Technical Memorandum #3

Minimum Detectable Change and Power Analysis

Introduction

Background

Documenting water quality improvements linked to best management practice (BMP) implementation is a critical aspect of many watershed studies. Challenges exist in targeting critical contaminants, dealing with timing lags, and shifting management strategies (Tomer and Locke 2011). Therefore, it is important to establish monitoring programs that can detect change (Schilling et al. 2013). The "minimum detectable change"—or MDC—is the smallest amount of change in a pollutant concentration or load during a specific time period required for the change to be considered statistically significant (Spooner et al. 2011). Practical uses for the MDC calculation include determining appropriate sampling frequencies and assessing whether a BMP implementation plan will create enough of a change to be measurable with the planned monitoring design. The same basic equations are used for both applications with the specific equations depending primarily on whether a gradual (linear) or step trend is anticipated. In simple terms, one can estimate the required sampling frequency based on the anticipated change in pollutant concentration or load, or turn the analysis around and estimate the change in pollutant concentration or load that is needed for detection with a monitoring design at a specified sampling frequency.

The process of conducting MDC analysis is described in _Tech Notes 7_ (Spooner et al. 2011) and includes the following steps:

1. Define the monitoring goal and choose the appropriate statistical trend test approach.
2. Perform exploratory data analyses.
3. Perform data transformations.
4. Test for autocorrelation.
5. Calculate the estimated standard error.
6. Calculate the MDC.
7. Express MDC as a percent change.

Sample size determination is often performed by selecting a significance level[1], power of the test, minimum change one wants to detect, monitoring duration, and type of statistical test. MDC is calculated similarly, except that the sample size (i.e., number of samples), significance level, and

---

[1] Significance level and power are defined under "Hypothesis Testing"

power are fixed and the minimum detectable change is computed. _Tech Notes 7_ includes the specific equations to use in performing an MDC analysis and provides guidance on evaluating explanatory variables to reduce the standard error (Spooner et al. 2011).

Purpose and Audience

Other authors have reviewed and examined the procedures for computing MDCs and determining sample sizes (Ward et al. 1990; Loftis et al. 2001; USEPA 1997a, 1997b, 2002). Generally, they recommend procedures similar to those presented in _Tech Notes 7_. These authors, however, also recommend that, for most applications of MDC calculations for sample size estimation, statistical powers other than 0.5 (i.e., the default power used in _Tech Notes 7_) should be considered. This technical memorandum extends _Tech Notes 7_ to include evaluation of minimum detectable changes using powers other than 0.5 for step-trend analysis with no explanatory variables. It has been developed for analysts both looking for a basic understanding of integrating power into MDC analyses and framing sample size selection for water resource managers.

Basic Principles

The data analyst usually summarizes a data set with a few descriptive statistics rather than presenting every observation collected. "Descriptive statistics" include characteristics designed to summarize important features of a data set such as range, central tendency, and variability. A "point estimate" is a single number that represents a descriptive statistic. Statistics typically used to summarize water quality data associated with BMP implementation include proportions, means, medians, totals, and variance. When estimating parameters of a population, such as the proportion or mean, it is useful to estimate the "confidence interval," which indicates the probable range in which the true value lies. For example, if the average total nitrogen (TN) concentration is estimated to be 1.2 mg/L and the 90 percent confidence limit is ±0.2 mg/L, there is a 90 percent chance that the true value is between 1.0 and 1.4 mg/L.

Hypothesis Testing

Hypothesis testing should be used to determine whether a change has occurred over time. The "null hypothesis" (H_o) is the root of hypothesis testing. Traditionally, H_o is a statement of no change, no effect, or no difference; for example, "the average TN concentration after the BMP implementation program is equal to the average TN concentration before the BMP implementation program." The "alternative hypothesis" (H_a) is counter to H_o, traditionally being a statement that change, effect, or difference has occurred. If H_o is rejected, H_a is accepted. Regardless of the statistical test selected for analyzing the data, the analyst must select the acceptable error levels for the test. There are two types of errors in hypothesis testing:

- Type I error: H_o is rejected when H_o is really true
- Type II error: H_o is accepted when H_o is really false

Table 1 and Figure 1 depict these error types, with the magnitude of Type I errors represented by α (the significance level or probability of committing a Type I error) and the magnitude of Type II errors represented by β. The probability of making a Type I error is equal to the α of the test and is selected

by the data analyst. In most cases, the manager or analyst will define *1-α* to be in the range of 0.90–0.99 (i.e., a confidence level of 90–99 percent), although there have been applications in which *1-α* has been set to as low as 0.80. Selecting a 90-percent confidence level implies that the analyst will reject the H_o when H_o is true (i.e., a false positive) 10 percent of the time. The same notion applies to the confidence interval for point estimates described above: *α* is set to 0.10, and there is a 10 percent chance that the true average TN concentration is outside the 1.0–1.4 mg/L range.

Table 1. Errors in Hypothesis Testing

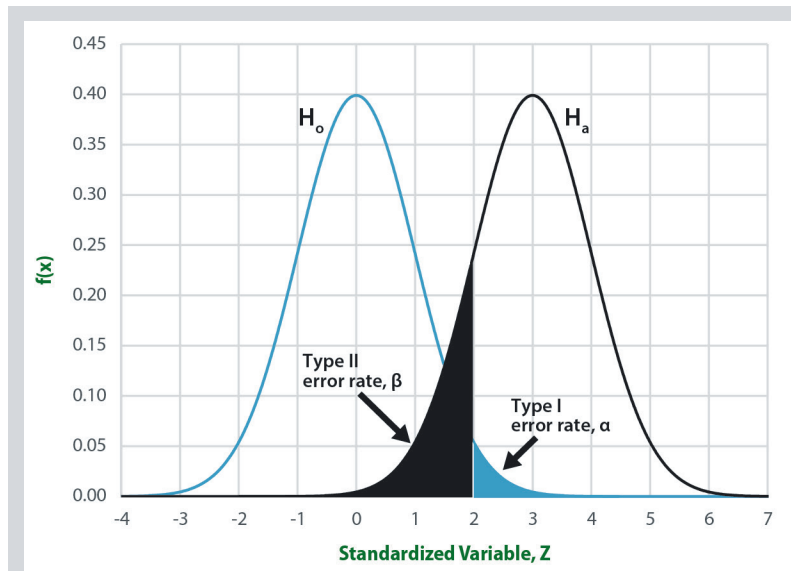| Decision	State of Affairs in the Population	
	H_o is True	H_o is False
Accept H_o	**1-α** (Confidence level)	**β** (Type II error)
Reject H_o	**α** (Significance level) (Type I error)	**1-β** (Power)

Figure 1. Type I and Type II Errors.

Type II errors (β) depend on the significance level, sample size, and data variability. In general, for a fixed sample size, *α* and *β* vary inversely. And similarly, for a fixed *α*, *β* can be reduced by increasing the sample size (Remington and Schork 1970). Power (1-β) is defined as the probability of correctly rejecting H_o when H_o is false and is discussed further in the next section.

