-.0389282	.0247339	-.0009011	.0917661
8	-.0568542	.0750842	-.0017514	.0489875
9	.1818816	-.1129118	-.0014983	.2094371
10	.0221156	.0065147	.0001999	-.1330108
11	-.0050027	.0178330	.0014406	-.1960418
12	-.0074468	.0117881	.0011129	-.1619267
13	-.0354960	.0157579	.0001338	-.0011925

$a_0^{i,j}$	i=1	2	3	4	5
j=1	-.0440592	.0845455	-.0281890	-.0000165	-.0122806
2	.0118519	-.0267414	.0142582	.0057321	-.0051008
3	-.0077459	-.0174273	-.0154556	-.0121859	.0528149
4	.0124397	-.0525950	.0295033	.0183969	-.0077449
5	-.0651127	-.0000165	.0620558	.0309861	-.0279126
6	.0226174	.0057321	-.0185955	-.0123494	.0025953
7	-.0132688	-.0121859	-.0185786	.0237008	.0203325
8	.0405650	.0183969	-.0223310	-.0470110	.0103801
9	-.0000165	.0015806	-.0746189	.0335222	.0395326
10	.0057321	.0051968	-.0030003	-.0134068	.0054781
11	-.0121859	-.0126304	.0941785	.0204385	-.0898005
12	.0183969	.0168396	.0018804	-.0470963	.0099792
13	.0307861	-.0106948	-.0211072	-.0007106	.0017265

Table B.1b General linear model coefficients used for rmse estimation (standard deviation, single censoring).

Technique code order ABCDEFGHIKLNPQ

a_0	a_1	a_2	a_3	a_4
.8408905	.0000000	-.0130588	-1.7194731	.0000000

a_5	a_6	a_7	a_8	a_9
-.5853457	.0242020	.2066096	.0000000	2.2075150

i	a_0^i	a_1^i	a_2^i	a_3^i
1	.1231546	-.2169104	-.0012703	-.1756169
2	.0009810	-.0281898	-.0002037	.0981583
3	.4294112	.1395655	-.0076060	-.6736098
4	-.0962218	-.0016720	.0029722	.4407964
5	-.4573250	.1072067	.0061079	.3102720

j	a_0^j	a_1^j	a_2^j	a_3^j
1	.0161183	.0366649	-.0019401	-.5213416
2	.5674953	-.5016166	-.0002598	.0578457
3	.0895734	-.0116850	-.0016693	-.4108071
4	.0244924	-.0223013	-.0008561	-.3072366
5	.1866094	-.1043964	-.0012027	-.3886945
6	.1714640	-.0919996	-.0009946	-.3988239
7	.1144323	-.0842524	-.0016407	-.2649115
8	.0605052	-.0038391	-.0021610	-.3570477
9	-2.2710173	.1035642	.0297118	4.3214199
10	1.4774011	.7050090	-.0210289	-2.4813513
11	-.7627444	.2512468	.0046572	1.3471846
12	.1272327	-.0672064	-.0005031	-.4136908
13	.0720530	-.1680416	.0002995	.2280081
14	.1263841	-.0411462	-.0024119	-.4105531

$a_0^{i,j}$	i=1	2	3	4	5
j=1	.0108632	.0140594	-.0961778	-.3846935	.4559486
2	.0319162	-.1079353	.0300248	-.3199987	.3659930
3	-.0984903	-.0423209	.0960984	.1181209	-.0734081
4	.0012484	.0927203	.0123034	-.0937113	-.0125609
5	.0912708	.0486280	-.1202926	.2227413	-.2423476
6	.0735733	-.0736557	-.0169084	.0905762	-.0735853
7	-.0302411	-.0143158	-.4457697	.0434055	.4469211
8	-.0330475	.0935191	-.1360013	-.0804923	.1560221
9	.0953467	.0435523	-.3550061	-.0135512	.2296582
10	.0231778	-.0778114	.4277996	.1229234	-.4960894
11	-.1000057	-.0167060	-.1329215	-.0459094	.2955427
12	-.0231015	.0218879	-.2858487	-.0563340	.3433964
13	.1154655	.0413219	1.4059025	-.1174235	-1.4452665
14	-.1579760	-.0229439	-.3832023	.5143465	.0497757

Table B.1c General linear model coefficients used for rmse estimation (median, single censoring).

Technique code order ABCDEFGHJKMN

a_0	a_1	a_2	a_3	a_4
.3887300	.0000000	-.0086313	-1.0173485	-.0082976

a_5	a_6	a_7	a_8	a_9
.6153879	.0000000	.1123042	.0001992	1.1895217

i	a_0^i	a_1^i	a_2^i	a_3^i
1	.1237978	-.1546858	.0015482	-.1848549
2	.0993362	-.1390957	.0008475	-.0951695
3	-.4227927	.5993154	-.0064439	.6007221
4	.1039448	-.1968648	.0020257	-.1274433
5	.0957136	-.1086690	.0020223	-.1932543

j	a_0^j	a_1^j	a_2^j	a_3^j
1	.2874493	-.0467393	-.0006526	-.7815379
2	.1115138	-.2230251	.0022774	.5615463
3	.2313774	-.1089751	.0013451	-.4443735
4	-.5772417	.2959216	.0017093	.6205721
5	-.0278390	.0637680	.0003075	-.2503930
6	.0902630	-.0728401	.0015878	-.1044448
7	-.1764944	.1940338	-.0041742	.3752879
8	-.1196379	.2049376	-.0061009	.1708430
9	.0459968	-.0532727	.0003076	-.2878211
10	.1710447	-.2742422	.0021007	.5940044
11	-.0535192	-.0152181	.0006846	-.1396165
12	.0170872	.0356517	.0006076	-.3140668

$a_0^{i,j}$	i=1	2	3	4	5
j=1	.0673341	.0085585	.0130747	.0088256	-.0977930
2	.0020890	-.0245511	-.0014144	-.0785766	.1024532
3	-.0816218	-.0325250	-.1068990	.0233407	.1977052
4	-.0031239	.1999305	.0288745	.0376612	-.2633424
5	.0290517	-.0823599	.0516164	.0399263	-.0382346
6	.0257689	-.0111532	.0295747	-.2094032	.1652127
7	-.1480952	-.0187959	-.0401034	.0661704	.1408241
8	.0464361	.0799014	-.0170316	.0308197	-.1401256
9	-.0020883	-.0294962	.1347548	-.0041485	-.0990216
10	-.0090800	-.0040960	-.0358249	-.0367287	.0857298
11	.0059023	-.0062524	.0346847	.0020085	-.0363433
12	.0674269	-.0791604	-.0913065	.1201044	-.0170644

Table B.1d General linear model coefficients used for rmse estimation (interquartile range, single censoring).

Technique code order ABCDEFGHJKMN

a_0	a_1	a_2	a_3	a_4
2.7056133	-3.0309439	-.0283407	3.0847608	-.0065492

a_5	a_6	a_7	a_8	a_9
-.3662960	.0043540	1.1412402	.0003904	-2.5254763

i	a_0^i	a_1^i	a_2^i	a_3^i
1	-.2496417	.0217303	.0054589	-.2803075
2	-.3204974	.2646025	-.0014890	.0121535
3	-.3999839	.2044177	.0004503	-.0496013
4	.6035518	-.3707734	-.0027867	.3853388
5	.3665713	-.1199771	-.0016335	-.0675834

j	a_0^j	a_1^j	a_2^j	a_3^j
1	-.5161096	.5855609	-.0063571	-.3176489
2	1.9676815	-.9995144	-.0029791	.1555308
3	.4313566	-.0127422	-.0017947	-.6337528
4	-1.0704680	.7617439	.0003161	-.5373368
5	.0401712	-.2919238	-.0059330	1.1494709
6	.4048313	-.2035033	-.0004608	.1093043
7	-.3942261	.1746421	-.0013368	.7852722
8	-.6481353	.5079943	-.0029218	-.3613596
9	-.9322408	.4086909	.0044451	-.4093707
10	1.4857113	-1.0792482	.0161174	-.1548606
11	-.9541997	.4045420	.0051310	-.4094727
12	.1856275	-.2562421	-.0042261	.6242239

$a_0^{i,j}$	i=1	2	3	4	5
j=1	.2499017	-.0777720	.0100247	.0534516	-.2356061
2	-.0553098	.1259666	-.0327982	.0146986	-.0525571
3	.0280090	.0731462	-.2572043	.0492116	.1068374
4	-.4521670	.1681427	.0573342	.0288771	.1978129
5	.0011755	-.3393704	-.0336385	-.1093171	.4811506
6	-.0970172	.0159129	.2931744	-.1613882	-.0506818
7	-.1311184	-.0178954	-.1343065	.2589958	.0243245
8	.2179145	-.0005219	.0794410	-.0931392	-.2036944
9	.0587707	-.0143890	.0589837	.0585046	-.1618699
10	-.0056876	.0475650	.0334341	.0074936	-.0828052
11	.0348738	-.0201030	-.0631069	.0836667	-.0353306
12	.1506547	.0393183	-.0113377	-.1910552	.0124199

Table B.1e General linear model coefficients used for rmse estimation (mean, multiple censoring).

Technique code order ACEFGH

a_0	a_1	a_2	a_3	a_4
.1496996	.0986265	-.0029488	-.1559616	-.0020938

a_5	a_6	a_7	a_8	a_9
-.0239551	.0017617	.0689191	.0000286	.3933840

i	a_0^i	a_1^i	a_2^i	a_3^i
1	-.1055179	.0934519	.0000618	.1118698
2	.0290678	-.0265451	-.0000686	-.0313235
3	.0316919	-.0239412	-.0005605	-.0061672
4	-.0039922	-.0240044	.0012088	-.0266285
5	.0487504	-.0189611	-.0006414	-.0477505

j	a_0^j	a_1^j	a_2^j	a_3^j
1	.0601225	.0167172	-.0005314	-.2694637
2	.0350089	.0055463	.0000420	-.1335416
3	.0453085	.0021548	-.0000449	-.1404135
4	.0350089	.0055463	.0000420	-.1335416
5	-.0065007	.0228872	-.0003608	-.0641881
6	-.0020574	.0173214	-.0001835	-.0821608

$a_0^{i,j}$	i=1	2	3	4	5
j=1	-.0466609	.0319563	-.0029004	.0025520	.0150529
2	.0078217	.0025520	.0009106	.0088691	-.0201536
3	-.0588465	.0088691	.0024332	-.0034283	.0509724
4	.0192104	.0931474	.0102507	-.0074725	-.1151362
5	.0005381	-.0317756	.0021340	.0003713	.0287322
6	.0636355	-.1092481	.0498279	-.0111522	.0069368

Table B.1f General linear model coefficients used for rmse estimation (standard deviation, multiple censoring).

Technique code order ABCDEFGH

a_0	a_1	a_2	a_3	a_4
.4602068	-.5061884	.0022649	.3205383	.0026516

a_5	a_6	a_7	a_8	a_9
-.1581597	.0000000	.2651278	-.0001078	.0000000

i	a_0^i	a_1^i	a_2^i	a_3^i
1	-.0048242	-.1697686	-.0003058	.0910221
2	.0037278	.0517460	.0001281	-.0134125
3	.0909749	-.1107311	-.0022578	.0064249
4	.0033111	.1139036	.0026615	-.0192362
5	-.0931896	.1148500	-.0002259	-.0647983

j	a_0^j	a_1^j	a_2^j	a_3^j
1	-.1334891	.1213068	-.0003316	-.1699511
2	.4195373	-.3113714	.0009658	.0793532
3	-.0687922	.0640501	.0003273	-.0643385
4	-.1694753	.0514459	.0003417	.2456230
5	.1487335	-.0825526	-.0010543	-.0012044
6	-.0421648	.0134524	.0011588	-.0385717
7	-.0877841	.0757835	-.0006540	-.0020816
8	-.0665650	.0678852	-.0007538	-.0488288

$a_0^{i,j}$	i=1	2	3	4	5
j=1	-.0632146	.0020634	-.0337870	.0160444	.0788937
2	.0037798	.0048140	-.0559827	.0143878	.0330008
3	-.0023717	-.0022255	.0090876	.0123978	-.0168882
4	.0210150	-.0035320	.0284187	-.0061444	-.0397573
5	.0064764	.0046040	.0313890	-.0228472	-.0196223
6	.0420920	.0570344	.0102883	-.0360532	-.0733615
7	.0562535	-.0447041	.0016736	.0032960	-.0165190
8	-.0640306	-.0180542	.0089123	.0189186	.0542538

Table B.1g General linear model coefficients used for rmse estimation (median, multiple censoring).

