Data-driven Planning for TMDL Effectiveness Monitoring & Statistical Test Selection for Analysis of Monitoring Data

Sponsored by EPA Region 10

Presenters:
Laura Blake, Corey Godfrey, and Andy Somor
The Cadmus Group, Inc.
Overview of Presentation

- Recommended steps for developing a TMDL effectiveness monitoring plan.
- Data-driven planning for TMDL effectiveness monitoring.
- Data exploration methods for selecting statistical tests for analyzing TMDL effectiveness monitoring data.
- Demonstration of Excel-based TMDL effectiveness monitoring planning tool.
Data-driven Planning for TMDL Effectiveness Monitoring & Statistical Test Selection for Analysis of Monitoring Data

TMDL EFFECTIVENESS MONITORING PLANNING
“Effectiveness Monitoring”

- The process of measuring improvements in the water quality of a water body.
- Not to be confused with BMP effectiveness, which measures the success or effectiveness of the BMPs themselves.
- The primary goal of TMDL effectiveness monitoring is to identify water quality improvements (or lack thereof) that result from TMDL implementation.
Need for Effectiveness Monitoring

- Provides a quantitative measure of progress towards attainment with WQS.
- Allows for demonstration of incremental improvements in water quality.
- Supports adaptive management approach to implementation and restoration.
- Provides data for use in making SP-12 and WQ-10 determinations.
- Provides documentation of water quality improvements, which can be used to communicate results and justify the need for funding to support water programs.
Recommended Steps for TMDL Effectiveness Monitoring Planning

1. Review existing data and information.
2. Select monitoring sites, parameters, and study design.
3. Estimate sample size.
Review Existing Data and Information

- Begin with a thorough review of all available information that may direct the process.

- Existing data and information will provide an understanding of:
  - Historical and current water quality conditions.
  - TMDL implementation activities.

- Involve stakeholders early.
Select Monitoring Sites, Parameters, and Study Design

- Design effectiveness monitoring projects at the watershed scale.
  - Specific project scale should be decided upon using information on the number and extent of impaired or threatened waters, project resources, and project partners.

- Watershed scale effectiveness monitoring:
  - Pour point method
  - Distributed sampling method
Site Selection Approach: Pour Point Method

Monitoring Site
Site Selection Approach: Distributed Sampling Method
Site Selection

- Locate monitoring sites where TMDL implementation is expected to have discernible water quality effects.
- Examples include sites on impaired/degraded water bodies that are downstream of:
  - WWTFs with new or revised WLAs.
  - Discontinued illicit discharges.
  - NPS that are managed through BMPs.
  - Stream channel restoration projects.
  - Improved onsite wastewater management or expansion of sanitary sewer service.
  - Other TMDL-specific pollution control measures.
- Evaluate any potential existing monitoring network or sites.
Parameters

- At a minimum, monitor the pollutants for which the TMDL was developed.
- If resources allow, monitor for stressor and/or response variables, which provide additional information about the condition of a water body.
- If resources allow, monitor parameters that may be covariates to the primary pollutants of interest.
  - Stream flow is a common covariate for pollutants in streams and rivers.
Study Design

- Outlines how water quality improvements will be demonstrated.
- Critical to ensuring the collection of the specific data needed to answer the study questions/goals.
- Selection is dependent on many factors, including:
  - Types of TMDL implementation actions.
  - Implementation schedule.
  - Availability and quality of previously collected data.
  - Resources.
  - Existence of suitable reference sites.
Potential study designs for TMDL effectiveness monitoring include:

- Before & After
- Upstream/downstream
- Paired watersheds
- Trend monitoring
Study Design: Before/After Study
Study Design:
Upstream/Downstream Study
Study Design: Paired Watersheds Study
Study Design: Trend Monitoring
Water quality data are often collected without considering the number of samples needed to demonstrate statistically significant changes.

Objective and informed sample size decisions can be made using a statistical method known as power analysis.

Power analysis uses information from pilot data to determine the optimal number of samples needed to identify statistically significant changes or trends.
Develop TMDL Effectiveness Monitoring Plan

Steps for planning a TMDL effectiveness monitoring project:

1. Review existing data and information.
2. Select monitoring sites, parameters, and study design.
3. Estimate sample size.
The planning document should include:
  - Relevant background information.
  - Project goals and objectives.
  - Where and when monitoring will occur.
  - List of parameters to be monitored.
  - Preliminary discussion of intended data analysis methods, including selected level of significance.

