


The influence of socioeconomic and spatial variables on total maximum daily load progress in the United States

Charitha Gunawardana and Walter McDonald *

Department of Civil, Construction & Environmental Engineering, Marquette University, Milwaukee, WI, USA

*Corresponding author. E-mail: walter.mcdonald@marquette.edu

 WM, 0000-0002-9217-7908

ABSTRACT

The total maximum daily load (TMDL) program requires each state in the United States to assess their water bodies, list those that are impaired, and develop TMDL plans or programs to restore them. However, the impact that spatial and socioeconomic variables have on TMDL progress is unknown. This study seeks to fill this gap through a nation-wide analysis of the influence of spatial and socioeconomic variables on indicators of TMDL progress. To do so, data were collected and analyzed for each state, including indicators of TMDL progress, spatial variables, and socioeconomic data. Then, these data were applied to identify overall trends and to define the relationships between indicators of TMDL progress and spatial and socioeconomic variables. Results indicate that the size of a state, the length of total streams, and median household income are related to the percentage of streams that are assessed within a state. In addition, states largely followed similar patterns in TMDL progress based on the US Environmental Protection Agency region that they were within, indicating that location plays a large role. Overall, this study helps to contextualize progress in TMDL development and aid in our understanding of factors that influence the implementation of water quality programs.

Key words: Pollutants, Socioeconomics, Streams, Total maximum daily load, Water quality

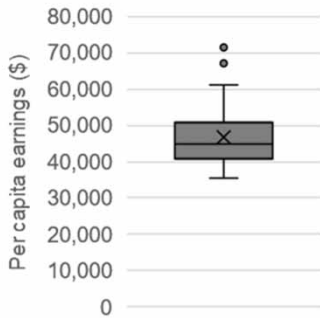
HIGHLIGHTS

- Progress of the total maximum daily load (TMDL) program across all US states is summarized.
- Regression analysis identifies what spatial or socioeconomic factors are related to indicators of TMDL progress.
- Size of a state, stream lengths, and per capita earnings are related to the percent of streams that are assessed.
- Indicators of TMDL progress in each state are influenced by their Environmental Protection Agency region.

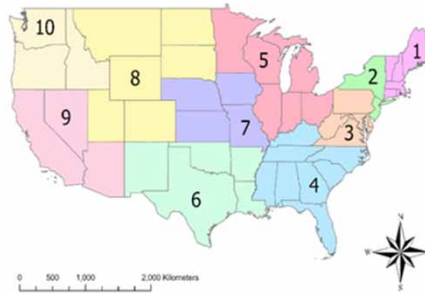
This is an Open Access article distributed under the terms of the Creative Commons Attribution Licence (CC BY 4.0), which permits copying, adaptation and redistribution, provided the original work is properly cited (<http://creativecommons.org/licenses/by/4.0/>).

GRAPHICAL ABSTRACT

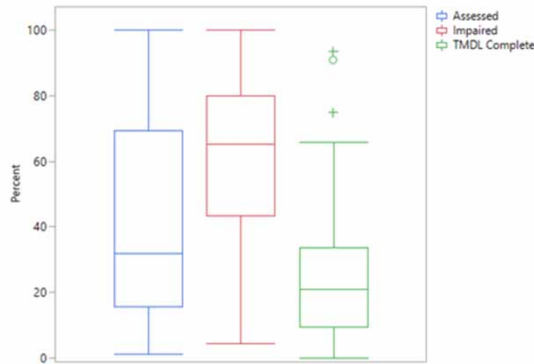
Socioeconomic indicators



U.S. EPA regions



Predictions of TMDL progress



1. INTRODUCTION

Nonpoint source pollution is one of the greatest threats to water quality in streams and rivers across the world (Brown & Froemke, 2012; du Plessis, 2022). Within the United States, the quality of most freshwater bodies is regulated through the Clean Water Act (CWA). Section 303(d) of the CWA requires states to assess their water bodies and, for those that are identified as impaired, establish total maximum daily loads (TMDLs) (US Congress, 1972). A TMDL is defined as the maximum amount of a pollutant allowed to enter a water body so that it will meet water quality standards for that particular pollutant and designated use (i.e., fishable, swimmable, and/or drinkable) (US Congress, 1972). This program is important as it is the statutory method to address nonpoint source pollution, which the US Environmental Protection Agency (EPA) claims is the leading remaining cause of water quality problems (USEPA, 2022b), with over 610,000 miles of rivers and streams listed as impaired (USEPA, 2017). To address nonpoint source pollution, the TMDL program relies on a system of policy implementation where the federal government delegates primary authority to agencies at the state level to execute the TMDL development process. The TMDL development process has three distinct steps: (1) assessment of surface water bodies; (2) listing impaired water bodies; and (3) development of TMDLs to impaired water bodies.

Once TMDLs are developed, they could be included in watershed management plans and implemented to meet these criteria by reducing both point source and nonpoint source pollution. Point source pollution is limited through National Pollutant Discharge Elimination System (NPDES) permits, which are regulated by Section 402 of the CWA. However, no federal regulation or authority exists to enforce pollutant reductions against nonpoint sources, including the enforcement of implementing TMDLs to achieve these target loads. Rather, since the implementation of pollutant controls is outside of the EPA's statutorily defined authority, the implementation of TMDLs is done by the states, often through incentives to polluters (Copeland, 2014; Lichtenberg, 2019). For example, the CWA section 319 grant programs provide funding for nonpoint source pollution controls when the control of point source pollution through NPDES alone does not achieve TMDL goals (Jones, 2014). TMDL development is the responsibility of state agencies; therefore, the extent to which states assess water bodies, identify impairments, and develop TMDLs can vary from state to state. It has been over 30 years since the EPA published regulations establishing TMDL requirements in 1992 (Houck, 1997) and therefore presents an opportune time to reflect on the progress of the TMDL programs across the United States.

This reflection is valuable in understanding what factors may contribute to advancing or prohibiting the progress of TMDLs. The US states vary in their populations, geography, demographics, and economic output, which may influence both the approach and extent to which they enact water quality programs. To that end, there is a diversity of approaches to assessing and implementing TMDLs through monitoring and modeling technologies. This includes assessment methods to monitor water quality parameters that range from detailed monitoring requiring scientific expertise, to those that can be carried out by volunteer citizens (Loperfido *et al.*, 2010; Nation & Johnson, 2016; Brett, 2017; Webster & Dennison, 2022). While technical approaches to assessing water bodies and applying models to develop TMDLs are well defined and available for use (Zhang & Quinn, 2019), it is less clear the extent to which different states have implemented TMDL programs and the factors that contribute to their progress. This is important because understanding the factors that impact the progress of protecting the nations' water bodies can help to improve TMDL methods and ultimately reach the designated goal of obtaining fishable, swimmable, and/or drinkable waters.

However, the extent to which socioeconomics or the geography of a state influences the implementation of TMDLs is unclear. This gap in understanding is critical as socioeconomics, land use, and environmental dynamics have been shown to be important factors in developing effective TMDL plans (Mirchi & Watkins, 2012) and in implementing other aspects of the CWA, such as the NPDES permit program (McDonald & Naughton, 2019) that seeks to address water quality impairments. Water quality impairments have also been shown to have a relationship with socioeconomic factors, such as a negative impact correlation to housing prices (Papenfus, 2019), as well as significant correlations with education, ethnic composition, age structure, and population density (Farzin & Grogan, 2013). In addition, the economics of a region could influence the degree to which the public influences management decisions as per capita income is positively correlated with environmental concerns (Andrew *et al.*, 2019).

