A Statistical Framework for the Onondaga Lake Ambient Monitoring Program

prepared for

Onondaga County, Department of Drainage & Sanitation

by

William W. Walker, Jr., Ph.D., Environmental Engineer

Draft, July 13, 1998

Introduction

One of the primary purposes of the Ambient Monitoring Program is to provide information for supporting future decisions on wastewater and watershed management. Future decisions may be based in part upon changes detected in Onondaga Lake and Seneca River over the next several years. Decisions may also rely upon comparisons of monitored conditions with water quality standards or management goals. The ability to detect such changes and the reliability of such comparisons depends in part upon the design of the monitoring program. Decisions should not be made based upon the monitoring results without an adequate understanding of the sources and magnitudes of variability in the data.

This section describes and demonstrates a statistical framework (Figure 1) that is an integral part of the monitoring program. The framework has been designed to provide the following functions:

- identifying and quantifying sources of variability in the data;
- evaluating uncertainty associated with summary statistics;
- formulating and testing specific hypotheses; and
- refining monitoring program designs;

Continuous implementation of this framework over the course of the monitoring program will help to ensure that data-collection efforts are cost-effective and that the resulting data base is adequate to support future management decisions.

To some extent, elements of the framework are already in place under the existing lake monitoring program. Similar statistical concepts and procedures were used in evaluating of lake monitoring data collected through 1990 (Walker, 1991b). Routine trend analyses have become a standard component of annual lake monitoring reports (Stearns & Wheler, 1997). The framework is demonstrated below using data from the historical lake monitoring program. Steps required to implement the framework are also described. Methodologies will be refined and applied to key variables tracked under the expanded monitoring effort described in the Ambient Monitoring Program, including both water quality and biological measurements.

Type I and Type II Errors

Applications of monitoring data and statistical methods to address management questions generally involves the following procedure:

1. Identify the management question;
2. Formulate a null hypothesis (H_0);
3. Identify an appropriate statistical procedure to test H_0
4. Apply procedure to estimate an outcome probability 'p', assuming that H_0 is true;
5. Accept or reject H_0 based upon the value of p relative to a pre-selected significance level, α;
6. Interpret the test results, i.e., translate them back into language that can be understood by decision-makers in the context of the question that has been asked;

While the following discussion focuses on Steps 2-5, Steps 1 and 6 are also essential if the effort is to be successful.

A typical management question is whether concentrations of a particular water quality component "changed" over a specified time period. This is typically translated into a null hypothesis that there is no "trend" in the data (i.e., that the long-term mean or central tendency of the measurements has not changed). The Seasonal Kendall Test (Hirsch et al., 1982; Hirsch & Slack, 1984) is a statistical method that is well-suited for testing this hypothesis. The test estimates the probability "p" of observing a particular set of data if the underlying central tendency of lake concentrations is stable. The outcome probability is compared with a pre-selected "significance level" or "α". If $p>\alpha$, then the null hypothesis is accepted. This does not "prove" that there is no trend in the data; it merely indicates that any trend, if present, is too small to be detectable in the presence of background variability over this tested time interval. If $p<\alpha$, then the null hypothesis is rejected. This indicates that a trend is "likely" with a "confidence level" of 1-p.

The selection of a significance level (α) for the test is critical in this process. Essentially, α represents the assumed maximum risk of falsely identifying a trend (incorrectly rejecting the null hypothesis). This it termed a "Type I" error (Snedocor & Cochran, 1989) . Although α levels of .1, .05, and .01 are often used in scientific investigations, the choice of α is ultimately subjective. Decision- makers relying on test results may have different risk tolerances (Berryman et al., 1988) For this reason, it is important to report the outcome probability p, so that decision-makers can form their own conclusions depending on their own risk tolerances. This is more realistic that just reporting a true/false outcome, which inappropriately conveys the notion that there is no uncertainty in the conclusion.

The choice of α has important implications because it determines the maximum risk of a "Type II" error (failing to detect a real trend or incorrectly accepting the null hypothesis). Possible test outcomes are illustrated in the following table (Lettenmaier, 1975,1976):

	Reality	
Test Outcome	**No Trend Exists**	**Trend Exists**
H_0 Accepted **Trend Unlikely**	Correct	Type II Error max. prob = β = 1-α
H_0 Rejected **Trend Likely**	Type I Error max. prob = α	Correct

The maximum risk of a Type I Error is selected when the choice of α is made. The maximum risk of a Type II error, often designated as "β", is equal to 1 – α. Selecting lower levels of α will increase β and decrease the probability of detecting real trends in the data, particularly small ones. As noted by Berryman et al. (1988), Type II errors (failing to detect real trends) may in fact be of more concern to resource managers and decision-makers than Type I errors (false trends). The inverse relationship between β and α should be recognized in selecting an appropriate α level for use in testing hypotheses.

A significance level of α=0.10 has been used routinely in Onondaga Lake and tributary water quality data for trends (Walker, 1991b; Stearns & Wheler et al, 1997). This has been used in the context of null hypotheses formulated without regard to trend direction (two-tailed tests). These tests would be essentially identical to one-tailed tests (designed to address questions like "is the lake improving") conducted at a significance level of 0.05. It is proposed that a significance level of 0.10 be used routinely in testing two-tailed hypotheses under the Ambient Monitoring Plan. As discussed above, outcome probabilities (p), will also be routinely reported, so that readers can interpret results according to their own risk tolerances.

As stated above, the maximum risk of a Type II error ("β") is equal to 1 – α. This maximum risk is encountered only when trend magnitudes are extremely small and/or variance in the data is extremely high. The actual risk of Type II error in the context of a particular data set and test depends upon the following factors:

- magnitude of the trend or change to be detected
- duration or length of the data set
- variability in the data
- statistical methods employed

Trends are relatively easy to detect (low β), when the "signal-to-noise" ratio is high (large trends, long data sets, low variability) and the best statistical methods are used. Trends are relatively difficult to detect (high β), when the ratio is low (weak trends, short data sets, high variability) and/or inappropriate statistical methods are used. Statistical theory and simulation techniques can be used to quantify these relationships (Lettenmaier, 1975, 1976; Hirsch et al., 1982; Loftis et al., 1989; Thas et al., 1998).

