Compilation and Analysis of Long-term Nitrogen and Phosphorus Monitoring Data in Tennessee

Final Report

Prepared By
Grace McClellan, Paul J. Tinker, Tania Datta

Center for the Management, Utilization and
Protection of Water Resources

Tennessee Tech University

ACKNOWLEDGEMENT

This study was funded by the Tennessee Department of Environment and Conservation (TDEC) and the Center for the Management, Utilization and Protection of Water Resources at Tennessee Tech University. We appreciate TDEC's constant support and partnership on this study, and express our gratitude to all organizations who shared water quality data with us to create the Tennessee Nutrient Database

Abstract

This study aimed to compile and analyze long-term nitrogen and phosphorus monitoring data in Tennessee, specifically in surface waters. To meet this goal, four main objectives were established: *1. identify potential nitrogen and phosphorus data sources across the state of Tennessee; 2. compile and standardize the acquired nutrient data; develop a central database to house the re-formatted data; 3. identify the data gaps; and 4. analyze the data for possible nutrient trends present in the watersheds of Tennessee.* Water quality data acquired from multiple organizations across the state were wrangled and harmonized for storage in the Tennessee Nutrient Database. This database was developed as a part of this study. Analyzing the data compiled in the database revealed that many watersheds lacked sufficient monitoring sites and sample collection frequency for a thorough statistical understanding of how nutrients have changed over time. The types of nutrients monitored by different organizations, across sites and watersheds were also not consistent. Nevertheless, the Regional Seasonal Mann Kendall statistical test was utilized for trend analysis, and detected significant increases and decreases in nutrient concentrations in some watersheds. However, gaps in the data may have impacted these results, making the results useful as a starting point, but not for drawing conclusions regarding nutrient trends. Through the aforementioned data processing and statistical analysis, this study revealed the need for a standardized reporting format and sampling methodology for the collection of water quality data. Improving the quantity and quality of nutrient data in the database will allow for more robust statistical analyses, resulting in a better understanding of how nutrients are changing in the surface waters of Tennessee.

TABLE OF CONTENTS

1. Background

Nutrient pollution has been attributed to declining surface water quality on a global scale (Howarth et al., 2002; Smith, 2003). Land-use changes, point and non-point source discharges, and increasing volumes of stormwater runoff prompt rapid transport of nutrients from landscapes to downstream surface waterbodies, resulting in devastating effects on human, animal and ecosystem health (Heisler et al., 2008; Woodward et al., 2012). One of the most widely known consequences of nutrient pollution is human-induced eutrophication, which may cause harmful algal blooms (HABs), fish kills and loss of biodiversity in aquatic ecosystems (Heisler et al., 2008; Dodds et al., 2009; Woodward et al., 2012). A study published in 2008 reported that nutrient pollution has a drastic economic impact on the waters of the United States. An estimated annual $2.2 billion can be incurred in revenue losses from reduced recreational water usage, losses in waterfront real estate value, increased potable water treatment cost, and expenses for recovering threatened or endangered wildlife (Dodds et al., 2009). The environmental and economic impacts of nutrient pollution demand a path toward better nutrient management.

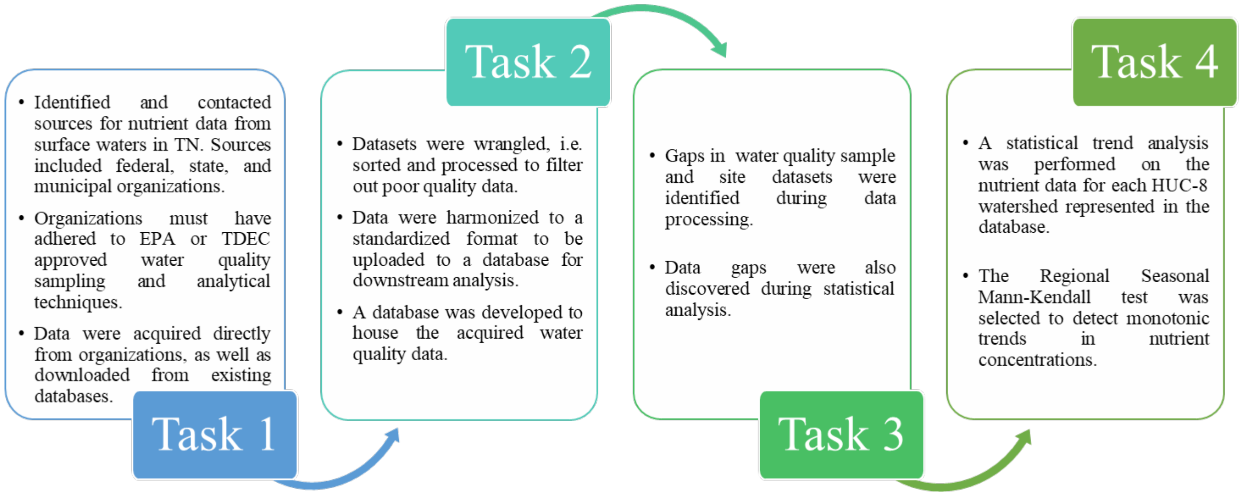
Tennessee's surface water bodies are not immune to nutrient pollution. The USGS estimates that 5.5% of the total nitrogen flux and 5.3% of the total phosphorus flux delivered to the Northern Gulf of Mexico are contributed by sources in Tennessee (Alexander et al., 2008). To address this problem, the Tennessee Department of Environment and Conservation (TDEC) developed a draft nutrient reduction framework in 2015 (TDEC, 2015). The framework serves as a blueprint by which point and nonpoint source reductions would contribute to fewer nutrients reaching rivers and streams, and acts as a consideration for NPDES permit writers to incorporate nutrient limits in permits until numeric standards are developed (TDEC, n.d.). In 2019, TDEC began a process of engaging stakeholders from multiple sectors to help refine and improve the framework. Many stakeholder workgroups were formed to focus on the various aspects of the nutrient reduction framework. This created the Tennessee Nutrient Reduction Taskforce (https://www.tn.gov/environment/program-areas/wr-water-resources/nutrient-management-in-tennessee/partnerships/tennessee-nutrient-reduction-task-force.html).

Since the formation of the Taskforce, some key questions have emerged, such as: how a baseline can be established for nutrients in Tennessee, and how nutrient loads/concentrations have changed over the years. Understanding the time-series trends in nutrients will help stakeholders better understand the current state and future needs toward nutrient reduction. Therefore, this project aims to compile and analyze long-term nitrogen and phosphorus monitoring data in Tennessee. We approached this exploratory study with four main objectives:

1. Identify potential nitrogen and phosphorus data sources across the state of Tennessee.
2. Compile and standardize the acquired nutrient data; develop a central database to house the re-formatted data.
3. Identify the data gaps.
4. Analyze the data for possible nutrient trends present in the watersheds of Tennessee.

2. Methodology

The above objectives were addressed through four specific tasks summarized in Figure 1. Details of each task are described below.



Task 1

- Identified and contacted sources for nutrient data from surface waters in TN. Sources included federal, state, and municipal organizations.
- Organizations must have adhered to EPA or TDEC approved water quality sampling and analytical techniques.
- Data were acquired directly from organizations, as well as downloaded from existing databases.

Task 2

- Datasets were wrangled, i.e. sorted and processed to filter out poor quality data.
- Data were harmonized to a standardized format to be uploaded to a database for downstream analysis.
- A database was developed to house the acquired water quality data.

Task 3

- Gaps in water quality sample and site datasets were identified during data processing.
- Data gaps were also discovered during statistical analysis.

Task 4

- A statistical trend analysis was performed on the nutrient data for each HUC-8 watershed represented in the database.
- The Regional Seasonal Mann-Kendall test was selected to detect monotonic trends in nutrient concentrations.

Figure 1: Flowchart of the tasks performed to address project objectives.

2.1 Nitrogen and Phosphorus Data Sources for the State of Tennessee

Many federal, state, and municipal organizations were contacted to obtain in-stream nitrogen and phosphorus data for the state of Tennessee. In-stream total suspended solids, turbidity, chlorophyll, dissolved oxygen, temperature, pH, and conductivity data were also requested to be curated in the database for any future analysis.

