Trend detection with *Escherichia coli* samples

Michael Schramm 1,a, Lucas Gregory 2

1 Texas Water Resources Institute, Texas AM AgriLife Research

a Corresponding author, [michael.schramm@ag.tamu.edu](mailto:michael.schramm@ag.tamu.edu)

2020-11-13

 This is the abstract.  
  
 It consists of two paragraphs.

**Keywords** trend detection; *E. coli*, statistical power

**Abbreviations**

|  |  |
| --- | --- |
| Abbreviation | Definition |
| *E. coli* | *Escherichia coli* |
| FIB | fecal indicator bacteria |
| GAM | generalized additive model |
| mL | milliliter |
| MPN | most probable number |
| SWQM | Surface Water Quality Monitoring |
| TCEQ | Texas Commission on Environmental Quality |

# Introduction

Fecal indicator bacteria (FIB) are used to assess the sanitary quality of water for recreational and water supply purposes. FIBs themselves are not dangerous but are utilized as an indicator of potential health risks associated with exposure to pathogens associated with fecal matter. *Escherichia coli* (*E. coli*) is used as an FIB in Texas to assess if streams and other freshwater bodies meet numeric criteria for contact recreation. *E. coli* is a non-host specifc bacteria found in the gut of warm-blooded animals, the presence of *E. coli* is used to indicate the potential for recent fecal contamination.

In-stream FIB concentrations typically follow a lognormal distribution (Novotny, 2004). As a result, the Texas Commission on Environmental Quality (TCEQ) biennially evaluates compliance with the in-stream criterion of 126 most probable number (MPN)/ 100 milliliters (mL) using the geometric mean over a seven-year assessment period. The geometric mean is simply a measure of central tendancy calculated as the exponential of the arithmetic mean of logarithms, when :

Simplified, the geometric mean computes the arithmetic mean of *log(y)* and exponentiation returns the mean to the original scale. The current assessment approach requires a sample size of 20 over the previous 7-years with an 80% confidence interval that exceeds the 126 MPN/100 mL criterion at the lower bound in order to be determined impaired (TCEQ, 2019a). Delistings require 20 samples and the geometric mean below the 126/100 mL criterion. TCEQ (2019a) does not specify how the confidence interval should be calculated. Traditional methods multiply a critical value (obtained from the standard normal distribution or Student’s t-distribution) by the standard error. Alternatively, the confidence intervals can be obtained by parametric bootstrap methods (Wilcox, 2013). At 20 samples the method is fairly robust for estimating exceedance of the water quality criterion.

As of 2018, TCEQ identifed 237 water bodies impaired due to elevated FIB (TCEQ, 2019b). Total Maximum Daily Loads and Implementation Plans or Watershed Protection Plans are developed for these impaired water bodies to address potential FIB sources. As part of these plans, trend analysis is typically conducted to assess if bacterial concentrations have increased or decreased over time. Two common methods for assessing statistical signficance of trends are the modified Mann-Kendall test and linear regression on log transformed FIB concentration values (Helsel and Hirsch, 2002; Yue and Wang, 2002).

Yue and Wang (2002) described the calculation of the Mann-Kendall test and the modifications for correlated data. In short, when the Mann-Kendall test statistic, *S* is negative, newer values tend to be smaller than older values and indicate a downward trend. A small absolute value of *S* indicates no trend. The *P* value of the test statistic is estimated using the normal cumulative distribution function. The null hypothesis of the Mann-Kendall test is that there is no trend.

Simple linear regression can also be applied for trend detection in the form:

where *y* is *E. coli* concentration, *β0* is the intercept, *β1* is the coeffcient of time variable *x*, and ε are model residuals assumed normally distrbuted around mean zero. If linear regression is utilized to assess trends in the *E. coli*, assumptions of normal distribution almost always require that *E. coli* concentrations are log transformed prior to fitting the model. The parameteric assumptions associated with linear regression requires the analyst inspect model residuals to ensure model validity. However, the advantages of using linear regression include flexibility to include additional covariates such as seasonality, precipition, or discharge.

Both the Mann-Kendall test and linear regression are relatively easy methods for water quality analysts to apply and assess trends in *E. coli* concentrations. They are welll accepted, implementable in Microsoft Excel, and have routines available in most statistical software. However, general guidance is not available for the number of samples required to detect given effect sizes. The 20 samples over 7-years guidance for assessment of the water quality criterion is adequate given the ability to estimate confidence intervals for the geometric mean calculation. Many monitoring programs across the state utilize quarterly routine sampling regimes, which equate to approximately 4 samples per year or 28 samples over a 7-year assessment period. Reporting the results of trend detection test implies the test has the statistical power to detect trends of certain magnitudes. However, that information is rarely reported and unlikely that it is routinely calculated by water quality analysts. Therefore, there is considerable uncertaintity if monitoring schedules used across the state are adequate for detecting trends in FIB.