Power Curves

The above principles are demonstrated in the hypothetical power curves shown in Figure 2. The green lines in Figure 2 represent hypothesis tests with α = 0.05 and the orange lines represent hypothesis tests with α = 0.10. When there is no change in water quality between pre- and post-BMP implementation (i.e., MDC = 0% on the x-axis), the plotted value in Figure 2 is equal to the selected significance level (i.e., α) of the hypothesis test. As the MDC increases (i.e., the difference between
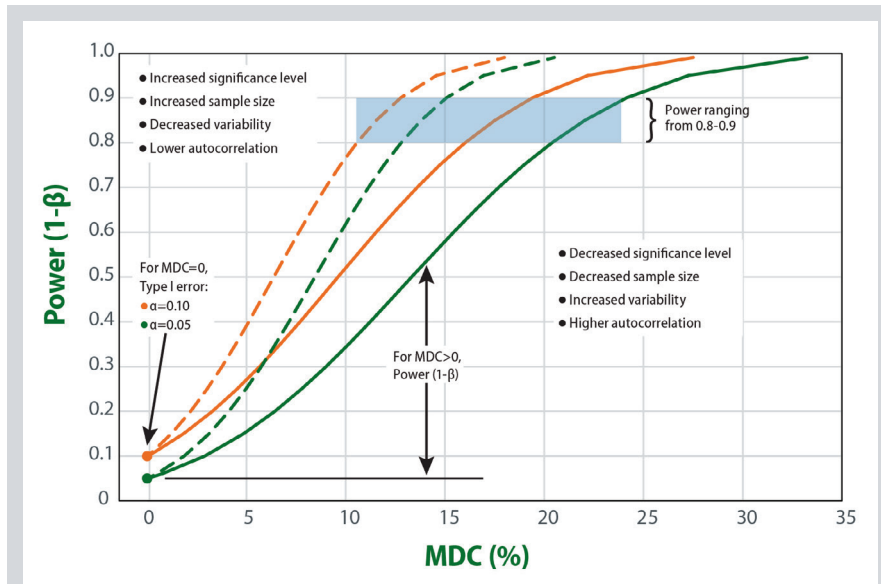
Figure 2. Hypothetical Power Curves.

pre- and post-BMP implementation water quality becomes larger), the power increases. Ideally, the power curve starts at α where the MDC = 0 and rapidly rises to a power of 1.0 as the MDC increases. The rate at which the power increases is controlled by the significance level, sample size, and variability. Power also is affected by the amount of autocorrelation in the collected data (Spooner et al. 2011). As illustrated by the dashed versus solid lines in Figure 2, increasing the sample size, decreasing the variability of the data set, or lower autocorrelation generally will improve the power.

Analysts and managers often rely on preliminary data or data from previous studies to help establish the likely variability of a data set for estimating sample size. As stated in Spooner et al. (2011), incorporating explanatory variables into the calculation increases the probability of detecting significant changes by reducing data set variability and produces statistical trend analysis results that better represent true changes due to BMP implementation rather than to hydrologic and meteorological variability. Commonly used explanatory variables for hydrologic and meteorological variability include streamflow and total precipitation (Spooner et al. 2011). Based on a numerical study, Loftis et al. (2001) suggest that including poorly correlated explanatory variables (i.e., $\rho<0.3$) is not helpful while correlations greater than 0.6 can yield significant improvements in MDC calculations.

Practical Considerations

Establishing an appropriate level of power when designing a monitoring program is as important as establishing the significance level and often is tied to a management or risk trade-off. In many cases, a water resource manager is interested in detecting changes in the monitored resource as a result of a BMP implementation program. Selecting a power from 0.8 to 0.9 might be an appropriate choice that will detect a resulting change with an 80 to 90 percent probability if a change really happened.

The choice of power in combination with other factors including significance level, data variability, sample size, and autocorrelation can be weighed against the actual changes that could result from BMP implementation. For example, the four power curves presented in Figure 2 represent a

range of conditions considered in the power analysis; the gray shading highlights the portion of the curves that intersects with a power from 0.8 to 0.9. The solid green line (with α = 0.05) indicates that a change of 24 percent or greater can be detected with a 90 percent probability (i.e., power = 0.9). This example is not particularly helpful, however, if BMP implementation is expected to cause only a 15–20-percent change. Detecting this smaller change would require taking more samples or accepting different probabilities of Type I and Type II errors. Selecting a significance level of α = 0.10 (solid orange line) is more appropriate as an MDC of 19 percent now can be achieved with a power of 0.9. By also reducing the power to 0.8, a change of 16 percent can be detected (solid orange line). Both dashed lines, on the other hand, allow for MDCs smaller than the change expected from BMP implementation. The results can then be tied to monitoring costs and communicated to managers. Because these types of analyses are developed with incomplete and imperfect information, some level of safety needs to be built into the range of conditions being evaluated and the resulting design recommendations to ensure that the monitoring program meets its intended data quality objectives.