Beyond socioeconomic effects, there may also be regional or spatial similarities among states in their approach to managing water quality due to common geomorphologic or climatic similarities, or their presence in similar EPA regions. The US EPA is divided into 10 different regions with a regional office associated with each that is responsible for enforcing implementation of TMDL programs within their states. This regionalization of enforcement offices may therefore lead to different levels of program development, as these regions have largely developed independently due to different factors such as water body priorities, litigation, and resource availability (Neilson & Stevens, 2002). However, no studies have evaluated the relationship between socioeconomics, regionalization, and TMDL implementation in the United States. Therefore, efforts to manage water quality through the TMDL program may be impacted by the socioeconomic and spatial factors of a watershed. Understanding

these relationships could allow watershed managers and planners to better allocate resources and guide efforts to improve upon TMDL development programs.

The objectives of this study are to analyse the state-level TMDL progress across the United States and evaluate its relationship to socioeconomic and spatial factors. To do so, data were gathered from the EPA Assessment and Total Maximum Daily Load Tracking and Implementation System (ATTAINS) database for each state, including TMDL progress and impairment causes, and then analyzed to identify spatial trends or other clusters in state-level TMDL progress, including the percentage of streams that have been assessed, the percent of assessed streams that are impaired, and those for which a TMDL is complete. In addition, US Census (gross domestic product (GDP), educational attainment levels, etc.) and regionalization (EPA region, climatic region, etc.) data were collected for each state. These data were then used to explore how socioeconomic factors and spatial relationships correlate to markers of TMDL progress. Ultimately, this study can help to contextualize progress in TMDL development and help in our understanding of the influence that socioeconomic or regionalization may have on the implementation of water quality programs.

2. METHODOLOGY

This section presents the methodology to analyse the state-level TMDL progress across the United States and evaluate its relationship to socioeconomic and spatial factors. Section 2.1 presents the methods and approaches of data collection, Section 2.2 presents data preparation, and Section 2.1 presents data analysis.

2.1. Data collection

TMDL assessment data were extracted from EPA ATTAINS archived data (US EPA, 2022a). For each state, ATTAINS includes data on the status of TMDL for surface water bodies, including the miles of water body that are assessed, miles unassessed, specific impairment, miles impaired, and miles with a completed TMDL. From these data, we derived indicators of TMDL progress including (1) the percent of streams and rivers that have been assessed; (2) the percentage of the assessed streams and rivers that are listed as impaired; and (3) the percentage of impaired streams and rivers for which a TMDL has been completed.

In addition, US Census data were obtained for each state including the population, GDP, educational attainment, and per capita earnings (U.S. Census Bureau, 2019). Finally, other geographic information including state land area, elevation (highest and lowest elevation), and average annual rainfall were derived using GIS analysis and datasets listed in Table SI-1. From these data, elevation gain and average geographical slopes were derived as they have been shown to be a significant factor impacting stream water quality (Connolly *et al.*, 2018). Remote land percentage was derived from data from the Economic Research Service (Cromartie & Nulph, 2019), which defines zip-code level land areas as frontier and remote (FAR) based on population and travel times from population centers. Specifically, a FAR area is defined as an area that is 60 min or more from an urban area of 50,000 or more people (Cromartie & Nulph, 2019). The FAR area is meant to represent the population's access to services, and in the case of assessing streams, it may be that the more remote areas the state has, the more effort and resources it takes to travel to and assess stream quality. Furthermore, remoteness may indicate the level of interest in better water quality among citizens, as water quality perceptions have been linked to income level and location (Andrew *et al.*, 2019). All datasets and sources are listed in Table SI-1.

2.2. Data preparation

Once data were obtained from ATTAINS, they were processed to ensure consistency across each state. Because TMDLs are implemented at the state level, each state operates independently and may have different ways in

which they classify pollutants, perform monitoring and testing, and report their results. This can make it challenging to summarize data across all states due to differences in methodology and reporting. For example, in evaluating the suspended matter in water, states may list this as total suspended solids or turbidity. Therefore, to enable comparisons among states, the data were aggregated into common categories as shown in Table SI-2.

2.3. Data analysis

To analyze the data, three types of analyses were performed: summary statistics, clustering, and linear regression. To summarize data across all states, descriptive statistics (mean, median, and standard deviation) were performed on the TMDL and socioeconomic data. Two types of clustering were used: k-means clustering for TMDL progress and hierarchical clustering among socioeconomic variables. Clustering is a common method to allocate objects (e.g., states) with multiple datasets (e.g., water quality parameters of each state) into groups that have similar attributes across datasets (Javadi *et al.*, 2017). For TMDL progress, k-means clustering (Hartigan & Wong, 1979) was used to identify if states had any common groupings considering the percent of streams that were assessed, those that were impaired, and those in which a TMDL was completed. K-means clustering is one of the most commonly used clustering methods, and it can be used to generate an optimally defined number of clusters (Ali & Kadhum, 2017). The hypothesis for k-means cluster analysis was that, given an optimal k-value (number of clusters), the dataset would reveal meaningful groupings that could represent the progress of the TMDL process. By using k-means clustering, we sought to identify states whose TMDL progress varied between high and low values across all three variables (i.e., $k = 3$).

Alternatively, socioeconomic data included numerous variables, and therefore, a hierarchical clustering was applied based on Euclidean distance and complete linkage method (Murtagh & Contreras, 2012). Unlike the k-means method, hierarchical clustering does not require prior knowledge of the number of clusters. It is a stepwise clustering method that merges the most similar data points together into groups at each level. One objective of using this method to analyse socioeconomic data between states is to determine how clusters might be associated geographically or spatially. Once clustering was complete, we observed that, spatially, many of the k-means and hierarchical clusters appeared to mirror the boundaries of the EPA regions or a collection of them. Therefore, we performed descriptive statistics on both the clusters described earlier as well as the specific EPA regions a state fell within and performed analysis of variance (ANOVA) tests to determine the statistical significance of these groupings.

Finally, linear regressions, both multiple and simple, were performed to find statistical relationships between dependent (percentage of TMDL completed, percentage of assessed streams, etc.) and independent variables (socioeconomic variables and regional clusters) (Table 1). The relationship generated by the regression can be represented as follows:

$$y = b_0 + b_1x_1 + \dots + b_nx_n + e \quad (1)$$

where y represents the dependent variable, b_0 represents the intercept, and $b_{(1-n)}$ represents the independent variables represented by $x_{(1-n)}$. This was performed across all data, as well as subsets of the data based on the specific impairment type (e.g., nutrients, sediment, bacteria). All regression models were checked for assumptions of linear regression including normality of the residuals, multicollinearity among independent variables, and homoscedasticity and normality of the residuals. Stepwise model selection using bidirectional elimination was carried out for multiple linear regression based on a significance level to entry of 0.10 and a significance level to stay of 0.05. The parameters used for regression analysis are given in Table 2. The hypothesis of the regression analysis is that the socioeconomic parameters have a significant relationship to the TMDL progress parameters.