Because they influence to some extent the amount of variability in the data, monitoring program designs can influence Type II error. In the context of designing monitoring programs, its useful to consider the concept of "Power", or the probability of detecting a real trend (Lettenmaier, 1976). Numerically, power is equal to 1-β. A "good" or "cost-effective" monitoring program is one that has the most power (lowest risk of Type II error) for a given investment.

Once the sources and magnitudes of variability in the data have been characterized, relationships between power and monitoring program designs (sampling frequency, etc.) can be investigated in the context of testing specific hypotheses. This, in turn, provides a basis for evaluating and refining monitoring program designs and is a key element of the statistical framework described below (Figure 1). Although discussed above in the context of a trend analysis, the concepts apply to other types of hypothesis tests.

Variance Component Analysis

Variance component models (Snedocor & Cochran, 1989) explicitly represent the sources and magnitudes of variability in monitoring data. They provide a basis for estimating the uncertainty associated with yearly and long-term summary statistics and for estimating the power of trend tests or other hypothesis tests. The following example (Walker, 1991b) illustrates the structure, calibration, and applications of such models in evaluating monitoring program designs. Calculations are performed using a version of LRSD.WK1 ("Lake & Reservoir Sampling Design") spreadsheet (Walker, 1988a, 1988b), which incorporates procedures described by Smeltzer et al. (1989), Knowlton et al. (1984), Walker (1980), Lettenmaier (1976), and Loftis et al. (1991).

A nested analysis of variance model can be used to describe concentration variations in samples collected in the mixed surface layer of the lake:

$$c_{ijk} = \mu + y_i + d_{ij} + z_{ijk}$$

where,

c_{ijk}	=	concentration measured in sample k collected on date j in year i
μ	=	long-term mean
y_i	=	effect of year i (mean = 0, standard deviation = S_y), i = 1 to n_y
d_{ij}	=	effect of date j in year i (mean = 0, standard deviation = S_d), j = 1 to n_d
e_{ijk}	=	depth effect (mean = 0, standard deviation = S_z), k = 1 to n_z

The samples are assumed to be taken from a homogenous stratum (e.g. surface mixed layer).

The term μ reflects the long-term mean concentration. This value cannot be directly measured, but it can be estimated based upon data from a specific time interval.

Theoretically, μ is independent of random variations (e.g., climate, hydrology, seasons, sampling error) that can influence measured concentrations and that are represented by the remaining terms in the equation. Management actions or other anthropogenic factors may cause changes in μ. A key objective of the monitoring program is to provide data for detecting such changes in the presence of other sources of variability. The monitoring data and variance component model can also be used to estimate the probability of attaining lake-management goals expressed as annual or long-term average concentrations.

The yearly term reflects random year-to-year variations in climate or other factors influencing lake conditions (e.g., wet-year vs. dry-year variations). The sampling date term reflects random variations from one sampling date to the next within a given year. This term can be further partitioned into fixed seasonal effects and random effects. The depth term reflects variations within the mixed layer at the sampling station and variations associated with the measurement process (random sampling & analytical error). Replicate sampling data can be used to partition the depth term into "real" and measurement-related variations.

Assuming that the variance terms are independent, the total variance in the measurement, S^2, can be calculated from the following equation:

$$S^2 = S^2_y + S^2_d + S^2_z$$

The terms of the equation represent yearly, daily, and depth variance components. Basic elements of the monitoring program design are represented by the following variables:

n_y = number of years of monitoring
n_d = number of sampling dates per year
n_z = number of sampled depths per date

Calibration of the model involves estimation of parameters S_y, S_d, and S_z. This can be accomplished by applying a nested analysis of variance (Snedecor & Cochran, 1989) to data derived from several years of monitoring. Fixed seasonal effects and long-term trend are removed from the data prior to estimating variance components by subtracting monthly medians (computed from all years) from each sample. Long-term trend is removed by computing the median of the deseasoned values within each year, regressing yearly medians against year, and applying the regression slope to the year associated with each sample. The serial correlation of detrended, deseasoned, daily-median values is also calculated for the purpose of estimating its effects on the precision of yearly and long-term means (Loftis et al., 1991; Muskens & Kateman, 1978).

Because of skewness in the distributions of many lake measurements (nutrient concentrations, algae, etc.), analyses are often conducted on log-transformed concentration data, although the same methodology can be applied to un-transformed data. When the analysis is performed on log-transformed data, results can be used to estimate the long-term geometric mean (e^{μ}).

The following parameter estimates are derived from log-transformed total inorganic phosphorus data collected in Onondaga Lake between 1981 and 1990 (Walker, 1991b). The data set is restricted to samples collected in the mixed layer between April and September. Samples were collected biweekly from the mixed layer at depths of 0, 3, and 6 meters. The sampling program is characterized by the following

n_y = 10 (10 years of data)
n_d = 13 (180 days per season / 14 days between sampling events)
n_z = 3 (3 depths sampled on each date within the mixed layer)

Calibrated standard deviations are as follows:

S_y = 0.206
S_d = 0.377
S_z = 0.300

According to the above equation, the total variance in the concentration data is given by:

$$S^2 = S^2_y + S^2_d + S^2_z$$

$$= (0.206)^2 + (0.377)^2 + (0.300)^2$$

$$= 0.042 + 0.142 + 0.090 = 0.275$$

Results indicate that yearly, daily, and depth variations account for 15%, 52%, and 33% of the total concentration variance, respectively. Potential uses of the calibrated model in evaluating monitoring program designs are described below.

Precision in Yearly and Long-Term Means

Monitoring objectives may include collection of data for estimating yearly-average or long-term-average lake conditions. These statistics are potentially important because they may be used to classify the water body according to trophic state, for example. Management goals or standards may be expressed in terms of yearly or long-term average conditions. The variance component model described above can be used to estimate the uncertainty in yearly and long-term means calculated from the data. The dependence of that uncertainty on sampling design parameters (n_d, n_z) can also be examined.

According to the nested analysis of variance model (Snedocor & Cochran, 1989), the following equations can be used to estimate the uncertainty associated with a yearly mean or long-term mean calculated from a given set of data:

$$E^2_i = [\ S^2_d / n_d + S^2_z / (n_d\, n_z)\ ]\ F_r$$

$$E^2_\mu = S^2_y / n_y + E^2_i / n_y$$

where,

E_i	=	standard error of the mean for year i
E_μ	=	standard error of long-term mean
F_r	=	factor accounting for serial correlation between sampling dates (Muskens & Kateman, 1978)

For the lake total inorganic phosphorus time series described above, the serial correlation adjustment factor for a biweekly sampling frequency is computed at F_r = 0.28. When the analysis is performed on logarithmic scales, uncertainty can be expressed as a coefficient of variation or relative standard error (= standard error / geometric mean) using the following equation:

$$CV(e^u) = E_\mu$$

where E_μ is computed using log-transformed data.