Statistical power refers to the probability that a statistical test rejects the null hypothesis when the alternative hypothesis is actually true. In the case of the discussed trend tests, power is the probability that the null hypothesis of no trend is rejected when there is in fact a trend in the data. Statistical power is a function of pre-assigned signifcance level (α), effect size, sample size, and variance within the time series (Yue et al., 2002). Effect size and variance are out of the control of the analyst. The significance level should be predetermined at the time of the assessment. Sample size must be determined before a sampling program begins. The purpose of this article is to provide some guidance and context in determining monitoring frequency for trend analysis of FIB, specifically *E. coli*. First, we will estimate the statistical power of Mann-Kendall and regression trend tests at sampling sites across the state using Monte Carlo simulation. Second, we will utilize generalized additive models (GAMs) to describe functional relationships between the statistical power of the two trends tests, effect size, variance, and number of samples across sampling sites in Texas.

# Methods

## Data

TCEQ Surface Water Quality Monitoring (SWQM) site information and associated *E. coli* samples collected during the 7-year period from January 2012 through Decemeber 2019 were obtained from the Water Quality Portal using the dataRetrieval package in R (De Cicco et al., 2018; R Core Team, 2019). Data was restricted to river or stream sampling sites, and SWQM sites with fewer than 1 sample per year were removed from analysis. In total, *E. coli* data was assessed from 984 SWQM sites.

## Statistical Power Computation

The signficance level, α, is the probability of rejecting the null hypothesis when it is true (Type I error). The probability of accepting the null hypothesis when it is false is a Type II error (*β*). The statistical power of a test if the probability of rejecting the null hypothesis with the alternative hypothesis is true and is equal to 1 - *β*. If sampling from a population where the null hypothesis is false, power is calculated as:

where *N* is the total number of tests and *Nrejected* are the total number of times the test rejected the null hypothesis.

For each SWQM site, Monte Carlo simulation was used to observe the statistical power of the Mann-Kendall and linear regression test for detecting trends. The simulation generates 800 independent lognormal distributed time series samples per evaluated effect size for every SWQM site using the site specific log-transformed mean and standard deviation. Effect sizes were induced by reducing the annual log-transformed mean over the 7-year sampling period by 5, 10, 15, 20, and 25 percent. In total, 3.9 million simulations were run per trend detection method. Signifcance level, α, was set at 0.10. The Mann Kendall test and linear regression is applied to each simulation sample and the number of times the tests correctly reject the null hypothesis (*Nrejected*) are tabulated.

## Describing Functional Relationships

The functional observed relationships between variance, sample size, and effect size at SWQM sites across Texas were described and visualized using GAMs. GAMs are a flexible semi-parametric approach for modeling non-linear relationships between variables (Wood, 2011, p. @wood\_2017). GAM were fit with the mgcv package in R (Wood, 2017). The general form of the GAM equation was:

where *s* is a spline based smoothing function applied to each covariate, *mu* is the log transformed *E. coli* mean, *sd* is the log transformed site standard deviation, *sample size* is the annual number of samples collected at the site, and *effect size* is the specified percent decrease to detect in the Monte Carlo simulation. The response variable *Power* is the statistical power within the interval [0,1]. GAMs were fit using the beta regression family and the logit link function.