Table 1 | Parameters used in regression analysis of the TMDL study.

Dependent variable	Independent variable
Assessed (%)	Total stream miles
Impaired (%)	Stream density
TMDL complete (%)	State land area (miles)
Assessed miles	Average area rainfall
Impaired miles	Elevation gain feet
TMDL complete miles	Average slope
	Population
	Population density
	GDP (billion USD)
	No high school diploma percentage
	Bachelor's degree or higher percentage
	Per capita earnings
	Median household income
	Remote area percentage

Table 2 | Single variable regression analysis for all relationships with a $p < 0.1$.

Dependent (y)	Independent (x)	Equation	p-Value	R ²
Assessed (%)	Total stream length (mi)	$y = 64.3 - 0.0003 \times x$	0.002	0.19
Assessed (%)	State land area (km ²)	$y = 59.1 - 0.0003 \times x$	0.004	0.16
Assessed (%)	Median household income (\$)	$y = -31.4 + 0.0011 \times x$	0.012	0.13
Assessed (%)	Population density (pop/mi ²)	$y = 33.3 + 0.0410 \times x$	0.017	0.12
Assessed (%)	Per capita earnings (\$)	$y = -18.6 + 0.0013 \times x$	0.050	0.08
Assessed (%)	Bachelor or higher degree (%)	$y = -10.1 + 1.85 \times x$	0.053	0.08
Assessed (%)	No high school diploma (%)	$y = 73.2 - 2.50 \times x$	0.088	0.06
TMDL complete (%)	No high school diploma (%)	$y = 4.8 + 1.687 \times x$	0.099	0.06

3. RESULTS AND DISCUSSION

The results are presented on the TMDL data itself in Section 3.1, including a summary of indicators of TMDL progress and impairment types, K-means clustering of these indicators, and the impact of regionalization on indicators of TMDL progress. In addition, results are presented in Section 3.2 that summarizes the socioeconomic data through summary statistics and hierarchical clustering. Finally, the relationship between indicators of TMDL progress and socioeconomic factors is presented through single and multivariable regression in Section 3.3.

3.1. TMDL data

3.1.1. Indicators of TMDL progress

Within the ATTAINS database, TMDL data for each state are reported as follows: (1) the percent of streams and rivers that have been assessed; (2) the percentage of the assessed rivers that are listed as impaired; and (3) the percentage of impaired rivers for which a TMDL has been completed. A distribution of these statistics across all states is illustrated in Figure 1. The wide range of the whiskers for all three of these parameters highlights the extent of the variation in the progress of states in implementing the TMDL program. For example, the mean percent of streams with assessments complete is 43% with a standard deviation of 33.1%, illustrating a vast range in the percentage of streams that have been assessed. Of those assessed water bodies, on average, 61% were impaired, and of those impaired, on average, 25% had a TMDL completed.

3.1.2. Impairment types

We also sought to determine the extent to which the impairment type might affect the statistical relationship between indicators of TMDL progress (e.g., percent of streams that are impaired) and socioeconomic and spatial variables. Therefore, the data were categorized based on the impairment of the streams themselves. The ATTAINS database ranks the top impairments within each state and lists the total miles of streams impaired for each pollutant or impairment cause. The categorization of the top four impairments across all states is illustrated in Figure 2. In this figure, a first-level impairment is the largest impairment cause in terms of miles of rivers and streams that are impaired within an individual state. As illustrated, bacteria are the most frequent first-level and second-level impairment across the nation and are followed as the top impairment by temperature, sediments, and nutrients.

Figure 3(a) further illustrates the top impairments across the United States and represents the number of states for which an impairment appears at any level. As illustrated, bacteria, which were the most frequent level-one impairment, are also the most frequent impairment overall at any level. Interestingly, while temperature is a

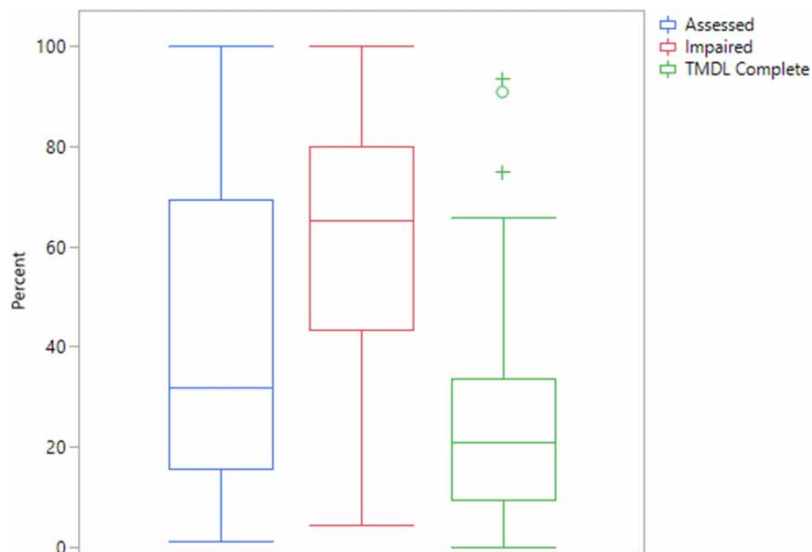


Fig. 1 | Distribution of the percent of streams within a state that have been assessed, listed as impaired, and have had a completed TMDL.