The TMDL effectiveness monitoring plan can be incorporated into a QAPP.
Data-driven Planning for TMDL Effectiveness Monitoring & Statistical Test Selection for Analysis of Monitoring Data

DATA-DRIVEN PLANNING
Common Example of post-TMDL Monitoring

- Ambient monitoring reveals that the Fox Run Watershed is impaired for total phosphorus.
- TMDL developed for total phosphorus.
- Post-TMDL monitoring completed through the routine monthly ambient monitoring of:
  - TP
  - TN
  - TSS
  - E. Coli
  - DO, conductivity, pH, Temp, Turbidity
Explore Pilot Data

- Pilot data are previously collected data
  - Ambient monitoring program
  - TMDL data
  - Special studies
- Helps to informs sample size estimation (power analysis)
- Informs selection of parameters to monitor for (correlation analysis)
Power Analysis

\[ \beta \approx \text{MDC} \times \alpha \times \sqrt{n/\sigma} \]

Where:
\( \beta \)=power
\( \text{MDC} \)=minimum detectable change
\( \alpha \)=significance
\( n \)=sample size
\( \sigma \)=standard deviation (variability)
Estimate Sample Size

Minimum Detectable Change E. coli (MPN/100 mL) per Year vs. Sample Size (N)
Correlation Analyses

![Graph showing correlation between Total Phosphorus (mg/L) and Streamflow (cfs)]
## Select Parameters

<table>
<thead>
<tr>
<th></th>
<th>Chloride</th>
<th>Conductivity</th>
<th>TN</th>
<th>TP</th>
<th>TSS</th>
<th>Turbidity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chloride</td>
<td>1</td>
<td>0.88</td>
<td>0.62</td>
<td>0.31</td>
<td>0.26</td>
<td>0.31</td>
</tr>
<tr>
<td>Conductivity</td>
<td>0.88</td>
<td>1</td>
<td>0.74</td>
<td>0.18</td>
<td>0.3</td>
<td>0.32</td>
</tr>
<tr>
<td>TN</td>
<td>0.62</td>
<td>0.74</td>
<td>1</td>
<td>0.28</td>
<td>0.34</td>
<td>0.36</td>
</tr>
<tr>
<td>TP</td>
<td>0.31</td>
<td>0.18</td>
<td>0.28</td>
<td>1</td>
<td>0.52</td>
<td>0.56</td>
</tr>
<tr>
<td>TSS</td>
<td>0.26</td>
<td>0.3</td>
<td>0.34</td>
<td>0.52</td>
<td>1</td>
<td>0.88</td>
</tr>
<tr>
<td>Turbidity</td>
<td>0.31</td>
<td>0.32</td>
<td>0.36</td>
<td>0.56</td>
<td>0.88</td>
<td>1</td>
</tr>
</tbody>
</table>
Data-driven Planning for TMDL Effectiveness Monitoring & Statistical Test Selection for Analysis of Monitoring Data

SELECTING STATISTICAL TESTS FOR ANALYZING EFFECTIVENESS MONITORING DATA
Common Types of Statistical Tests

- **Parametric:** Assumes the data are normally distributed. Examples include:
  - t-Test
  - Multiple linear regression

- **Nonparametric:** Makes no assumption about data distribution. Examples include:
  - Signed rank test
  - Mann-Kendall test for trend
Data Distributions
Histograms (cont.)