Nutrient data were either downloaded from EPA's Water Quality Portal (WQP), or received from US Army Corps of Engineers Nashville District (USACOE), City of Cookeville Water Quality Department, City of Murfreesboro Water Resources Recovery Facility, City of Nashville Stormwater Division, and the Oak Ridge National Lab. Contact information and the type of data downloaded or received from the abovementioned sources are summarized in Table 1. Some organizations provided more water characteristics than were slated for the database. Data outside the parameters set for the database were saved and stored with the raw data files. For a dataset to be accepted, it had to include at least the desired water quality characteristic names with appropriate measurement units, monitoring locations with coordinates, dates when samples were collected, and the laboratory and methods used for chemical analysis. The water quality characteristics that were most sought after for downstream analysis purposes, included total nitrogen, ammonia, nitrate/nitrite, total phosphorus, and orthophosphate.

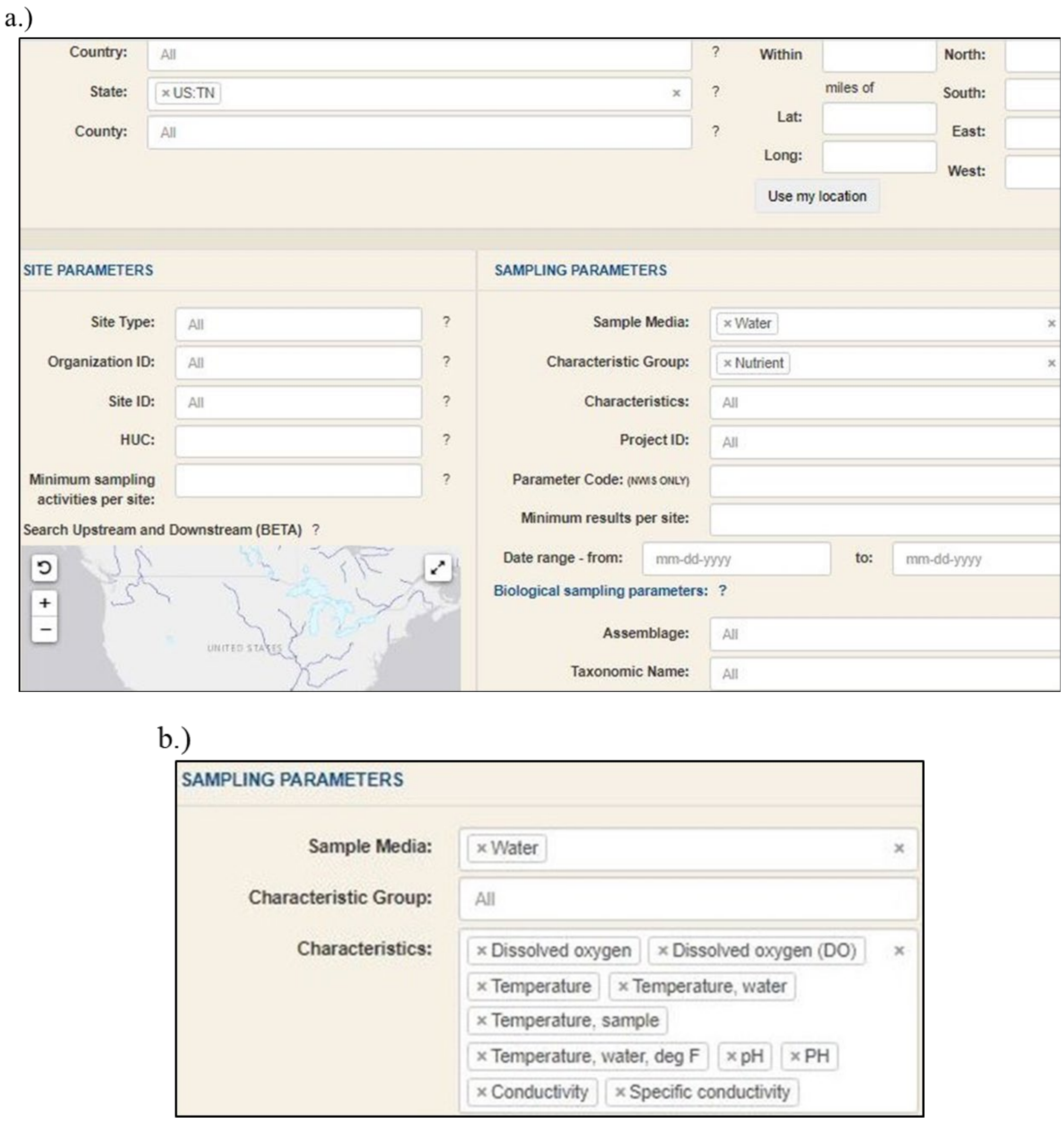
Table 1: Summary of nutrient data received.

Organization	Contact Information	Summary of Data Received
EPA Water Quality Portal	General Info Email: WQX@epa.gov Regional WQX Coordinator: Elizabeth Smith, Region 4 Atlanta, GA Email: smith.elizabeth@epa.gov Phone: (404) 562-8721	Downloaded on 04/12/2021 Data Range 05/09/1907-02/17/2021 Parameters: -Ammonia/Ammonium -Inorganic nitrogen (nitrate and nitrite) -Nitrate -Nitrite -Nitrogen -Nitrogen, mixed forms (NH_3), (NH_4), organic, (NO_2) and (NO_3) -Organic Nitrogen -Total Kjeldahl nitrogen -Organic Phosphorus -Orthophosphate -Phosphorus -Phosphorus, hydrolysable -Conductivity -Dissolved oxygen -pH -Temperature -Total suspended solids -Turbidity -Chlorophyll a, b, c

Organization	Contact Information	Summary of Data Received
US Army Corps of Engineers – Nashville District	Ashley Fuentes, Biologist, Water Management Email: Ashley.A.Fuentes@usace.army.mil Phone: 615-390-2146	Rec'd on 02/08/2021 Data Range 02/08/1994-12/15/2020 Parameters: -Total Ammonia -Total Nitrate-Nitrite -Total Kjeldahl Nitrogen -Total Nitrogen -Dissolved Phosphorus -Total Phosphorus -Conductivity -Dissolved Oxygen -pH -Secchi Disk -Temperature -Chlorophyll a
City of Cookeville Water Quality Department	Barry Turner, Director Email: bturner@cookeville-tn.gov Phone: 931-520-5259	Rec'd on 01/12/2021 Data Range 01/03/1997-04/04/2019 Parameters: -Ammonia -Nitrate -Total Kjeldahl Nitrogen -Total Nitrogen -Phosphorus -Dissolved Oxygen -Temperature -Total Suspended Solids -Chlorophyll a
City of Nashville – Stormwater Division	Mary Bruce, Watershed Evaluation Coordinator Email: Mary.Bruce@nashville.gov Phone: 615-862-4720	Rec'd on 02/08/2021 & 03/04/2021 Data Range 02/04/1998–09/28/2020 Parameters: -Ammonia -Nitrate-Nitrite -Total Kjeldahl Nitrogen -Total Nitrogen (1998-2009) -Dissolved Phosphorus -Phosphorus -Conductivity -Dissolved Oxygen -pH -Temperature -Total Dissolved Solids -Total Suspended Solids

Organization	Contact Information	Summary of Data Received
City of Murfreesboro Water Resources Recovery Facility	John Strickland, Plant Manager, Sinking Creek Email: jstrickland@murfreesborotn.gov Phone: 615-848-3225, ext. 3401	Rec'd on 03/03/2021 from TDEC Data Range 06/05/2014-06/06/2019 Parameters: -Ammonia -Nitrate-Nitrite -Organic Nitrogen -Total Kjeldahl Nitrogen -Organic Phosphorus -Orthophosphate -Total Phosphorus -Conductivity -Dissolved Oxygen -pH -Temperature
Oak Ridge National Laboratory	Taylor Frye, Water Quality Specialist Email: fryetb@ornl.gov Phone 865-576-7209	Rec'd on 06/24/2021 Data Range 03/10/2009-12/10/2020 Parameters: -Ammonia -Nitrate-Nitrite -Total Kjeldahl Nitrogen -Total Phosphorus -Conductivity -Dissolved Oxygen -Flow -pH -Temperature -Turbidity

Some data were received directly from individual organizations in excel and pdf formats, however, majority of the data was acquired from EPA's WQP. When querying the WQP database, all counties across Tennessee were selected. Unless otherwise noted, all parameters remained in the default selection "all" and the sample media selection was set to "water". For the nutrient dataset, the characteristic group was set to "nutrients" (Figure 2a). For other water quality data, total suspended solids, turbidity, turbidity field, dissolved oxygen, temperature, pH, conductivity, and all forms of chlorophyll, were selected under characteristics (an example is provided in Figure 2b). Water quality characteristics were downloaded in different files due to large file sizes. For the monitoring location dataset, the same parameters were set as the nutrient dataset (Figure 2a), but the "site data" option was selected in place of "sample results".



a.)