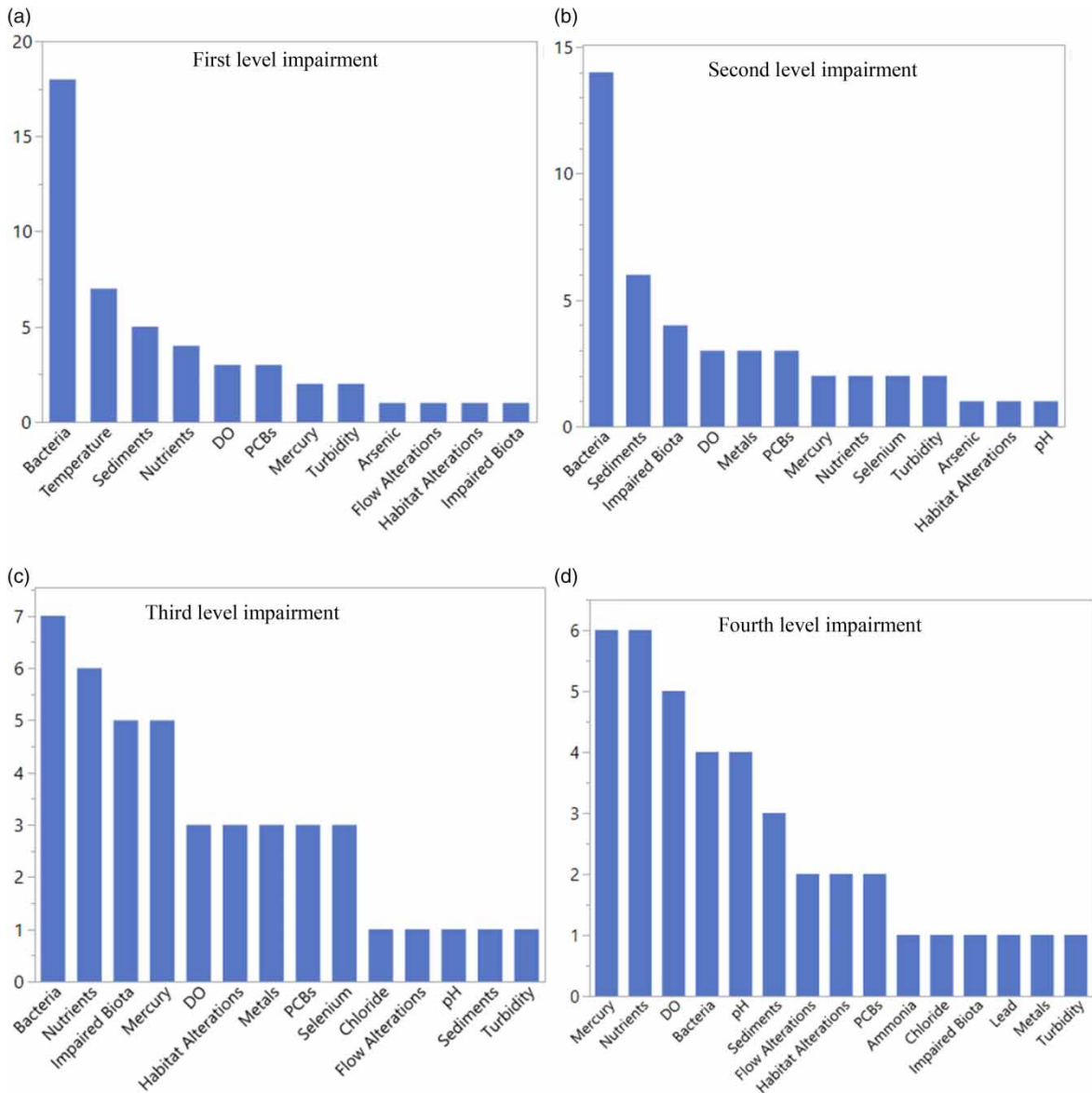


Fig. 2 | Summary of the top four impairments across each state with the number of states on the y-axis.

level 1 impairment in 7 out of 49 states, it only shows up in 15 states in total as an impairment. Finally, Figure 3(b) represents the distribution of the percent of assessed streams that are listed as impaired for each individual pollutant. As illustrated, bacteria are the largest impairment with a median of 20% of assessed streams having a bacteria impairment cause.

3.1.3. K-means clustering

To determine if there were any groupings in the progress of TMDLs among states, we performed K-means clustering, considering the three variables in Figure 1, and three distinct groups were identified. Figure 4(a) illustrates

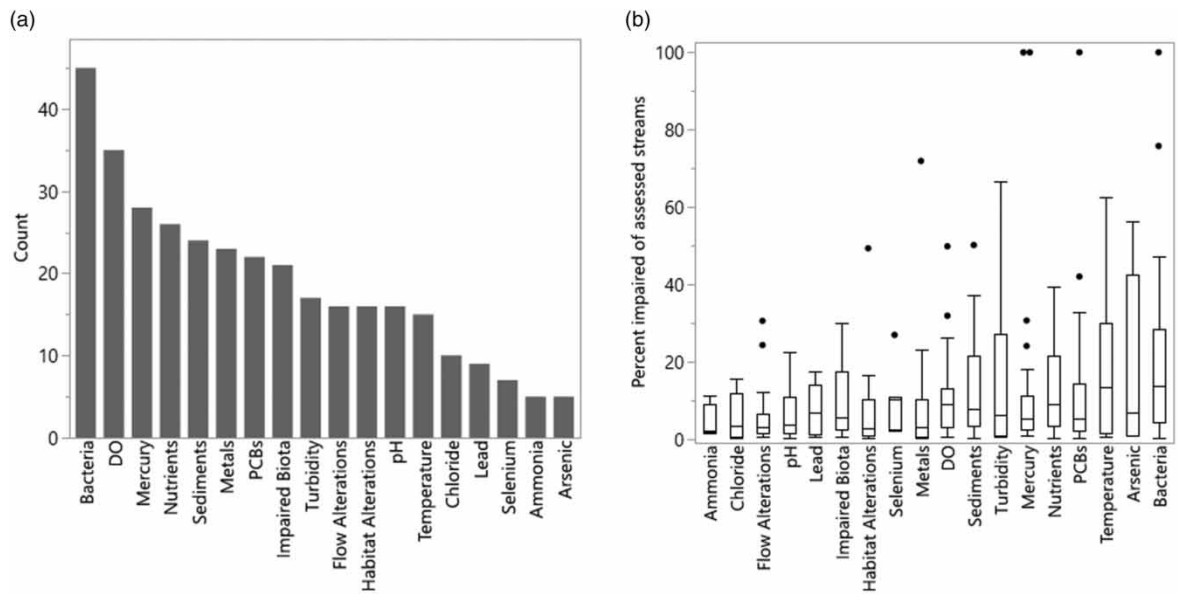


Fig. 3 | (a) Total number of states with each impairment cause and (b) distribution of the percent of assessed streams that are listed as impaired within each state for each impairment cause.

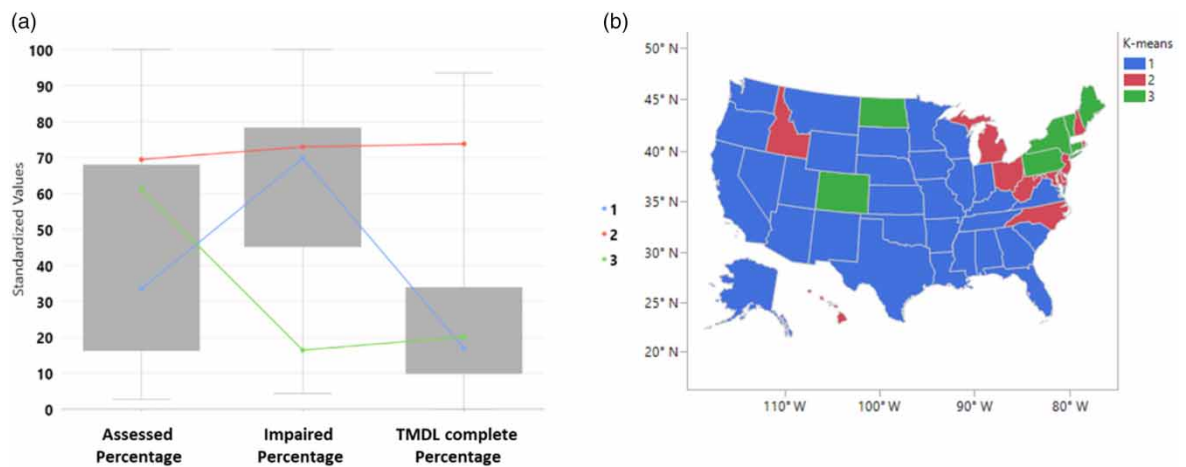


Fig. 4 | (a) K-means clustering of assessed, impaired, and TMDL complete percentages, and (b) spatial representation of clusters. Groups 1 (blue), 2 (red), and 3 (green) represent clusters of states that behave similarly in terms of TMDL progress.

the distribution of the indicators of TMDL progress with three distinct clusters identified using k-means clustering. As illustrated, there is a cluster of states that have assessed a large percentage of streams, but have found those streams to be unimpaired (green). The second cluster represents a group of states that have assessed most of their streams and found them to be impaired (red). The remaining clusters are those to which the percentage of streams that are assessed is largely below 50%, but for which their impairments are varied (blue). There are several states

that are outliers, such as New Hampshire, which with 16,962 miles of streams assessed, has assessed 100% of their streams, found 100% of them to be impaired, and completed TMDLs on 75% of their streams. Alaska has assessed 4,409 miles of streams representing only 1.2% of their total streams (the lowest percentage of all states), finding 9.4% to be impaired, and completing a TMDL on none of their streams. A spatial distribution of these clusters is illustrated in Figure 4(b), which demonstrates that there are some states, especially within the northeast, that appear to cluster together.