# Results

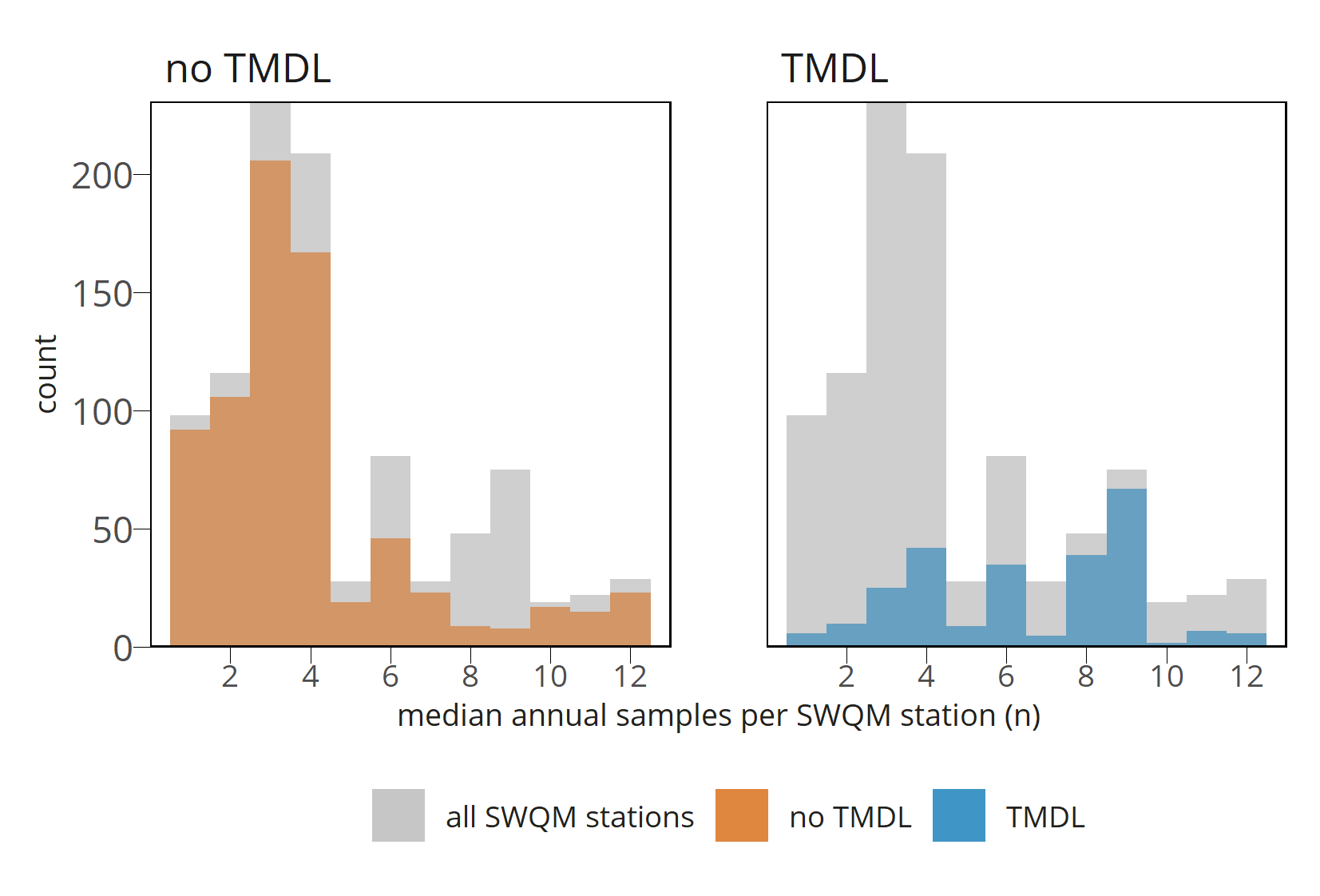


Figure : Annual *E. coli* sampling distribution for SWQM sites across Texas.

