



Texas Water Journal

Volume 12 Number 1 | 2021





Texas Water Journal

Volume 12, Number 1

2021

ISSN 2160-5319

texaswaterjournal.org

THE TEXAS WATER JOURNAL is an online, peer-reviewed journal devoted to the timely consideration of Texas water resources management, research, and policy issues. The journal provides in-depth analysis of Texas water resources management and policies from a multidisciplinary perspective that integrates science, engineering, law, planning, and other disciplines. It also provides updates on key state legislation and policy changes by Texas administrative agencies.

For more information on TWJ as well as TWJ policies and submission guidelines, please visit texaswaterjournal.org. As a 501(c)(3) nonprofit organization, the Texas Water Journal needs your support to provide Texas with an open-accessed, peer-reviewed publication that focuses on Texas water. Please consider [donating](#).

Editorial Board

Todd H. Votteler, Ph.D.

Editor-in-Chief

Collaborative Water Resolution LLC

Kathy A. Alexander, Ph.D.

Gabriel Collins, J.D.

Center for Energy Studies

Baker Institute for Public Policy

Ken A. Rainwater, Ph.D.

Texas Tech University

Rosario Sanchez, Ph.D.

Texas Water Resources Institute

Jude A. Benavides, Ph.D.

University of Texas, Rio Grande Valley

Michael Young, Ph.D.

University of Texas, Austin

Managing Editor

Chantal Cough-Schulze

Texas Water Resources Institute

Layout Editor

Sarah Richardson

Texas Water Resources Institute

Staff Editor

September Martin

Texas Water Resources Institute

Kristina J. Trevino, Ph.D.

Trinity University

The Texas Water Journal is published in cooperation with the Texas Water Resources Institute, part of Texas A&M AgriLife Research, the Texas A&M AgriLife Extension Service, and the College of Agriculture and Life Sciences at Texas A&M University.


**Texas Water
Resources Institute**
make every drop count

Cover photo: Llano River with fisherman. ©2018 Ray Uherek.

Estimating statistical power for detecting long term trends in surface water *Escherichia coli* concentrations

Michael Schramm^{1*} 

Abstract: Water quality monitoring programs commonly use the Mann-Kendall test or linear regression to identify statistically significant monotonic trends in fecal indicator bacteria concentrations (typically *Escherichia coli* [*E. coli*]). The statistical power of these tests to detect trends of different magnitudes (effect size) is rarely communicated to stakeholders, and it is unlikely they are considered when designing monitoring schedules. The statistical power for detecting trends in surface water *E. coli* bacteria concentrations using Mann-Kendall and linear regression at water quality monitoring sites across Texas was estimated using Monte Carlo simulation. The probability that an individual water quality monitoring site in Texas had adequate statistical power was also estimated using logistic regression.

Mann-Kendall and linear regression trend tests show similar statistical power. Both tests are unlikely to achieve adequate statistical power when *E. coli* concentrations decrease by 20% or less over 7 years under most sampling frequencies. To adequately detect concentration decreases of 30% to 40% over 7 years, monthly sampling is required. Because many sites across Texas are sampled quarterly, monotonic trends tests will not be powerful enough to detect trends of moderate magnitudes. To better facilitate stakeholder decision-making, it is important to communicate the relative power of statistical tests and detectible magnitudes of changes. I suggest data analysts conduct power analyses to improve monitoring program designs and improve communication of trend test limitations. Software and training for water quality analysts could facilitate communication of power and effect sizes. Alternative trend assessment methods may be more reliable for describing changes in fecal indicator bacteria concentrations.

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

¹ Research Specialist, Texas Water Resources Institute, Texas A&M AgriLife Research

* Corresponding author: michael.schramm@ag.tamu.edu

Received 10 December 2020, Accepted 7 May 2021, Published online 26 October 2021.

Citation: Schramm M. 2021. Estimating statistical power for detecting long term trends in surface water *Escherichia coli* concentrations. Texas Water Journal. 12(1):140-150. Available from: <https://doi.org/10.21423/twj.v12i1.7126>.

© 2021 Michael Schramm. This work is licensed under the Creative Commons Attribution 4.0 International License. To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/> or visit the TWJ [website](#).

Terms used in paper

Acronym/Initialism	Descriptive Name
CV	coefficient of variation
EPA	U.S. Environmental Protection Agency
<i>E. coli</i>	<i>Escherichia coli</i>
GLM	generalized linear model
LOADEST	Load Estimator
mL	milliliter
MPN	most probable number
SWAT	Soil and Water Assessment Tool
SWQM	Surface Water Quality Monitoring
TCEQ	Texas Commission on Environmental Quality
TMDL	Total Maximum Daily Load
WRTDS	Weighted Regressions on Time, Discharge, and Season

INTRODUCTION

Excessive concentrations of fecal indicator bacteria are one of the primary sources of surface water quality impairment in the state of Texas. Fecal indicator bacteria trends are often assessed for significant downward or upward trends using statistical tests. The number of samples and the statistical variance directly impact the magnitude (or effect size) that a statistical test can reliably detect. This is typically referred to as statistical power. Because fecal indicator bacteria concentrations are often characterized by high variance, there is considerable likelihood that common trend tests are not powerful enough to detect trends with the magnitude of interest to stakeholders or decision-makers under typical monitoring frequencies. The primary purpose of this article is to provide an improved understanding of and guidance for determining monitoring frequencies for trend analyses of fecal indicator bacteria in Texas.

Fecal indicator bacteria are used to assess the sanitary quality of water for recreational and water supply purposes. Fecal indicator bacteria 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 a non-host specific bacteria found in the gut of warm-blooded animals and used as a fecal indicator bacteria in Texas to assess if streams and other freshwater bod-

ies meet numeric water quality criteria for contact recreation. The numeric criterion for *E. coli* concentrations is based on U.S. Environmental Protection Agency (EPA) epidemiological studies correlating risks of illness with concentrations of *E. coli* at recreational beaches with point source sewage discharges (Dufour 1984; Fujioka et al. 2015). Typical sources of *E. coli* include sewage, domestic livestock, wildlife, and pets, although *E. coli* has also been documented to naturalize in sediment and algae (Ishii and Sadowsky 2008).