Technique code order ACEFGHJ

a_0	a_1	a_2	a_3	a_4
.3683246	-.1045873	-.0084920	-.6464120	-.0049956

a_5	a_6	a_7	a_8	a_9
.5586421	.0043282	.1316690	.0001229	.6643843

i	a_0^i	a_1^i	a_2^i	a_3^i
1	.1037759	-.1494354	.0005444	-.1125652
2	.0827532	-.1328215	.0007798	-.0797634
3	-.4198583	.5727731	-.0038015	.5150790
4	.0802093	-.1818777	.0013885	-.0923200
5	.1531197	-.1086384	.0010887	-.2304302

j	a_0^j	a_1^j	a_2^j	a_3^j
1	.3134511	-.0761073	-.0011573	-.7038873
2	.0807529	-.0277775	.0010942	-.1952838
3	-.0627254	.0840159	-.0008730	-.0894037
4	.0174888	.0330346	-.0001172	-.2175472
5	.1150849	-.0762182	-.0012644	-.1431201
6	.1350022	-.0919481	-.0007708	-.1931697
7	.1823581	-.1210657	-.0003087	-.3746352

$a_0^{i,j}$	i=1	2	3	4	5
j=1	.0742275	-.0231395	-.0114901	-.0062640	-.0333338
2	-.1049757	.1057976	.0920951	-.1060837	.0131667
3	.0296072	-.0706035	-.0133553	-.0027258	.0570775
4	.0353947	-.0551402	-.0173650	.0314273	.0056832
5	-.1761218	.3757531	.0804852	-.1176955	-.1624209
6	.0593199	-.1318575	-.0215204	.0539606	.0400974
7	.0654978	-.1588095	-.0604791	.0613939	.0923968

Table B.1h General linear model coefficients used for rmse estimation (interquartile range, multiple censoring).

Technique code order ACDEFGHJ

a_0	a_1	a_2	a_3	a_4
2.7007163	-2.2255022	-.0366739	.8305174	-.0028236

a_5	a_6	a_7	a_8	a_9
-.6197755	.0000000	.8785374	.0004697	.0000000

i	a_0^i	a_1^i	a_2^i	a_3^i
1	-.2856246	.0599664	.0041937	-.1517067
2	-.3509710	.2231128	-.0007244	.0898660
3	-.3884065	.1857330	.0010709	-.0208598
4	.2650430	-.2505140	-.0034769	.6455665
5	.7599591	-.2182983	-.0010633	-.5628660

j	a_0^j	a_1^j	a_2^j	a_3^j
1	-.3323192	.3909936	-.0012457	-.2566763
2	.8268709	-.2762565	.0015775	-.8721538
3	-.6217549	.5797965	.0020234	-1.1356002
4	.6989393	-.6324707	-.0159471	1.8197873
5	.4409875	-.4526168	.0013218	.5212956
6	-.4764656	.1378299	.0024723	.4845317
7	-.3243981	.1015420	.0021915	.3716455
8	-.2118598	.1511819	.0076061	-.9328299

$a_0^{i,j}$	i=1	2	3	4	5
j=1	.2539450	.0576492	-.0443421	.0616189	-.3288711
2	-.0473738	.1100042	.1386311	.0151576	-.2164191
3	.0162691	.0797333	.0159696	.1370234	-.2489956
4	-.3449075	.1518214	-.0227345	-.1244777	.3402984
5	.0549698	-.2307446	-.0176665	.0234615	.1699797
6	-.0193851	-.0107204	.0899711	-.0171316	-.0427338
7	.0034009	-.0362614	-.1139036	.1061253	.0406387
8	.0830815	-.1214817	-.0459249	-.2017776	.2861028

Table B.2a Refitted general linear model coefficients used for rmse estimation (mean, single censoring).

Technique code order CFGHOPQ

a_0	a_1	a_2	a_3	a_4
.1510890	.1993221	-.0061090	-.2362415	-.0033320

a_5	a_6	a_7	a_8	a_9
.0000000	-.0014914	.0499911	.0001181	.4593365

i	a_0^i	a_1^i	a_2^i	a_3^i
1	-.1149966	.1131777	-.0007180	.1628219
2	.0297823	-.0354748	.0002048	-.0439621
3	.0227618	.0064076	-.0007158	-.0122807
4	.0188874	-.0533993	.0012989	-.0408245
5	.0435650	-.0307112	-.0000698	-.0657544

j	a_0^j	a_1^j	a_2^j	a_3^j
1	.0347268	-.0177722	.0002341	-.0626204
2	.0383159	-.0213962	.0000504	-.0406781
3	-.0263170	.0004469	-.0008669	.1621565
4	-.0442431	.0507973	-.0017172	.1193780
5	.0114231	.0012490	.0012296	-.1407744
6	.0089790	-.0047958	.0009019	-.1066594
7	-.0228848	-.0085290	.0001680	.0691978

$a_0^{i,j}$	i=1	2	3	4	5
j=1	-.0149662	-.0094529	.0471061	.0314392	-.0541262
2	.0085425	.0228964	-.0157851	-.0426110	.0269571
3	-.0090084	-.0431386	-.0154010	.0212570	.0462912
4	.0244537	.0170686	-.0162742	-.0123369	-.0129113
5	-.0133689	-.0122781	.0187209	.0281768	-.0212507
6	.0080072	.0355601	-.0112795	-.0426963	.0104084
7	-.0036599	-.0106555	-.0070871	.0167711	.0046314

Table B.2b Refitted general linear model coefficients used for rmse estimation (standard deviation, single censoring).

Technique code order CDEFGHNQ

a_0	a_1	a_2	a_3	a_4
.4692932	-.3761024	-.0045601	.3902981	.0020552

a_5	a_6	a_7	a_8	a_9
-.2396035	.0000000	.2459454	.0000000	.0000000

i	a_0^i	a_1^i	a_2^i	a_3^i
1	-.0309221	-.1240746	.0001347	.0375383
2	-.0035898	.0320067	.0001552	.0403320
3	.0607574	-.0964824	-.0024797	.0540348
4	.0590004	.0661385	.0019855	.0278147
5	-.0852459	.1224118	.0002042	-.1597200

j	a_0^j	a_1^j	a_2^j	a_3^j
1	-.0291411	.0416682	.0000000	-.0418364
2	-.0734039	.0310520	.0000000	.0617340
3	.0798401	-.0510431	.0000000	-.0197238
4	.0700203	-.0386463	.0000000	-.0298532
5	-.0035494	-.0308990	.0000000	.1040591
6	-.0707963	.0495141	.0000000	.0119229
7	.0383724	-.0138530	.0000000	-.0447201
8	-.0113420	.0122071	.0000000	-.0415824

$a_0^{i,j}$	i=1	2	3	4	5
j=1	.0111427	-.0277428	.0187035	-.0182151	.0161117
2	-.0073407	.0085163	-.0021279	-.0237777	.0247301
3	-.0034907	.0181094	-.0623161	-.0023303	.0500278
4	-.0085234	.0228591	.0108033	.0063722	-.0315113
5	.0312615	.0002622	.0003370	.0128869	-.0447478
6	-.0164591	.0093151	.0446029	.0160226	-.0534815
7	-.0114204	.0130336	.0118944	.0010268	-.0145345
8	.0048303	-.0443531	-.0218971	.0080146	.0534054

Table B.2c Refitted general linear model coefficients used for rmse estimation (median, single censoring).

Technique code order CEJMN

a_0	a_1	a_2	a_3	a_4
.5579097	-.1984227	-.0083117	-1.4876181	-.0065485

a_5	a_6	a_7	a_8	a_9
.6453387	0.0000000	.1635204	.0001742	1.3394007

i	a_0^i	a_1^i	a_2^i	a_3^i
1	.1277168	-.1484401	.0012792	-.1774216
2	.1237581	-.1441697	.0005586	-.1406127
3	-.5430540	.6209200	-.0047485	.7252665
4	.1301842	-.2011099	.0013950	-.1520154
5	.1613947	-.1272002	.0015155	-.2552167

j	a_0^j	a_1^j	a_2^j	a_3^j
1	.2065398	-.0933659	0.0000000	-.1571192
2	-.0792407	.0793772	0.0000000	.0368611
3	-.0054023	-.0376634	0.0000000	-.0005669
4	-.0952663	.0003911	0.0000000	.1476377
5	-.0266304	.0512609	0.0000000	-.0268126

$a_0^{i,j}$	i=1	2	3	4	5
j=1	-.0104089	-.0194737	.0263642	.0224991	-.0189808
2	-.0009100	-.0106260	.0169955	.0040214	-.0094809
3	-.0020463	.0719527	-.0865254	-.0446774	.0612965
4	.0124501	-.0256046	.0272323	.0059001	-.0199781
5	.0009151	-.0162483	.0159332	.0122566	-.0128567

Table B.2d Refitted general linear model coefficients used for rmse estimation (interquartile range, single censoring).

Technique code order DHJM

a_0	a_1	a_2	a_3	a_4
2.0261195	-.7953677	-.0268056	-1.6143531	-.0063598

a_5	a_6	a_7	a_8	a_9
-.0805426	.0075895	.3542731	.0003620	1.3860652

i	a_0^i	a_1^i	a_2^i	a_3^i
1	-.1221213	-.0091721	.0042980	-.4491667
2	-.1696638	.1655537	-.0013948	.0786303
3	-.2325272	.1166352	.0002037	-.0276558
4	.5025133	-.3305325	-.0023446	.4577816
5	.0217990	.0575157	-.0007623	-.0595893

j	a_0^j	a_1^j	a_2^j	a_3^j
1	-.1690473	.2419896	-.0014076	-.1104064
2	.2530724	-.0130780	-.0046707	.0688435
3	-.0310330	-.1123813	.0026961	.0208324
4	-.0529919	-.1165302	.0033821	.0207304

$a_0^{i,j}$	i=1	2	3	4	5
j=1	.1687129	-.2964493	-.0032315	-.0125698	.1435377
2	.0066670	-.0936932	.0768131	-.0475166	.0577297
3	.1054697	.0134196	-.0224937	.0925905	-.1889861
4	-.2808497	.3767230	-.0510878	-.0325040	-.0122813

Table B.2e Refitted general linear model coefficients used for rmse estimation (mean, multiple censoring).

Technique code order CEGH

a_0	a_1	a_2	a_3	a_4
.1659255	.0949296	-.0031306	-.2271387	-.0021865

a_5	a_6	a_7	a_8	a_9
-.0563492	.0017671	.0851162	.0000306	.3821035

i	a_0^i	a_1^i	a_2^i	a_3^i
1	-.0883374	.1243963	-.0001041	.0468303
2	.0225140	-.0334848	-.0000342	-.0112569
3	.0307538	-.0287313	-.0005866	-.0017399
4	-.0153636	-.0392066	.0013753	.0208536
5	.0504332	-.0229736	-.0006503	-.0546871

j	a_0^j	a_1^j	a_2^j	a_3^j
1	.0224347	-.0064311	0.0000000	-.0284655
2	.0301261	-.0098226	0.0000000	-.0353375
3	-.0311617	.0109097	0.0000000	.0408879
4	-.0213991	.0053439	0.0000000	.0229151

$a_0^{i,j}$	i=1	2	3	4	5
j=1	-.0004224	.0037160	.0043892	-.0034763	-.0042064
2	.0025040	-.0016458	.0050976	-.0055164	-.0004393
3	.0043207	.0032143	.0053392	-.0047818	-.0080926
4	-.0064023	-.0052846	-.0148261	.0137746	.0127384

Table B.2f Refitted general linear model coefficients used for rmse estimation (standard deviation, multiple censoring).

Technique code order CDEFGH

	a_0	a_1	a_2	a_3	a_4
	.3947077	-.3513421	0.0000000	.3417239	.0029143

	a_5	a_6	a_7	a_8	a_9
	-.1686456	0.0000000	.2114886	-.0000768	0.0000000

i	a_0^i	a_1^i	a_2^i	a_3^i
1	-.0334226	-.1399439	-.0001197	.1069051
2	.0005883	.0504820	.0000577	-.0171466
3	.0830574	-.1111812	-.0022830	.0053400
4	.0066625	.1093768	.0025911	-.0222381
5	-.0568857	.0912664	-.0002460	-.0728604

j	a_0^j	a_1^j	a_2^j	a_3^j
1	-.0209381	.0326688	.0004351	-.0801781
2	-.1226978	.0182877	.0004369	.2342230
3	.1965876	-.1139339	-.0009465	-.0170440
4	.0056892	-.0179288	.0012666	-.0544112
5	-.0399300	.0444022	-.0005462	-.0179211
6	-.0187108	.0365039	-.0006460	-.0646684

$a_0^{i,j}$	i=1	2	3	4	5
j=1	-.0039747	-.0363092	.0402194	.0160943	-.0160297
2	.0052698	-.0240570	.0139848	-.0149331	.0197355
3	.0052982	-.0513866	-.0071151	-.0153518	.0685553
4	.0083005	.0002988	.0235398	-.0272228	-.0049164
5	.0442401	.0359141	.0232182	.0069925	-.1103650
6	-.0591341	.0755398	-.0938471	.0344210	.0430204

Table B.2g Refitted general linear model coefficients used for rmse estimation (median, multiple censoring).