Country: All
State: × US:TN
County: All
Within
miles of
Lat:
Long:
North:
South:
East:
West:

SITE PARAMETERS

Site Type: All
Organization ID: All
Site ID: All
HUC:
Minimum sampling activities per site:

SAMPLING PARAMETERS

Sample Media: × Water
Characteristic Group: × Nutrient
Characteristics: All
Project ID: All
Parameter Code: (NWIS ONLY)
Minimum results per site:
Date range - from: mm-dd-yyyy to: mm-dd-yyyy
Assemblage: All
Taxonomic Name: All

b.)

SAMPLING PARAMETERS

Sample Media: × Water
Characteristic Group: All
Characteristics: × Dissolved oxygen × Dissolved oxygen (DO) × Temperature × Temperature, water × Temperature, sample × Temperature, water, deg F × pH × PH × Conductivity × Specific conductivity

Figure 2: Parameters selected when downloading data from WQP: a.) for nutrient characteristics, b.) for other water quality characteristics.

2.2 Compiling the Water Quality Data into a Database

Compiling the water quality data into a database initiated with data wrangling and harmonization. Data wrangling involved cleaning and filtering the raw data, while harmonization involved formatting the data and data structure to adhere to the database schema. The majority of data processing was conducted using RStudio (R Core Team, 2022; RStudio Team, 2022). R scripts and templates for data wrangling and harmonization were based on the layout of the WQP datasets. Therefore, datasets received from organizations that were not part of the WQP had to be semi-harmonized before wrangling could begin. These datasets were received in various structures and file formats, including emails, pdfs, and excel workbooks. Consequently, the semi-harmonization process included changing data orientation and column headings. Using R, the melt() function from the data.table package was utilized to transpose columns and rows, and the separate() function from the tidyr package was utilized to separate data into the appropriate columns (Dowle, 2021; Wickham and Girlich, 2022).

2.2.1 Data Wrangling

Water quality and monitoring location data were wrangled separately for each organization. Once the water quality data was wrangled, monitoring location datasets were filtered based on the sites in the wrangled water quality datasets. Data from most organizations were processed using the same general workflow (Figure 3). The USACOE and TDEC are two examples where the workflow was modified to accommodate special circumstances. The USACOE data was acquired from two different sources and had duplicate sample and site data due to overlap between the datasets. Each dataset was wrangled separately, following the workflow laid out in Figure 3. However, after the two datasets were combined during harmonization, the combined dataset had to be wrangled again to remove duplicate samples and sites. The workflow for the TDEC dataset was modified due to sample duplicates and discrepancies in site data, which will be further explained in the following paragraphs.

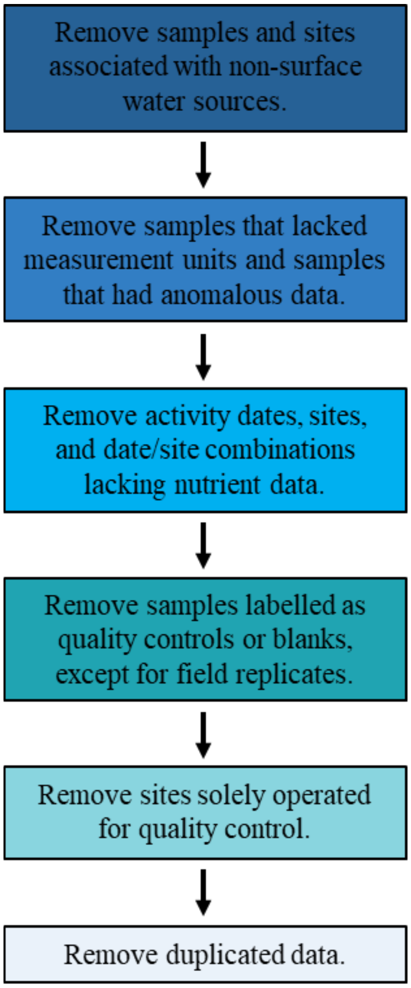


Figure 3: General workflow for data wrangling

When wrangling the water quality data, only surface water samples were retained. In some cases, samples were not labelled as surface or ground water. For those samples, the site dataset was examined for the type of water source the samples were collected from. Samples were kept if collected from a lake, reservoir, river, stream, and sometimes a spring. For samples collected from a spring, the site name and description were examined to determine whether it was associated with a surface water. Sample measurements that had numerical values with no measurement units were removed, except for pH. Samples with atypical values such as erroneous dates or text as the result measurement value were also removed. While many different water quality characteristics were accepted, the main focus of this project was to acquire and analyze nutrient data. Therefore, when wrangling the datasets, physical characteristic data were kept only if measured along with nutrients. Using dplyr and stringr R packages (Wickham, 2019; Wickham et al., 2022), water quality datasets were filtered based on activity date, monitoring location, or a combination of the two, in order to identify and remove samples that were not associated with nutrient data. Additionally, samples labeled as quality control or blanks were removed, with the exception of field replicates.

When wrangling the monitoring location datasets, sites were removed if they were not in the wrangled water quality datasets. Additionally, sites solely operated for quality control were removed. When a quality control site was identified, the associated water quality dataset was then double-checked for samples that might have been collected from that site. If such samples were detected, they were also removed. Throughout the wrangling process, it was noticed that some datasets contained multiple sites with similar GPS coordinates (Table 2). Since this could skew results during downstream statistical analysis, a column was created that combined the latitude and longitude for each site and was titled "TTU Analysis Site Identifier" (Table 2). This column would allow data to be grouped together based on GPS coordinates instead of monitoring location identifier. The GPS was truncated to four decimal places because when plotted on a map, five or more decimal places did not noticeably alter the location of the site.

Table 2: Examples of sites with similar GPS coordinates

Organization ID	Monitoring Location Identifier	Monitoring Location Name	Latitude	Longitude	TTU Analysis Site Identifier
TDECWPC	TDECWPC-CONAS054.4PO	Conasauga River	34.9906	-84.7747	34.9906 -84.7747
TDECWR_WQX	TDECWR_WQX-TNW000001376	Conasauga River	34.9906	-84.7747	34.9906 -84.7747
TDECWPC	TDECWPC-WEATH000.6BR	Weatherly Branch	34.99205	-84.89419	34.992 -84.8941
TDECWR_WQX	TDECWR_WQX-TNW000006681	Weatherly Branch	34.99205	-84.89419	34.992 -84.8941
TDECWPC	TDECWPC-MILL000.1BR	Mill Creek	34.9926	-84.7758	34.9926 -84.7758
TDECWR_WQX	TDECWR_WQX-TNW000004123	Mill Creek	34.9926	-84.7758	34.9926 -84.7758
11NPSWRD_WQX	11NPSWRD_WQX-OBED_7	Daddys Creek at Devils Breakfast Table	36.0592	-84.79238	36.0592 -84.7923