The above equations can be used to examine the sensitivity of uncertainty in yearly means (E_i) and long-term means (E_μ) to sampling frequencies, as represented by n_d and n_z. Figure 2 shows relative standard errors of yearly and long-term means based upon 5 years of growing-season data with assumed sampling frequencies ranging from yearly (one sampling date per year) to daily. The shaded areas in each bar indicate the relative contribution of each variance component to uncertainty in the mean. The importance of daily and depth variations decrease substantially as sampling frequency is increased from yearly to monthly. After that, the response is relatively flat because uncertainty is controlled by year-to-year variations. The analysis demonstrates that increasing the sampling frequency from biweekly to weekly would provide little benefit in terms of reducing uncertainty in the long-term geometric mean (CV = 9.6% vs. 9.3%).

The CV of the annual geometric mean would decrease from 6.2% for biweekly sampling to 3.1% for weekly sampling. This improvement relies heavily on variations in the serial correlation adjustment factor (F_r). Without the adjustment, sensitivity to sampling frequency would be lower. Refinements to the statistical framework will include further testing of the adjustment procedure. Although increasing the sampling frequency from biweekly to weekly may improve the precision of yearly geometric mean estimates, the CV's for both sampling frequencies are relatively low. Increasing the sampling frequency would be justified if a CV of 6.2% in the annual geometric mean were determined to be unacceptable. This level of uncertainty is substantially lower than relative standard errors characteristic of empirical phosphorus mass-balance models (CV = 27%, Walker, 1985). The biweekly sampling frequency seems adequate to support application of such models.

Once the variance components have been defined, theoretical equations or simulation techniques can be used to estimate the uncertainty in any yearly or long-term-average summary statistic (e.g., arithmetic mean, median, 90th percentile, exceedence frequency, stream pollutant loads). Sensitivity to sampling frequency can be evaluated in a manner similar to that demonstrated above. These results, in turn, can be used as a basis for evaluating the adequacy of sampling frequency for any water quality component.

Power for Detecting Step Changes in the Long-Term Mean

As discussed above, detection of changes in the long-term mean is a primary monitoring objective. Variance component estimates can also be used to estimate the probability of detecting changes in the long-term mean using data from different monitoring periods. When a t-test (Montgomery & Loftis, 1987) is used for this purpose, the following equations are involved:

$$t = (m_1 - m_2) / E_{12}$$

$$E_{12} = (E^2_{\mu 1} + E^2_{\mu 2})^{1/2}$$

$$dof = n_{y1} + n_{y2} - 2$$

Null Hypothesis: $\mu_1 = \mu_2$, accepted if $|t| < \alpha$

where,

m_1 = measured mean in period 1

m_2 = measured mean in period 2

E_{12} = standard error of difference in long-term means between period 1 and 2

dof = degrees of freedom

α = significance level (two-tailed test)

$t_{\alpha,dof}$ = value of student's t at significance level α and dof degrees of freedom

The power or probability of detecting an actual change in the mean ($\mu_1 - \mu_2$) can be estimated from the following equations (Lettenmaier, 1976):

$$N_t = | \mu_1 - \mu_2 | / S_{12}$$

$$Power = F (N_t - t_{\alpha,dof} , dof)$$

where,

N_t = dimensionless trend number

F = cumulative distribution of Student's t with dof degrees of freedom

Power = probability of detecting change
= probability that t will exceed the critical value $t_{\alpha,dof}$

Generally, the probability of detecting a change in the long-term mean increases with the magnitude of the change ($| \mu_1 - \mu_2 |$) and decreases with the magnitude of the standard error term (E_{12}).

Figure 3 plots power for detecting a 25% change in total inorganic phosphorus as a function of duration (number of monitoring years before and after the step change) and sampling frequency (ranging from yearly to daily) for a two-tailed t-test conducted at a significance level α=0.1. Power increases substantially as the sampling frequency increases from yearly to biweekly. Further increases in sampling frequency would provide little benefit. This reflects the fact that uncertainty in the mean within each period is controlled by the yearly variance component when the sampling frequency is biweekly (Figure 2).

Figure 3 also shows power as a function of duration for step changes ranging from 10% to 80% and a biweekly sampling frequency. For durations ranging from 1 to 20 years, power for detecting a 10% change ranges from ~10 to ~40%. This demonstrates the difficulty of detecting small changes in the long-term mean, even with a daily sampling regime. Step changes >50%, however, are detectable with >90% probability for durations of 5 years or longer.

Lettenmaier (1976) and Loftis et al (1989) show that power dependence on the "trend number" is similar when non-parametric procedures are used for detecting changes in the mean or median. With ideal data sets (normally distributed, serially independent), non-parametric methods have slightly less power than the t-test. The value of the non-parametric methods is that they are more robust to outliers and deviations from normality (Helsel & Hirsch, 1988). The power of a t-test decreases significantly (relative to that shown in Figure 3) when the data are skewed or contain outliers. For this reason, non-parametric methods, such as the Wilcoxon rank-sum test (Snedecor & Cochran, 1989) are generally preferred in water quality applications. Because the 'trend number' is also a good power predictor for nonparametric tests, dependence on sampling frequency is not likely to be substantially different from that depicted in Figure 3. Once variance components have been estimated, simulation methods can be used to test power dependence on sampling frequency for any parametric or non-parametric procedure used in testing hypotheses.

Power for Detecting Linear Trends

Similar equations have been developed to estimate power for detecting linear trends in the long-term mean. When a regression analysis of yearly means is used to test for trend, the dimensionless trend number can be calculated from (Lettenmaier, 1976; Loftis et al., 1989):

$$N_t = b\,[\,n_y\,(n_y-1)\,(n_y+1)\,]^{1/2} \;/\; [\,12^{1/2}\,\sigma\,]$$

$$dof = n_y - 2$$

$$\text{Power} = F\,[\,N_t - t_{\alpha,dof},\ dof\,]$$

where,

b	=	trend magnitude (concentration units / year)
σ	=	standard deviation of yearly means in absence of trend
n_y	=	number of years tested

Using the variance component model, the standard deviation of yearly means can be calculated from:

$$\sigma^2 = S^2_y + [\,S^2_d / n_d + S^2_z / (n_d\, n_z)\,]\ F_r$$

The effects of monitoring frequencies (n_d, n_z) on power for detecting trends are reflected in these equations. Essentially, the dimensionless trend number represents a signal-to-noise ratio. The probability of detecting a trend increases with magnitude of the trend (b), increases with the number of years examined (n_y), and decreases with the standard deviation of the yearly means (σ).