Technique code order CEFGHJ

	a_0	a_1	a_2	a_3	a_4
	.4774785	-.1267160	-.0086340	-1.0658681	-.0054042

	a_5	a_6	a_7	a_8	a_9
	.4903879	.0030558	.1472811	.0001362	.9793554

i	a_0^i	a_1^i	a_2^i	a_3^i
1	.1042502	-.1433410	.0004522	-.1257768
2	.0722293	-.1223046	.0007951	-.0773754
3	-.4095634	.5494336	-.0039348	.5128087
4	.0622480	-.1754284	.0014956	-.0555838
5	.1708358	-.1083594	.0011918	-.2540726

j	a_0^j	a_1^j	a_2^j	a_3^j
1	.0467878	.0055489	0.0000000	.0069094
2	-.1557111	.1173425	0.0000000	.1127895
3	-.0528207	.0663611	0.0000000	-.0153539
4	.0103591	-.0428917	0.0000000	.0590731
5	.0450825	-.0586215	0.0000000	.0090236
6	.1063024	-.0877392	0.0000000	-.1724418

$a_0^{i,j}$	i=1	2	3	4	5
j=1	-.0176009	-.0196049	.0962611	.0427902	-.1018455
2	-.0275655	.1078710	-.0321277	.0028127	-.0509905
3	.1215735	-.0413040	-.0007255	.0294956	-.1090396
4	-.0269734	-.0078168	.0306581	-.1019196	.1060517
5	-.0059516	-.0192966	-.0903077	.0433534	.0722026
6	-.0434819	-.0198485	-.0037583	-.0165324	.0836212

Table B.2h Refitted general linear model coefficients used for rmse estimation (interquartile range, multiple censoring).

Technique code order CDFGHJ

a_0	a_1	a_2	a_3	a_4
2.8500521	-2.2716073	-.0328404	0.0000000	0.0000000

a_5	a_6	a_7	a_8	a_9
-.3528500	.0096900	.8226335	.0003275	0.0000000

i	a_0^i	a_1^i	a_2^i	a_3^i
1	-.3107297	.0429988	.0045033	-.1704193
2	-.3547985	.2212690	-.0010523	.1572934
3	-.3988363	.1893070	.0004752	.0471616
4	.2827026	-.2804752	-.0027879	.7026065
5	.7816620	-.1730996	-.0011383	-.7366422

j	a_0^j	a_1^j	a_2^j	a_3^j
1	.8875421	-.3171066	-.0012807	-.6104085
2	-.5584906	.5425703	-.0008780	-.8812158
3	.5016587	-.4934670	-.0015364	.7830409
4	-.4157945	.0969797	-.0003859	.7462770
5	-.2637270	.0606918	-.0006667	.6333909
6	-.1511886	.1103317	.0047479	-.6710846

$a_0^{i,j}$	i=1	2	3	4	5
j=1	.0946916	.0649475	-.0221760	.1028133	-.2402765
2	-.0331551	.1462960	.0557610	-.0847559	-.0841460
3	-.0011085	-.2659706	-.0741818	.0096916	.3315693
4	.0234391	.0556914	.0478490	-.0216411	-.1053385
5	.1546199	-.0365044	.0106481	.0719153	-.2006788
6	-.2384871	.0355400	-.0179002	-.0780232	.2988706

APPENDIX C

TREND DETECTION IN THE PRESENCE OF SERIAL CORRELATION

This appendix contains the actual significance levels (slope = 0.0) and power (slope > 0.0) from the Monte Carlo simulation used to investigate the effect of autocorrelation on trend tests (Chapter V). The Mann-Kendall tau test on deaseasonalized data (MKD), the Seasonal Kendall tau tests (SK), and the Seasonal Kendall tau test with correction for autocorrelation (SKC) were applied to monthly data. Monthly data were collapsed to quarterly values by computing the mean and median and selecting the middle value of each quarter. The three tests were then reapplied to the three sets of quarterly collapsed data. Finally, quarterly means and medians were collapsed to yearly means and medians, respectively. The standard Mann-Kendall tau test (MK) was then applied to this data. The values in the following tables were scaled by 1000 for convenience.

Table C.1 Comparison of significance level (slope = 0) and power (slope > 0) for trend tests. The errors are from a normal distribution with ρ = 0.0 and σ_r = 1.0. Values in table are scaled by 1000.

	MONTHLY			QUARTERLY									YEARLY	
				MEAN			MEDIAN			MIDDLE			MEAN	MED
slope	MKD	SK	SKC	MKD	SK	SKC	MKD	SK	SKC	MKD	SK	SKC	MK	MK
5 years of record														
0.000	88	42	10	70	34	8	74	36	6	60	30	8	16	10
0.002	74	40	6	62	40	10	58	34	8	56	46	12	34	14
0.005	122	68	10	114	64	6	94	64	4	50	36	4	26	20
0.020	756	600	100	684	540	108	584	442	70	284	186	38	176	138
0.050	1000	1000	578	1000	1000	734	1000	994	608	906	838	238	742	606
0.200	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.500	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
10 years of record														
0.000	64	44	18	58	40	30	56	40	32	44	34	30	48	34
0.002	110	90	60	104	88	62	86	76	54	74	58	36	76	76
0.005	490	422	294	460	390	296	356	292	226	184	140	110	324	244
0.020	1000	1000	1000	1000	1000	1000	1000	1000	1000	970	952	894	1000	992
0.050	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.200	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.500	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
15 years of record														
0.000	54	42	32	70	48	46	54	38	34	44	38	44	44	32
0.002	306	260	222	292	274	224	242	180	176	132	122	96	250	192
0.005	922	896	826	898	850	830	812	766	706	488	426	378	832	668
0.020	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.050	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.200	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.500	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000

Table C.2 Comparison of significance level (slope = 0) and power (slope > 0) for trend tests. The errors are from a normal distribution with ρ = 0.0 and σ_r = 5.0. Values in table are scaled by 1000.

	MONTHLY			QUARTERLY									YEARLY	
				MEAN			MEDIAN			MIDDLE			MEAN	MED
slope	MKD	SK	SKC	MKD	SK	SKC	MKD	SK	SKC	MKD	SK	SKC	MK	MK
5 years of record														
0.000	80	46	14	82	44	6	74	40	2	66	44	6	24	16
0.002	68	32	16	78	42	6	60	30	14	52	18	6	14	18
0.005	80	52	20	70	48	6	86	40	6	64	42	6	16	30
0.020	332	200	14	272	172	24	230	158	32	124	88	14	18	30
0.050	930	850	256	790	678	244	726	602	164	508	354	76	134	188
0.200	1000	1000	1000	1000	1000	974	1000	1000	960	990	986	800	854	870
0.500	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	994	1000	996
10 years of record														
0.000	68	56	38	48	36	26	50	36	48	42	30	38	40	52
0.002	74	56	58	72	50	54	82	66	56	66	56	54	44	40
0.005	224	184	106	186	140	104	166	114	76	110	82	68	58	66
0.020	984	980	930	962	936	906	922	874	762	650	586	506	396	558
0.050	1000	1000	1000	1000	1000	1000	1000	1000	1000	998	1000	994	990	994
0.200	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.500	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
15 years of record														
0.000	62	56	42	72	64	40	70	48	46	56	40	40	58	62
0.002	136	108	80	122	104	108	102	88	88	72	60	66	52	76
0.005	530	456	378	476	410	386	388	324	284	228	190	164	116	182
0.020	1000	1000	1000	1000	1000	1000	1000	998	998	984	978	970	926	970
0.050	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.200	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.500	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000

Table C.3 Comparison of significance level (slope = 0) and power (slope > 0) for trend tests. The errors are from a lognormal distribution with ρ = 0.0 and σ_r = 1.0. Values in table are scaled by 1000.

	MONTHLY			QUARTERLY									YEARLY	
				MEAN			MEDIAN			MIDDLE			MEAN	MED
slope	MKD	SK	SKC	MKD	SK	SKC	MKD	SK	SKC	MKD	SK	SKC	MK	MK
5 years of record														
0.000	78	50	6	64	36	10	48	26	8	46	32	6	8	8
0.002	96	58	12	68	34	4	74	36	16	66	34	10	26	20
0.005	218	134	24	122	72	8	146	92	20	114	70	14	36	32
0.020	978	962	414	808	696	214	896	814	290	634	530	134	200	304
0.050	1000	1000	976	1000	1000	874	1000	1000	950	972	960	668	750	858
0.200	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	998	998	1000
0.500	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
10 years of record														
0.000	76	70	60	80	56	44	76	56	36	68	66	50	48	50
0.002	292	270	200	166	140	106	206	186	146	124	112	80	96	162
0.005	928	918	786	686	624	476	804	730	616	494	454	340	398	542
0.020	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	998	1000	1000
0.050	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.200	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.500	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
15 years of record														
0.000	62	56	48	64	46	52	54	36	48	66	48	48	54	38
0.002	730	730	644	446	386	342	574	524	462	328	278	272	260	360
0.005	1000	1000	1000	986	982	966	1000	1000	994	902	904	862	884	968
0.020	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.050	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.200	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.500	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000

Table C.4 Comparison of significance level (slope = 0) and power (slope > 0) for trend tests. The errors are from a lognormal distribution with ρ = 0.0 and σ_r = 5.0. Values in table are scaled by 1000.

	MONTHLY			QUARTERLY									YEARLY	
				MEAN			MEDIAN			MIDDLE			MEAN	MED
slope	MKD	SK	SKC	MKD	SK	SKC	MKD	SK	SKC	MKD	SK	SKC	MK	MK
5 years of record														
0.000	84	40	6	50	32	6	74	38	8	46	40	8	30	20
0.002	74	40	22	70	52	16	68	46	8	54	32	10	16	22
0.005	124	88	10	74	44	12	102	58	14	76	42	14	10	22
0.020	696	598	136	416	288	52	462	330	66	262	184	38	32	70
0.050	990	990	626	834	782	376	888	856	452	698	634	250	112	246
0.200	1000	1000	998	1000	1000	996	1000	1000	998	996	998	926	816	966
0.500	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	988	998
10 years of record														
0.000	54	46	42	52	42	34	58	48	44	38	40	34	50	40
0.002	178	136	104	102	60	48	118	92	76	96	60	64	54	68
0.005	502	462	366	268	236	216	370	318	230	182	146	114	76	114
0.020	1000	1000	1000	986	990	968	986	982	980	916	904	846	482	806
0.050	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	984	1000
0.200	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.500	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
15 years of record														
0.000	48	34	32	42	36	38	52	46	46	58	44	48	38	52
0.002	388	362	326	240	198	182	256	222	214	180	168	144	68	98
0.005	920	916	874	686	654	596	780	764	712	496	484	448	162	314
0.020	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	958	1000
0.050	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.200	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.500	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000

Table C.5 Comparison of significance level (slope = 0) and power (slope > 0) for trend tests. The errors are from a normal distribution with ρ = 0.2 and σ_r = 1.0. Values in table are scaled by 1000.

	MONTHLY			QUARTERLY									YEARLY	
				MEAN			MEDIAN			MIDDLE			MEAN	MED
slope	MKD	SK	SKC	MKD	SK	SKC	MKD	SK	SKC	MKD	SK	SKC	MK	MK
5 years of record														
0.000	116	84	10	80	42	10	82	38	6	64	46	12	14	20
0.002	128	70	2	92	44	6	78	32	12	64	26	2	20	10
0.005	150	110	10	98	54	10	88	66	10	82	66	12	16	20
0.020	708	598	82	566	452	58	508	386	60	306	194	30	140	98
0.050	998	996	536	996	990	588	990	986	486	938	842	242	600	446
0.200	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	998
0.500	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
10 years of record														
0.000	92	70	34	44	36	32	50	42	28	48	34	34	38	44
0.002	206	186	68	130	120	68	112	92	50	108	100	50	96	82
0.005	466	402	222	362	280	210	296	228	184	192	164	132	272	192
0.020	1000	1000	1000	1000	1000	1000	1000	1000	996	984	976	922	998	984
0.050	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.200	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.500	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
15 years of record														
0.000	116	102	52	68	46	44	74	58	50	52	32	50	50	44
0.002	290	272	152	250	204	164	206	180	138	114	94	70	164	130
0.005	878	854	684	824	780	660	776	722	614	546	488	424	694	566
0.020	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.050	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.200	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.500	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000

Table C.6 Comparison of significance level (slope = 0) and power (slope > 0) for trend tests. The errors are from a normal distribution with ρ = 0.2 and σ_r = 5.0. Values in table are scaled by 1000.