The Texas Commission on Environmental Quality (TCEQ) biennially assesses water quality across the state as part of the requirements with the federal Clean Water Act. Water bodies that exceed water quality standards are placed on the 303(d) list that is provided to EPA. The state is required to develop total maximum daily loads (TMDLs) that calculate allowable pollutant loads and allocate the loads between different sources that discharge to a water body when a water body fails to achieve improved water quality and removal or delisting from the 303(d) list.

In-stream fecal indicator bacteria concentrations typically follow a log-normal distribution (Novotny 2004). As a result, TCEQ biennially evaluates compliance with the in-stream criterion of 126 most probable number (MPN)/100 milliliters (mL) using the geometric mean over a 7-year assessment period. The geometric mean is simply a measure of central ten-

gency calculated as the exponential of the arithmetic mean of logarithms:

$$\left(\prod_{i=1}^n y_i\right)^{\frac{1}{n}} = \exp\left[\frac{1}{n}\sum_{i=1}^n \log y_i\right], \quad \text{when } y_1, y_2, \dots, y_n > 0 \quad (1)$$

Simplified, the geometric mean computes the arithmetic mean of $\log(y)$ and exponentiation returns the mean to the original scale. An alternative approach is to take the n th root of the product of y_i . The current assessment approach requires a minimum sample size of 20 over a 7-year assessment period with an 80% confidence interval that exceeds the 126 MPN/100 mL criterion at the lower bound for a stream to be considered impaired and added to the 303(d) list of impaired water bodies (TCEQ 2019a). Delisting requires 20 samples and a 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. Confidence intervals can also be obtained by parametric bootstrap methods (Wilcox 2013).

As of 2018, TCEQ identified 237 impaired water bodies based on elevated fecal indicator bacteria (TCEQ 2019b). TMDLs and implementation plans or watershed protection plans are developed for these impaired water bodies to address potential fecal indicator bacteria 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 significance of monotonic trends are the Mann-Kendall test and linear regression on fecal indicator bacteria 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 regressions on log-transformed *E. coli* concentrations are also suitable for identifying trends. In order to assess presence of a trend, the following linear regression is used:

$$\log(y) = \beta_0 + \beta_1 x + \varepsilon \quad (2)$$

where y is *E. coli* concentration, β_0 is the intercept, β_1 is the coefficient of time variable x , and ε is the error term assumed normally distributed around mean zero. If linear regressions are utilized to assess *E. coli* trends, the analyst should assess model

residuals to ensure the regression model meets assumptions of heterogeneity and normal distribution.

Both the Mann-Kendall test and linear regression are straightforward methods for water quality analysts to apply and assess trends in *E. coli* concentrations. They are well accepted 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. Current assessment guidance for attainment of the water quality criterion (20 samples over 7 years) is adequate given the ability to estimate confidence intervals for the geometric mean calculation. As a result, many monitoring programs across the state utilize quarterly sampling regimes, which equate to approximately four samples per year or 28 samples over a 7-year assessment period.

Often, the results of trends tests are simply communicated as, "the Mann-Kendall trend test detected a significant trend ($p < 0.05$)." On its own, the presence or lack of statistical significance does not provide meaningful information for decision-making. The p -value is simply a threshold for the researcher to reject the null hypothesis. More bluntly, the researcher infers that an effect exists from the p -value, but the p -value does not communicate the magnitude of the effect or whether it is meaningful. Reporting a model coefficient or Sen slope with the p -value provides context of effect size. For water quality, this is typically described in units of total change or more commonly percent change over the time period of interest, such as by saying "a statistically significant 35% decrease in fecal indicator bacteria was observed." Such reporting of effect sizes to stakeholders is important because it provides context of environmental change that is useful for decision-making.

Reporting the results of a trend detection test implies the test has the statistical power to detect trends of certain magnitudes or effect size. However, that information is rarely reported, and it is unlikely that it is routinely calculated by water quality analysts. Therefore, there is considerable uncertainty if monitoring schedules (especially those designed around quarterly monitoring) used across the state are adequate for detecting trends in fecal indicator bacteria.

Statistical power refers to the probability that a statistical test rejects the null hypothesis when the alternative hypothesis is true. In the case of the discussed trend tests, power is the probability that the null hypothesis (that no trend is present) is rejected when there is in fact a trend in the data. Statistical power is a function of pre-assigned significance level (α), effect size, sample size, and variance within the time series (Yue et al. 2002). First, a meaningful effect size must be determined. The effect size might be biologically meaningful or informed by stakeholder input. Statistical power can be determined for a range of sample size, significance levels, effect sizes and sample variance. Using this information, a monitoring program can be designed that balances sample size and ability to detect meaningful effect sizes with a trend test.