Organization ID	Monitoring Location Identifier	Monitoring Location Name	Latitude	Longitude	TTU Analysis Site Identifier
11NPSWRD_WQX	11NPSWRD_WQX-OBED_DEC_DC	Daddys Creek at Devils Breakfast Table	36.05925	-84.7923056	36.0592 -84.7923
11NPSWRD_WQX	11NPSWRD_WQX-OBED_NPS_D1	Daddys Creek at Devils Breakfast Table	36.05925	-84.7923056	36.0592 -84.7923
11NPSWRD_WQX	11NPSWRD_WQX-OBED_NPS_DC-2	Daddys Creek at Devils Breakfast Table	36.05925	-84.7923056	36.0592 -84.7923

Using activity dates, monitoring location identifiers, GPS coordinates, and characteristic names, each dataset was then analyzed for duplicate data. For clarity, replicates were not removed. A duplicate was defined as a sample that was reported or documented more than once, typically with the same result measure value. Replicates were defined as more than one sample collected from the same site on the same day, perhaps at different times or depths, and typically had different result measure values. For example, in some instances, USGS reported nutrient measurements twice, once as the elemental form and again as the molecular form. The elemental forms were retained and the matching molecular versions were removed. A more complex example of addressing duplicates is how the TDEC dataset was wrangled. Occasionally identical data was uploaded to the WQP by more than one TDEC division (Table 3). Since some monitoring locations had multiple names assigned to a single GPS coordinate (Tables 2 and 3) and TDEC had a large amount of sample and site data (Table 6), additional R scripts were developed to wrangle TDEC's data. The first step was to address the monitoring locations that shared similar or exact GPS coordinates. TDEC has a database under the Division of Water Resources (DWR) that includes monitoring station information that is more updated than the WQP. A station file was downloaded from the DWR database for comparison with the WQP site data. The WQP site information for TDEC did not always match site information from the DWR database. Some sites shared the same monitoring ID but had different coordinates and some sites had the same coordinates and different monitoring IDs. Since the DWR database houses the most current data, site information in the WQP dataset was updated with the DWR monitoring location names and GPS coordinates. The DWR file had two names associated with a single site. One was labelled "DWR Station ID" and followed a format that incorporated the name of the water source; the other name was labelled "Monitoring Location ID" and followed a format with the TNW prefix followed by numbers. In the cases where incorrect information for a WQP site was detected, the version of the DWR name that matched the format of the original WQP name was selected. The descriptions of these sites were also compared to double-check that the modifications were correct. Once site data were updated, the TDEC water quality sample file was re-analyzed to remove duplicate data.

Table 3: Examples of sample duplicates in the TDEC dataset

Organization ID	Date	Monitoring Location Identifier	Characteristic	Result Measure Value	Unit	TTU Analysis Site Identifier
TDECDOE_WQX	4/22/2009	TDECDOE_WQX-RACCO000.4RO	Dissolved oxygen	11.8	mg/l	35.9054 -84.3491
TDECWR_WQX	4/22/2009	TDECWR_WQX-TNW000005005	Dissolved oxygen	11.8	mg/l	35.9054 -84.3491
TDECDOE_WQX	4/22/2009	TDECDOE_WQX-RACCO000.4RO	pH	7.9		35.9054 -84.3491
TDECWR_WQX	4/22/2009	TDECWR_WQX-TNW000005005	pH	7.9		35.9054 -84.3491
TDECDOE_WQX	4/22/2009	TDECDOE_WQX-RACCO000.4RO	Phosphorus	0.029	mg/l	35.9054 -84.3491
TDECWR_WQX	4/22/2009	TDECWR_WQX-TNW000005005	Phosphorus	0.029	mg/l	35.9054 -84.3491
TDECDOE_WQX	4/22/2009	TDECDOE_WQX-RACCO000.4RO	Temperature	9.4	deg C	35.9054 -84.3491
TDECWR_WQX	4/22/2009	TDECWR_WQX-TNW000005005	Temperature	9.4	deg C	35.9054 -84.3491
TDECDOE_WQX	4/22/2009	TDECDOE_WQX-MCCOY000.9AN	Dissolved oxygen	11.2	mg/l	35.9654 -84.2483
TDECWR_WQX	4/22/2009	TDECWR_WQX-TNW000004023	Dissolved oxygen	11.2	mg/l	35.9654 -84.2483
TDECDOE_WQX	4/22/2009	TDECDOE_WQX-MCCOY000.9AN	Inorganic nitrogen	0.099	mg/l	35.9654 -84.2483
TDECWR_WQX	4/22/2009	TDECWR_WQX-TNW000004023	Inorganic nitrogen	0.099	mg/l	35.9654 -84.2483
TDECDOE_WQX	4/22/2009	TDECDOE_WQX-MCCOY000.9AN	Kjeldahl nitrogen	0.22	mg/l	35.9654 -84.2483

Organization ID	Date	Monitoring Location Identifier	Characteristic	Result Measure Value	Unit	TTU Analysis Site Identifier
TDECWR_WQX	4/22/2009	TDECWR_WQX-TNW000004023	Kjeldahl nitrogen	0.22	mg/l	35.9654 -84.2483
TDECDOE_WQX	4/22/2009	TDECDOE_WQX-MCCOY000.9AN	pH	7.5		35.9654 -84.2483
TDECWR_WQX	4/22/2009	TDECWR_WQX-TNW000004023	pH	7.5		35.9654 -84.2483
TDECDOE_WQX	4/22/2009	TDECDOE_WQX-MCCOY000.9AN	Phosphorus	0.008	mg/l	35.9654 -84.2483
TDECWR_WQX	4/22/2009	TDECWR_WQX-TNW000004023	Phosphorus	0.008	mg/l	35.9654 -84.2483
TDECDOE_WQX	4/22/2009	TDECDOE_WQX-MCCOY000.9AN	Temperature	9.6	deg C	35.9654 -84.2483
TDECWR_WQX	4/22/2009	TDECWR_WQX-TNW000004023	Temperature	9.6	deg C	35.9654 -84.2483

2.2.2 Data Harmonization

The database required all data from various organizations to be uploaded in a standardized format. Therefore, during the harmonization process, templates were developed for formatting the water quality and monitoring location data consistently, as shown in Tables 4a and 4b, respectively. The column headers in each template matched the database table column names, and a format for each column, such as numeric, free form, word bank, decimal degrees, military time, etc., was established. If a dataset had extra columns that did not match the templates, they were removed. Sometimes essential information, such as organization identifier, activity identifier, and monitoring location identifier, were not provided by an organization, but were required for the dataset to be uploaded to the database. In these cases, information was created following a standard procedure. An organization identifier was simply a shortened form of the organization name. An activity identifier was generated by combining the organization identifier with a number, which would increment every time the date, the time, the monitoring location, or the depth changed. Finally, a monitoring location identifier was created by combining the organization ID with the site name.

Table 4a: Database template for water quality data

TN Nutrient Database Column Names	WQP Column Names	Column Notes
activity_identifier	ActivityIdentifier	free form
activity_type_id	ActivityTypeCode	select from reference table
start_date	ActivityStartDate	mm/dd/yyyy
start_time	ActivityStartTime.Time	military HH:MM:SS
start_time_zone_id	ActivityStartTime.TimeZoneCode	select from reference table
end_date	ActivityEndDate	mm/dd/yyyy
end_time	ActivityEndTime.Time	military HH:MM:SS
end_time_zone_id	ActivityEndTime.TimeZoneCode	select from reference table
measureable_type : activity_depth	ActivityDepthHeightMeasure.MeasureValue	numerical
measureable_unit_id : activity_depth	ActivityDepthHeightMeasure.MeasureUnitCode	select from reference table
project_identifier	ProjectIdentifier	free form
monitoring_location_id	MonitoringLocationIdentifier	provide same name used in monitoring loc table; will be converted to an index
activity.legacy_comments	ActivityCommentText	only add if already provided
activity.comments	a place holder to add new activity comments (new column)	optional
sample_collection_method_id	SampleCollectionMethod.MethodIdentifier	select from reference table
sample_collection_method_context_id	SampleCollectionMethod.MethodIdentifierContext	select from reference table
sample_collection_equipment_id	SampleCollectionEquipmentName	select from reference table
result_detection_condition_id	ResultDetectionConditionText	select from reference table
name (characteristic)	CharacteristicName	select from reference table/word bank
sample_fraction_id	ResultSampleFractionText	select from a reference table
measureable_type : result_measure	ResultMeasureValue	numerical
measureable_unit_id : result_measure	ResultMeasure.MeasureUnitCode	select from a reference table
result_speciation_id	this is a new column	select from a reference table
measure_qualifier.code	MeasureQualifierCode	select from ref table: if more than one code, use semi-colon