Figure 4 plots power for detecting a 5%/yr trend in total inorganic phosphorus as a function of monitoring duration and sampling frequency when the tests are conducted at significance level α=0.1. Power increases substantially as the sampling frequency increases from yearly to biweekly. Further increases in sampling frequency would provide little benefit. For a biweekly frequency, power exceeds ~60% for durations of 10 years or longer.

Figure 4 also shows power as a function of duration for trend magnitudes ranging from 2% to 10%/year and a biweekly sampling frequency. For durations ranging from 1 to 20 years, power for detecting a 2%/year trend ranges from ~5% to ~75%. For a 10%/year trend, power exceeds 90% for durations of ~8 years or longer. This demonstrates the importance of maintaining long-term data sets for trend detection

As discussed above for step changes, non-parametric methods are generally preferred over linear regression in analyzing water quality data for trends because their power is less sensitive to outliers and deviations from normality (Helsel & Hirsch, 1992). The Seasonal Kendall test will be used in analyzing data from the Ambient Monitoring Program. Simulations performed by Hirsch et al. (1982), Loftis et al (1989) and Thas et al. (1998) indicate that the power of this test is also correlated with the 'trend number', as defined above. Thus, it is expected that dependence on sampling frequency is similar to that depicted in Figure 4. This hypothesis will be tested by simulation.

Role of Modeling

Figures 2-4 demonstrate that, with a biweekly sampling frequency, power for detecting changes or trends in total inorganic phosphorus concentration is controlled largely by the random year-to-year variance component. Increasing sampling frequency would not increase power for trend detection. Power can be potentially increased, however, by constructing models that "explain " a portion of the year-to-year variability (Hirsch et al., 1982; Walker, 1991a,1998). For example, a portion of year-to-year variability may be related to wet-year vs. dry-year influences. Such a relationship might be reflected by correlating yearly variations with precipitation or watershed runoff. If the correlation explained 50% of the year-to-year variations, the power for detecting a 5%/year trend with 10 years of data would increase from ~60% to ~83%. Similarly, variance in bacteria levels might be reduced by correlating measured values with antecedent rainfall.

The above concept applies to models with various degrees of complexity (simple regressions to complex simulations). At some point, investment in modeling becomes more cost-effective than investment in additional data collection if the objective is to increase precision in the long-term means or to increase power for detection of trends. While development of such models will not occur within the statistical framework, it is a potential recommendation. Relevant models that become available as the monitoring plan is implemented will be factored into the trend analyses.

Implementation

Implementation of the framework (Figure 1) involves the following tasks:

1) Compilation & validation of recent water quality data (1991-1997)
2) Refinement and testing of statistical methods
3) Estimation of variance components for water quality and biological variables
4) Estimation of uncertainty in yearly and long-term means
5) Formulation of hypotheses
    a) Long-term trends
    b) Comparisons with management goals
    c) Other
6) Hypothesis testing
7) Evaluation of power of hypothesis tests as a function of sampling frequency
8) Recommendation of improvements to the monitoring program design

Existing long-term data sets (required to estimate variance components) will provide a basis for initial implementation of the framework. Implementation will occur in two phases.

The first phase will include the following variables measured in the Lake and tributaries: total phosphorus, nitrogen (total, Kjeldahl, and ammonia N), chlorophyll-a, and transparency. The monitoring program design will be evaluated based upon criteria discussed above (uncertainty in yearly and long-term means, power for trend detection).

The second phase will include other water quality measurements and preliminary evaluations of biological measurements in the Lake, Tributaries, and Seneca River. The water quality and biological measurements to be evaluated in the second phase will be limited to those variables that are likely to influence management decisions. The Onondaga Lake Advisors Group will assist in identifying those measurements and in developing specific hypotheses to be tested in the framework (aside from trend detection). Once the second implementation phase is complete, it is anticipated that the analysis will be refined and updated on an annual basis.

Biological measurements recently added to the monitoring plan will be brought into the framework as sufficient data become available. While at least 5 years of data are desired for estimating year-to-year variance components, preliminary evaluations of within-year and sampling variations should be possible after the first year of data collection. To the extent possible based upon available information, literature values or sampling data from other locations can provide initial estimates of variance components for biological measurements. These can be used in preliminary evaluations of the monitoring plan and refined as site-specific data become available.

References

Berryman, D., B. Bobee, D. Cluis, & J. Haemmerli, "Nonparametric Tests for Trend Detection in Water Quality Time Series", Water Resources Bulletin, Vol. 24, No. 3, pp. 545-556, June 1988.

Helsel, D.R. & R.M. Hirsch, Discussion of "Applicability of the t-Test for Detecting Trends in Water Quality Variables", by R.H. Montgomery & J.C. Loftis, Water Resources Bulletin, Volume 24, No. 1, pp. 201-204, February 1988.

Helsel, D.R. and R.M. Hirsch, Statistical Methods in Water Resources, Elsevier, 1992.

Hirsch R.M., J.R. Slack, & R.A. Smith, "Techniques of Trend Analysis for Monthly Water Quality Data", Water Resources Research, Vol. 18, No. 1, pp. 107-121, February 1982.

Hirsch R.M. & J.R. Slack, "A Nonparametric Trend Test for Seasonal Data with Serial Dependence", Water Resources Research, Vol. 20, No. 6, pp. 727-732, June 1984.

Knowlton, M.F., M.V. Hoyer, & J.R. Jones, "Sources of Variability in Phosphorus and Chlorophyll and Their Effects on Use of Lake Survey Data, Water Resources Bulletin, Volume 20, pp. 397-407, 1984.

Lettenmaier, D.P., "Design of Monitoring Programs for Detection of Trends in Stream Quality", prepared for Washington State Department of Ecology, University of Washington Hydraulics Lab, NTIS PB-272-814, August 1975.