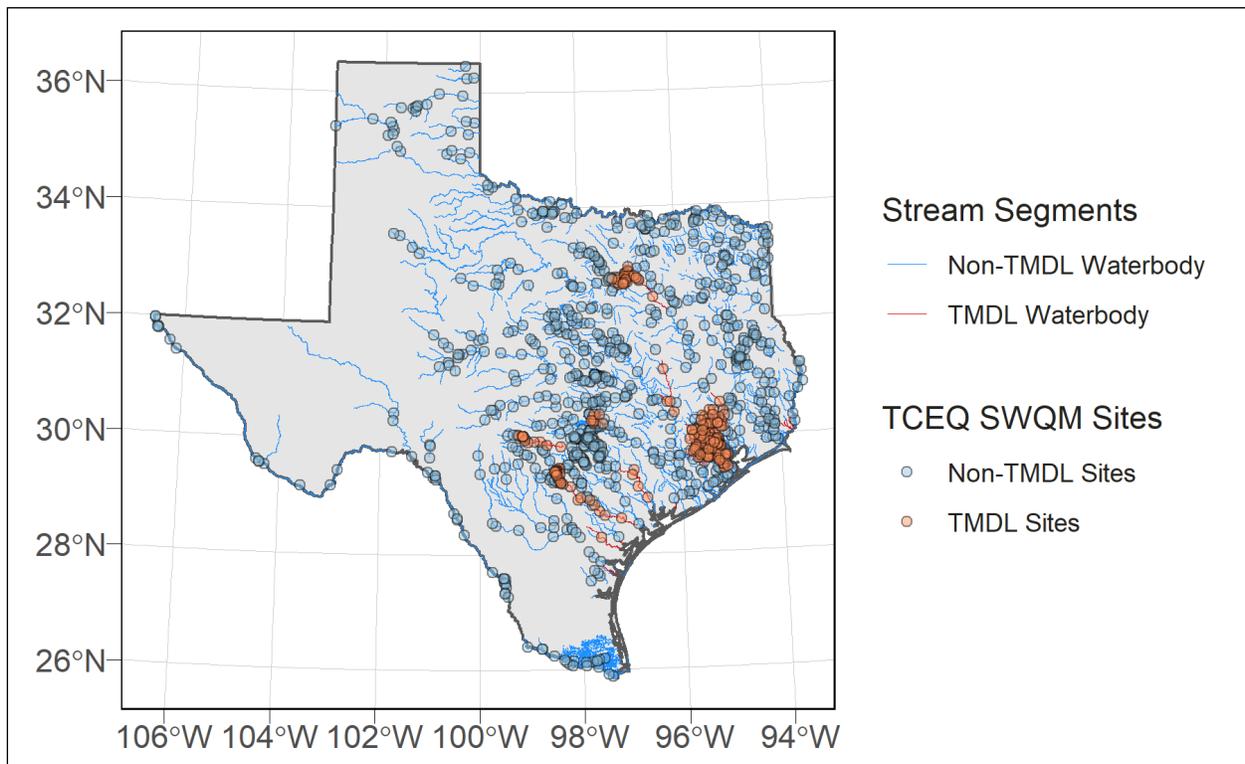


Figure 1. Map of TMDL and non-TMDL stream and SWQM site locations.

Two examples are described to provide practical context of statistical power and effect sizes. In the first scenario, a watershed group is interested in monitoring *E. coli* concentration trends following the installation of a large best management practice in the watershed. The best management practice is expected to result in a 10% reduction in bacteria concentrations over 5 years. The monitoring plan will need to determine how many samples are needed annually to confidently detect the hypothesized trend. If too few samples are collected, a 10% change may never be detected by the hypothesis test. If too many samples are collected, trends of smaller effect size can be detected. However, the group is not interested in detecting a small effect and the money could be better used elsewhere by the group.

In the second scenario, a watershed group is analyzing *E. coli* data collected over the last 7 years. In this case, the number of samples is already established. The hypothesis test fails to reject the null hypothesis that there is no trend in the bacteria data. The watershed group is disappointed because they worked hard to address bacteria sources and expected at least a small improvement in bacteria concentrations. In this case, a post-hoc power analysis indicates that the statistical power is 0.80 when there is an 40% change in bacteria concentrations over 7 years. Power analysis also indicates that the statistical power drops to 0.40 if there is only a 30% change in bacteria concentrations. In this case, it is important to tell the group that based on the number of samples, it was unlikely that a trend

could be detected at a given significance level. Using different experimental design (such as pre-, post-testing) or improved sampling procedures would provide more meaningful insight for the stakeholders.

The purpose of this article is to provide some guidance and context in determining monitoring frequency for trend analysis of fecal indicator bacteria, specifically *E. coli*. First, I estimate the statistical power of Mann-Kendall and linear regression trend tests at sampling sites across the state using Monte Carlo simulation. Second, I provide statistical power plots at different effect sizes for a range of observed variance values. Finally, I model the likelihood of adequate statistical power for *E. coli* trend detection at sampling sites across 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 December 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 one sample per year were removed from analysis. In total, *E. coli* data was assessed from 984 SWQM sites (Figure 1). Stations were also divided

into groups based on the presence or absence of an indicator bacteria TMDL. Water body locations and TMDL classification were spatially linked to a SWQM station location layer to classify SWQM stations as located within or outside a TMDL water body (TCEQ 2020). Although stations could have been split by a number of different variables (for example, watershed protection plans, impairment status, or region), TMDLs were used due to the simplified association with specific water bodies and relatively high number of indicator bacteria TMDLs to ease interpretation of results.

Statistical power computation

The significance 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 is the probability of rejecting the null hypothesis when the alternative hypothesis is true and is equal to $1 - \beta$. A power of 0.80 is typically considered appropriate, which equates to a 20% chance of encountering a Type II error. If sampling from a population where the null hypothesis is false, power is calculated as:

$$Power = \frac{N_{rejected}}{N} \quad (3)$$

where N is the total number of tests and $N_{rejected}$ 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 (Sigal and Chalmers 2016). The simulation generates 1,000 independent log-normal distributed time series samples per evaluated effect size for each 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 10%, 20%, 40%, and 80%. Over 3.93 million simulations were run per trend detection method. Significance level, α , was set at 0.10. The Mann-Kendall test and linear regression were applied to each simulation sample and the number of times the tests correctly rejected the null hypothesis ($N_{rejected}$) were tabulated. Statistical power plots were also generated using Monte Carlo simulation on sample datasets generated using the quartiles (lower, median, and upper) of the observed coefficient of variation (CV) of *E. coli* from SWQM sites. CV is a method of measuring the spread of a distribution relative to the size of the mean; specifically, it is ratio of the standard deviation to the mean. These power plots provide a general idea of the expected statistical power of characteristic *E. coli* datasets in the state using typical sampling intervals. They are not intended to be a replacement for conducting a statistical power test using site-specific data.

Likelihood of adequate statistical power

I modeled the likelihood that a SWQM site would have adequate statistical power (≥ 0.80) as a function of sample size, variance, and effect size using generalized linear models (GLMs). GLMs are an extension of linear regression that allows for response variable with non-normal error distributions through the use of a link function. GLMs were setup as a logistic regression model of form:

$$\log \left[\frac{P(Y)}{1 - P(Y)} \right] = \quad (4)$$

$$\beta_0 + \beta_1 cv + \beta_2 sample\ size + \beta_3 effect\ size + \varepsilon$$

where the probability of adequate statistical power is response on the right-hand side of the equation and is a function of the sum of the dependent variables with their corresponding coefficients (β) and random errors (ε). GLMs were fit using the “glm” function in R with the binomial family and logit link function.