	MONTHLY			QUARTERLY									YEARLY	
				MEAN			MEDIAN			MIDDLE			MEAN	MED
slope	MKD	SK	SKC	MKD	SK	SKC	MKD	SK	SKC	MKD	SK	SKC	MK	MK
5 years of record														
0.000	116	72	6	80	42	6	74	42	4	78	38	8	16	16
0.002	144	82	22	70	42	10	60	40	6	72	48	8	20	20
0.005	176	118	20	106	58	4	124	66	18	88	48	14	10	18
0.020	404	292	22	262	152	24	210	102	16	136	78	10	22	34
0.050	910	840	172	702	594	198	674	532	144	454	338	76	70	124
0.200	1000	1000	996	1000	1000	954	998	1000	928	988	990	788	674	812
0.500	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	998	1000	1000
10 years of record														
0.000	82	66	26	58	54	30	62	42	34	46	38	32	36	46
0.002	154	122	66	100	82	60	88	76	66	76	62	30	60	44
0.005	244	200	88	164	128	82	156	114	84	98	84	80	54	84
0.020	958	942	814	894	870	804	832	818	740	670	606	482	306	552
0.050	1000	1000	1000	1000	1000	1000	1000	1000	1000	996	996	996	950	988
0.200	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.500	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
15 years of record														
0.000	104	90	30	80	70	32	68	58	32	48	46	40	36	38
0.002	172	140	80	112	108	90	102	90	74	86	70	64	30	58
0.005	540	486	312	422	370	304	382	330	250	226	194	160	106	170
0.020	1000	1000	1000	1000	1000	1000	1000	1000	1000	982	978	972	814	940
0.050	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.200	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.500	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000

Table C.7 Comparison of significance level (slope = 0) and power (slope > 0) for trend tests. The errors are from a lognormal distribution with ρ = 0.2 and σ_r = 1.0. Values in table are scaled by 1000.

	MONTHLY			QUARTERLY									YEARLY	
				MEAN			MEDIAN			MIDDLE			MEAN	MED
slope	MKD	SK	SKC	MKD	SK	SKC	MKD	SK	SKC	MKD	SK	SKC	MK	MK
5 years of record														
0.000	130	74	12	64	30	0	70	42	4	50	22	4	18	14
0.002	160	116	18	78	48	14	66	58	8	72	50	12	16	12
0.005	272	216	20	146	94	8	160	106	6	106	68	10	24	22
0.020	954	926	260	720	576	146	802	694	210	578	426	86	152	222
0.050	1000	1000	926	998	998	820	1000	996	866	958	942	620	650	758
0.200	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	998	1000
0.500	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
10 years of record														
0.000	122	114	58	66	48	38	66	58	40	60	34	30	46	58
0.002	306	288	106	156	124	76	166	130	88	120	90	72	92	66
0.005	758	758	518	508	460	336	604	550	388	412	384	280	270	370
0.020	1000	1000	1000	1000	1000	1000	1000	1000	1000	998	998	998	986	998
0.050	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.200	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.500	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
15 years of record														
0.000	124	114	40	74	62	50	68	50	42	48	42	36	46	42
0.002	626	622	428	382	324	256	446	398	336	326	284	242	218	254
0.005	996	992	978	956	952	916	968	962	944	910	876	836	784	852
0.020	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.050	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.200	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.500	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000

Table C.8 Comparison of significance level (slope = 0) and power (slope > 0) for trend tests. The errors are from a lognormal distribution with ρ = 0.2 and σ_r = 5.0. Values in table are scaled by 1000.

	MONTHLY			QUARTERLY									YEARLY	
				MEAN			MEDIAN			MIDDLE			MEAN	MED
slope	MKD	SK	SKC	MKD	SK	SKC	MKD	SK	SKC	MKD	SK	SKC	MK	MK
5 years of record														
0.000	132	106	18	68	32	6	88	34	10	58	34	8	10	28
0.002	172	124	10	96	54	8	78	52	6	104	58	12	12	16
0.005	212	142	8	110	62	2	128	72	16	92	46	18	26	40
0.020	640	552	78	328	252	46	354	262	66	262	200	40	32	52
0.050	986	974	506	788	734	298	830	796	374	682	612	202	100	234
0.200	1000	1000	998	1000	1000	966	1000	1000	986	996	998	910	760	928
0.500	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	988	998
10 years of record														
0.000	146	104	42	60	54	36	78	54	38	46	28	26	44	52
0.002	192	172	84	108	72	66	104	84	66	82	74	58	48	62
0.005	460	438	236	240	216	170	270	254	186	174	164	128	80	114
0.020	996	998	986	944	934	894	974	966	932	872	858	790	382	708
0.050	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	998	946	998
0.200	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.500	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
15 years of record														
0.000	136	110	54	74	56	46	60	52	44	68	56	48	50	48
0.002	336	316	190	184	160	130	222	182	140	150	128	100	56	66
0.005	878	844	672	602	562	474	688	650	566	446	410	398	126	294
0.020	1000	1000	1000	1000	1000	1000	1000	1000	1000	998	998	1000	880	996
0.050	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.200	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.500	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000

Table C.9 Comparison of significance level (slope = 0) and power (slope > 0) for trend tests. The errors are from a normal distribution with $\rho = 0.4$ and $\sigma_r = 1.0$. Values in table are scaled by 1000.

	MONTHLY			QUARTERLY									YEARLY	
				MEAN			MEDIAN			MIDDLE			MEAN	MED
slope	MKD	SK	SKC	MKD	SK	SKC	MKD	SK	SKC	MKD	SK	SKC	MK	MK
5 years of record														
0.000	212	140	8	108	74	6	96	56	12	92	56	12	18	22
0.002	192	130	12	110	60	4	102	66	6	64	44	4	16	12
0.005	240	172	8	138	82	10	132	66	8	96	58	10	26	26
0.020	684	582	60	478	358	56	458	326	48	318	204	24	106	78
0.050	998	992	406	988	970	438	982	924	406	906	830	216	462	358
0.200	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	998
0.500	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
10 years of record														
0.000	244	194	58	130	86	60	106	86	56	68	52	40	56	66
0.002	286	246	72	164	140	62	152	124	60	114	88	58	66	70
0.005	496	462	200	358	290	200	322	264	168	232	186	120	212	166
0.020	1000	1000	988	1000	998	974	998	990	956	984	966	886	970	932
0.050	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.200	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.500	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
15 years of record														
0.000	188	158	44	98	90	40	86	64	38	56	48	54	56	48
0.002	374	344	176	268	252	166	240	202	150	176	136	106	162	144
0.005	824	804	540	704	664	508	648	592	448	490	436	330	530	422
0.020	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.050	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.200	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.500	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000

Table C.10 Comparison of significance level (slope = 0) and power (slope > 0) for trend tests. The errors are from a normal distribution with $\rho = 0.4$ and $\sigma_r = 5.0$. Values in table are scaled by 1000.

	MONTHLY			QUARTERLY									YEARLY	
				MEAN			MEDIAN			MIDDLE			MEAN	MED
slope	MKD	SK	SKC	MKD	SK	SKC	MKD	SK	SKC	MKD	SK	SKC	MK	MK
5 years of record														
0.000	220	134	16	102	58	8	80	50	6	86	46	6	18	22
0.002	216	142	6	102	52	12	86	44	8	80	48	10	22	14
0.005	224	146	18	120	72	18	120	68	10	92	46	10	20	16
0.020	422	300	36	254	172	28	220	144	28	162	84	12	26	42
0.050	854	784	160	644	540	154	586	456	132	458	344	76	64	112
0.200	1000	1000	988	1000	1000	920	998	998	864	996	996	770	564	714
0.500	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	994	1000
10 years of record														
0.000	208	184	30	114	76	30	86	76	24	82	68	42	50	44
0.002	222	178	58	100	82	56	104	74	56	80	62	54	64	48
0.005	324	266	102	178	154	92	156	128	74	120	98	76	62	86
0.020	932	906	660	822	776	612	780	732	600	634	560	440	248	408
0.050	1000	1000	998	1000	1000	996	1000	1000	998	998	998	996	868	946
0.200	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.500	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
15 years of record														
0.000	208	190	60	98	100	50	102	82	48	76	62	48	50	46
0.002	270	244	72	148	126	72	138	108	64	88	74	64	50	62
0.005	508	472	210	346	300	230	330	256	190	208	188	162	86	132
0.020	1000	1000	1000	998	1000	994	994	998	988	984	972	962	690	890
0.050	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.200	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.500	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000

Table C.11 Comparison of significance level (slope = 0) and power (slope > 0) for trend tests. The errors are from a lognormal distribution with ρ = 0.4 and σ_r = 1.0. Values in table are scaled by 1000.

	MONTHLY			QUARTERLY									YEARLY	
				MEAN			MEDIAN			MIDDLE			MEAN	MED
slope	MKD	SK	SKC	MKD	SK	SKC	MKD	SK	SKC	MKD	SK	SKC	MK	MK
5 years of record														
0.000	236	184	18	128	76	10	98	60	16	82	64	8	24	26
0.002	268	174	14	122	76	4	134	66	10	100	54	12	10	18
0.005	322	260	18	162	124	14	174	126	16	146	72	6	28	38
0.020	862	836	154	624	510	74	660	558	110	486	382	68	110	132
0.050	1000	1000	826	988	984	680	998	988	708	966	950	560	536	608
0.200	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.500	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
10 years of record														
0.000	246	214	54	118	92	42	110	82	44	66	44	42	56	56
0.002	350	312	108	184	158	88	180	146	84	126	104	64	98	90
0.005	690	690	352	460	414	286	494	470	308	392	350	254	242	266
0.020	1000	1000	1000	1000	1000	996	1000	1000	994	1000	998	990	972	994
0.050	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.200	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.500	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
15 years of record														
0.000	224	204	58	102	94	48	90	78	54	94	76	54	66	56
0.002	548	552	210	306	298	194	320	284	194	242	244	170	134	182
0.005	974	976	860	870	838	756	912	880	800	808	774	688	628	704
0.020	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.050	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.200	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.500	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000

Table C.12 Comparison of significance level (slope = 0) and power (slope > 0) for trend tests. The errors are from a lognormal distribution with ρ = 0.4 and σ_r = 5.0. Values in table are scaled by 1000.

	MONTHLY			QUARTERLY									YEARLY	
				MEAN			MEDIAN			MIDDLE			MEAN	MED
slope	MKD	SK	SKC	MKD	SK	SKC	MKD	SK	SKC	MKD	SK	SKC	MK	MK
5 years of record														
0.000	234	164	8	112	84	6	104	66	8	78	46	6	16	26
0.002	206	156	8	74	52	10	86	50	8	72	36	10	10	12
0.005	268	192	12	136	76	22	124	86	16	116	62	10	28	24
0.020	618	562	60	362	276	44	368	262	46	248	186	28	32	42
0.050	930	920	340	702	622	218	722	632	238	606	506	168	72	202
0.200	1000	1000	998	1000	998	954	1000	1000	968	998	1000	930	682	882
0.500	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	978	996
10 years of record														
0.000	224	192	46	86	76	42	94	64	38	88	70	38	38	46
0.002	280	248	78	144	106	76	132	108	82	82	64	44	48	78
0.005	448	412	134	244	214	128	244	216	140	184	148	120	68	110
0.020	986	986	906	916	910	806	926	908	814	860	838	740	312	584
0.050	1000	1000	1000	1000	1000	1000	1000	1000	1000	998	998	998	878	990
0.200	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.500	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
15 years of record														
0.000	192	156	40	80	68	34	82	80	42	68	56	36	48	52
0.002	382	366	166	224	192	126	236	208	130	154	150	120	70	78
0.005	734	720	416	486	444	330	542	474	360	410	378	302	124	208
0.020	1000	1000	1000	1000	1000	1000	1000	1000	998	998	996	996	774	966
0.050	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.200	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.500	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000

Table C.13 Comparison of significance level (slope = 0) and power (slope > 0) for trend tests. The errors are from a normal distribution with $\rho = 0.6$ and $\sigma_r = 1.0$. Values in table are scaled by 1000.

	MONTHLY			QUARTERLY									YEARLY	
				MEAN			MEDIAN			MIDDLE			MEAN	MED
slope	MKD	SK	SKC	MKD	SK	SKC	MKD	SK	SKC	MKD	SK	SKC	MK	MK
5 years of record														
0.000	344	266	6	174	124	14	150	112	12	130	82	6	20	16
0.002	302	218	12	132	70	12	128	66	4	84	56	4	22	8
0.005	382	284	24	200	154	22	176	108	16	154	102	24	20	40
0.020	640	558	54	446	344	54	400	278	42	338	254	32	80	62
0.050	988	978	368	942	902	376	934	880	344	900	814	232	352	290
0.200	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	998	988
0.500	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
10 years of record														
0.000	346	306	64	174	154	52	174	134	48	120	94	52	50	60
0.002	348	322	72	200	170	76	166	156	64	158	120	62	80	72
0.005	478	434	122	290	248	106	274	218	104	204	178	96	120	118
0.020	1000	998	922	990	982	906	988	974	880	972	960	838	918	854
0.050	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.200	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.500	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
15 years of record														
0.000	364	324	64	170	144	68	138	130	58	112	84	50	64	54
0.002	452	408	112	266	228	108	232	206	102	182	154	92	128	106
0.005	768	756	390	620	572	380	592	544	370	484	448	308	372	308
0.020	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.050	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.200	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.500	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000

Table C.14 Comparison of significance level (slope = 0) and power (slope > 0) for trend tests. The errors are from a normal distribution with $\rho = 0.6$ and $\sigma_r = 5.0$. Values in table are scaled by 1000.