Lettenmaier, D.P. "Detection of Trends in Water Quality Data from Records with Dependent Observations", Water Resources Research, Vol. 12, No. 5, pp. 1037-1046, October 1976.

Loftis, J.C., R.C. Ward, R.D. Phillips, & C. Taylor, "An Evaluation of Trend Detection Techniques for Use in Water Quality Monitoring Programs", U.S. Environmental

Protection Agency, Office of Acid Deposition Environmental Monitoring and Quality Assurance, Washington, DC, EPA/600/3-89/037, March 1989.

Loftis, J.C., G.C. McBride, & J.C. Ellis, "Considerations of Scale in Water Quality Monitoring and Data Analysis", Water Resources Bulletin, Vol. 27, No. 2, pp. 255-264, April 1991.

Montgomery, R.H. & J.C. Loftis "Applicability of the t-Test for Detecting Trends in Water Quality Variables", Water Resources Bulletin, Volume 23, No. 4, pp. 653-662, August 1987.

Muskens, P.J.W.M. and G. Kateman, "Sampling of Internally Correlated Lots. The Reproducibility of Gross Samples as a Function of Sample Size, Lot Size, and Number of Samples", Analytical Chimica Acta, Vol. 103, pp. 1-9, 1978.

Smeltzer, E., W.W. Walker, & V. Garrison, "Eleven Years of Lake Eutrophication Monitoring in Vermont: A Critical Evaluation", in "Enhancing States' Lake Management Programs", U.S. Environmental Protection Agency, pp. 53-62, 1989.

Snedecor, G.W. & W.G. Cochran, Statistical Methods, 8th Edition, The Iowa State University Press, 1989.

Stearns & Wheler, Inc., R. Canale, E. Mills, Onondaga County Department of Drainage and Sanitation, C. Driscoll, "Report and Appendix A – Onondaga Lake Monitoring Program 1996 Annual Report, Onondaga County, NewYork, September 1997.

Thas, O., L. Van Vooren, & J.P. Ottoy, "Nonparametric Test Performance for Trends in Water Quality with Sampling Applications", Journal of the American Water Resources Association, Volume 34, Number 2, pp. 347-357, April 1998.

Van Belle, G. & J.P. Hughes, "Nonparametric Tests for Trend in Water Quality", Water Resources Research, Vol. 20, No. 1, pp. 127-136, January 1984.

Walker, W.W., "Analysis of Water Quality Variations in Reservoirs: Implications for Monitoring and Modeling Efforts", in Stefan, H.G., ed., Surface Water Impoundments, American Society of Civil Engineers, New York, June 1980.

Walker ,W. W., "Empirical Methods for Predicting Eutrophication in Impoundments, Report 3: Model Refinements", prepared for U.S. Army Corps of Engineers, Waterways Experiment Station, Vicksburg, Mississippi, Technical Report E-81-9, March 1985.

Walker, W.W., "Evaluating Watershed Monitoring Programs", prepared for Water Quality Management Office, Bureau of Water and Wastewater, Department of Public Works, City of Baltimore, Maryland, December 1988a.

Walker, W.W., "LRSD.WK1 - Lake/Reservoir Sampling Design Worksheet", Version 1.0, North American Lake Management Society, December 1988b.

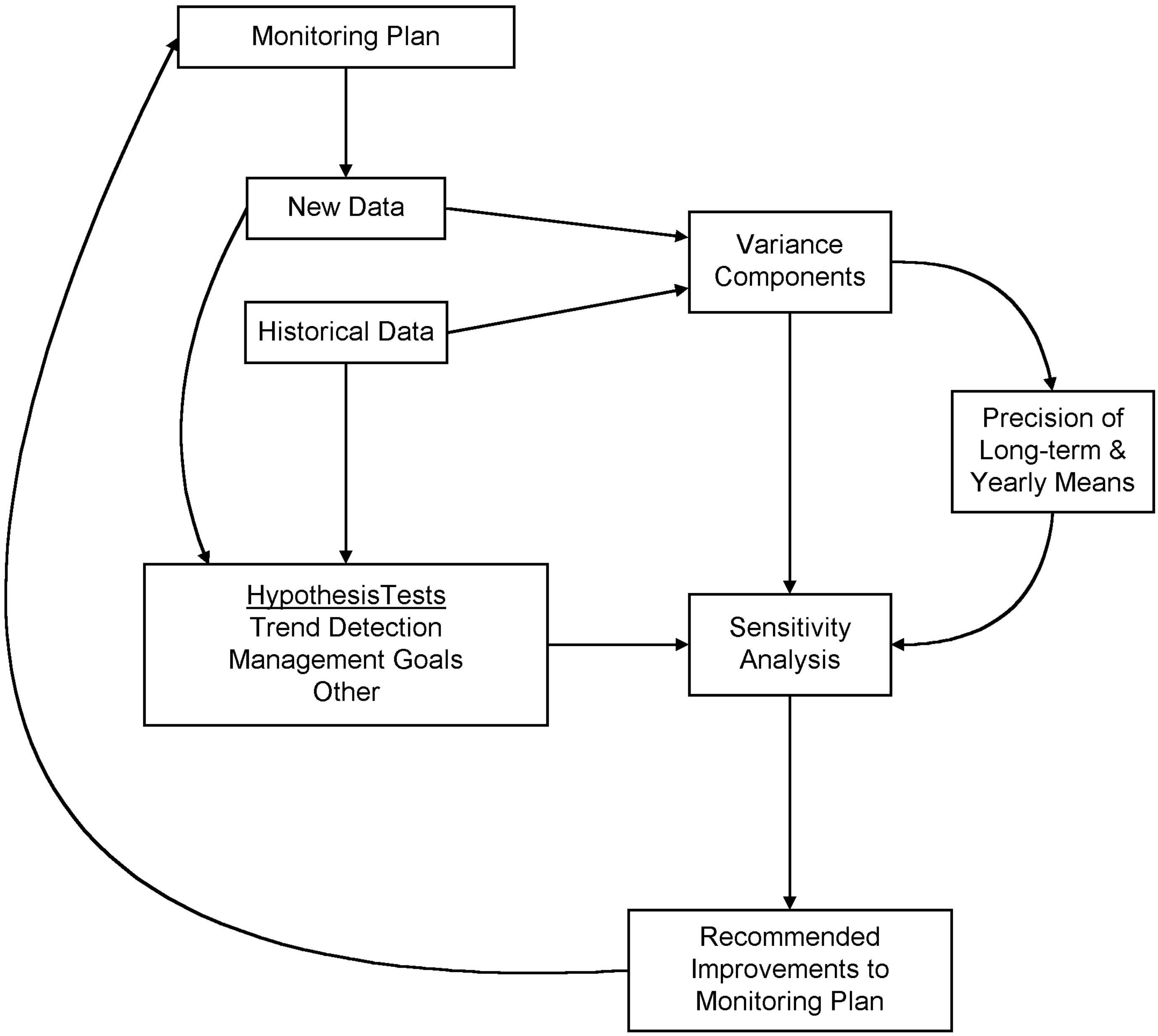
Walker, W.W., "Water Quality Trends at Inflows to Everglades National Park", Water Resources Bulletin, Vol. 27, No. 1, pp. 59-72, February 1991a.