Frequency

pH

1.0  1.2  1.4  1.6  1.8  2.0  2.2  2.4  2.6  2.8
33  

Frequency
Data Transformations

Total Phosphorous (mg/L)

Frequency
Boxplots

log (Fecal Coliform (#/100 mL))

Spring/Summer

Fall/Winter

Minimum
75%
Median
25%
Maximum

37
Autocorrelation & Lag Plots

\[ R^2 = 0.4172 \]

Log (Total Phosphorus (mg/L))_{t} vs. Log (Total Phosphorus (mg/L))_{t-1}
Censored Data

<table>
<thead>
<tr>
<th>Date</th>
<th>Fecal Coliform (#/100 mL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1/13/2012</td>
<td>345</td>
</tr>
<tr>
<td>2/1/2012</td>
<td>280</td>
</tr>
<tr>
<td>2/29/2012</td>
<td>&gt;2,400</td>
</tr>
<tr>
<td>3/17/2012</td>
<td>279</td>
</tr>
<tr>
<td>4/2/2012</td>
<td>240</td>
</tr>
</tbody>
</table>

Substituting >2,400 with the value of 2,400 gives a calculated mean of 709

If that value were really 10,000, the mean would be 2,229
Selecting an Appropriate Test

- Objectives
- Study Design
- Characteristics of our data:
  - Normal distribution
  - Outliers
  - Censored data points
  - Seasonality
  - Autocorrelation
  - Missing data
Selecting an Appropriate Test (cont.)

- Study Objective: Compare two independent data groups
- Relevant Study Designs:
  - Before/after;
  - Upstream/downstream
- Parametric test: Two sample t-test
- Nonparametric test: Rank sum test
Selecting an Appropriate Test (cont.)

- Study Objective: Compare data groups with matched pairs
- Relevant Study Designs
  - Paired watersheds;
  - Upstream/downstream
- Parametric test: Paired t-test
- Nonparametric test: Signed rank test
Selecting an Appropriate Test (cont.)

- Study Objective: Compare two data groups while adjusting for covariates
- Relevant Study Designs: Paired watersheds
- Parametric test: Analysis of covariance (ANCOVA)
Selecting an Appropriate Test (cont.)

- **Study Objective:** Evaluate the relationship between one data group and time (without seasonality)
- **Relevant Study Designs:** Trend monitoring
- **Parametric test:** Linear regression
- **Nonparametric test:** Mann-Kendall test
Selecting an Appropriate Test (cont.)

- Study Objective: Evaluate the relationship between one data group and time (with seasonality)
- Relevant Study Designs: Trend Monitoring
- Parametric test: Linear regression with seasonal term
- Nonparametric test: Seasonal Kendall test
Selecting an Appropriate Test (cont.)

- Study Objective: Evaluate the relationship between one data group, time, and other variables (without seasonality)
- Relevant Study Designs: Trend Monitoring
- Parametric test: Multiple linear regression
- Nonparametric test: Mann-Kendall test with LOWESS (“locally weighted scatterplot smoothing”)
Selecting an Appropriate Test (cont.)

- **Study Objective:** Evaluate the relationship between one data group, time, and other variables (with seasonality)

- **Relevant Study Designs:** Trend monitoring

- **Parametric test:** Multiple linear regression with seasonal term

- **Nonparametric test:** Seasonal Kendall test with LOWESS
Data-driven Planning for TMDL Effectiveness Monitoring & Statistical Test Selection for Analysis of Monitoring Data

DEMO OF TMDL EFFECTIVENESS MONITORING PLANNING TOOL
Data Exploration Software

- Spreadsheet software (Excel)
  - Simple user-interface
  - Limited analysis options
- Statistical packages
  - Open source (R) or proprietary (SAS)
  - Steep learning curve
- Web-based tools
A Customized Tool

- TMDL Effectiveness Monitoring Planning Tool (EMTool.xlsm)
  - Excel-based
  - Enhances basic Excel features with VBA code
- Objective: Facilitate *data-driven* planning.
  - Create exploratory plots
  - Estimate sample size using power analysis
  - Estimate monitoring costs
Data Exploration Checklist

Inputs:
- Pollutants of concern
- Pollution control types & implementation schedule
- Monitoring locations
- Study design
- Water quality targets

Outputs:
- Covariates
- Sample frequency, timing, and size
  - Project costs
- Statistical methods
Example Dataset: Bacteria in Bear Creek, OR

- TMDLs for Bear Creek & tributaries (1992 & 2007):
  - Ammonia
  - BOD
  - Total phosphorus
  - Sediment
  - Bacteria
  - Temperature

- 2008-2009 summer *E. coli* concentrations at the watershed outlet.
Excel Tool - Pilot Data Worksheet

Pilot Data Worksheet - Enter pilot data for the water quality parameters included in TMDL effectiveness monitoring.