RESULTS

Monitoring frequency

Out of the 987 evaluated SWQM sites, 329 were located in water bodies with a TMDL. A total of 22,766 samples were collected at the 658 non-TMDL SWQM sites compared to 13,008 collected at the 329 TMDL SWQM sites. SWQM sites located on water bodies without a TMDL were generally sampled three to four times per year (Figure 2). SWQM sites with a TMDL skewed higher, with a peak at nine times per year and smaller peaks at four and six times per year. This suggests that increased monitoring efforts are targeted towards sites with TMDLs. Similarly, the *E. coli* geometric mean skewed higher at sites with a TMDL (Figure 3). This is expected as TMDL sites are impaired for bacteria, although there are non-TMDL sites that are also impaired and a TMDL has not been developed yet.

Estimated statistical power at SWQM sites

At current annual sampling frequencies, all SWQM sites fell below 0.80 power for detecting effect sizes of 10% (Figure 4). At 20% effect size, all non-TMDL sites had less than 0.80 power. The majority of TMDL SWQM sites fail to detect a 20% change. However, there is large observed variance in statistical power for TMDL sites at 20% effect size. At 40% and 80% effect sizes, the majority of TMDL SWQM sites had power above 0.80. Non-TMDL SWQM sites exhibit high variance at 40% effect sizes and sufficient statistical power at most sites

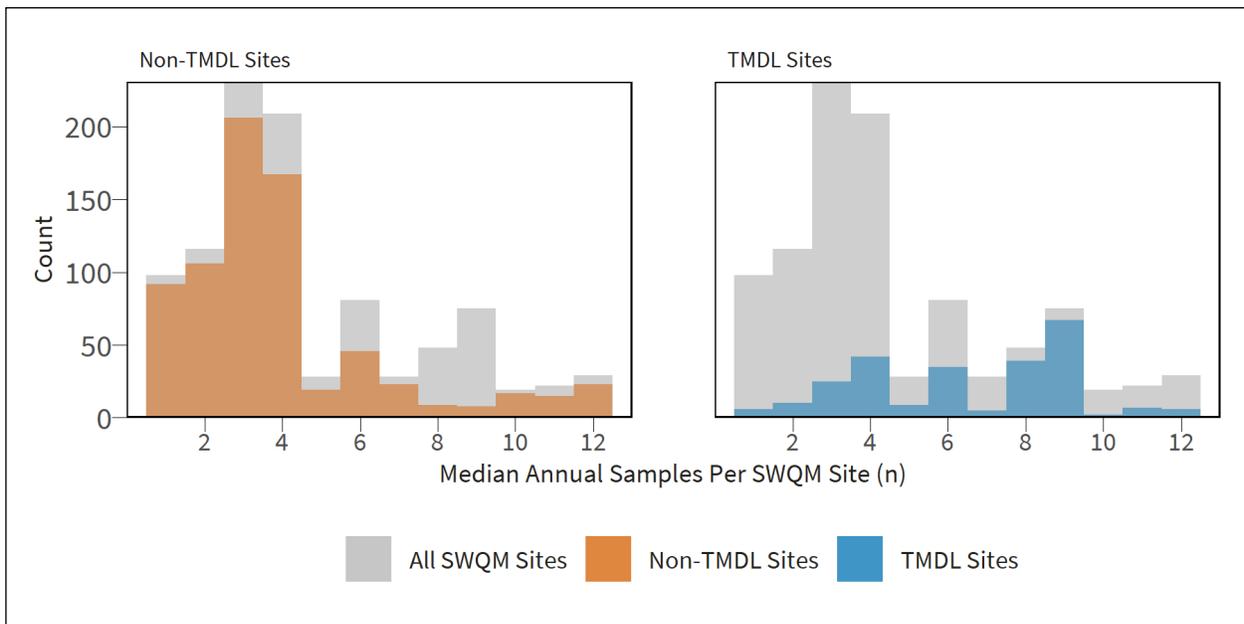


Figure 2. Histograms of annual *E. coli* sampling distribution for TMDL and non-TMDL SWQM sites across Texas (January 2012 through December 2019).

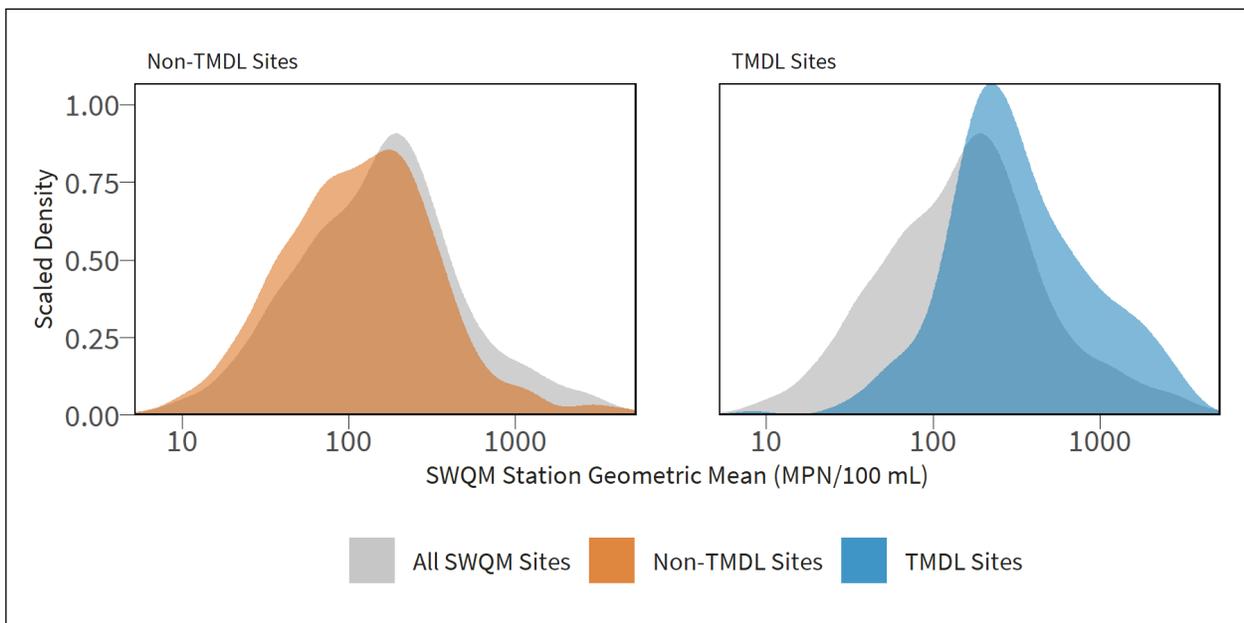


Figure 3. Scaled density plots of *E. coli* geometric mean distribution for TMDL and non-TMDL SWQM sites across Texas.