Application to MDC

Tech Notes 7 presents equations for computing the MDC assuming a power of 0.5. The equations are generally applicable in this context except as noted below to allow for alternative levels of power (see *Tech Notes 7* for more details about the applicable equations). A Student's t-test or Analysis of Covariance (ANCOVA) is used for step-trend analysis. For log-normal data, this analysis is conducted on log-transformed data (Spooner et al. 2011).

The MDC calculation requires an estimate of the standard deviation of the difference between the mean values of pre-BMP data vs. post-BMP data ($s_{\bar{x}_{pre}-\bar{x}_{post}}$). Use the following formula to obtain this estimate:

$$s_{\bar{x}_{pre}-\bar{x}_{post}} = \sqrt{\frac{MSE}{n_{pre}} + \frac{MSE}{n_{post}}}$$

where:

$s_{\bar{x}_{pre}-\bar{x}_{post}}$ = the estimated standard error of the difference between the mean values of the pre- and post-BMP periods.

MSE = s_p^2 = the estimate of the pooled Mean Square Error (MSE) or, equivalently, the pooled variance of the two time periods.

n_{pre} and n_{post} = the number of samples in the pre- and post-BMP periods.

The variance of pre-BMP data (s_{pre}^2) can be used to estimate MSE—or s_p^2—for both pre- and post-BMP periods if post-BMP data are not available. If the variability of the post-BMP data set is expected to be different from the pre-BMP data set, s_p^2 can be estimated as:

$$s_p^2 = \frac{(n_{pre}-1)s_{pre}^2 + (n_{post}-1)s_{post}^2}{(n_{pre}-1) + (n_{post}-1)}$$

If autocorrelation is present, then generalized least squares (GLSs) with Yule Walker methods should be used to estimate the standard deviation. If ordinary least squares (OLSs) are used, then an estimate of the true standard deviation can be estimated using the following large sample approximation formula:

$$s_b = s_b' \sqrt{\frac{1+\rho}{1-\rho}}$$

where:

s_b = the true standard deviation of the trend (difference between two means) estimate (e.g., calculated using GLS).

s_b' = the incorrect variance of the trend estimate calculated without regard to autocorrelation using OLS (e.g., using a statistical linear regression procedure that does not take into account autocorrelation).

ρ = the autocorrelation coefficient for autoregressive lag 1, AR(1).

In practice, an estimate of the MDC for a step trend is obtained by using the following formula:

$$\text{MDC} = (t_{\alpha,n_{pre}+n_{post}-2} + t_{2\beta,n_{pre}+n_{post}-2}) \sqrt{\frac{\text{MSE}}{n_{pre}} + \frac{\text{MSE}}{n_{post}}}$$

where:

$t_{\alpha,n_{pre}+n_{post}-2}$ = Student's t-value with $(n_{pre} + n_{post} -2)$ degrees of freedom corresponding to α.

$t_{2\beta,n_{pre}+n_{post}-2}$ = Student's t-value with $(n_{pre} + n_{post} -2)$ degrees of freedom corresponding to 2β.

This MDC equation includes an additional Student's t-value term ($t_{2\beta,n_{pre}+n_{post}-2}$) relative to the equation presented in _Tech Notes 7_ to allow for alternative values of power. If a power of 0.5 is selected, the value of $t_{2\beta,n_{pre}+n_{post}-2}$ is equal to zero and reduces to the equation shown in _Tech Notes 7_ (Spooner et al. 2011).

To convert the above MDC to percent MDC, the following two formulas are used depending on whether raw or log-transformed values were used in the analysis:

- Raw data: MDC% = $100*(\text{MDC}/\bar{x}_{pre})$
- Log-transformed data: MDC% = $(1 - 10^{-MDC}) * 100$

Example Power Curve Calculations

The calculations are illustrated below with the following assumptions (following the example initiated in _Tech Notes 7_):