**Step 1.**
Define Water Quality Parameters - Select a water quality parameter from the drop-down menu OR type the parameter name in the drop-down box. Enter up to ten parameters and measurement units for pilot data.

<table>
<thead>
<tr>
<th>Parameter 1:</th>
<th>E. Coli</th>
<th>Measurement Units:</th>
<th>MPN/100 mL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parameter 2:</td>
<td>Streamflow</td>
<td>Measurement Units:</td>
<td>cfs</td>
</tr>
<tr>
<td>Parameter 3:</td>
<td>Ammonia</td>
<td>Measurement Units:</td>
<td></td>
</tr>
<tr>
<td>Parameter 4:</td>
<td>Chlorophyll a</td>
<td>Measurement Units:</td>
<td></td>
</tr>
<tr>
<td>Parameter 5:</td>
<td>Fecal Coliform</td>
<td>Measurement Units:</td>
<td></td>
</tr>
<tr>
<td>Parameter 6:</td>
<td>Macrinovertebrate</td>
<td>Measurement Units:</td>
<td></td>
</tr>
<tr>
<td>Parameter 7:</td>
<td>Nitrate/Nitrite</td>
<td>Measurement Units:</td>
<td></td>
</tr>
<tr>
<td>Parameter 8:</td>
<td>Periphtyon</td>
<td>Measurement Units:</td>
<td></td>
</tr>
<tr>
<td>Parameter 9:</td>
<td>Total Nitrogen</td>
<td>Measurement Units:</td>
<td></td>
</tr>
<tr>
<td>Parameter 10:</td>
<td>Total Organic Carbon</td>
<td>Measurement Units:</td>
<td></td>
</tr>
</tbody>
</table>

**Enter Parameter Names & Units**
Enter Pilot Data - Enter sample dates and values in the appropriate column. Notes:
- 500 sample dates/values can be entered for each parameter.
- Do not leave cells empty to indicate missing data. If a sample date is specified, values must be entered for all parameters.
- If pasting data from another spreadsheet, select *Values and Number Formats* from the *Paste Special* menu.
- Enter a numeric season code (1-12) if seasonality will be explored. Up to 12 seasons can be defined.
- Separate project files should be created for parameters collected at varied frequencies/dates (e.g., monthly nitrate and weekly TSS).
- Users with continuous data may wish to enter daily mean values (or daily minimum/maximum/etc.).

When finished, click on the Data Exploration worksheet tab to explore pilot data.

<table>
<thead>
<tr>
<th>Sample Date</th>
<th>Sample Season</th>
<th>E. Coli (MPN/100 mL)</th>
<th>Streamflow (cfs)</th>
<th>Parameter 3</th>
<th>Parameter 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>5/15/2006</td>
<td>5</td>
<td>150</td>
<td>164</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5/24/2006</td>
<td>5</td>
<td>1203</td>
<td>216</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6/20/2006</td>
<td>6</td>
<td>179</td>
<td>65</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6/28/2006</td>
<td>6</td>
<td>649</td>
<td>38</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7/11/2006</td>
<td>7</td>
<td>248</td>
<td>33</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7/19/2006</td>
<td>7</td>
<td>139</td>
<td>34</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7/26/2006</td>
<td>7</td>
<td>272</td>
<td>31</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8/8/2006</td>
<td>8</td>
<td>517</td>
<td>53</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8/23/2006</td>
<td>8</td>
<td>579</td>
<td>41</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9/13/2006</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9/20/2006</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9/27/2006</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10/5/2006</td>
<td>10</td>
<td>236</td>
<td>58</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10/18/2006</td>
<td>10</td>
<td>328</td>
<td>48</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5/4/2007</td>
<td>5</td>
<td>133</td>
<td>212</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5/16/2007</td>
<td>5</td>
<td>146</td>
<td>36</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6/1/2007</td>
<td>6</td>
<td>196</td>
<td>30</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6/21/2007</td>
<td>6</td>
<td>162</td>
<td>27</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7/19/2007</td>
<td>7</td>
<td>1047</td>
<td>111</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7/25/2007</td>
<td>7</td>
<td>291</td>
<td>43</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8/6/2007</td>
<td>8</td>
<td>649</td>
<td>53</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8/27/2007</td>
<td>8</td>
<td>147</td>
<td>51</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9/6/2007</td>
<td>9</td>
<td>987</td>
<td>35</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9/11/2007</td>
<td>9</td>
<td>260</td>
<td>26</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Excel Tool - Data Exploration Worksheet

**Step 1.**
Select a Water Quality Parameter & Data Transformation Option - Summary statistics and a time series plot for the selected parameter are provided below.