3.1.4. EPA regionalization

To further explore potential spatial relationships, the TMDL data were categorized into EPA regions (Figure 5) to identify any common trends in TMDL progress among these regions. Figure 5 illustrates the distribution of the



Fig. 5 | (a) Assessed percentage, (b) impaired percentage, (c) TMDL completed percentage for streams by EPA region, and (d) EPA regions.

percent of streams assessed, streams impaired, and those streams with a TMDL complete. As illustrated, there is noticeable variance among all categories within EPA regions; therefore, we applied a grouped ANOVA and found a p -value of 0.005 for assessed streams, suggesting that some EPA regions are significantly different than others based upon the percent of streams they have assessed. Furthermore, using paired ANOVA, it was observed that Region 6 is significantly different from Regions 1 and 3 ($p < 0.05$), and Region 7 is significantly different from Region 3 ($p < 0.05$). This may be because both Regions 6 and 7 have assessed less than 25% of their streams.

3.2. Socioeconomic data

3.2.1. Summary of socioeconomic data

Socioeconomic data were collected from the US Census for each state, and the distribution of selected variables is illustrated in Figure 6. This figure demonstrates the variation within the land area (Figure 6a), population (Figure 6b), GDP (Figure 6c), percent with no high school diploma (Figure 6d), percent of population with a bachelor's degree or higher (Figure 6e), and per capita earnings (Figure 6f). As illustrated, there is a large degree of variation in socioeconomic variables with a few large outliers for several of the parameters. A complete table of socioeconomic data distributions is provided in Table SI-2.

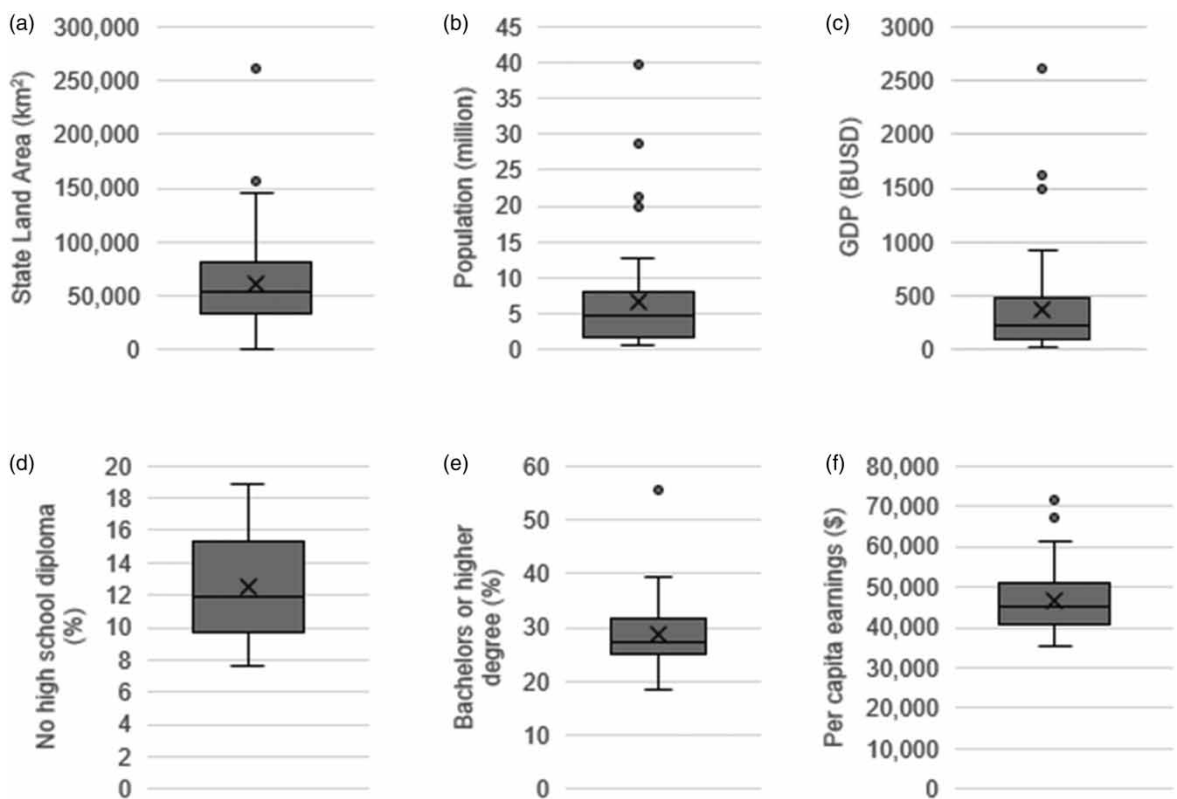


Fig. 6 | Boxplot distribution of socioeconomic data: (a) land area, (b) population, (c) GDP, (d) percent with no high school diploma, (e) percent of population with a bachelors degree or higher, and (f) per capita earnings.

3.2.2. Hierarchical clustering

To determine if there were any groupings among states with common socioeconomic variables, we performed hierarchical clustering as illustrated in Figure 7. This figure represents various levels of clusters, with the states color coded based on their EPA region. For simplicity, the most northeast EPA regions (1, 2, and 3), west EPA regions (8, 9, and 10), and southwest/great plains regions (6,7) are coded the same color. As illustrated, Texas and California are grouped together as a single cluster, as they have the largest land areas, population, GDP, and percent population without high school diplomas. The District of Columbia is the smallest region considered and has the highest population with bachelor’s degrees or higher (55.4%); therefore, it stands apart as its own second-level cluster. Finally, notably, most of the northeastern states are also clustered together. Given these apparent regional clusters, we sought to identify the influence of socioeconomic factors on TMDL compliance as outlined in the following section.