> One-sided, two-sample t-test ($H_o: \bar{x}_{pre} = \bar{x}_{post}$, $H_a: \bar{x}_{pre} > \bar{x}_{post}$)
>
> Significance level (α) = 0.05
>
> Power (1-β) = 0.5 and 0.8
>
> ρ = 0 (no autocorrelation)
>
> n_{pre} = 52 samples in the pre-BMP period
>
> n_{post} = 52 samples in the post-BMP period
>
> $\bar{x}_{pre}$ = 36.9 mg/L, mean of the 52 samples in the pre-BMP period
>
> s_p = 21.2 mg/L = standard deviation of the 52 pre-BMP samples
>
> MSE = s_p^2 = 449.44

Table 2 presents sample Excel formulas that can be used to calculate $t_{\alpha,n_{pre}+n_{post}-2}$ and $t_{2\beta,n_{pre}+n_{post}-2}$ as 1.6599 and 0.8452, respectively. These equations can be adopted for general use.

$$\text{MDC (for power = 0.5)} = (1.6599 + 0)\sqrt{\frac{449}{52} + \frac{449}{52}} = 6.9 \text{ mg/L}$$

Percent change required = MDC% = 100*(6.9/36.9) = 19%.

$$\text{MDC (for power = 0.8)} = (1.6599 + 0.8452)\sqrt{\frac{449}{52} + \frac{449}{52}} = 10.4 \text{ mg/L}$$

Percent change required = MDC% = 100*(10.4/36.9) = 28%.

There is a 50 percent probability that a decrease of 6.9 mg/L (or 19 percent) or an 80 percent probability that a decrease of 10.4 mg/L (or 28 percent) can be detected (see Figure 3, points A and B).

Table 2. Sample Excel© Formulas for Computing $t_{2\beta} + t_\alpha$ Assuming a One-Sided t-test, 52 Samples in Both Pre- and Post-BMP Data Sets, a Significance Level of 0.05 and Power of 0.80

Row	Column A	Column B	Column C
1	1- or 2-sided test (set as 1-sided)	1	1
2	Pre-BMP n	52	52
3	Post-BMP n	52	52
4	Significance Level (α)	0.05	0.05
5	Power (1-β)	0.80	0.80
6	2β	0.4	= 2*(1-B5)
7	$t_{2\beta}$	0.8452	= T.INV(1-B6/2,(B2+B3-2))
8	t_α	1.6599	= T.INV(1-B4/B1,(B2+B3-2))
9	$t_{2\beta} + t_\alpha$	2.5051	= B7+B8

Power Curve Interpretation and Use in Monitoring Design

Figure 3 displays a series of power curves using a variety of sample sizes ($n_{pre} = n_{post} = 8, 12, 24, 52,$ 104) and the above assumptions. Each plotted line in Figure 3 was created by selecting an alternative value of power (1-β), computing $t_{2\beta,n_{pre}+n_{post}-2}$ and $t_{\alpha,n_{pre}+n_{post}-2}$, and replicating the above sample calculations for MDC(%). When there is no change (i.e., MDC = 0%), the plotted value corresponds to the chosen significance level. Power increases with increased MDC. If we presume that developing a monitoring program with a power of 0.8–0.9 is desirable and the expected benefit of a BMP program might result in a 30–50-percent reduction in the measured water quality parameter, then it is logical that sampling programs with just 8 and 12 pre- and post-BMP samples will not likely detect the expected changes and have little value for documenting the effectiveness of BMPs. Increasing the sample size to 52 would allow detection of a 30–50-percent reduction with a power of 0.8. Sampling programs using a pre- and post-BMP sample size of 24 fall between clearly having little value (n= 8 or 12) and meeting the objective (n= 52 and 104).
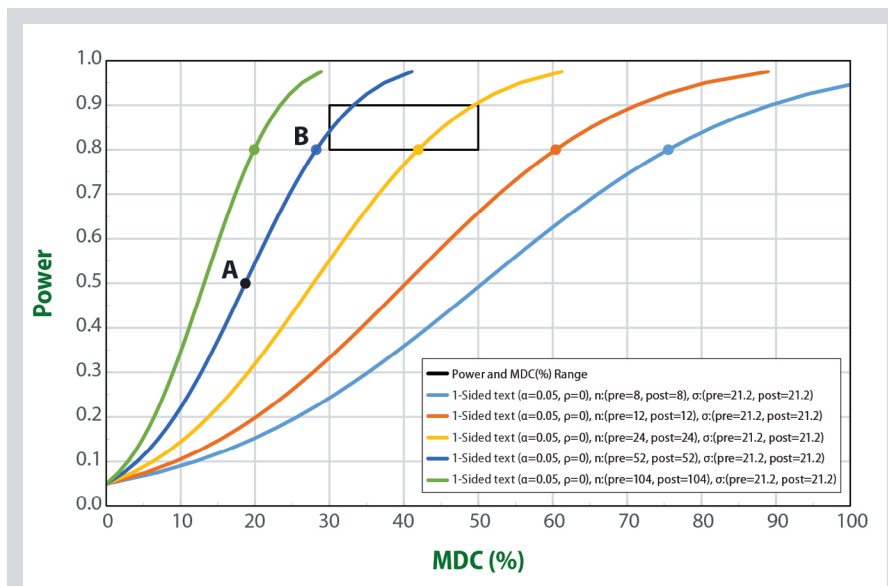