Walker, W.W., "Compilation and Review of Onondaga Lake Water Quality Data", prepared for Onondaga County, Department of Drainage & Sanitation, December 1991b.

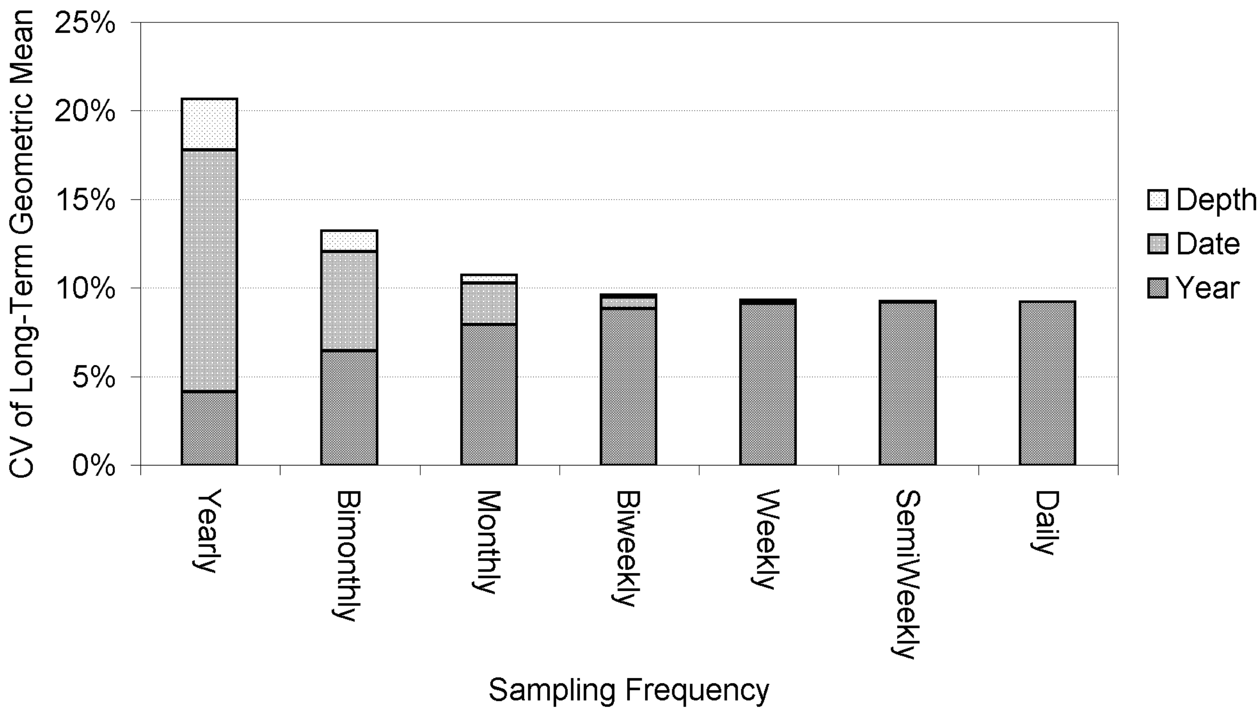
Walker, W.W., "Long-term Water Quality Trends in the Everglades", in Phosphorus Biogeochemisty in Florida Ecosystems, Lewis Publishers, in press, 1998.