3.3. Relationship between TMDLs and socioeconomic

Response screening was performed to determine if there are any correlations among TMDL progress and socioeconomic variables across all states. The results in Table 2 indicate that the percent of streams that are assessed is negatively correlated to total stream length ($R^2 = 0.19$) and state land area ($R^2 = 0.16$), both of which explain the largest degree of variability in percent of stream that are assessed. This suggests that the larger the land area and length of streams to assess, the lower the percentage of streams that are actually assessed, likely due to the sheer size and the length of streams and rivers within the state. In addition, the percent of streams that are assessed is positively correlated to percent of population with a bachelor’s degree or higher, percent of population with no high school diploma, and per capita earnings ($p < 0.05$). These positive correlations with median household income and bachelor’s degree or higher suggest that the economic output of a state has an influence on the percentage of streams that are assessed. This implies that state economic resources are a contributing factor to TMDL progress. Funding for the development of TMDLs mainly comes from state taxes and to a certain

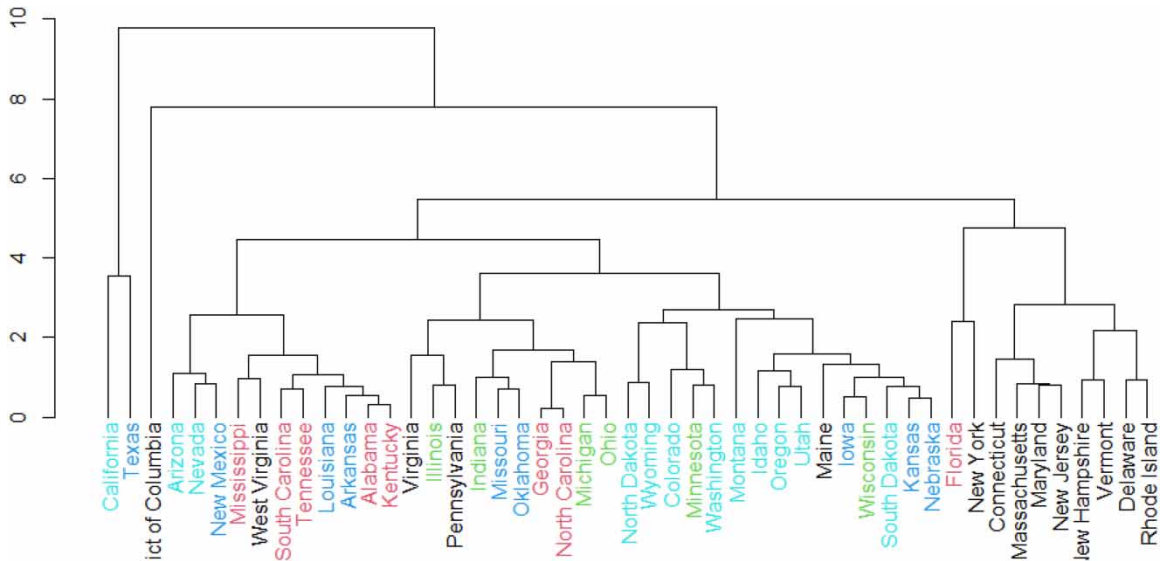


Fig. 7 | Cluster analysis of socioeconomic data color coded by EPA regions: black, 1, 2, 3; red, 4; green, 5; blue, 6, 7; light blue, 8, 9, 10.

degree, from nonprofit organizations directly involved with the water quality concerns of the regions. Therefore, per capita earnings have a direct relationship to these processes.

Finally, the percent of TMDLs that are complete has a low positive correlation ($p < 0.1$) to percentage of population without a high school diploma. These trends are opposite to what is observed with the percent that are assessed and is unclear how these variables would act in such a way to influence the completion of TMDLs. However, despite having a moderate significance of the slope, it also has the lowest model fits ($R^2 = 0.06$), suggesting that it only explains a small portion of the variability in the percent of TMDLs that are complete. Finally, there were no correlations between any socioeconomic and spatial variables to the percentage of streams that are impaired when categorized based on the specific impairments themselves.

To understand how multiple variables might better predict TMDL progress, we performed multivariable linear regression using a stepwise approach. Results demonstrated that the prediction of the percentage of assessed streams could be improved by combining the total stream length and median household income, with an adjusted R^2 of 0.29 (Table 3). No multivariable equations were statistically significant for the percent of impairments or TMDLs completed.

To evaluate how these particular variables might be influenced by the EPA region, we plotted the mean TMDL progress variables (assessed, impaired, and TMDL complete percentages) against the mean socioeconomic variables for each EPA region. In doing so, we found a relationship between the mean percent of assessed streams and per capita earnings of each EPA region as illustrated in Figure 8(a) and 8(b). Figure 8(a) is plotted in a

Table 3 | Multiple linear regression model output for percent of streams that are assessed.

Variables	Estimate	p-Value
Intercept	0.00027	–
Total stream length (mi)	–0.00030	0.003
Median household income (USD)	0.00097	0.025

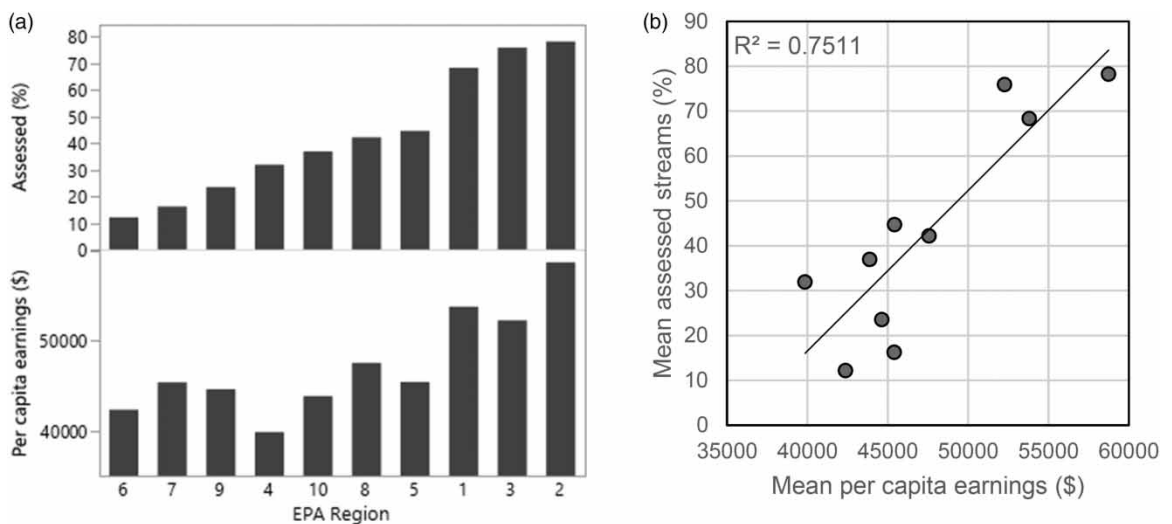


Fig. 8 | (a) Comparison of mean assessed miles and per capita earnings in each EPA region and (b) their correlation.

descending order based on the mean per capita earnings, and as illustrated, the percentage of assessed streams follows a similar trend. This indicates that the resources available within each region may have an influence on the progress of TMDL implementation within the states. This could be due to the distribution of resources at the regional level. Federal funding is dispersed to regions for the implementation of TMDL programs and is dependent on revenue from point source discharges within the region, among other sources (Neilson & Stevens, 2002).

4. DISCUSSION

This study presents TMDL progress across the country and the influence that spatial and socioeconomic factors have on the percent of streams that are assessed, those that are impaired, and those for which a TMDL is complete. The outcomes of this work demonstrate that not all states have implemented their TMDL programs to the same degree. While some of this can be attributed to the difference in state land areas and stream miles that must be assessed, there are other socioeconomic factors that have a similar degree of explanatory power.