Figure 3. Power Curves Associated with Alternative Sample Sizes. (Point A: 50% probability that a decrease of 19% can be detected with a monitoring program that includes 52 samples in the pre- and post-BMP periods. Point B: 80% probability that a decrease of 28% can be detected with the same monitoring program. Black box: Targeted power [0.8–0.9] and expected reductions from BMP implementation [30–50%]. Other unlabeled markers: MDC(%) that can be detected with a 0.8 power for alternative sampling designs.)

Another approach for evaluating alternative monitoring designs is to fix the significance level and power and select alternative monitoring strategies. This approach might be necessary, for example, if pre-BMP data already have been collected and the only option left is to collect an increased level of post-BMP data. Figure 4 displays MDC(%) with a power equal to 0.8 as a function of post-BMP sample size for fixed pre-BMP sample sizes. The unlabeled markers in Figure 4 indicate locations where pre- and post-BMP sample sizes are equal (and correspond to the similar markers in Figure 3). Continuing with the scenario in which the pre-BMP data set has 24 samples (yellow line), it is clear

Figure 4. Comparison of MDC(%) with Power = 0.8 as a Function of Increasing Post-BMP
Sample Size for Fixed Pre-BMP Sample Sizes. (Unlabeled markers indicate
locations where pre- and post-BMP sample sizes are equal—see Figure 2.
Black box: Expected reductions from BMP implementation [30–50%].)
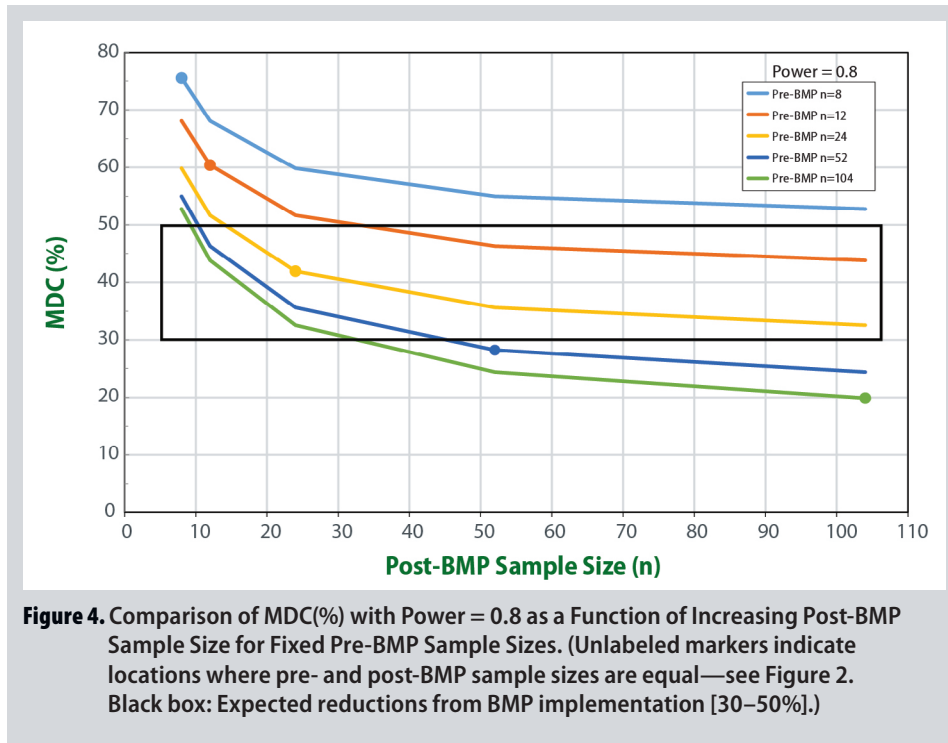
that if the post-BMP sampling effort is increased from 24 to 52 samples, the MDC decreases from 42 percent to 36 percent. Increasing the post-BMP data set to 104 reduces the MDC to 33 percent, for a total of 128 (24+104) samples—not as sensitive as if a total of 104 collected samples were split evenly between the pre- and post-BMP monitoring period (i.e., the MDC for 52 samples in the pre- and post-BMP periods is 28 percent). The latter observation is confirmed by Loftis et al. (2001) in demonstrating that for a fixed cost, the best allocation of resources is to split the sampling effort evenly between pre- and post-BMP monitoring periods.