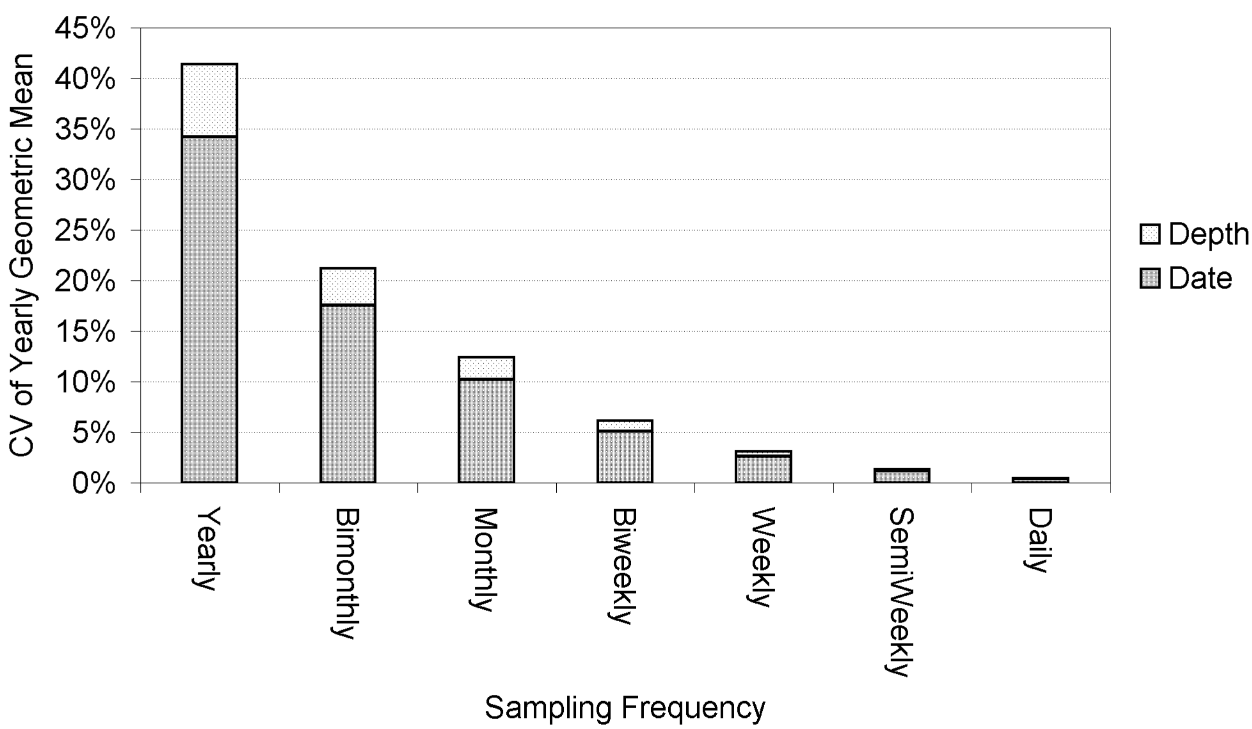
List of Figures

Statistical Framework for Ambient Monitoring Plan



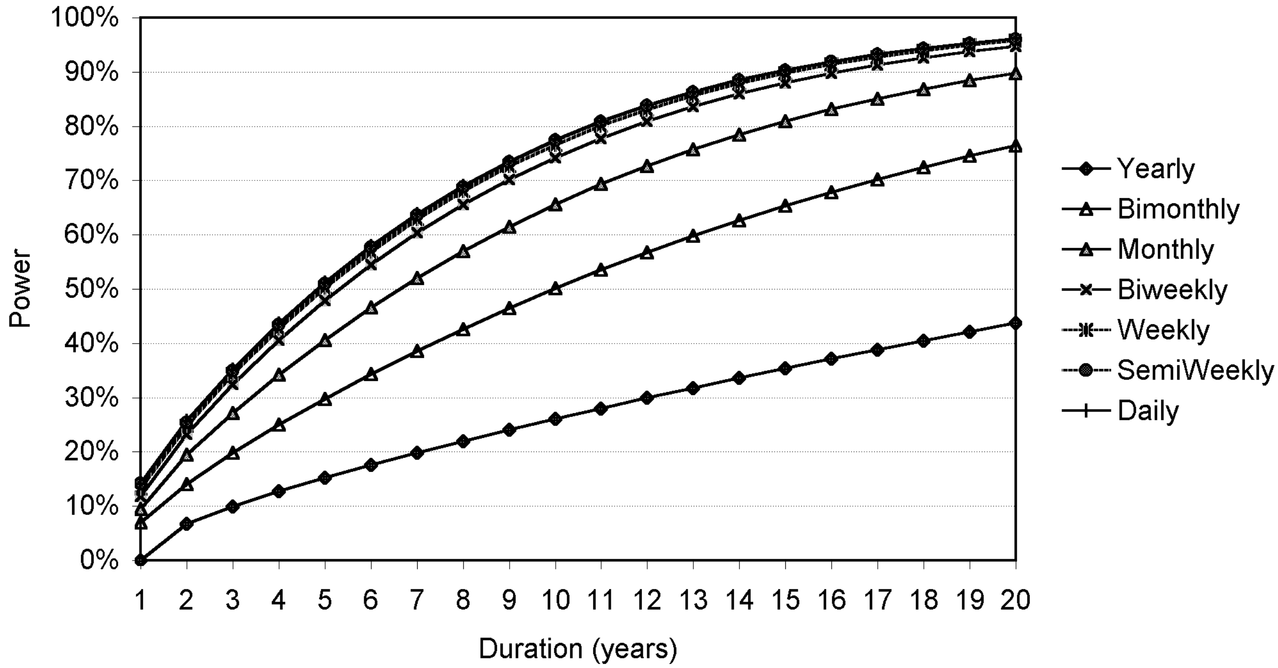
Precision in Long-term & Yearly Geometric Means



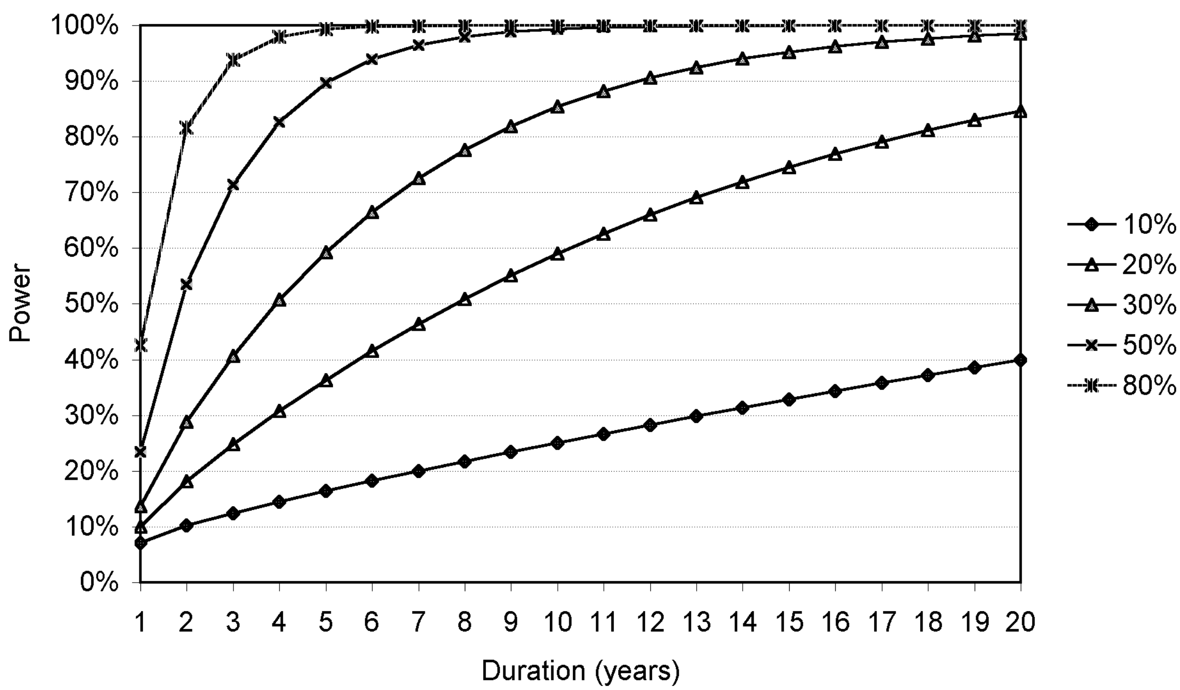


Shaded areas in each bar reflect percent of variance attributed to yearly, daily, or depth variation
Variable: Total Inorganic P
Duration = 5 years