### Data Exploration Worksheet - Review pilot data summary statistics, histograms, seasonality, autocorrelation, and potential covariates.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Mean</th>
<th>Median</th>
<th>25th Percentile</th>
<th>75th Percentile</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>E. Coli</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Streamflow</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- **Select a Parameter**
- **Review Time Series Plot & Summary Statistics**
Excel Tool - Data Exploration Worksheet

**Step 2**

**Review the Histogram** - Does it take on the characteristic “bell” shape of normally-distributed data or appear highly-skewed? How does the skewness change when a data transformation is applied? If the data appear normally distributed, a parametric statistical test may be appropriate for evaluating post-TMDL water quality change.

- **Histogram showing normal distribution**
- **Histogram showing right-skewed distribution**
- **Histogram showing left-skewed distribution**

![Histograms with frequency and value on axes](image)

- **Frequency**
- **Value**

![Bins and Frequency](image)

- **Bins (Upper Limit)**
  - 335.793
  - 610.913
  - 885.027
  - 1101.142
  - 1436.256
  - 1711.371
  - 2066.485

**Review Histogram**
Step 3.

Review the Seasonal Boxplot - Are observed values similar for each season or does the parameter display strong seasonality? If a strong seasonal pattern is evident, planners may wish to monitor and analyze single-season data only (e.g., the season with the highest risk for water quality impairment) or apply advanced statistical methods that account for the effects of seasonality on trend detection or comparisons of two data groups.

The boxplot displays 5 key summary statistics for each season: the minimum, 25th percentile, median, 75th percentile, and maximum values.

Note that the seasonal boxplot will not be populated unless two or more numeric seasonal codes (1-12) are entered in the Pilot Data Worksheet (up to 12 seasons can be defined).

<table>
<thead>
<tr>
<th>Season 1</th>
<th>Season 2</th>
<th>Season 3</th>
<th>Season 4</th>
<th>Season 5</th>
<th>Season 6</th>
<th>Season 7</th>
<th>Season 8</th>
<th>Season 9</th>
<th>Season 10</th>
<th>Season 11</th>
<th>Season 12</th>
</tr>
</thead>
<tbody>
<tr>
<td>25th Percentile</td>
<td>147.252</td>
<td>102.500</td>
<td>254.109</td>
<td>146.652</td>
<td>259.693</td>
<td>259.032</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Minimum</td>
<td>139.408</td>
<td>80.684</td>
<td>125.888</td>
<td>118.592</td>
<td>82.268</td>
<td>112.572</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td>214.954</td>
<td>170.284</td>
<td>283.302</td>
<td>205.866</td>
<td>328.309</td>
<td>572.290</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maximum</td>
<td>1203.396</td>
<td>648.721</td>
<td>1985.495</td>
<td>648.721</td>
<td>987.216</td>
<td>1413.760</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>75th Percentile</td>
<td>594.065</td>
<td>228.150</td>
<td>541.063</td>
<td>552.572</td>
<td>448.338</td>
<td>959.288</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Review the Lag Plot - Does the parameter display lag-1 autocorrelation (correlation between observed values and observations lagged by one sampling interval)? The presence of strong lag-1 autocorrelation is noteworthy since the assumption of independent/random samples included in many statistical tests is violated. Autocorrelated data require the use of specialized statistical tests for evaluating TMDL effectiveness.

The lag plot displays observed values on the y-axis and lagged values on the x-axis. Note that the lag plot is valid for parameters sampled at regular intervals only (e.g., monthly or weekly) with complete data records.

Step 4:

- Lag plot showing strong autocorrelation
- Lag plot showing moderate autocorrelation
- Lag plot showing weak autocorrelation

Review Lag Plot
Explore Covariates - Review the scatterplot and correlation statistics between parameters. Does the scatterplot appear random or do the parameters "vary together"? The addition of covariates in monitoring and statistical analysis can increase the detectability of water quality changes.