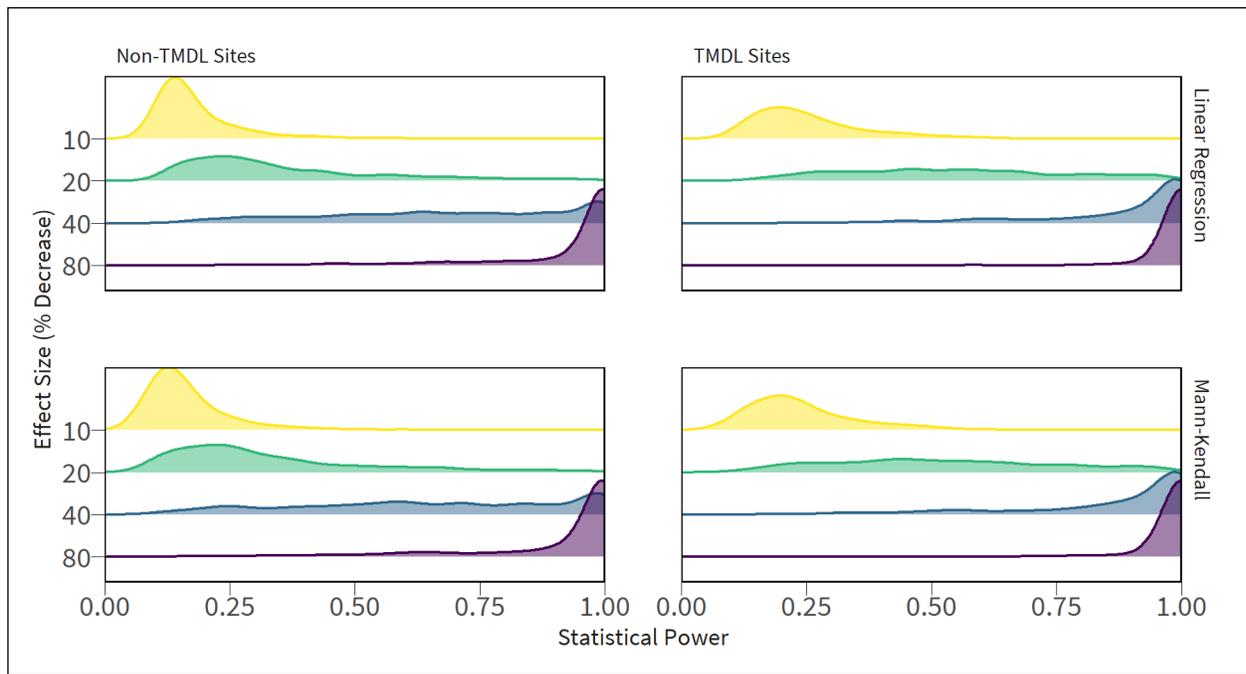


Figure 4. Scaled density plots of Mann-Kendall and linear regression statistical power distribution for TMDL and non-TMDL SWQM sites as a function of effect size at $\alpha = 0.1$. Individual curves represent the scaled density estimate of statistical power values calculated for SWQM sites at a given effect size (y-axis values).

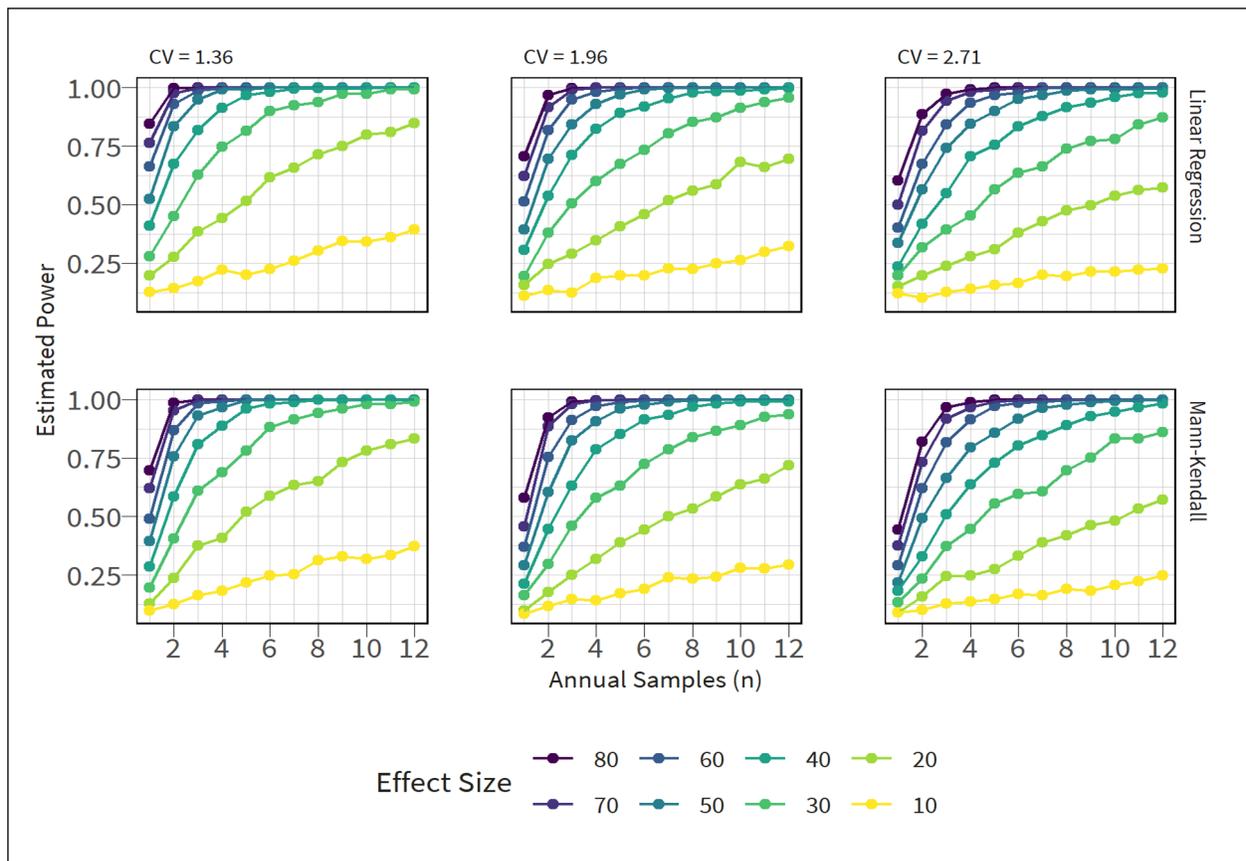


Figure 5. Estimated statistical power of Mann-Kendall and linear regression trend tests at upper, middle, and lower quartiles of observed station *E. coli* variance.