	MONTHLY			QUARTERLY									YEARLY	
				MEAN			MEDIAN			MIDDLE			MEAN	MED
slope	MKD	SK	SKC	MKD	SK	SKC	MKD	SK	SKC	MKD	SK	SKC	MK	MK
5 years of record														
0.000	354	284	20	166	106	24	156	98	18	120	74	10	20	20
0.002	350	254	18	164	104	16	160	108	18	118	74	14	14	16
0.005	340	234	10	158	106	6	146	104	8	118	68	6	12	16
0.020	442	366	34	260	182	28	222	158	14	176	142	16	24	40
0.050	776	718	116	586	482	134	556	434	110	464	366	88	54	80
0.200	1000	1000	946	996	998	872	998	996	824	988	978	786	502	642
0.500	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	976	974
10 years of record														
0.000	312	278	48	170	122	40	140	126	46	148	96	52	46	56
0.002	330	294	64	190	138	56	172	138	72	136	110	56	50	64
0.005	420	374	90	224	186	86	212	182	78	162	132	70	72	94
0.020	872	850	486	726	690	454	712	638	444	622	562	388	166	326
0.050	1000	1000	998	998	996	986	998	992	976	990	986	962	736	886
0.200	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.500	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
15 years of record														
0.000	328	296	54	148	126	52	134	114	56	100	82	44	40	42
0.002	358	332	70	194	158	76	158	144	64	138	110	70	46	60
0.005	566	528	196	342	308	180	326	280	156	278	224	170	80	138
0.020	1000	1000	958	990	986	944	984	978	936	968	964	922	522	792
0.050	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.200	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.500	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000

Table C.15 Comparison of significance level (slope = 0) and power (slope > 0) for trend tests. The errors are from a lognormal distribution with ρ = 0.6 and σ_r = 1.0. Values in table are scaled by 1000.

	MONTHLY			QUARTERLY									YEARLY	
				MEAN			MEDIAN			MIDDLE			MEAN	MED
slope	MKD	SK	SKC	MKD	SK	SKC	MKD	SK	SKC	MKD	SK	SKC	MK	MK
5 years of record														
0.000	406	332	10	204	126	10	182	118	12	128	88	8	16	18
0.002	354	256	16	156	80	14	138	88	8	116	84	8	18	18
0.005	354	274	18	160	102	12	158	92	8	146	104	8	26	12
0.020	766	736	104	560	462	76	564	458	80	500	408	56	92	70
0.050	998	998	610	984	968	490	988	966	528	964	932	436	398	420
0.200	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	994	986
0.500	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
10 years of record														
0.000	344	298	58	150	124	42	134	112	44	128	92	48	54	44
0.002	430	426	106	236	216	104	240	216	102	204	174	96	108	100
0.005	642	630	230	408	392	186	428	362	192	390	348	184	168	200
0.020	1000	1000	992	998	996	964	998	996	978	1000	996	960	928	956
0.050	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.200	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.500	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
15 years of record														
0.000	358	326	52	164	152	54	162	146	54	124	98	50	70	60
0.002	534	516	170	320	280	140	314	286	158	278	256	142	142	128
0.005	912	904	606	762	736	520	774	748	558	718	670	510	428	454
0.020	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.050	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.200	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.500	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000

Table C.16 Comparison of significance level (slope = 0) and power (slope > 0) for trend tests. The errors are from a lognormal distribution with ρ = 0.6 and σ_r = 5.0. Values in table are scaled by 1000.

	MONTHLY			QUARTERLY									YEARLY	
				MEAN			MEDIAN			MIDDLE			MEAN	MED
slope	MKD	SK	SKC	MKD	SK	SKC	MKD	SK	SKC	MKD	SK	SKC	MK	MK
5 years of record														
0.000	328	256	8	138	82	6	142	98	0	102	60	4	10	10
0.002	354	264	14	164	108	14	116	78	14	92	60	12	14	18
0.005	384	302	16	182	118	10	190	120	10	150	94	12	18	30
0.020	562	500	52	308	216	46	306	206	32	276	188	30	36	48
0.050	902	848	220	644	588	156	652	578	142	622	548	108	46	118
0.200	1000	1000	966	998	998	884	998	998	876	994	994	848	528	736
0.500	1000	1000	1000	1000	1000	998	1000	1000	1000	1000	1000	994	964	982
10 years of record														
0.000	362	320	74	156	128	66	142	118	74	114	98	62	44	60
0.002	370	342	70	192	172	66	174	154	78	140	112	54	52	56
0.005	468	438	130	290	248	106	254	232	116	220	176	102	96	92
0.020	950	956	690	830	804	606	826	802	606	776	740	566	232	388
0.050	1000	1000	998	998	998	996	1000	1000	996	1000	996	990	800	950
0.200	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.500	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
15 years of record														
0.000	336	320	66	174	150	64	148	122	50	134	114	56	62	66
0.002	430	412	104	230	214	102	216	194	102	184	150	88	68	78
0.005	650	650	292	450	416	236	434	424	268	396	368	246	102	158
0.020	1000	1000	998	998	992	980	996	994	986	990	988	970	612	888
0.050	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	998	1000
0.200	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.500	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000

Table C.17 Comparison of significance level (slope = 0) and power (slope > 0) for trend tests. The errors are from a normal distribution with ρ = 0.8 and σ_r = 1.0. Values in table are scaled by 1000.

	MONTHLY			QUARTERLY									YEARLY	
				MEAN			MEDIAN			MIDDLE			MEAN	MED
slope	MKD	SK	SKC	MKD	SK	SKC	MKD	SK	SKC	MKD	SK	SKC	MK	MK
5 years of record														
0.000	512	448	30	282	214	24	260	202	26	230	168	20	36	30
0.002	508	432	24	278	198	26	270	198	20	230	182	18	22	22
0.005	540	442	36	294	206	24	284	202	24	246	182	22	30	38
0.020	702	620	58	458	388	60	454	370	68	422	332	54	54	54
0.050	954	938	308	888	840	286	886	818	268	864	790	236	230	210
0.200	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	988	966
0.500	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
10 years of record														
0.000	480	440	78	230	200	72	234	202	66	208	158	68	80	64
0.002	496	444	68	290	244	78	286	252	76	260	222	84	88	80
0.005	552	516	144	354	328	134	342	304	132	330	274	126	124	126
0.020	984	982	760	944	934	746	942	924	728	928	908	714	704	660
0.050	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	998
0.200	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.500	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
15 years of record														
0.000	494	464	92	302	262	84	274	246	82	250	196	98	82	82
0.002	556	512	176	368	352	164	364	340	170	336	312	134	148	136
0.005	728	704	276	568	520	258	548	506	268	514	462	240	258	246
0.020	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	998	996
0.050	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.200	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.500	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000

Table C.18 Comparison of significance level (slope = 0) and power (slope > 0) for trend tests. The errors are from a normal distribution with ρ = 0.8 and σ_r = 5.0. Values in table are scaled by 1000.

	MONTHLY			QUARTERLY									YEARLY	
				MEAN			MEDIAN			MIDDLE			MEAN	MED
slope	MKD	SK	SKC	MKD	SK	SKC	MKD	SK	SKC	MKD	SK	SKC	MK	MK
5 years of record														
0.000	506	416	22	296	210	20	272	198	20	256	148	12	18	28
0.002	514	444	18	268	202	24	262	180	12	230	148	4	22	20
0.005	550	456	14	308	232	24	292	214	14	258	192	24	28	26
0.020	570	500	48	360	286	48	338	280	44	322	242	34	22	26
0.050	766	722	110	518	456	108	520	440	92	492	412	76	46	64
0.200	998	1000	866	980	974	764	978	974	756	976	974	726	370	496
0.500	1000	1000	1000	1000	1000	996	1000	1000	996	1000	1000	996	934	946
10 years of record														
0.000	522	478	106	292	256	104	268	224	96	246	206	88	86	86
0.002	510	476	64	260	220	62	268	218	72	228	202	60	74	76
0.005	506	478	76	286	254	64	240	222	74	226	212	76	74	80
0.020	848	822	374	700	650	346	694	638	340	658	610	334	172	262
0.050	998	998	936	982	970	906	980	972	906	974	978	890	548	734
0.200	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.500	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
15 years of record														
0.000	530	494	90	312	278	94	306	270	86	276	232	90	100	94
0.002	522	484	100	330	300	102	320	296	94	278	244	98	82	88
0.005	594	574	162	374	354	164	366	342	156	338	322	134	92	120
0.020	982	972	812	928	924	784	924	918	776	914	894	754	356	586
0.050	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	978	994
0.200	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.500	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000

Table C.19 Comparison of significance level (slope = 0) and power (slope > 0) for trend tests. The errors are from a lognormal distribution with ρ = 0.8 and σ_r = 1.0. Values in table are scaled by 1000.

	MONTHLY			QUARTERLY									YEARLY	
				MEAN			MEDIAN			MIDDLE			MEAN	MED
slope	MKD	SK	SKC	MKD	SK	SKC	MKD	SK	SKC	MKD	SK	SKC	MK	MK
5 years of record														
0.000	516	454	44	312	248	36	300	248	28	254	212	30	20	34
0.002	544	446	30	294	216	28	278	208	18	252	184	18	24	26
0.005	546	486	36	312	248	40	304	232	30	278	212	24	34	28
0.020	698	654	92	496	424	80	480	406	82	456	386	70	66	60
0.050	970	966	444	906	878	402	908	872	364	898	850	334	296	268
0.200	1000	1000	998	1000	1000	996	1000	1000	998	1000	1000	998	980	968
0.500	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
10 years of record														
0.000	530	500	106	308	274	96	304	276	82	290	264	94	104	96
0.002	488	472	98	292	264	106	278	250	106	250	230	92	124	94
0.005	618	574	130	360	316	114	336	304	110	336	298	120	118	114
0.020	990	990	830	946	938	794	948	946	806	938	932	782	736	740
0.050	1000	1000	998	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.200	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.500	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
15 years of record														
0.000	518	494	102	294	270	98	284	256	102	272	246	104	104	90
0.002	568	560	134	328	318	130	328	308	134	288	284	120	132	112
0.005	772	772	388	620	602	352	612	598	362	566	558	342	324	306
0.020	1000	1000	1000	1000	1000	996	1000	1000	998	1000	1000	1000	994	996
0.050	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.200	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.500	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000

Table C.20 Comparison of significance level (slope = 0) and power (slope > 0) for trend tests. The errors are from a lognormal distribution with ρ = 0.8 and σ_r = 5.0. Values in table are scaled by 1000.

	MONTHLY			QUARTERLY									YEARLY	
				MEAN			MEDIAN			MIDDLE			MEAN	MED
slope	MKD	SK	SKC	MKD	SK	SKC	MKD	SK	SKC	MKD	SK	SKC	MK	MK
5 years of record														
0.000	490	408	20	256	168	10	230	176	20	216	156	14	30	20
0.002	538	452	22	290	216	16	284	202	26	242	178	22	34	22
0.005	514	442	36	272	214	32	246	196	36	230	164	22	38	30
0.020	552	502	32	332	256	26	314	262	38	294	224	26	22	28
0.050	812	784	196	584	532	154	590	534	156	570	494	146	70	76
0.200	998	998	878	978	988	794	978	984	794	980	980	758	454	588
0.500	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	998	954	954
10 years of record														
0.000	504	472	88	286	252	88	278	240	86	260	228	82	98	92
0.002	470	422	84	268	240	82	266	224	82	240	218	72	74	70
0.005	584	552	102	330	290	84	328	288	80	298	264	80	92	98
0.020	848	852	424	716	708	400	694	684	404	694	678	388	206	278
0.050	998	996	930	978	980	910	978	978	910	974	970	908	574	790
0.200	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.500	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
15 years of record														
0.000	496	482	84	290	254	76	278	244	86	258	238	78	68	84
0.002	566	548	118	344	312	122	336	318	124	316	288	124	126	94
0.005	624	602	202	424	398	190	418	402	186	404	390	172	100	128
0.020	990	992	886	954	956	854	956	960	876	950	944	862	416	664
0.050	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	976	1000
0.200	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.500	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000

APPENDIX D

TREND DETECTION IN THE PRESENCE OF MISSING VALUES

This appendix contains the actual significance levels (slope = 0.0) and power (slope > 0.0) from the Monte Carlo simulation used to investigate the effect of missing values on trend tests (Chapter VI). The Mann-Kendall tau test on deaseasonalized data (MKD), the Seasonal Kendall tau tests (SK), and the Seasonal Kendall tau test with correction for autocorrelation (SKC) were applied to monthly data. Monthly data were collapsed to quarterly values by computing the quarterly median. The three tests were then reapplied to the three sets of quarterly collapsed data. The values in the following tables were scaled by 1000 for convenience.