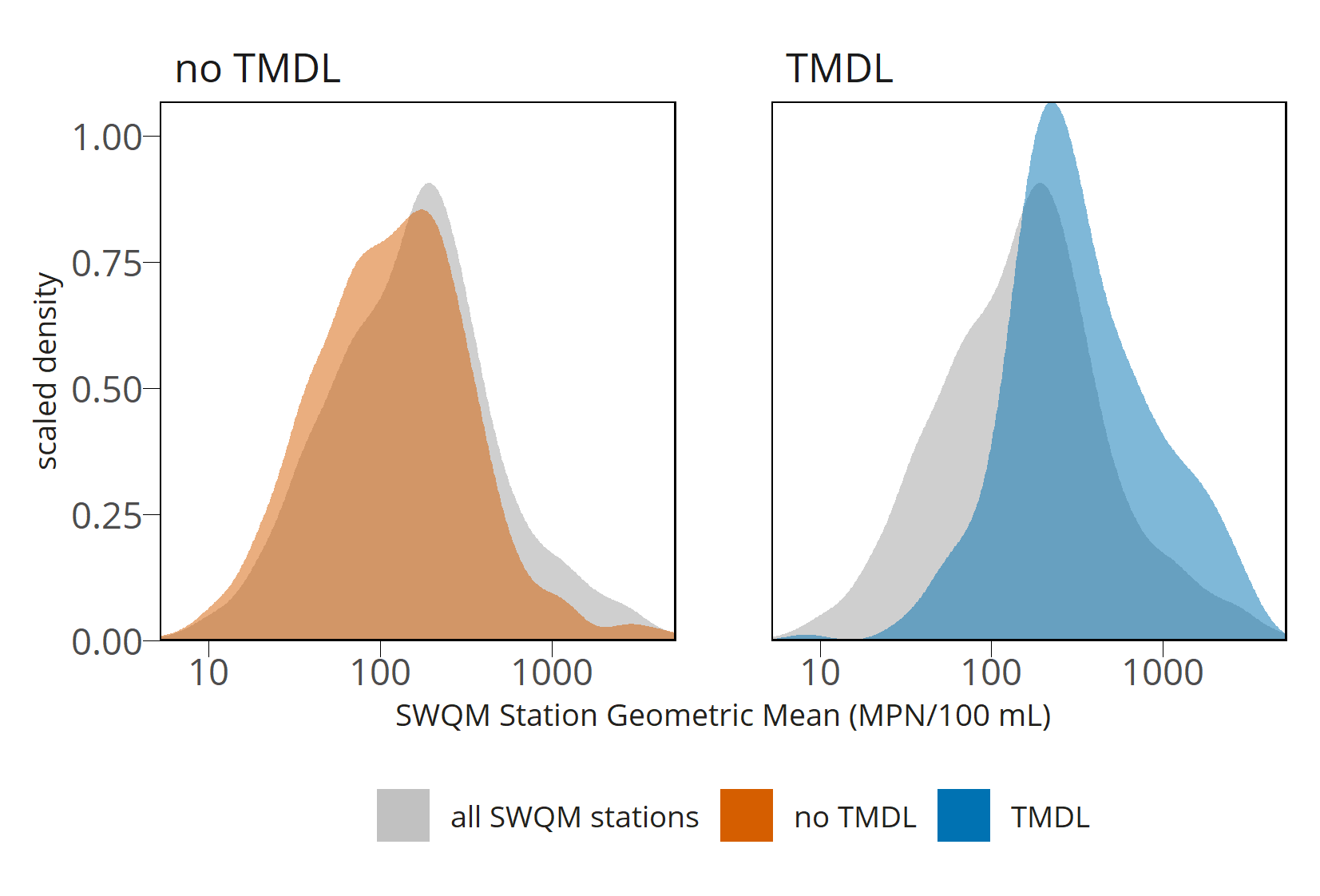


Figure : Distribution of *E. coli* geometric means for SWQM sites across Texas.

# Discussion

# Bibliography

De Cicco, L.A., Lorenz, D., Hirsch, R.M., Watkins, W., 2018. DataRetrieval: R packages for discovering and retrieving water data available from u.s. Federal hydrologic web services. U.S. Geological Survey; U.S. Geological Survey, Reston, VA. <https://doi.org/10.5066/P9X4L3GE>

Helsel, D., Hirsch, R., 2002. Statistical methods in water resources, Techniques of water-resources investigations of the united states geologic survey. U.S. Geological Survey.

Novotny, V., 2004. Simplified databased total maximum daily loads, or the world is log-normal. Journal of Environmental Engineering 130, 674–683. <https://doi.org/10.1061/(ASCE)0733-9372(2004)130:6(674)>

R Core Team, 2019. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

TCEQ, 2019a. 2016 guidance for assessing and reporting surface water quality in texas.

TCEQ, 2019b. Executive summary 2018 texas integrated report for clean water act sections 305(b) and 303(d). TCEQ, Austin, TX.

Wilcox, R., 2013. Introduction to robust estimation and hypothesis testing, Third. ed. Academic Press Elsevier.

Wood, S.N., 2017. Generalized additive models: An introduction with r, 2nd ed. Chapman; Hall/CRC.

Wood, S.N., 2011. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models: Estimation of semiparametric generalized linear models. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 73, 3–36. <https://doi.org/10.1111/j.1467-9868.2010.00749.x>

Yue, S., Pilon, P., Cavadias, G., 2002. Power of the mann–kendall and spearman’s rho tests for detecting monotonic trends in hydrological series. Journal of Hydrology 259, 254–271. <https://doi.org/10.1016/S0022-1694(01)00594-7>

Yue, S., Wang, C.Y., 2002. Regional streamflow trend detection with consideration of both temporal and spatial correlation. International Journal of Climatology 22, 933–946. <https://doi.org/10.1002/joc.781>