Covariate: Streamflow

Covariate Transformation:
- None
- Log (x+1)
- Square Root
- Logarithm
- Square
- Reciprocal Root

Correlation Coefficient (R) 0.08
Coefficient of Determination (R^2) 0.01

Scatterplot showing strong correlation →

Scatterplot showing moderate correlation →

Scatterplot showing weak correlation →

E. Coli (MPN/100 mL)

Streamflow (cfs)
Excel Tool - Data Exploration Worksheet

**Data Exploration Worksheet** - Review plot data summary statistics, histograms, seasonality, autocorrelation, and potential covariates.

### Step 1
Select a Water Quality Parameter & Data Transformation Option - Summary statistics and a time series plot for the selected parameter are provided below.

- **Parameter:** E. Coli
- **Data Transformation:** None, Log (x+1), Square Root, Logarithm, Square, Reciprocal Root

<table>
<thead>
<tr>
<th>Summary Statistics</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Samples</td>
<td>52</td>
</tr>
<tr>
<td>Minimum</td>
<td>1.783</td>
</tr>
<tr>
<td>Maximum</td>
<td>3.298</td>
</tr>
<tr>
<td>Mean</td>
<td>2.437</td>
</tr>
<tr>
<td>Median</td>
<td>2.426</td>
</tr>
<tr>
<td>25th Percentile</td>
<td>2.173</td>
</tr>
<tr>
<td>75th Percentile</td>
<td>2.725</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>0.3452</td>
</tr>
</tbody>
</table>

- **Time Series Plot:** Log E. Coli (MPN/100 mL) vs. Date

### Step 2
Review the Histogram - Does it take on the characteristic bell-shaped cross-sectional normal distribution? If so, a parametric test for evaluating post-TMDL water quality changes may be appropriate.

- **Histograms**:
  - Normal distribution
  - Right-skewed distribution
  - Left-skewed distribution

---

Explore Data Transformation Options
How many samples are needed to detect a statistically significant water quality change?

Depends on:

- Statistical test applied
- Confidence level
- Statistical power
- Data variability
- Minimum detectable change
Select a Parameter & Data Transformation Option

Key Assumption: A Parametric Statistical Test Will be Applied
Excel Tool - Sample Size Worksheet

- Sample size depends on:
  1) Statistical test applied

Select Study Design
Excel Tool - Sample Size Worksheet

- Sample size depends on:
  1. Statistical power
  2. Confidence level
  3. Data variability

---

**Step 3.**
Enter Desired Power and Confidence Level:
- Sample size increases with increased statistical power and confidence level.

<table>
<thead>
<tr>
<th>Statistical Power</th>
<th>0.80</th>
</tr>
</thead>
<tbody>
<tr>
<td>Confidence Level</td>
<td>0.90</td>
</tr>
</tbody>
</table>

- Statistical power is the probability (0 to 1) that a water quality change will be detected given that a change has actually occurred. The minimum recommended value is 0.8.
- The confidence level is the probability (0 to 1) that a water quality change that is detected has actually occurred. The minimum recommended value is 0.9.

---

**Step 4.**
Estimate Data Variability (Standard Deviation):
- Sample size increases with increased data variability.

| Standard Deviation (from pilot data) | 0.345 |

- The standard deviation of the selected parameter is calculated from data entered in the Pilot Data worksheet. If a transformation was applied in Step 1, the standard deviation of the transformed dataset is used for estimating sample size and is displayed here.

---

**Review Statistical Power & Confidence Level**

**Review Standard Deviation**
Excel Tool - Sample Size Worksheet

- Sample size depends on:
  5) Minimum detectable change

---

**Step 5**

Enter Minimum Detectable Change - Water quality changes less than the minimum detectable change cannot be detected with statistical significance. Sample size increases with decreased minimum detectable change.