Table D.1 Comparison of significance level and power for trend tests. The errors are from a normal distribution with σ_r = 1.0 and missing value percentage = 0.0. Values in table are scaled by 1000.

SLOPE	ρ=0.0 MONTHLY			QUARTERLY			ρ=0.4 MONTHLY			QUARTERLY		
	MKD	SK	SKC	MKD	SK	SKC	MKD	SK	SKC	MKD	SK	SKC
	5 years of record											
0.000	88	42	10	74	36	6	186	132	20	78	42	6
0.002	74	40	6	58	34	8	220	130	8	98	46	8
0.005	122	68	10	94	64	4	218	150	14	104	64	8
0.020	756	600	100	584	442	70	674	594	58	470	358	60
0.050	1000	1000	578	1000	994	608	996	992	456	976	946	416
0.200	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.500	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
	10 years of record											
0.000	64	44	18	56	40	32	200	136	34	64	52	28
0.002	110	90	60	86	76	54	270	242	70	134	114	62
0.005	490	422	294	356	292	226	470	416	172	282	210	128
0.020	1000	1000	1000	1000	1000	1000	998	998	986	1000	998	960
0.050	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.200	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.500	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
	15 years of record											
0.000	54	42	32	54	38	34	216	190	54	94	84	52
0.002	306	260	222	242	180	176	362	320	120	204	160	100
0.005	922	896	826	812	766	706	840	824	530	660	622	478
0.020	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.050	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.200	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.500	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000

Table D.2 Comparison of significance level and power for trend tests. The errors are from a normal distribution with σ_r = 5.0 and missing value percentage = 0.0. Values in table are scaled by 1000.

SLOPE	ρ=0.0 MONTHLY			QUARTERLY			ρ=0.4 MONTHLY			QUARTERLY		
	MKD	SK	SKC	MKD	SK	SKC	MKD	SK	SKC	MKD	SK	SKC
	5 years of record											
0.000	80	46	14	74	40	2	242	154	10	98	60	8
0.002	68	32	16	60	30	14	228	160	12	78	52	4
0.005	80	52	20	86	40	6	262	184	10	122	88	8
0.020	332	200	14	230	158	32	438	328	20	208	118	14
0.050	930	850	256	726	602	164	868	784	148	602	494	134
0.200	1000	1000	1000	1000	1000	960	1000	1000	984	994	998	896
0.500	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
	10 years of record											
0.000	68	56	38	50	36	48	160	148	36	72	56	38
0.002	74	56	58	82	66	56	232	200	54	114	90	54
0.005	224	184	106	166	114	76	302	260	86	164	134	84
0.020	984	980	930	922	874	762	926	916	696	788	736	618
0.050	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	996
0.200	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.500	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
	15 years of record											
0.000	62	56	42	70	48	46	192	160	38	92	80	44
0.002	136	108	80	102	88	88	258	220	76	118	100	64
0.005	530	456	378	388	324	284	532	486	240	332	288	198
0.020	1000	1000	1000	1000	998	998	1000	1000	996	996	996	988
0.050	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.200	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.500	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000

Table D.3 Comparison of significance level and power for trend tests. The errors are from a lognormal distribution with σ_r = 1.0 and missing value percentage = 0.0. Values in table are scaled by 1000.

SLOPE	ρ=0.0 MONTHLY			QUARTERLY			ρ=0.4 MONTHLY			QUARTERLY		
	MKD	SK	SKC	MKD	SK	SKC	MKD	SK	SKC	MKD	SK	SKC
	5 years of record											
0.000	78	50	6	48	26	8	224	156	16	92	48	2
0.002	96	58	12	74	36	16	208	164	10	102	68	6
0.005	218	134	24	146	92	18	314	240	24	162	116	6
0.020	978	962	414	896	814	290	866	818	162	652	540	120
0.050	1000	1000	976	1000	1000	950	1000	1000	822	994	982	690
0.200	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.500	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
	10 years of record											
0.000	76	70	60	76	56	36	210	174	58	94	62	40
0.002	292	270	200	206	186	146	336	330	90	164	142	84
0.005	928	918	786	804	730	616	662	652	322	444	402	258
0.020	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	994
0.050	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.200	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.500	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
	15 years of record											
0.000	62	56	48	54	36	48	246	210	46	98	72	38
0.002	730	730	644	574	524	462	546	538	264	346	320	230
0.005	1000	1000	1000	1000	1000	994	976	972	864	916	904	804
0.020	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.050	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.200	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.500	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000

Table D.4 Comparison of significance level and power for trend tests. The errors are from a lognormal distribution with σ_r = 5.0 and missing value percentage = 0.0. Values in table are scaled by 1000.

SLOPE	ρ=0.0 MONTHLY			QUARTERLY			ρ=0.4 MONTHLY			QUARTERLY		
	MKD	SK	SKC	MKD	SK	SKC	MKD	SK	SKC	MKD	SK	SKC
	5 years of record											
0.000	84	40	6	74	38	8	210	150	8	98	46	4
0.002	74	40	22	68	46	8	278	182	10	118	70	4
0.005	124	88	10	102	58	14	270	202	8	146	88	10
0.020	696	598	136	462	330	66	558	500	62	322	226	28
0.050	990	990	626	888	856	452	946	924	354	746	676	272
0.200	1000	1000	998	1000	1000	998	1000	1000	994	1000	1000	964
0.500	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
	10 years of record											
0.000	54	46	42	58	48	44	232	190	36	94	66	44
0.002	178	136	104	118	92	76	258	246	82	104	96	58
0.005	502	462	366	370	318	230	428	402	136	236	202	120
0.020	1000	1000	1000	986	982	980	990	988	898	918	894	826
0.050	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.200	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.500	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
	15 years of record											
0.000	48	34	32	52	46	46	254	224	60	96	76	50
0.002	388	362	326	256	222	214	342	336	124	190	156	102
0.005	920	916	874	780	764	712	738	750	448	548	508	390
0.020	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.050	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.200	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.500	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000

Table D.5 Comparison of significance level and power for trend tests. The errors are from a normal distribution with σ_r = 1.0 and missing value percentage = 0.1. Values in table are scaled by 1000.

SLOPE	ρ=0.0 MONTHLY			QUARTERLY			ρ=0.4 MONTHLY			QUARTERLY		
	MKD	SK	SKC	MKD	SK	SKC	MKD	SK	SKC	MKD	SK	SKC
	5 years of record											
0.000	72	38	8	80	40	8	166	110	20	96	54	12
0.002	62	40	10	60	32	8	182	114	12	90	32	6
0.005	120	58	14	96	60	8	198	134	14	104	76	10
0.020	682	524	88	578	440	74	650	568	42	472	358	66
0.050	1000	998	488	1000	996	576	990	986	396	980	936	396
0.200	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.500	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
	10 years of record											
0.000	52	32	18	40	28	30	158	108	36	60	48	16
0.002	94	82	54	76	72	70	254	216	62	136	110	64
0.005	436	356	238	334	274	230	468	400	160	300	226	140
0.020	1000	1000	1000	1000	1000	996	998	998	980	998	990	968
0.050	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.200	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.500	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
	15 years of record											
0.000	42	36	40	42	40	28	224	186	50	112	92	46
0.002	286	242	214	242	196	172	318	296	130	186	158	116
0.005	886	836	768	792	746	690	816	784	516	662	590	468
0.020	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.050	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.200	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.500	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000

Table D.6 Comparison of significance level and power for trend tests. The errors are from a normal distribution with σ_r = 5.0 and missing value percentage = 0.1. Values in table are scaled by 1000.

SLOPE	ρ=0.0 MONTHLY			QUARTERLY			ρ=0.4 MONTHLY			QUARTERLY		
	MKD	SK	SKC	MKD	SK	SKC	MKD	SK	SKC	MKD	SK	SKC
	5 years of record											
0.000	78	52	14	62	36	6	210	148	10	94	56	10
0.002	86	42	12	70	44	10	208	144	26	88	48	6
0.005	78	54	10	96	48	10	236	154	14	114	82	12
0.020	284	164	26	216	132	36	390	280	22	222	128	14
0.050	884	804	214	708	590	140	836	724	152	592	498	124
0.200	1000	1000	996	1000	1000	942	1000	1000	948	998	996	884
0.500	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
	10 years of record											
0.000	60	48	38	56	46	40	170	136	40	64	54	34
0.002	72	70	60	70	64	48	202	150	42	126	88	62
0.005	188	138	92	152	136	82	290	234	80	162	124	78
0.020	978	964	882	896	862	756	910	898	654	786	738	600
0.050	1000	1000	1000	1000	1000	1000	1000	1000	998	1000	1000	998
0.200	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.500	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
	15 years of record											
0.000	52	42	46	58	38	46	180	152	44	84	80	44
0.002	134	104	84	102	76	72	216	194	80	106	96	74
0.005	472	400	338	378	314	286	490	446	224	328	294	204
0.020	1000	1000	1000	998	1000	1000	1000	1000	994	994	996	984
0.050	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.200	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.500	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000

Table D.7 Comparison of significance level and power for trend tests. The errors are from a lognormal distribution with σ_r = 1.0 and missing value percentage = 0.1. Values in table are scaled by 1000.

SLOPE	ρ=0.0 MONTHLY			QUARTERLY			ρ=0.4 MONTHLY			QUARTERLY		
	MKD	SK	SKC	MKD	SK	SKC	MKD	SK	SKC	MKD	SK	SKC
	5 years of record											
0.000	84	48	20	68	54	8	204	158	12	96	48	4
0.002	106	58	12	68	58	12	202	164	18	92	64	4
0.005	186	130	26	126	94	24	300	220	14	164	102	8
0.020	960	928	336	838	746	228	838	768	120	638	512	140
0.050	1000	1000	938	1000	1000	920	1000	1000	774	990	980	668
0.200	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.500	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
	10 years of record											
0.000	68	62	46	70	56	40	206	158	68	92	72	46
0.002	256	200	180	192	154	112	318	286	92	156	130	78
0.005	904	852	712	710	662	548	626	608	296	426	386	236
0.020	1000	1000	1000	1000	1000	1000	1000	1000	998	998	998	992
0.050	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.200	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.500	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
	15 years of record											
0.000	62	50	46	50	40	60	220	198	44	104	86	50
0.002	672	664	596	510	460	400	516	504	236	332	304	222
0.005	1000	1000	998	996	994	980	966	960	848	894	874	780
0.020	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.050	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.200	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.500	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000

Table D.8 Comparison of significance level and power for trend tests. The errors are from a lognormal distribution with σ_r = 5.0 and missing value percentage = 0.1. Values in table are scaled by 1000.

SLOPE	ρ=0.0 MONTHLY			QUARTERLY			ρ=0.4 MONTHLY			QUARTERLY		
	MKD	SK	SKC	MKD	SK	SKC	MKD	SK	SKC	MKD	SK	SKC
	5 years of record											
0.000	66	38	20	60	28	10	208	148	16	100	50	4
0.002	66	44	18	66	30	6	232	156	10	112	70	8
0.005	122	86	10	84	56	10	250	186	2	160	94	10
0.020	614	500	118	414	302	60	526	446	56	306	244	26
0.050	978	962	524	854	832	418	922	888	286	710	636	242
0.200	1000	1000	996	1000	1000	990	1000	1000	986	1000	1000	954
0.500	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
	10 years of record											
0.000	60	40	36	58	52	48	212	174	52	98	76	48
0.002	148	128	94	90	76	66	252	234	76	108	88	56
0.005	454	402	322	304	286	218	408	376	122	224	206	114
0.020	1000	1000	996	976	976	964	976	980	868	910	880	796
0.050	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.200	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.500	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
	15 years of record											
0.000	46	34	24	48	40	28	224	188	48	98	66	38
0.002	360	348	302	250	220	202	312	312	108	180	156	92
0.005	892	882	822	718	680	666	704	696	430	546	506	398
0.020	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.050	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.200	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.500	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000

Table D.9 Comparison of significance level and power for trend tests. The errors are from a normal distribution with σ_r = 1.0 and missing value percentage = 0.3. Values in table are scaled by 1000.