To that end, median household income had a significant relationship with the percentage of streams that are assessed within a state. Without financial resources, states may not have the necessary personnel to collect the data that are needed for assessing streams. It could be that this lack of funding inhibits progress and should be addressed by permitting states to calculate TMDLs with alternative and less costly assessment methods, such as using proxies (e.g., impervious cover or stormwater volume) that are correlated with water quality (DeGioia, 2019). Other policy changes to advance TMDL progress could include more direct obligations to limit pollution with discretion and flexibility on how to do so, rather than whether and how much to do so (Stephenson *et al.*, 2022). In addition, there may be an economic case that there are returns to investments to improve water quality. These returns can be quantified through economic models to determine the cost and benefits of water quality improvements; however, uncertainties in these models regarding pollution damages and economic benefits make them difficult to apply (Bosch *et al.*, 2006).

Public participation, which is mandated through the CWA, is an important part of the TMDL development process as it can provide information regarding impairments, collect water quality data, and review and comment on impairment lists and TMDL drafts. This study shows that education, both population with college degrees and population without high school diplomas, has a considerable impact on the assessment of streams and rivers. Efforts to improve public awareness of water quality issues and their implications through improved communication methods and technologies may improve their participation in various stages of the TMDL process (Quinn *et al.*, 2022). Furthermore, it has been shown that citizen scientist programs not only allow states to collect water quality data but also to educate socioeconomically underprivileged communities through information dissemination (Webster & Dennison, 2022).

It is also clear that EPA regions themselves tend to have a similar level of TMDL progress. Some of this can be attributed to geographic similarities as demonstrated in the clustering analysis; however, as Figure 8 indicates, there could be economic factors common to certain regions that influence TMDL progress. EPA regions having a significant influence on the implementation of programs are not surprising as the regional offices are charged with enforcing and overseeing state adherence to the CWA. The findings suggest that progress towards assessing streams and completing TMDLs varies depending on the EPA region, with those with higher amounts of per capita earnings having a larger percentage of their streams assessed. This is aligned with other studies that have found similar differences among EPA regions in implementing elements of the CWA, including the NPDES program (Woods, 2021).

One limitation of this study is that we do not evaluate the methodology for which streams are assessed, which could also vary depending on what resources are available. For example, a robust assessment would have large

amounts of data that demonstrate an impairment, while others could have a 'drive by' assessment (Neilson & Stevens, 2002). Furthermore, some states may list a large number of impairments due to a large amount of data, while others could have impaired streams that are not listed due to a lack of assessment data. As a matter of practical implementation, states with more streams have a greater workload, and therefore, the results indicate that these states have not assessed as many streams as a percentage of their overall stream lengths.

This study demonstrates the variability in TMDL progress across states and what factors may contribute to it; however, how to bridge the gap between states is unclear. As the TMDL programs continue to mature, modeling approaches and technologies may accelerate assessment and TMDL development for waters of the United States. For example, modeling approaches have evolved over time, and different impairment types may require different modeling tools (Quinn *et al.*, 2019b), which can be selected based on the unique technical criteria and management constraints of a watershed (Sridharan *et al.*, 2021). Furthermore, advancements in remote sensing and geospatial analysis can be used to support TMDL assessment and modeling (Quinn *et al.*, 2019a; Sridharan *et al.*, 2022). In addition, there may be ways to further improve the process of developing TMDLs with modelers, stakeholders, and regulatory entities. This is important as in many cases, the TMDL serves as the framework for contextualizing watershed science and regulatory policies toward stakeholders and the general public through collaboration and coordination of watershed management (Slota, 2021).

Finally, the outcomes of this study can be used by water resource decision makers to improve their approach to the TMDL development process mandated by the federal government. For example, for states or regions with higher economic capacities but low education levels, it might be efficient to reallocate resources to educate communities about the need for public participation in the TMDL process. Furthermore, while the TMDL is a federally mandated national requirement, this study demonstrates that progress toward meeting TMDL requirements differs among EPA regions. Therefore, it may be beneficial to investigate how to reallocate resources to EPA regions such that these discrepancies could be alleviated. To that extent, the use of new federal funding methods and regulations could be explored.

5. CONCLUSIONS

This study evaluated the relationship that socioeconomics and regionalization have on TMDL progress and implementation. Outcomes indicate that TMDL progress and impairments had a large degree of variation, some of which could be explained by EPA region, spatial clustering, and socioeconomic variables. To that end, results suggest that the size of a state, the length of total streams, and the economic output are related to the percentage of streams that are assessed within a state. In addition, states largely followed similar patterns based on the EPA region that they were within, indicating that regions play a large role in TMDL progress. The outcomes of this study can be used by water resource decision makers to improve their approach to the TMDL development process mandated by the federal government, including the targeted allocation of resources among and within states. Overall, this study highlights the diversity in the implementation of TMDLs across states and highlights some of the factors that may explain variation in TMDL approaches to date.

DATA AVAILABILITY STATEMENT

All relevant data are included in the paper or its Supplementary Information.

CONFLICT OF INTEREST

The authors declare there is no conflict.