Summary

The calculation of MDC has several practical uses, including determining appropriate sampling frequencies and assessing whether a BMP implementation plan will be sufficient for creating change that is measurable with the planned monitoring design. Evaluation of MDC should consider power other than 0.5 for step trends with no explanatory variables.

References

Loftis, J.C., L.H. MacDonald, S. Streett, H.K. Iyer, and K. Bunte. 2001. Detecting cumulative watershed effects: The statistical power of pairing. *Journal of Hydrology* 251(1-2):49–64.

Remington, R.D., and M.A. Schork. 1970. *Statistics with Applications to the Biological and Health Sciences*. Prentice-Hall, Englewood Cliffs, New Jersey.

Schilling, K.E., C.S. Jones, and A. Seeman. 2013. How paired is paired? Comparing nitrate concentrations in three Iowa drainage districts. *Journal of Environmental Quality* 42(5):1412–1421.

Spooner, J., S.A. Dressing, and D.W. Meals. 2011. *Minimum Detectable Change Analysis*. Tech Notes 7, December 2011. Prepared for U.S. Environmental Protection Agency by Tetra Tech, Inc., Fairfax, VA. Accessed October 1, 2015.
https://www.epa.gov/polluted-runoff-nonpoint-source-pollution/
nonpoint-source-monitoring-technical-notes

Tomer, M.D., and M.A. Locke. 2011. The challenge of documenting water quality benefits of conservation practices: A review of USDA-ARS's conservation effects assessment project watershed studies. *Water Science and Technology* 64(1):300–310.

USEPA (U.S. Environmental Protection Agency). 1997a. *Techniques for Tracking, Evaluating, and Reporting the Implementation of Nonpoint Source Control Measures: Agriculture*. EPA-841-B-97-010. U.S. Environmental Protection Agency, Office of Water, Washington, DC.

USEPA (U.S. Environmental Protection Agency). 1997b. *Techniques for Tracking, Evaluating, and Reporting the Implementation of Nonpoint Source Control Measures: Forestry*. EPA-841-B-97-009. U.S. Environmental Protection Agency, Office of Water, Washington, DC.

USEPA (U.S. Environmental Protection Agency). 2002. *Urban Stormwater BMP Performance Monitoring: A Guidance Manual for Meeting the National Stormwater BMP Database Requirements*. EPA-821-B-02-001. U.S. Environmental Protection Agency, Office of Water, Washington, DC.

Ward, R.C., J.C. Loftis, and G.B. McBride. 1990. *Design of Water Quality Monitoring Systems*. Van Nostrand Reinhold, New York, NY.