TN Nutrient Database Column Names	WQP Column Names	Column Notes
result_status_id	ResultStatusIdentifier	select from a reference table
result_value_type_id	ResultValueTypeName	select from a reference table
result.legacy_comments	ResultCommentText	only add if already provided
result.comments	a place holder to add new result comments (new column)	optional
analytical_method_id	ResultAnalyticalMethod/MethodIdentifier	select from a reference table
analytical_method_context_id	ResultAnalyticalMethod.MethodIdentifierContext	select from a reference table
laboratory_id	LaboratoryName	use the lab names from the lab table
quantitation_limit_type_id	DetectionQuantitationLimitTypeName	select from a reference table
measureable_type : result_ql	DetectionQuantitationLimitMeasure.MeasureValue	numeric
measureable_unit_id : result_ql	DetectionQuantitationLimitMeasure.MeasureUnitCode	select from a reference table
result_ql_speciation_id	this is a new column	select from a reference table
result_provider	ProviderName	free form

Table 4b: Database template for monitoring location data

TN Nutrient Database Column Names	WQP Column Names	Column Notes
organization_id	OrganizationIdentifier	unique name already assigned to the associated organization
monitoring_location_identifier	MonitoringLocationIdentifier	needs to be unique
monitoring_location_name	MonitoringLocationName	display name
monitoring_location_type_id	MonitoringLocationTypeName	select from a reference table
horizontal_reference_datum_id	HorizontalCoordinateReferenceSystemDatumName	select from a reference table
county_code	CountyCode	county fips code
state_code	StateCode	state fips code

TN Nutrient Database Column Names	**WQP Column Names**	**Column Notes**
coordinates	LatitudeMeasure	latitude as decimal degrees
	LongitudeMeasure	longitude as decimal degrees
huc_code_eight	HUCEightDigitCode	huc 8 code
eco_region_level_three	Level III Eco	ecoregion 3 (2 digits)
eco_region_level_four	Level IV Eco	ecoregion 4 (3 digits)
description	MonitoringLocationDescriptionText	optional
drainage_area (measure value)	DrainageAreaMeasure	numeric
drainage_area_unit	DrainageAreaMeasureUnit	select from a reference table
horizontal_collection_method_id	HorizontalCollectionMethodName	select from a reference table
ttu_analysis_site_identifier	trunc_gpscombo	truncated GPS combo as a label
ttu_analysis_site_notes	New column, notes for the ttu analysis site label regarding sites that share GPS	examples: surface/depth, close together, separated by influent/effluent

Water quality characteristic terms were reported differently among organizations; therefore, a word bank was created to standardize the terminology (shown in Table 5). For nutrients specifically, name, fraction, and speciation were separated into three columns. Units for nutrient measurements were also standardized. Nutrient speciation, for example “as NH_3” or “as N”, was uploaded to the database in the form that it was received. In cases where there was no speciation and the original organization could not be contacted, a default species was determined based on how that nutrient is typically reported (Table 5). Temperature measurements reported as Fahrenheit were converted to Celsius prior to database upload. Any units reported for pH were removed. The format for activity date and time was also standardized.

Table 5: Harmonized characteristic names and default speciation

Harmonized Characteristic Name	Harmonized Unit	Default Speciation
Ammonia	mg/l	as NH_3
Ammonium	mg/l	as NH_4*
Chlorophyll a	as received	-
Chlorophyll b	as received	-
Chlorophyll c	as received	-
Conductivity	as received	-
Dissolved Oxygen (DO)	mg/l	-
Inorganic nitrogen (nitrate and nitrite)	mg/l	as N
Inorganic nitrogen (NO_2, NO_3, & NH_3)	mg/l	as N
Kjeldahl nitrogen	mg/l	as N
Nitrate	mg/l	as NO_3
Nitrite	mg/l	as NO_2
Nitrogen	mg/l	as N
Nitrogen, mixed forms (NH_3), (NH_4), organic, (NO_2) and (NO_3)	mg/l	as N
Organic Nitrogen	mg/l	as N
Phosphorus, hydrolyzable	mg/l	as P
Organic phosphorus	mg/l	as P
Orthophosphate	mg/l	as PO_4
Phosphate	mg/l	as PO_4
Phosphorus	mg/l	as P
Phosphorus, mixed forms	mg/l	as P
pH	no units	-
Temperature, water	deg C	-
Total suspended solids	as received	-
Turbidity	as received	-

*if received data was originally reported as "ammonia and ammonium", default speciation was "as N"

It was important to make the distinction between samples that were measured "below method detection limit (MDL)" from those that were classified as "not detected". This was because the MDL numerical values could be used for data analysis when the values were divided in half (Helsel, D. 2005; TDEC et al., 2021), while the "not detected" values could not be used. A sample measure was considered "not detected" if the reported value was 0 or labelled as "not detected" in the original dataset. However, in the case of USGS and TDEC, the "not detect" samples that also had a blank in the result value column, were re-labeled as "below MDL" because the EPA requires organizations to report to WQP all data that was below MDL as "not detected" (TDEC et al., 2021). Furthermore, regarding USGS and TDEC datasets, if there were samples with a 0 in the result value column, and the samples lacked any notation regarding MDL or not detected, those samples were labelled as "not detected".

For site data, the monitoring location type column had to be populated with a type of surface water, such as a river, stream, reservoir, lake, or spring. If the horizontal reference datum id or the horizontal collection method id was not provided, then those columns were populated as "unknown". The "TTU Analysis Site Identifier" column was populated with the truncated GPS coordinates. Monitoring sites were spatially mapped using the ArcGIS software along with watershed boundaries, and was utilized to populate FIPS and HUC-8 codes for organizations that did not report it. Data for ecoregion levels 3 and 4 were not reported by any organization; therefore, were also added to each dataset using ArcGIS. For site datasets that were received with FIPS and HUC-8 codes, GIS was employed to verify if the codes were correct. Some values in the original datasets did not match the values generated from the GIS verification process (Table 6). Before updating or changing any original site data, three resources were referenced to visualize and check the mismatched data. For TDEC sites, the GIS outputs were first compared to the DWR site file; if the values matched, it was considered correct and the GIS value was used. If the values did not match between the GIS output and the DWR file, two different web-based maps were checked to verify whether the GIS outputs were correct (Esri ArcGIS, 2022; Conservation Biology Institute, 2022). Note, these maps were also used to check mismatched site data from other organizations in addition to TDEC. Once the coordinates were plotted, the location point was visually inspected to determine which watershed, ecoregion, and county the point resided in. This secondary verification process supported the use of the GIS outputs. In cases where the point in question was located in the middle of a boundary line, the GIS data was chosen in order to be consistent and so that monitoring locations were not sorted into more than one region or county, which could have happened since there were sites that shared the same GPS.