SLOPE	ρ=0.0 MONTHLY			QUARTERLY			ρ=0.4 MONTHLY			QUARTERLY		
	MKD	SK	SKC	MKD	SK	SKC	MKD	SK	SKC	MKD	SK	SKC
	5 years of record											
0.000	56	40	14	60	38	6	136	100	10	96	50	8
0.002	82	40	8	66	40	8	158	92	12	80	40	8
0.005	110	66	8	84	40	10	170	110	18	96	54	6
0.020	536	372	58	510	322	46	506	412	44	432	292	36
0.050	996	972	318	992	958	470	980	954	276	960	908	344
0.200	1000	1000	998	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.500	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
	10 years of record											
0.000	54	40	38	46	36	20	132	90	34	62	40	18
0.002	88	68	52	80	74	52	194	152	46	120	102	56
0.005	356	286	188	320	264	192	378	306	118	276	194	132
0.020	998	994	990	1000	996	994	998	996	950	992	982	952
0.050	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.200	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.500	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
	15 years of record											
0.000	60	60	54	50	58	32	162	146	44	90	78	50
0.002	230	184	146	202	168	140	296	248	126	200	170	122
0.005	784	696	628	724	680	606	740	692	454	608	564	434
0.020	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.050	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.200	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.500	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000

Table D.10 Comparison of significance level and power for trend tests. The errors are from a normal distribution with σ_r = 5.0 and missing value percentage = 0.3. Values in table are scaled by 1000.

SLOPE	ρ=0.0 MONTHLY			QUARTERLY			ρ=0.4 MONTHLY			QUARTERLY		
	MKD	SK	SKC	MKD	SK	SKC	MKD	SK	SKC	MKD	SK	SKC
	5 years of record											
0.000	88	58	20	50	32	6	170	118	6	92	50	12
0.002	72	44	6	80	38	12	164	108	12	94	54	12
0.005	78	54	12	60	32	12	160	108	16	112	56	10
0.020	216	126	14	216	122	24	298	202	22	208	100	18
0.050	756	606	110	638	506	142	690	584	104	570	414	94
0.200	1000	998	870	998	996	912	1000	996	794	996	994	826
0.500	1000	1000	992	1000	1000	996	1000	1000	992	1000	1000	1000
	10 years of record											
0.000	58	34	38	66	48	44	158	100	32	70	44	18
0.002	76	52	52	74	70	48	188	120	50	104	86	60
0.005	122	100	72	108	88	86	230	184	68	134	88	74
0.020	934	856	714	822	754	668	834	778	526	740	682	562
0.050	1000	1000	1000	1000	1000	1000	1000	1000	994	998	1000	994
0.200	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.500	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
	15 years of record											
0.000	56	46	46	64	40	44	166	140	42	98	84	58
0.002	124	110	92	102	82	74	208	162	72	124	108	76
0.005	372	326	284	310	288	268	440	364	210	296	270	212
0.020	1000	1000	1000	1000	998	998	1000	998	982	996	994	980
0.050	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.200	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.500	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000

Table D.11 Comparison of significance level and power for trend tests. The errors are from a lognormal distribution with σ_r = 1.0 and missing value percentage = 0.3. Values in table are scaled by 1000.

SLOPE	ρ=0.0 MONTHLY			QUARTERLY			ρ=0.4 MONTHLY			QUARTERLY		
	MKD	SK	SKC	MKD	SK	SKC	MKD	SK	SKC	MKD	SK	SKC
	5 years of record											
0.000	70	40	18	64	34	4	164	98	14	88	26	8
0.002	100	64	12	80	42	18	170	124	12	82	58	6
0.005	180	110	28	134	86	18	238	148	12	122	78	10
0.020	856	772	214	756	620	180	722	632	100	574	444	122
0.050	1000	998	744	996	988	784	996	988	596	978	950	580
0.200	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.500	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
	10 years of record											
0.000	74	52	34	78	58	40	186	144	64	92	60	40
0.002	218	182	138	166	142	108	264	214	76	156	132	82
0.005	788	708	538	612	548	438	564	494	248	396	346	234
0.020	1000	1000	1000	1000	1000	1000	1000	1000	988	996	998	986
0.050	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.200	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.500	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
	15 years of record											
0.000	54	40	40	54	40	44	190	152	48	100	82	48
0.002	546	488	416	390	372	310	436	400	210	294	262	180
0.005	1000	998	982	974	970	948	942	930	774	848	826	712
0.020	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.050	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.200	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.500	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000

Table D.12 Comparison of significance level and power for trend tests. The errors are from a lognormal distribution with σ_r = 5.0 and missing value percentage = 0.3. Values in table are scaled by 1000.

SLOPE	ρ=0.0 MONTHLY			QUARTERLY			ρ=0.4 MONTHLY			QUARTERLY		
	MKD	SK	SKC	MKD	SK	SKC	MKD	SK	SKC	MKD	SK	SKC
	5 years of record											
0.000	56	30	8	52	16	10	160	102	16	106	46	6
0.002	66	40	18	66	34	4	194	128	16	110	56	4
0.005	92	52	14	94	60	10	204	116	6	106	48	8
0.020	496	342	88	378	260	54	432	320	62	306	204	44
0.050	918	842	332	780	712	302	822	746	220	684	554	192
0.200	1000	1000	956	1000	1000	956	1000	998	892	998	994	920
0.500	1000	1000	996	1000	1000	994	1000	1000	1000	1000	1000	1000
	10 years of record											
0.000	80	62	54	50	56	48	178	124	40	110	98	58
0.002	122	92	70	96	76	60	188	174	68	108	86	54
0.005	362	296	210	234	208	166	326	288	92	206	184	102
0.020	994	984	968	966	964	942	954	940	800	882	838	752
0.050	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.200	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.500	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
	15 years of record											
0.000	46	42	40	42	36	34	162	146	42	84	74	38
0.002	278	246	202	222	190	146	286	266	122	174	160	104
0.005	810	732	682	636	582	540	646	632	378	494	456	344
0.020	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.050	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.200	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.500	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000

Table D.13 Comparison of significance level and power for trend tests. The errors are from a normal distribution with σ_r = 1.0 and missing value percentage = 0.5. Values in table are scaled by 1000.

SLOPE	ρ=0.0 MONTHLY			QUARTERLY			ρ=0.4 MONTHLY			QUARTERLY		
	MKD	SK	SKC	MKD	SK	SKC	MKD	SK	SKC	MKD	SK	SKC
	5 years of record											
0.000	76	34	12	66	32	0	132	58	18	84	52	6
0.002	78	34	10	64	36	10	124	60	14	92	42	8
0.005	90	40	12	78	50	6	152	60	14	116	64	14
0.020	354	186	38	358	222	36	372	234	32	354	218	38
0.050	906	754	182	928	862	282	864	746	148	910	756	210
0.200	1000	1000	888	1000	1000	994	1000	1000	908	1000	1000	992
0.500	1000	1000	956	1000	1000	996	998	1000	944	1000	1000	1000
	10 years of record											
0.000	54	48	40	58	28	22	104	68	34	58	40	22
0.002	90	68	34	78	62	34	152	130	58	122	106	70
0.005	270	188	120	206	170	122	242	162	78	210	166	102
0.020	990	962	884	990	982	950	990	978	848	986	972	874
0.050	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.200	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.500	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
	15 years of record											
0.000	54	60	46	56	36	30	148	108	48	94	68	50
0.002	174	146	118	150	144	124	250	196	98	184	140	106
0.005	646	518	416	604	526	444	602	540	338	520	468	354
0.020	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.050	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.200	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.500	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000

Table D.14 Comparison of significance level and power for trend tests. The errors are from a normal distribution with σ_r = 5.0 and missing value percentage = 0.5. Values in table are scaled by 1000.

SLOPE	ρ=0.0 MONTHLY			QUARTERLY			ρ=0.4 MONTHLY			QUARTERLY		
	MKD	SK	SKC	MKD	SK	SKC	MKD	SK	SKC	MKD	SK	SKC
	5 years of record											
0.000	78	48	26	78	28	6	136	82	14	86	42	10
0.002	86	42	6	76	40	6	130	70	8	104	54	8
0.005	86	38	18	70	38	16	124	56	24	90	48	10
0.020	164	84	14	148	82	30	212	100	10	168	82	14
0.050	478	312	82	480	362	112	466	344	48	448	316	70
0.200	966	944	504	982	970	704	960	922	482	972	942	686
0.500	998	1000	782	1000	998	962	998	996	804	1000	998	964
	10 years of record											
0.000	66	44	18	60	42	22	124	72	14	84	50	22
0.002	80	46	38	66	58	32	146	98	50	94	68	48
0.005	100	70	52	100	80	72	166	126	52	128	76	48
0.020	746	604	446	688	620	494	664	524	324	600	526	406
0.050	998	982	950	996	988	970	992	986	932	988	978	970
0.200	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.500	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
	15 years of record											
0.000	58	56	48	54	44	40	138	96	40	98	78	52
0.002	108	56	56	86	76	80	168	116	72	100	98	58
0.005	288	230	192	278	220	196	354	286	170	254	232	174
0.020	1000	978	954	992	980	966	994	972	918	992	978	946
0.050	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.200	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.500	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000

Table D.15 Comparison of significance level and power for trend tests. The errors are from a lognormal distribution with σ_r = 1.0 and missing value percentage = 0.5. Values in table are scaled by 1000.

SLOPE	ρ=0.0 MONTHLY			QUARTERLY			ρ=0.4 MONTHLY			QUARTERLY		
	MKD	SK	SKC	MKD	SK	SKC	MKD	SK	SKC	MKD	SK	SKC
	5 years of record											
0.000	74	28	6	74	28	6	124	48	2	72	38	4
0.002	94	40	6	76	28	10	124	68	16	74	46	4
0.005	144	82	18	104	88	20	184	84	10	136	86	12
0.020	624	486	122	570	462	94	556	400	58	494	358	64
0.050	958	938	422	962	942	620	942	894	344	942	872	470
0.200	1000	1000	928	1000	1000	994	1000	1000	920	1000	1000	990
0.500	1000	1000	954	1000	1000	998	998	1000	944	1000	1000	994
	10 years of record											
0.000	74	52	48	54	50	30	146	102	36	88	72	42
0.002	164	142	128	138	120	86	226	170	70	158	104	72
0.005	606	492	336	502	424	308	464	392	194	334	314	200
0.020	1000	1000	998	1000	1000	998	998	992	964	994	992	978
0.050	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.200	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.500	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
	15 years of record											
0.000	58	46	36	48	36	42	110	90	40	82	72	52
0.002	386	308	268	288	254	218	342	310	170	268	254	170
0.005	962	938	894	912	898	858	860	816	652	762	750	646
0.020	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.050	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.200	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.500	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000

Table D.16 Comparison of significance level and power for trend tests. The errors are from a lognormal distribution with σ_r = 5.0 and missing value percentage = 0.5. Values in table are scaled by 1000.

SLOPE	ρ=0.0 MONTHLY			QUARTERLY			ρ=0.4 MONTHLY			QUARTERLY		
	MKD	SK	SKC	MKD	SK	SKC	MKD	SK	SKC	MKD	SK	SKC
	5 years of record											
0.000	58	36	14	62	40	6	124	48	10	74	46	2
0.002	74	26	12	46	32	8	154	84	12	100	40	10
0.005	94	46	18	64	38	12	146	68	8	94	54	16
0.020	344	212	46	330	196	40	308	190	30	262	158	36
0.050	664	546	144	668	576	226	618	480	116	562	464	142
0.200	994	986	670	990	992	830	986	968	592	978	976	772
0.500	1000	1000	842	1000	1000	970	996	998	868	1000	1000	972
	10 years of record											
0.000	70	54	46	56	50	44	170	110	40	104	74	40
0.002	92	64	52	90	78	62	170	118	38	112	84	44
0.005	240	178	114	176	138	116	236	170	70	166	152	86
0.020	942	890	790	892	872	792	850	800	610	830	794	670
0.050	1000	998	996	998	998	998	1000	1000	986	998	998	994
0.200	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.500	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
	15 years of record											
0.000	30	18	18	38	34	32	122	86	34	84	64	36
0.002	214	170	136	158	144	130	222	188	88	156	144	100
0.005	626	548	472	502	456	402	518	464	270	418	366	262
0.020	1000	1000	998	1000	998	998	1000	1000	984	998	994	992
0.050	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.200	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.500	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000

Table D.17 Comparison of significance level and power for trend tests. The errors are from a normal distribution with σ_r = 1.0 and missing value percentage = 0.6. Values in table are scaled by 1000.