REFERENCES

- Ali, H. H. & Kadhum, L. E. (2017). K-means clustering algorithm applications in data mining and pattern recognition. *International Journal of Science and Research (IJSR)* 6(8), 1577–1584.
- Andrew, R. G., Burns, R. C. & Allen, M. E. (2019). The influence of location on water quality perceptions across a geographic and socioeconomic gradient in Appalachia. *Water* 11(11), 2225. <https://doi.org/10.3390/W11112225>.
- Bosch, D. J., Ogg, C., Osei, E. & Stoecker, A. L. (2006). Economic models for TMDL assessment and implementation. *Transactions of the ASABE* 49(4), 1051–1065. <https://doi.org/10.13031/2013.21744>.
- Brett, A. E. (2017). Putting the public on trial: Can citizen science data be used in litigation and regulation? *Villanova Environmental Law Journal* 28(2), 1–44.
- Brown, T. C. & Froemke, P. (2012). Nationwide assessment of nonpoint source threats to water quality. *BioScience* 62(2), 136–146. <https://doi.org/10.1525/BIO.2012.62.2.7>.
- Connolly, C. T., Khosh, M. S., Burkart, G. A., Douglas, T. A., Holmes, R. M., Jacobson, A. D., Tank, S. E. & McClelland, J. W. (2018). Watershed slope as a predictor of fluvial dissolved organic matter and nitrate concentrations across geographical space and catchment size in the Arctic. *Environmental Research Letters* 13(10), 104015. <https://doi.org/10.1088/1748-9326/AAE35D>.
- Copeland, C. (2014). Clean Water Act and pollutant total maximum daily loads (TMDLs). *Congressional Research Services Report, R42752*, 1–28.
- Cromartie, J. & Nulph, D. (2019). *USDA ERS - 2010 Frontier and Remote (FAR) Area Codes Documentation*. USDA, Washington, DC. <https://www.ers.usda.gov/data-products/frontier-and-remote-area-codes/documentation/>. Accessed Nov. 2023.
- DeGioia, M. (2019). Overboard: The complexity of traditional TMDL calculations under the clean water Act. *Environmental Law Reporter News & Analysis* 49. Available at: <https://heinonline.org/HOL/Page?handle=hein.journals/elrna49&id=1206&div=&collection=>.
- du Plessis, A. (2022). Persistent degradation: Global water quality challenges and required actions. *One Earth* 5(2), 129–131.
- Farzin, Y. H. & Grogan, K. A. (2013). Socioeconomic factors and water quality in California. *Environmental Economics and Policy Studies* 15(1), 1–37. <https://doi.org/10.1007/s10018-012-0040-8>.
- Hartigan, J. A. & Wong, M. A. (1979). Algorithm AS 136: A k-means clustering algorithm. *Journal of the Royal Statistical Society. Series c (Applied Statistics)* 28(1), 100–108.
- Houck, O. A. (1997). TMDLs: The resurrection of water quality standards-based regulation under the Clean Water Act. *Envtl. L. Rep. News & Analysis* 27, 10329.
- Javadi, S., Hashemy, S. M., Mohammadi, K., Howard, K. W. F. & Neshat, A. (2017). Classification of aquifer vulnerability using K-means cluster analysis. *Journal of Hydrology* 549, 27–37. <https://doi.org/10.1016/J.JHYDROL.2017.03.060>.
- Jones, S. (2014). Making regional and local TMDLs work: The Chesapeake Bay TMDL and lessons from the Lynnhaven River. *William & Mary Environmental Law and Policy Review* 38(2), 277.
- Lichtenberg, E. (2019). Conservation and the environment in US farm legislation. *EuroChoices* 18(1), 49–55.
- Loperfido, J. V., Beyer, P., Just, C. L. & Schnoor, J. L. (2010). Uses and biases of volunteer water quality data. *Environmental Science and Technology* 44(19), 7193–7199. <https://doi.org/10.1021/ES100164C>.
- McDonald, W. M. & Naughton, J. B. (2019). Stormwater management actions under regulatory pressure: A case study of southeast Wisconsin. *Journal of Environmental Planning and Management* 0(0), 1–22. <https://doi.org/10.1080/09640568.2018.1539391>.
- Mirchi, A. & Watkins Jr., D. (2012). A systems approach to holistic total maximum daily load policy: Case of lake Allegan, Michigan. *Journal of Water Resources Planning and Management* 139(5), 544–553. [https://doi.org/10.1061/\(ASCE\)WR.1943-5452.0000292](https://doi.org/10.1061/(ASCE)WR.1943-5452.0000292).
- Murtagh, F. & Contreras, P. (2012). Algorithms for hierarchical clustering: An overview. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 2(1), 86–97.
- Nation, T. H. & Johnson, L. A. (2016). Use of a volunteer monitoring program to assess water quality in a TMDL watershed utilized for recreational Use, Pickens County, South Carolina. *Journal of South Carolina Water Resources* 2(1), 11. <https://doi.org/10.34068/JSCWR.02.02>.
- Neilson, B. T. & Stevens, D. K. (2002). Issues related to the success of the TMDL program. *Journal of Contemporary Water Research and Education* 122(1), 8.

- Papenfus, M. (2019). Do housing prices reflect water quality impairments? Evidence from the Puget Sound. *Water Resources and Economics* 27, 100133.
- Quinn, N. W. T., Kumar, S. & Imen, S. (2019a). Overview of remote sensing and GIS uses in watershed and TMDL analyses. *Journal of Hydrologic Engineering* 24(4), 02519002. [https://doi.org/10.1061/\(asce\)he.1943-5584.0001742](https://doi.org/10.1061/(asce)he.1943-5584.0001742).
- Quinn, N. W. T., Kumar, S., La Plante, R. & Cubas, F. (2019b). Tool for searching USEPA's TMDL reports repository to analyze TMDL modeling state of the practice. *Journal of Hydrologic Engineering* 24(9), 1–11. [https://doi.org/10.1061/\(asce\)he.1943-5584.0001805](https://doi.org/10.1061/(asce)he.1943-5584.0001805).
- Quinn, N. W. T., Sridharan, V., Ramirez-Avila, J., Imen, S., Gao, H., Talchabhadel, R., Kumar, S. & McDonald, W. (2022). Applications of GIS and remote sensing in public participation and stakeholder engagement for watershed management. *Socio-Environmental Systems Modelling* 4, 18149. <https://doi.org/10.18174/sesmo.18149>.
- Slota, S. C. (2021). Bootstrapping the boundary between research and environmental management: The TMDL as a point of engagement between science and governance. *Science Technology and Human Values* 47(4), 750–773. <https://doi.org/10.1177/016224392111026364>.
- Sridharan, V. K., Quinn, N. W. T., Kumar, S., McCutcheon, S. C., Ahmadisharaf, E., Fang, X., Zhang, H. X. & Parker, A. (2021). Selecting reliable models for total maximum daily load development: Holistic protocol. *Journal of Hydrologic Engineering* 26(10), 04021031. [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0002102](https://doi.org/10.1061/(ASCE)HE.1943-5584.0002102).
- Sridharan, V. K., Kumar, S. & Kumar, S. M. (2022). Can remote sensing fill the United States' monitoring gap for watershed management? *Water (Switzerland)* 14(13), 04021031. <https://doi.org/10.3390/w14131985>.
- Stephenson, K., Shabman, L., Shortle, J. & Easton, Z. (2022). Confronting our agricultural nonpoint source control policy problem. *Journal of the American Water Resources Association* 58(4), 496–501. <https://doi.org/10.1111/1752-1688.13010>.
- U.S. Census Bureau (2019). *American Community Survey 5-Year Estimates*. Available at: <http://factfinder2.census.gov>
- US Congress (1972). Federal Water Pollution Control Act.–33 USC § 1251 et seq. USA. *Search In*.
- USEPA (2017). *National Water Quality Inventory: Report to Congress, EPA 841-R-16-011*. August. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/2347274>.
- US EPA (2022a). *The Assessment, Total Maximum Daily Load (TMDL) Tracking and Implementation System (ATTAINS)*. Available at: <https://www.epa.gov/waterdata/attains>.
- USEPA (2022b). *Basic Information About Nonpoint Source (NPS) Pollution*. Available at: <https://www.epa.gov/nps/basic-information-about-nonpoint-source-nps-pollution>.
- Webster, S. E. & Dennison, W. C. (2022). Stakeholder perspectives on the roles of science and citizen science in Chesapeake Bay environmental management. *Estuaries and Coasts* 45(8), 2310–2326. <https://doi.org/10.1007/S12237-022-01106-5>/METRICS.
- Woods, N. D. (2021). Regulatory competition, administrative discretion, and environmental policy implementation. *Review of Policy Research, March*, 486–511. <https://doi.org/10.1111/ropr.12461>.
- Zhang, H. X. & Quinn, N. W. T. (2019). Simple models and analytical procedures for total maximum daily load assessment. *Journal of Hydrologic Engineering* 24(2), 02518002. [https://doi.org/10.1061/\(asce\)he.1943-5584.0001736](https://doi.org/10.1061/(asce)he.1943-5584.0001736).

First received 8 May 2023; accepted in revised form 4 November 2023. Available online 16 November 2023