Table 6: Examples of site data (highlighted in yellow) that did not match the output from the GIS analysis

Organization ID	Monitoring Location Identifier	County Code Original Dataset	County Code GIS Output	HUC-8 Original Dataset	HUC-8 GIS Output	Eco 3 Original Dataset	Eco 3 GIS Output	Eco 4 Original Dataset	Eco 4 GIS Output
USGS-TN	USGS-03407790	1	129	5130104	5130104	69	69	69d	69d
USGS-TN	USGS-03407850	13	1	5130104	5130104	69	69	69d	69d
USGS-TN	USGS-355307084180200	145	105	6010207	6010207	67	67	67f	67f
USGS-TN	USGS-355859084332901	145	129	6010208	6010208	68	68	68a	68a
USGS-TN	USGS-03532202	13	13	6010205	6010205	69	67	69e	67f
USACOEND	USACOEND-JPP10001	37	37	5130203	5130203	71	71	71h	71i
11NPSWRD_WQX	11NPSWRD_WQX-GRSM_F_0173	9	9	6010204	6010204	66	66	66f	66g
11NPSWRD_WQX	11NPSWRD_WQX-GRSM_F_0360	155	155	6010107	6010107	66	66	66i	66g
11NPSWRD_WQX	11NPSWRD_WQX-GRSM_F_0062	155	155	6010107	6010107	66	66	66g	66i
11NPSWRD_WQX	11NPSWRD_WQX-GRSM_F_0061	155	155	6010107	6010107	66	66	66g	66i
TDECWPC	TDECWPC-ECO65J05	71	71	6040001	6030005	65	65	65e	65j
TDECWPC	TDECWPC-CYPRE004.8SH	69	157	8010208	8010210	65	74	65e	74b
TDECWR_WQX	TDECWR_WQX-TNW000002017	115	115	6030001	6030001	68	68	68c	68b

A brief count summary of the processed data can be found in Table 7. The number of original observations and sites represent the raw data counts. The observation and site counts after harmonization represent the data that was uploaded to the database. Observations are the number of rows in a dataset, not necessarily the number of samples. The post-harmonization site count in Table 7 reflects the number of sites based on GPS coordinates, not site names. TDEC's data comprised the majority of the data in the database, followed by USGS TN and USACOE Nashville District. As of this report, the database consists of 17 organizations with data spanning from July 1946 to February 2021.

Table 7: Summary of the data after processing.

Organization	Date Range after Harmonization	Original Observations	Observations after Harmonization	Original Monitoring Sites	Monitoring Sites after Harmonization
Alabama Dept. of Environmental Management*	March 2005 - Dec. 2020	18,672	12,289	5	5
City of Cookeville Water and Sewer Department	Jan. 1997 - April 2019	69,544	44,196	14	8
City of Murfreesboro Water Resources Recovery Facility	June 2014 - June 2019	1,445	743	17	9
City of Nashville – Stormwater Division	Feb. 1998 - Sept. 2020	10,061	8,384	91	83
EPA Region 4 Athens Lab (Georgia)*	May 2017 - Sept. 2017	161	125	54	16
Georgia DNR Environmental Protection Division*	Oct. 2000 - Dec. 2013	155	132	3	2
Kentucky Division of Water*	Feb. 1999 - Feb. 2014	910	877	4	3
National Park Service Water Resources Division*	March 1979 - Jan. 2020	52,235	37,220	818	329

Organization	Date Range after Harmonization	Original Observations	Observations after Harmonization	Original Monitoring Sites	Monitoring Sites after Harmonization
North Carolina Department of Environmental Resources*	Oct. 1972 - Sept. 2019	4,713	3,325	5	3
Oak Ridge National Laboratory	April 2009 - Nov. 2020	3,330	2,525	11	10
TDEC Division of Water Resources*	April 2009 - Feb. 2021	369,624	289,247	7,915	2,431
Tennessee Department of Environment and Conservation*	Jan. 1993 - June 2019	202,765	192,317	5,544	2,608
US Army Corps of Engineers, Nashville District**	March 1994 - Oct. 2019	838,617	93,427	342	137
USGS Georgia Water Science Center*	Sept. 1957 - Dec. 2001	858	727	3	3
USGS Kentucky Water Science Center*	May 2011	54	18	14	1
USGS New York Water Science Center*	Nov. 2010 - Nov. 2011	88	66	12	12
USGS Tennessee Water Science Center*	July 1946 - Feb. 2021	165,904	110,558	1,967	692
Virginia Department of Environmental Quality*	Aug. 2001 - Nov. 2019	3,314	1,500	14	10

*Data was downloaded from WQP

**Army Corps data is a combination of data downloaded from WQP and data received from Nashville District

-Note: Per TDEC discussion, the two divisions of TDEC listed above, were not combined

2.2.3 Database Development

The project database, titled Tennessee Nutrient Database (TNNDB) was built using the MySQL Relational Database Management System (RDBMS). To facilitate ease of new data acquisition and ingestion, as well as to capitalize on the knowledge and experience of TDEC and the Water Quality Exchange portal (WQX), the TNNDB schema was designed to mirror the WQX portal data schema as closely as possible while focusing only on the subset of data deemed pertinent to this project. While table structure may not be identical to the WQX, the granularity of persisted information is equivalent including any legacy data that might prove useful to other agencies in the future. The only area of major deviation is in the inclusion of a polymorphic measure value type. This choice reduces the number of tables that would be required in a many-to-many approach while offering a great deal of flexibility in the type of information that can be stored. The TNNDB entity relationship (ER) diagram is shown in Figure 4.