Power Curves for Detecting a Step Change



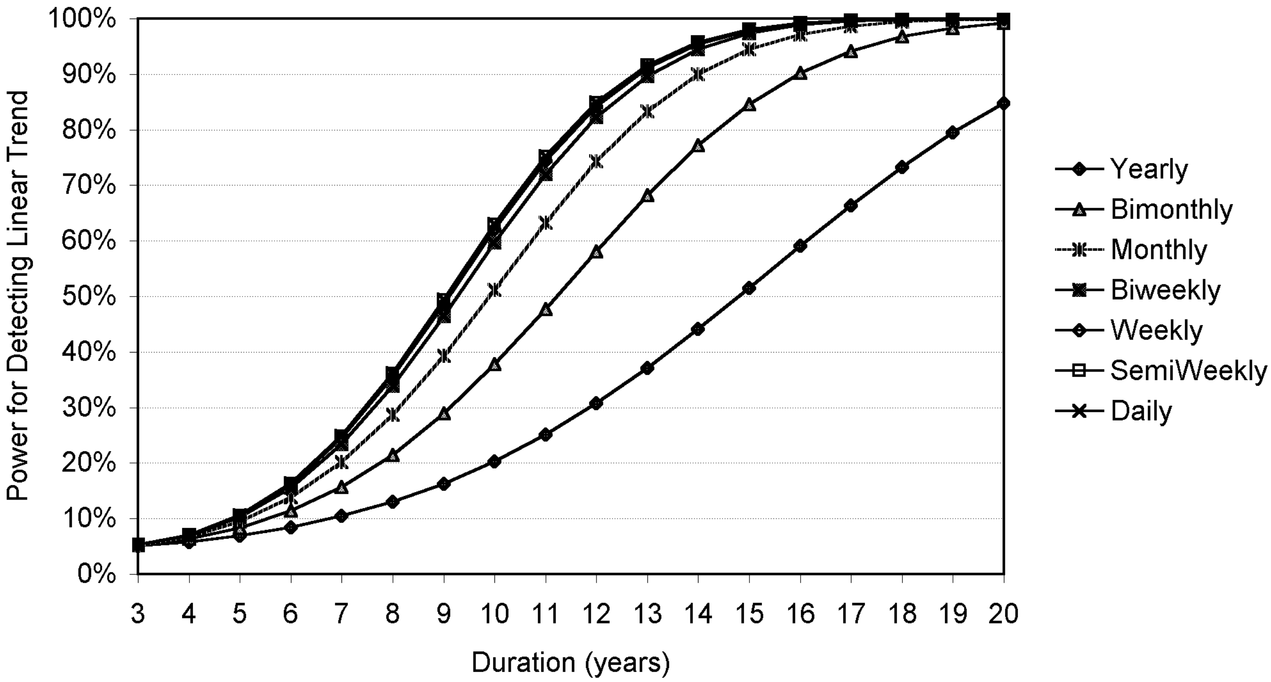
Sensitivity to Sampling Frequency
Step Change Magnitude = 25%
Duration = Number of Years of Monitoring Before & After Step Change



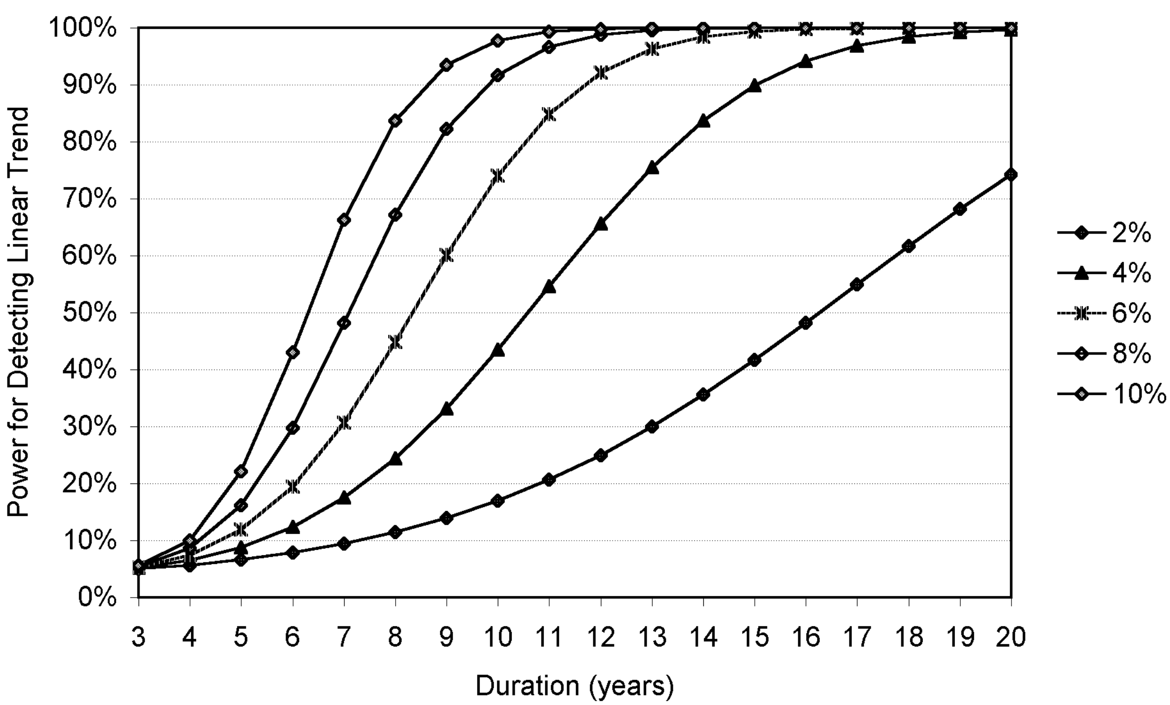
Sensitivity to Step Change Magnitude
Sampling frequency = Biweekly

Figu

Power Curves for Detecting a Linear Trend



Sensitivity to Sampling Frequency
Trend Magnitude = 5% /yr



Sensitivity to Trend Magnitude
Sampling Frequency = Biweekly