Table 1. GLMs for probability of adequate statistical power.

Variable	Mann-Kendall			Linear regression		
	OR ¹	95% CI ¹	p-value	OR ¹	95% CI ¹	p-value
CV	0.39	0.34, 0.45	<0.001	0.38	0.33, 0.43	<0.001
Sample size	1.74	1.66, 1.83	<0.001	1.72	1.63, 1.80	<0.001
Effect size	0.90	0.89, 0.90	<0.001	0.89	0.89, 0.90	<0.001

¹ OR = odds ratio, CI = confidence interval

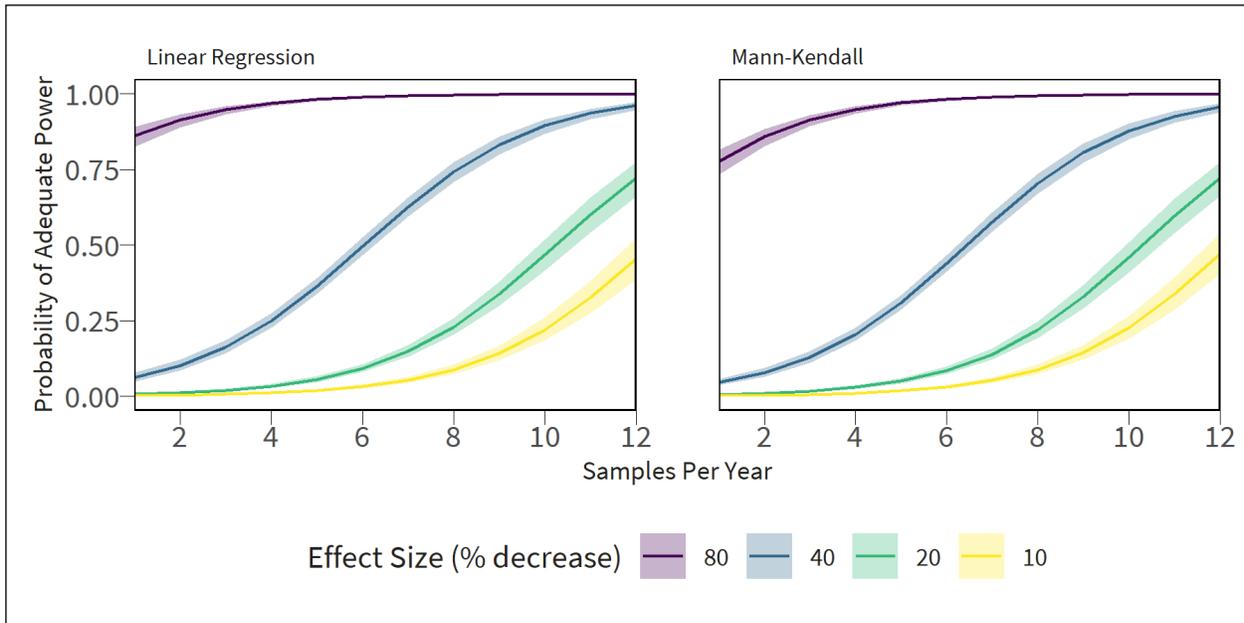


Figure 6. GLM marginal effects plots show the likelihood that a SWQM site has adequate statistical power for detecting trends as a function of the number of samples and desired detected effect size. CV is held constant at the mean.

at 80% effect size. These differences coincide with the higher sampling efforts devoted to TMDL SWQM sites.

The upper, middle, and lower quartiles of the CV across all sites was 2.71, 1.96, and 1.36. The CV values indicate the relatively high variance in *E. coli* concentrations within SWQM sites. Statistical power calculated for the Mann-Kendall and linear regression tests on simulated *E. coli* datasets at the identified CV quartiles is displayed in Figure 5. For each test, as CV increases, statistical power decreases at each given effect size. Overall, both methods show similar statistical power.

Neither method has adequate power to detect trends at 10% effect size. At median variance, both tests have marginal power to detect trends of 30% with 12 samples per year. At 40% effect size, Mann-Kendall and linear regression require five and four samples per year respectively to achieve greater than 0.8 power. At 50% and greater effect size, three or fewer samples per year are required to achieve adequate power. It is important to note that these figures are developed for typically expected *E. coli* distributions at SWQM sites. A site-specific power analysis conducted using existing sample sets would provide a more

accurate assessment of the expected sample distribution and estimated statistical power.

Likelihood of obtaining statistical power

Variance, sample size, and effect size are significant and substantial predictors of the probability that a SWQM site will have adequate power for detecting trends using linear regression or Mann-Kendall test methods (Table 1). Figure 6 displays the estimated effect of sample size and effect size on probability of adequate statistical power being obtained at a SWQM site. At mean variance values and large effect sizes, it is likely that adequate power will be obtained regardless of sample size. Probability decreases substantially as effect size and sample size decrease. Even with monthly sampling, there is only 0.5 probability that a SWQM site will obtain 0.80 power for detecting a 10% effect size.

The GLM models demonstrate the implications of sample design for identifying trends at SWQM sites. Non-TMDL sites often have four or fewer samples per year (Figure 2). The

likelihood of detecting all but the largest of changes in *E. coli* concentrations at non-TMDL sites are small. TMDL sites generally implement more sampling effort through the year and are more likely to obtain adequate power for identifying trends of smaller magnitude. In either case, the relative detectable effect size might seem high to stakeholders given the sampling effort expended.