SLOPE	ρ=0.0 MONTHLY			QUARTERLY			ρ=0.4 MONTHLY			QUARTERLY		
	MKD	SK	SKC	MKD	SK	SKC	MKD	SK	SKC	MKD	SK	SKC
	5 years of record											
0.000	82	30	12	58	32	10	116	54	4	80	34	4
0.002	58	28	10	64	40	12	88	36	8	92	38	20
0.005	64	26	4	66	34	4	128	44	10	96	46	10
0.020	236	120	28	306	182	44	278	168	14	290	166	30
0.050	728	538	114	856	686	194	716	518	84	822	646	150
0.200	996	984	650	1000	1000	922	982	984	680	1000	1000	926
0.500	992	990	798	1000	1000	964	992	994	784	1000	1000	984
	10 years of record											
0.000	56	46	28	74	38	40	72	52	22	52	46	30
0.002	90	44	48	68	60	50	132	96	46	112	100	58
0.005	224	140	92	174	124	92	182	112	54	164	142	98
0.020	958	874	740	978	946	878	966	884	702	952	934	830
0.050	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.200	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.500	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
	15 years of record											
0.000	70	56	50	58	42	44	134	98	54	96	86	62
0.002	162	100	98	140	130	108	178	140	78	136	98	72
0.005	532	388	310	532	456	394	506	422	292	476	414	326
0.020	1000	1000	998	1000	1000	998	1000	1000	998	1000	1000	1000
0.050	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.200	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.500	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000

Table D.18 Comparison of significance level and power for trend tests. The errors are from a normal distribution with σ_r = 5.0 and missing value percentage = 0.6. Values in table are scaled by 1000.

SLOPE	ρ=0.0 MONTHLY			QUARTERLY			ρ=0.4 MONTHLY			QUARTERLY		
	MKD	SK	SKC	MKD	SK	SKC	MKD	SK	SKC	MKD	SK	SKC
	5 years of record											
0.000	74	34	12	56	24	10	110	58	10	90	46	12
0.002	72	30	12	66	28	8	102	46	14	90	42	2
0.005	78	28	10	54	38	12	100	40	8	78	46	14
0.020	128	64	14	104	72	16	166	84	14	132	80	12
0.050	362	210	38	388	264	76	374	216	40	386	280	62
0.200	862	792	316	932	904	568	842	760	306	928	858	524
0.500	974	930	540	994	996	862	978	952	556	996	990	880
	10 years of record											
0.000	66	38	32	54	36	30	110	64	36	72	52	36
0.002	64	40	42	60	56	42	104	72	40	92	70	34
0.005	94	80	62	98	88	62	148	100	54	108	84	62
0.020	598	444	314	608	500	394	542	402	270	514	436	344
0.050	956	930	848	982	960	942	972	924	848	978	958	908
0.200	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.500	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
	15 years of record											
0.000	68	58	44	42	40	26	122	76	56	88	72	64
0.002	86	62	54	90	80	72	124	96	60	82	68	56
0.005	252	202	162	226	196	182	300	224	148	258	214	140
0.020	970	910	866	960	936	928	964	920	828	974	936	898
0.050	1000	1000	998	1000	1000	1000	998	1000	996	1000	1000	1000
0.200	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.500	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000

Table D.19 Comparison of significance level and power for trend tests. The errors are from a lognormal distribution with σ_r = 1.0 and missing value percentage = 0.6. Values in table are scaled by 1000.

SLOPE	ρ=0.0 MONTHLY			QUARTERLY			ρ=0.4 MONTHLY			QUARTERLY		
	MKD	SK	SKC	MKD	SK	SKC	MKD	SK	SKC	MKD	SK	SKC
	5 years of record											
0.000	78	30	10	74	32	6	106	36	4	72	34	8
0.002	68	30	8	68	30	12	106	48	6	92	44	4
0.005	128	66	16	84	40	14	140	58	14	118	68	6
0.020	462	336	78	494	380	84	430	280	24	416	284	50
0.050	844	746	220	928	856	446	822	692	200	894	800	374
0.200	980	980	696	996	998	952	994	986	680	998	998	944
0.500	990	998	748	1000	1000	980	978	988	760	998	998	966
	10 years of record											
0.000	76	46	30	54	38	26	124	80	38	80	58	28
0.002	128	108	86	120	96	70	170	122	56	130	112	64
0.005	470	352	230	434	358	240	402	292	160	328	272	170
0.020	998	996	966	1000	996	986	986	974	884	988	982	952
0.050	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.200	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.500	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
	15 years of record											
0.000	54	44	38	52	46	46	110	80	50	76	72	34
0.002	304	242	194	238	226	182	280	234	140	232	234	172
0.005	896	824	788	824	814	740	780	708	532	728	688	570
0.020	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.050	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.200	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.500	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000

Table D.20 Comparison of significance level and power for trend tests. The errors are from a lognormal distribution with σ_r = 5.0 and missing value percentage = 0.6. Values in table are scaled by 1000.

SLOPE	ρ=0.0 MONTHLY			QUARTERLY			ρ=0.4 MONTHLY			QUARTERLY		
	MKD	SK	SKC	MKD	SK	SKC	MKD	SK	SKC	MKD	SK	SKC
	5 years of record											
0.000	38	24	8	44	30	6	112	34	6	72	46	14
0.002	66	28	18	60	22	8	120	52	6	78	40	12
0.005	88	36	10	64	26	8	116	50	14	80	50	14
0.020	220	104	20	256	134	34	270	130	28	224	134	24
0.050	486	356	104	588	470	152	482	354	76	486	378	124
0.200	932	904	408	972	956	676	906	854	424	954	932	658
0.500	980	984	614	996	998	898	978	968	630	1000	992	906
	10 years of record											
0.000	54	48	40	54	36	42	128	78	32	82	54	30
0.002	84	64	46	90	80	56	130	86	36	84	72	60
0.005	206	112	92	168	138	110	200	140	62	162	120	84
0.020	832	782	652	828	782	714	746	688	472	750	696	558
0.050	994	990	970	994	994	988	998	986	934	992	992	974
0.200	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.500	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
	15 years of record											
0.000	38	32	36	34	24	22	84	70	32	56	66	42
0.002	176	126	116	142	116	110	178	148	76	144	124	90
0.005	516	456	376	436	394	364	438	368	234	342	324	258
0.020	998	996	980	998	996	996	996	984	958	994	986	960
0.050	1000	1000	998	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.200	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.500	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000

Table D.21 Comparison of significance level and power for trend tests. The errors are from a normal distribution with σ_r = 1.0 and missing value percentage = 0.7. Values in table are scaled by 1000.

SLOPE	ρ=0.0 MONTHLY			QUARTERLY			ρ=0.4 MONTHLY			QUARTERLY		
	MKD	SK	SKC	MKD	SK	SKC	MKD	SK	SKC	MKD	SK	SKC
	5 years of record											
0.000	68	22	6	72	24	6	78	18	4	74	24	8
0.002	70	16	0	58	36	14	68	24	6	58	20	14
0.005	52	16	0	76	30	6	90	34	12	74	42	14
0.020	160	64	10	234	134	24	194	80	10	226	116	18
0.050	494	260	54	690	442	106	454	250	44	654	432	102
0.200	858	786	270	974	972	716	836	794	360	968	960	726
0.500	832	814	388	982	980	796	846	832	396	980	982	816
	10 years of record											
0.000	60	38	30	66	44	50	68	42	28	56	30	24
0.002	78	42	40	80	74	50	106	74	42	88	70	44
0.005	148	88	66	148	102	88	150	78	46	150	108	62
0.020	816	676	516	904	850	718	852	692	506	902	836	652
0.050	998	996	966	1000	1000	998	1000	998	970	1000	1000	1000
0.200	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.500	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
	15 years of record											
0.000	52	36	40	56	60	52	112	86	52	88	86	66
0.002	138	82	80	134	102	96	156	120	66	110	88	66
0.005	372	254	192	424	312	298	376	264	188	396	314	230
0.020	1000	994	988	1000	1000	998	1000	996	976	1000	998	998
0.050	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.200	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.500	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000

Table D.22 Comparison of significance level and power for trend tests. The errors are from a normal distribution with σ_r = 5.0 and missing value percentage = 0.7. Values in table are scaled by 1000.

SLOPE	ρ=0.0 MONTHLY			QUARTERLY			ρ=0.4 MONTHLY			QUARTERLY		
	MKD	SK	SKC	MKD	SK	SKC	MKD	SK	SKC	MKD	SK	SKC
	5 years of record											
0.000	76	18	2	52	26	10	104	32	8	72	36	6
0.002	86	26	4	68	34	8	100	32	8	82	32	8
0.005	60	14	0	74	20	4	82	24	2	92	38	2
0.020	96	38	8	96	50	14	134	42	12	110	46	8
0.050	222	88	22	286	180	34	252	116	26	288	176	32
0.200	622	468	136	810	708	392	584	464	132	804	662	356
0.500	780	670	200	954	936	610	846	720	274	954	930	628
	10 years of record											
0.000	38	22	16	50	34	32	108	56	36	90	56	54
0.002	52	34	30	68	48	46	96	56	30	88	54	40
0.005	76	58	30	88	60	44	122	88	36	92	76	46
0.020	398	282	184	446	366	280	376	260	178	406	334	222
0.050	880	782	658	930	874	832	882	790	650	918	886	818
0.200	1000	1000	992	1000	1000	1000	1000	1000	994	1000	1000	1000
0.500	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
	15 years of record											
0.000	56	44	38	42	36	32	90	64	36	66	56	44
0.002	84	60	52	64	50	40	90	64	54	72	66	58
0.005	150	132	124	186	154	134	206	136	96	190	164	116
0.020	866	756	684	894	838	798	882	764	622	908	842	782
0.050	998	990	988	1000	1000	1000	998	992	980	1000	994	996
0.200	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.500	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000

Table D.23 Comparison of significance level and power for trend tests. The errors are from a lognormal distribution with σ_r = 1.0 and missing value percentage = 0.7. Values in table are scaled by 1000.

SLOPE	ρ=0.0 MONTHLY			QUARTERLY			ρ=0.4 MONTHLY			QUARTERLY		
	MKD	SK	SKC	MKD	SK	SKC	MKD	SK	SKC	MKD	SK	SKC
	5 years of record											
0.000	58	12	2	60	30	8	84	26	10	64	30	8
0.002	64	18	8	68	32	10	98	32	6	78	36	8
0.005	108	24	4	72	28	6	86	32	2	104	50	8
0.020	284	158	40	400	244	50	244	132	12	326	174	42
0.050	602	424	86	806	662	240	604	414	68	778	608	214
0.200	830	790	334	976	952	768	840	806	320	978	968	740
0.500	888	830	338	982	984	802	860	836	410	974	970	804
	10 years of record											
0.000	66	34	24	56	34	24	104	66	38	72	62	38
0.002	102	72	52	104	78	66	124	92	52	108	78	52
0.005	322	206	136	324	248	168	270	188	88	268	208	146
0.020	962	912	818	988	970	930	936	878	738	956	932	854
0.050	1000	998	998	1000	1000	1000	1000	998	990	1000	1000	998
0.200	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.500	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
	15 years of record											
0.000	64	40	40	40	32	42	96	68	44	76	62	34
0.002	228	160	122	226	174	140	230	164	104	200	178	122
0.005	734	652	556	738	680	620	640	526	388	614	568	478
0.020	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	998
0.050	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.200	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.500	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000

Table D.24 Comparison of significance level and power for trend tests. The errors are from a lognormal distribution with σ_r = 5.0 and missing value percentage = 0.7. Values in table are scaled by 1000.

SLOPE	ρ=0.0 MONTHLY			QUARTERLY			ρ=0.4 MONTHLY			QUARTERLY		
	MKD	SK	SKC	MKD	SK	SKC	MKD	SK	SKC	MKD	SK	SKC
	5 years of record											
0.000	36	22	4	44	24	6	86	22	4	78	28	8
0.002	80	16	6	70	20	4	94	24	0	80	40	10
0.005	74	16	2	64	26	6	96	30	2	76	34	6
0.020	132	50	12	192	98	28	186	62	12	162	78	14
0.050	326	200	48	452	310	82	312	166	30	384	266	94
0.200	730	576	174	860	792	450	688	558	186	852	776	438
0.500	816	746	288	958	948	686	820	752	276	956	946	682
	10 years of record											
0.000	54	32	24	54	36	20	100	58	30	86	58	36
0.002	78	44	30	78	58	52	120	72	48	104	72	64
0.005	112	74	56	140	96	90	146	106	60	138	110	60
0.020	644	532	430	732	646	556	592	478	300	648	570	438
0.050	936	920	840	982	974	936	926	902	782	960	954	926
0.200	1000	998	998	1000	1000	1000	1000	1000	998	1000	1000	1000
0.500	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
	15 years of record											
0.000	44	38	34	36	34	48	92	74	34	64	50	26
0.002	114	90	96	98	96	90	144	102	62	134	106	78
0.005	374	322	276	350	312	286	332	242	152	314	268	208
0.020	962	946	904	986	976	964	950	906	808	974	944	910
0.050	1000	998	998	1000	1000	1000	1000	996	994	1000	1000	1000
0.200	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.500	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000