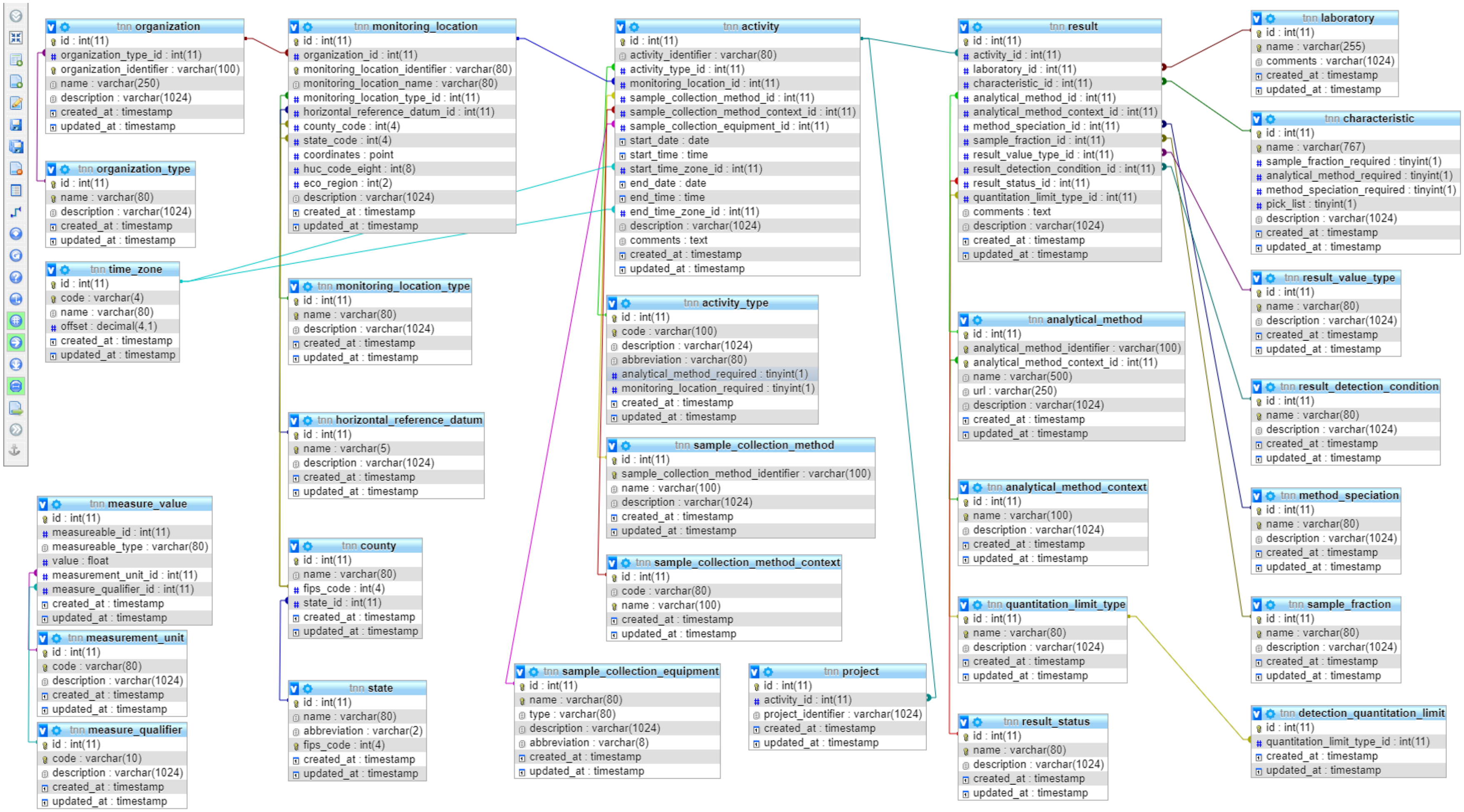


Figure 4: Tennessee Nutrient Database Entity Relationship Diagram

2.3 Identifying Data Gaps

Once the processed data was uploaded to the database, the database was ready to be queried and the data could be examined for any gaps. Data gaps were identified at various points throughout the project. First, the distribution of monitoring locations across the state was examined to establish watersheds that have been lacking in nutrient monitoring. This information could be informative when planning for future water quality monitoring projects. Next, the sample data was assessed for frequency of sampling events and characteristics that were commonly or uncommonly measured. Investigating gaps in time periods that samples were collected can also be informative for future monitoring and/or trend analysis projects. Evaluating which nutrients are commonly measured could be useful for development of a standardized protocol for state-wide nutrient collection. These gaps helped inform our decisions regarding which statistical test to perform and which data could be selected for analysis. Furthermore, during statistical analysis, additional data gaps were discovered, which will be discussed in the results section.

2.3.1 Identifying Data Gaps in Monitoring Sites

After the database was complete, data were examined to determine how nutrient monitoring sites were distributed across the state and to identify watersheds that were lacking in sites. Utilizing GIS, two different types of maps were generated to visualize the site coverage across each watershed. For an overall assessment of Tennessee, a heatmap was created to illustrate site distribution across the state. HUC-8 and HUC-10 boundaries overlaid the heatmap to delineate site counts per watershed. To examine the site distribution more closely within a watershed, individual maps were also created for each watershed depicting site location and the associated organization. The individual watershed maps can be found in Appendix A. Additionally, tables were generated from the database to numerically identify watersheds that may be lacking nutrient monitoring and those that might be heavily monitored. These tables included the first year a watershed was sampled to the last year it was sampled. These year ranges do not represent consecutive sampling years. For tables summarizing nutrient monitoring per ecoregion, refer to Appendix B.

Availability of stream flow data was also investigated to determine the statistical analysis strategy. Many nutrient trend analysis studies have assessed either flow-weighted concentrations or loads trends over a period of years (McIsaac et al., 2016; Oelsner and Stets, 2019). Such trends demonstrate the effects of streamflow, precipitation, and evapotranspiration changes on nutrients over time. Many studies have also analyzed flow-adjusted trends, which removes the hydrologic effects on nutrient trends, and displays just the anthropogenic causes of nutrient concentration changes over time (Sprague and Lorenz, 2009; Murphy and Sprague, 2019). However, for such analyses, nutrient concentrations and corresponding discharge or flow measurement data is required. In Tennessee, there are considerably more monitoring sites for measuring water quality than for measuring flow, as can be seen when comparing the sites listed in EPA's Water Quality Portal to the stream gages operated by USGS (Figure 5). Additionally, after identifying water quality sites located close to a stream gage, it was observed that some of these sites were not sampled when the stream gage was operational. Therefore, the trend analysis was specifically based on observed concentration trends due to lack of corresponding flow gaging stations or measurements near water quality monitoring sites.

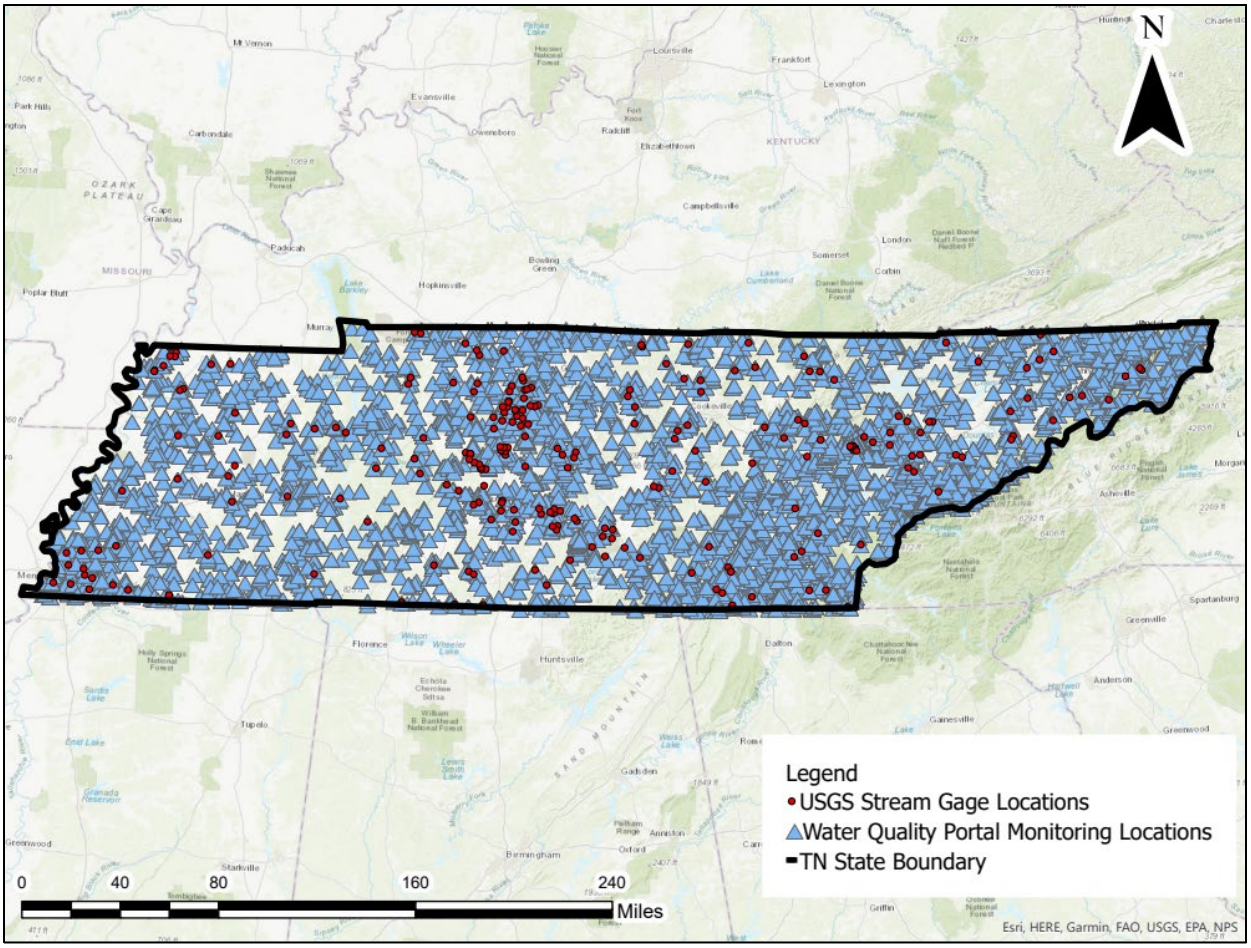


Figure 5: Monitoring sites from EPA's WQP and stream gages operated by USGS

2.3.2 Identifying Sample Data Gaps

For the sample data, potential gaps were investigated in the frequency that samples were collected and the different types of nutrients that were measured. The database was queried and a table was created summarizing the number of samples collected for each nutrient measured in each watershed (Appendix C). Sample results labeled as "below MDL" were included in these counts but those labeled as "not detected" were left out since they could not be used for statistical analysis. This table additionally grouped the sample counts into the years each watershed was actively sampled, illustrating the time periods when data was continuously collected. This table also served as a guide for nutrient selection for trend analysis since it displayed which nutrient characteristics potentially had sufficient data for statistical trend analysis. Nutrients that were selected for statistical analysis are noted with ** in Appendix C.