DISCUSSION

The primary objective of this exploratory analysis is to communicate the importance of considering effect sizes when utilizing hypothesis tests to identify trends in fecal indicator bacteria datasets. Given the high variance observed in *E. coli* samples, I observed relatively low power for detecting trends of 20% or less in magnitude. Logistic regression demonstrates there is low likelihood that SWQM sites will have the desired power for detecting up to a 20% change in *E. coli*. At 40% and larger effect sizes, various sampling regimes can be developed with sufficient power for detecting trends. The paper focuses on statistical power and effect size because effect size is a more useful metric than p-values and provides environmental or decision-making relevance ([Nakagawa and Cuthill 2007](#); [Hanel and Mehler 2019](#)). While statistical significance provides a metric to infer the presence of an effect, statistical power and effect sizes provide information about detectable magnitudes that can be used to make decisions.

Power calculations prior to development of monitoring schedules would allow improved estimation of the number of samples required for trend detection. Given the number of ongoing monitoring programs across the state, a more likely scenario is a post-hoc analysis to identify the minimum effect size that is likely to be detected by a trend test. The basis of identifiable effect sizes requires communication with stakeholders to determine meaningful changes in water quality. Conversely, power can be calculated after the data are collected to identify the statistical power achieved. Water quality management is an inherently stakeholder-driven process that requires substantial communication, trust, and knowledge-sharing ([Leach and Pelkey 2001](#)). Power analysis could be useful for communicating the anticipated or achieved statistical power of trend tests to stakeholders. By focusing discussion on effect sizes and not statistical significance, there is increased opportunity for communicating understandable results.

Although the Mann-Kendall and linear regression trend tests are relatively easy to conduct, statistical power tests are likely to be outside the expertise of a typical water quality analyst. Communication with a statistician is often recommended before sample design. Additionally, for such routinely designed water quality monitoring projects, an accessible software package for water quality analysts would be useful. For example,

the “*emon*” package in R provides accessible functions for estimating the statistical power of various hypothesis tests on environmental data sets ([Barry et al. 2017](#)). However, it does not include functions for evaluating typically log-normal data such as fecal indicator bacteria. There is an opportunity to provide simplified interfaces for routine power tests. For example, the Soil and Water Assessment Tool (SWAT) is now available online with a simplified user interface ([Yen et al. 2016](#)). SWAT is a fairly complex deterministic model used to simulate physical watershed processes such as streamflow and pollutant loading. Although it is used by hydrologists and researchers around the world, it requires substantial user investment to develop skills to properly use it. The simplified online interface provides calibrated SWAT model outputs that are much more accessible to watershed planners and non-modelers. With the increased availability of low-cost cloud computing and cloud based statistical platforms, similar implementation of simplified targeted statistical services should be feasible.

Alternative methods for evaluating indicator bacteria trends can also be utilized. Statistical models, such as generalized additive models, Load Estimator (LOADEST), or Weighted Regressions on Time, Discharge, and Season (WRTDS), can estimate monthly or annual average fecal indicator concentrations ([Runkel et al. 2004](#); [Hirsch et al. 2010](#); [Wood 2011](#)). Aggregated modeled values typically have less variance than sampled measurements, allowing for improved comparisons of year-to-year variations and trends. Furthermore, the marginal effect of the temporal component of these models can be assessed for periods of significant change using confidence intervals or decomposed to assess trends under different flow conditions ([Zhang et al. 2020](#)). It is likely that monthly sampling for at least several years is required to build an accurate statistical model. For example, WRTDS recommends 10 to 20 years of data and at least 100 samples to identify temporal trends with confidence. Even this recommendation might be low for log-normal data with such high variance. A second drawback is the difficulty fitting these models. Generalized additive models and WRTDS both rely on the R statistical software and an analyst who is proficient in statistical modeling and programming in R. LOADEST is available as a stand-alone executable; however, the program still requires some specified training.

Monotonic (and non-linear) trend analysis is not the only method to evaluate water quality. A plethora of statistical methods are available to analysts, and the appropriateness of those methods will vary based on the questions that stakeholders and decision-makers need answered. It is outside the scope of this article to discuss each method and scenario. Underlying the effective use of any empirical method is an understanding of appropriate sample sizes required to make informed decisions. In some cases, exact formulas are available to calculate required sample sizes to achieve adequate statistical power. In other cas-

es, as shown here, Monte Carlo simulation provides an effective way to estimate statistical power under various scenarios.

It is worth noting that despite the numerous TMDLs and watershed-based plans developed in Texas based on fecal indicator bacteria-based assessments, effort is being made toward developing risk-based assessments using quantitative microbial risk assessment and microbial source tracking (Goodwin et al. 2017). It is well established that pathogen sources (wildlife, raw sewage, or treated effluent for example) influence the infectivity of fecal pathogens, which directly influence the risk of infection associated with exposure to water with fecal contamination (Schoen and Ashbolt 2010; Soller et al. 2010; Gitter et al. 2020). Management based only on fecal indicator bacteria concentrations and not the makeup of the contributing sources results in overestimates of human health risk. As methods to assess water body compliance with potential future-risk-based pathogen exposure criteria develop, the methods to estimate and communicate trends and effect sizes with stakeholders will also need to evolve.

ACKNOWLEDGEMENTS

Thank you to the anonymous reviewers who spent time to review and make suggestions that improved this article.