<table>
<thead>
<tr>
<th>Change Type &amp; Direction</th>
<th>Pre-TMDL Mean (from pilot data) 390.03 MPN/100 mL</th>
<th>Minimum Detectable Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>percent decrease</td>
<td>The mean of the selected parameter is calculated from data entered in the Pilot Data worksheet. If a transformation is applied, the mean of transformed values is displayed in untransformed units.</td>
<td></td>
</tr>
<tr>
<td>percent increase</td>
<td>The minimum detectable change can be entered as a percent change (e.g., a 10% decrease) or absolute change (e.g., a 0.1 mg/L decrease).</td>
<td></td>
</tr>
</tbody>
</table>

Enter the desired minimum detectable change as a percent change or absolute change in untransformed units.

---

- **Enter Minimum Detectable Change**

---

The plot on the left displays three factors to take into account when selecting a minimum detectable change value:

1) The pre-TMDL condition (displayed here as the pre-TMDL mean);
2) The expected post-TMDL condition (displayed here as the pre-TMDL mean minus the minimum detectable change); and
3) The post-TMDL target.

Users may wish to designate the minimum detectable change as the change needed to achieve the water quality target (the difference between the pre-TMDL mean and the target). However, incremental changes which are less than this difference will not be detected as statistically significant. If a goal of the monitoring study is to identify incremental changes (i.e., the target is not expected to be met during the study period), a smaller minimum detectable change value should be entered.
**Excel Tool - Sample Size Worksheet**

**Pre-TMDL Mean (from pilot data)**: 200.002 MPN/100 mL

**Change Type & Direction**
- Percent
- Absolute
- Decrease
- Increase

**Minimum Detectable Change**
- 100.000 MPN/100 mL
- Enter the desired minimum detectable change as a percent change or absolute change in untransformed units.

**Minimum Length of Time**
- 10 years
- Enter the desired minimum length of time for detecting a statistically significant trend.

---

**Statistics**

Enter Minimum Detectable Change - Water quality changes less than the minimum detectable change cannot be detected with statistical significance. Sample size increases with decreased minimum detectable change.

The mean of the selected parameter is calculated from data entered in the Pilot Data worksheet. If a transformation is applied, the mean of transformed values is displayed in untransformed units.

The minimum detectable change can be entered as a percent change (e.g., a 10% decrease) or absolute change (e.g., a 0.1 mg/L decrease).
Excel Tool - Sample Size Worksheet

### Step 6.

#### Estimate Sample Size

Click the **Estimate Sample Size** button to calculate the minimum number of samples needed to satisfy the conditions specified in steps 2 - 5.

<table>
<thead>
<tr>
<th>Total Sample Size</th>
<th>68</th>
</tr>
</thead>
<tbody>
<tr>
<td>Samples per Period (Total Sample Size / 2)</td>
<td>34</td>
</tr>
</tbody>
</table>

**Save Sample Size**

Click the **Save Sample Size** button to add the sample size estimate to the **Cost Estimation** worksheet.

Before moving on, please note that this power analysis includes several assumptions:

- Data are normally distributed.
- An unpaired or paired t-test, or linear regression with time, will be used to evaluate water quality changes.
- Pilot data are representative of TMDL effectiveness monitoring data.
- Samples are independent/random (not autocorrelated).

If these assumptions are not met, sample size estimates will be skewed and should be viewed as general approximations only.

---

#### Estimate Sample Size

Click the **Estimate Sample Size** button to calculate the minimum number of samples needed to satisfy the conditions specified in steps 2 - 5.

<table>
<thead>
<tr>
<th>Total Sample Size</th>
<th>246</th>
</tr>
</thead>
<tbody>
<tr>
<td>Samples per Year (Total Sample Size / Number of Years)</td>
<td>25</td>
</tr>
</tbody>
</table>

**Save Sample Size**

Click the **Save Sample Size** button to add the sample size estimate to the **Cost Estimation** worksheet.

Before moving on, please note that this power analysis includes several assumptions:

- Data are normally distributed.
- An unpaired or paired t-test, or linear regression with time, will be used to evaluate water quality changes.
- Pilot data are representative of TMDL effectiveness monitoring data.
- Samples are independent/random (not autocorrelated).

If these assumptions are not met, sample size estimates will be skewed and should be viewed as general approximations only.
Excel Tool – Cost Estimation Worksheet

- Use information gained from data exploration to estimate project costs:
  - Lab/equipment
  - Labor/personnel
  - Travel
  - QAPP development
  - Data analysis

- Adjust monitoring decisions to maximize available resources
Question & Answer