REFERENCES

- Barry J, Maxwell D, Jennings S, Walker D, Murray J. 2017. Emon: An R-package to support the design of marine ecological and environmental studies, surveys and monitoring programmes. *Methods in Ecology and Evolution*. 8(10):1342-1346. Available from: <https://doi.org/10.1111/2041-210X.12748>.
- De Cicco LA, Lorenz D, Hirsch RM, Watkins W. 2018. dataRetrieval: R packages for discovering and retrieving water data available from U.S. Federal hydrologic web services. Reston, VA: U.S. Geological Survey; U.S. Geological Survey. Available from: <https://code.usgs.gov/water/dataRetrieval>.
- Dufour AP. 1984. Bacterial indicators of recreational water quality. *Canadian Journal of Public Health*. 75(1):9. Available from: <https://www.jstor.org/stable/41990233>.
- Fujioka R, Solo-Gabriele H, Byappanahalli M, Kirs M. 2015. U.S. Recreational water quality criteria: A vision for the future. *International Journal of Environmental Research and Public Health*. 12(7):7752-7776. Available from: <https://doi.org/10.3390/ijerph120707752>.
- Gitter A, Mena K, Wagner K, Boellstorff D, Borel K, Gregory L, Gentry T, Karthikeyan R. 2020. Human health risks associated with recreational waters: Preliminary approach of integrating quantitative microbial risk assessment with microbial source tracking. *Water*. 12(2):327. Available from: <https://doi.org/10.3390/w12020327>.
- Goodwin KD, Schriewer A, Jirik A, Curtis K, Crumpacker A. 2017. Consideration of natural sources in a bacteria TMDL—lines of evidence, including beach microbial source tracking. *Environmental Science & Technology*. 51(14):7775-7784. Available from: <https://doi.org/10.1021/acs.est.6b05886>.
- Hanel PH, Mehler DM. 2019. Beyond reporting statistical significance: Identifying informative effect sizes to improve scientific communication. *Public Understanding of Science*.:18. Available from: <https://doi.org/10.1177/0963662519834193>.
- Helsel D, Hirsch R. 2002. *Statistical methods in water resources*. U.S. Geological Survey (Techniques of water-resources investigations of the United States Geologic Survey). Available from: <http://water.usgs.gov/pubs/twri/twri4a3/>.
- Hirsch RM, Moyer DL, Archfield SA. 2010. Weighted regressions on time, discharge, and season (WRTDS), with an application to Chesapeake Bay river inputs. *JAWRA Journal of the American Water Resources Association*. 46(5):857-880. Available from: <https://doi.org/10.1111/j.1752-1688.2010.00482.x>.
- Ishii S, Sadowsky MJ. 2008. Escherichia coli in the environment: Implications for water quality and human health. *Microbes and Environments*. 23(2):101-108. Available from: <https://doi.org/10.1264/jsme2.23.101>.
- Leach WD, Pelkey NW. 2001. Making watershed partnerships work: A review of the empirical literature. *Journal of Water Resources Planning and Management*. 127(6):378-385. Available from: [https://doi.org/10.1061/\(ASCE\)0733-9496\(2001\)127:6\(378\)](https://doi.org/10.1061/(ASCE)0733-9496(2001)127:6(378)).
- Nakagawa S, Cuthill IC. 2007. Effect size, confidence interval and statistical significance: A practical guide for biologists. *Biological Reviews*. 82(4):591-605. Available from: <https://doi.org/10.1111/j.1469-185X.2007.00027.x>.
- Novotny V. 2004. Simplified databased total maximum daily loads, or the world is log-normal. *Journal of Environmental Engineering*. 130(6):674-683. Available from: [https://doi.org/10.1061/\(ASCE\)0733-9372\(2004\)130:6\(674\)](https://doi.org/10.1061/(ASCE)0733-9372(2004)130:6(674)).
- R Core Team. 2019. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Available from: <https://www.R-project.org/>.

- Runkel RL, Crawford CG, Cohn TA. 2004. Load estimator (LOADEST): A FORTRAN program for estimating constituent loads in streams and rivers. U.S. Geological Survey Techniques and methods Report No.: 4-A5. Available from: <https://doi.org/10.3133/tm4A5>.
- Schoen ME, Ashbolt NJ. 2010. Assessing pathogen risk to swimmers at non-sewage impacted recreational beaches. *Environmental Science & Technology*. 44(7):2286-2291. Available from: <https://doi.org/10.1021/es903523q>.
- Sigal MJ, Chalmers RP. 2016. Play it again: Teaching statistics with Monte Carlo simulation. *Journal of Statistics Education*. 24(3):136-156. Available from: <https://doi.org/10.1080/10691898.2016.1246953>.
- Soller JA, Schoen ME, Bartrand T, Ravenscroft JE, Ashbolt NJ. 2010. Estimated human health risks from exposure to recreational waters impacted by human and non-human sources of faecal contamination. *Water Research*. 44(16):4674-4691. Available from: <https://doi.org/10.1016/j.watres.2010.06.049>.
- [TCEQ] Texas Commission on Environmental Quality. 2019a. 2016 guidance for assessing and reporting surface water quality in Texas. Available from: https://www.tceq.texas.gov/assets/public/waterquality/swqm/assess/16txir/2016_guidance.pdf.
- [TCEQ] Texas Commission on Environmental Quality. 2019b. Executive summary 2018 Texas integrated report for Clean Water Act Sections 305(b) and 303(d). Austin, TX: TCEQ. Available from: https://www.tceq.texas.gov/assets/public/waterquality/swqm/assess/18txir/2018_exec_summ.pdf.
- [TCEQ] Texas Commission on Environmental Quality. 2020. List of segments with TMDLS. Segments with total maximum daily loads. [accessed 2020 Jul 6]. Available from: <https://www.tceq.texas.gov/waterquality/tmdl/nav/tmdlsegments>.
- Wilcox R. 2013. Introduction to robust estimation and hypothesis testing. Third. Academic Press Elsevier. Available from: <https://doi.org/10.1016/B978-0-12-386983-8.00015-9>.
- Wood SN. 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(1):3-36. Available from: <https://doi.org/10.1111/j.1467-9868.2010.00749.x>.
- Yen H, Daggupati P, White MJ, Srinivasan R, Gossel A, Wells D, Arnold JG. 2016. Application of large-scale, multi-resolution watershed modeling framework using the hydrologic and water quality system (HAWQS). *Water*. 8(4):164. Available from: <https://doi.org/10.3390/w8040164>.
- 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(1):254-271. Available from: [https://doi.org/10.1016/S0022-1694\(01\)00594-7](https://doi.org/10.1016/S0022-1694(01)00594-7).
- Yue S, Wang CY. 2002. Regional streamflow trend detection with consideration of both temporal and spatial correlation. *International Journal of Climatology*. 22(8):933-946. Available from: <https://doi.org/10.1002/joc.781>.
- Zhang Q, Webber JS, Moyer DL, Chanat JG. 2020. An approach for decomposing river water-quality trends into different flow classes. *Science of The Total Environment*. Available from: <https://doi.org/10.1016/j.scitotenv.2020.143562>.