2.4 Data Analysis

2.4.1 Statistical Test Selection for Trend Analysis

As previously mentioned, the lack of stream gages across TN supported the decision to analyze trends in nutrient concentrations rather than nutrient loads. After a literature review, it was determined that a Regional-Seasonal Mann Kendall (RSMK) trend analysis, a version of the Mann Kendall test, would be used for this study (Mann, 1945; Sprague and Lorenz, 2009). The Mann Kendall test is a nonparametric statistical test that has been used in other similar projects to detect monotonic (single-direction) trends in water quality data (Helsel and Frans, 2006; Sprague and Lorenz, 2009; Kuwayama, 2020). This test can analyze data over time, and allows for the data to be grouped, or blocked, into seasons and regions. For this study, a separate RSMK test was conducted for each HUC-8 watershed per nutrient characteristic. For each watershed and nutrient combination, data were grouped by monitoring location, which defined the regional portion of the test, and then grouped into four seasons, which defined the seasonal portion of the test. Several R packages are available for the Mann Kendall trend test. The *rkt* package was selected for this study because it included the RSMK version of the MK test, it could accommodate datasets with missing data, and the blocking parameter was user-defined, which made it flexible and easy to use (Marchetto, 2021). This package also had a parameter to include a correction for correlation between blocks of data, which was important because it accounts for serial dependence, or helps to correct data points from influencing each other (Hirsch & Slack 1984). If the correction for correlation was not implemented, statistical significance could be falsely detected. In regards to grouping the data, only one datapoint could represent a group or block for each year. If there was more than one datapoint in a block for a given year, the *rkt* package had the option to either average the data or calculate the median. The median was selected during this step because the median is less sensitive to outliers than the mean (Schuenemeyer and Drew, 2011). For this analysis, one block represented data at a single site in one season. The following is an example, accompanied with a visual in Figure 6, to further explain how blocking works. If sites A and B were sampled during the spring and the summer every year from 1995 to 2015, the dataset would be divided into four blocks. The maximum data points for each block would be 21, one for each year. As seen in Figure 6, site A had more than one sample collected in the spring of 2007, therefore, the median was calculated for those datapoints. That single datapoint would then represent the data in Block 1 for 2007.

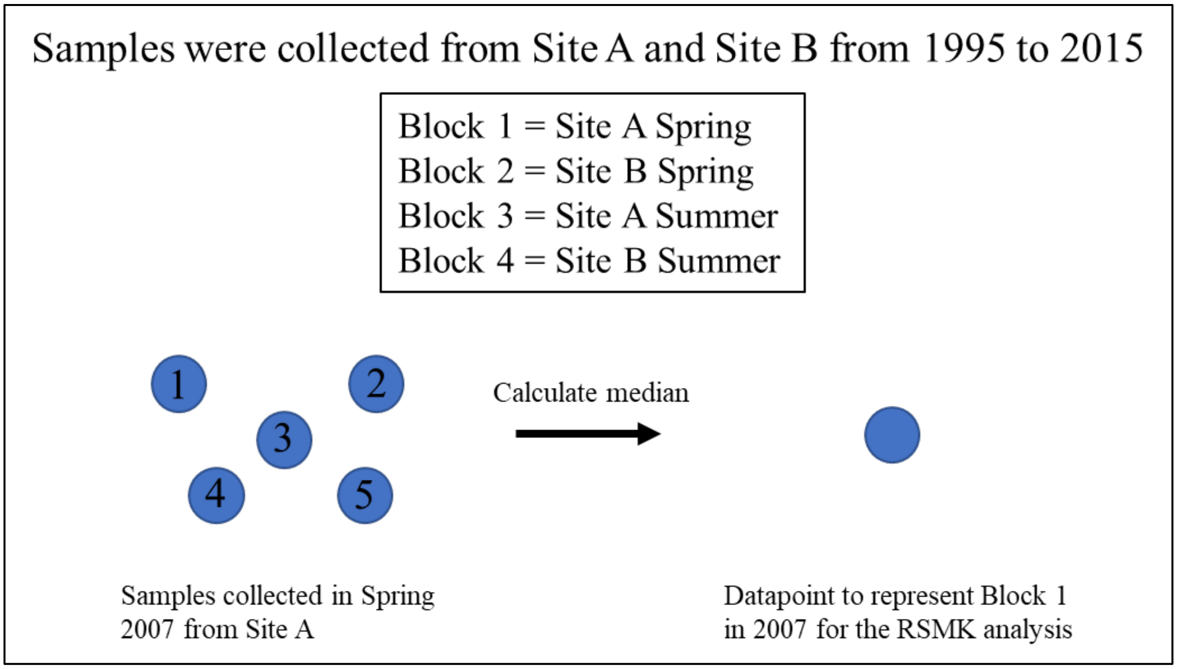


Figure 6: Example of how blocking works in a Regional Seasonal Mann Kendall test.

2.4.2 Database Query Parameters for RSMK Analysis

Nutrients for RSMK analysis were selected utilizing the table in Appendix C and are denoted by **. Datasets were then generated from the database, which was queried to include samples collected from 0 to 5 feet below the water surface and between 1990 and 2021 across all HUC-8 watersheds. It should be noted that data from 2021 was not included in any of the RSMK tests because data from 2021 only spanned into February. There were three main datasets exported for analysis: total ammonia and total phosphorus (total P), inorganic nitrogen (nitrate plus nitrite) and orthophosphate, and total nitrogen (total N) (Table 8). All fractions of inorganic nitrogen and orthophosphate were selected (total, dissolved, suspended, and NULL) for analysis, since both these nutrient species have high water solubility. Moreover, orthophosphate had a relatively low sample count when only selecting for the "total" fraction. Characteristics in the total N dataset included total N, total N mixed forms (NH_3, NH_4, organic, NO_2, and NO_3), inorganic N (all fractions), and total Kjeldahl nitrogen (TKN). Inorganic N and TKN were later added together in R to produce a calculated total N value. Measurements that were below method detection limits were included in the query, but those labelled as "not-detected" were left out. Calculations to divide the MDL values in half were also included and performed in the query. The TTU analysis site identifier was queried for use in the analysis, whereas the monitoring location ID was also queried but only for reference. The “Result Comments” column was later added to the inorganic N and total N queries due to a discrepancy discovered in the datasets during statistical analysis; this discrepancy will be further explained below.

Table 8: Nutrients selected for statistical trend analysis

	Nutrient	**Fraction**
Dataset 1	Ammonia	Total
	Phosphorus	Total
Dataset 2	Inorganic Nitrogen (nitrate and nitrite)	Total, Dissolved, Suspended, NULL
	Orthophosphate	Total, Dissolved, Suspended, NULL
Dataset 3	Nitrogen	Total
	Nitrogen, mixed forms (NH_3, NH_4, organic, NO_2, and NO_3)	Total
	Inorganic Nitrogen (nitrate and nitrite)	Total, Dissolved, Suspended, NULL
	Kjeldahl Nitrogen	Total

2.4.3 General RSMK Workflow

The datasets exported from the database were filtered and formatted in R before the RSMK test was performed. This process was conducted in two separate scripts and utilized the R packages lubridate, dplyr, ggplot2, and purr (Grolemund et al., 2011; Wickham et al., 2022; Wickham 2016; Henry et al., 2020). The first R script processed data from all the organizations together and removed data that did not meet the following criteria. Each site was checked to be associated with only one watershed. Sites with less than four years of data and samples with “null” result values were removed. Speciation was standardized as either N or P. Specifically, ammonia measurements reported as NH_3 were converted to N and orthophosphate measurements reported as PO_4 were converted to P. All other characteristics were previously reported as N or as P. For inorganic N and total N, samples were removed that had <, U, or "undetected" noted in the result legacy comment column. The majority of these samples had values that were magnitudes less than the MDL and were most likely quality control blanks. All samples with those notations were removed because they could not be differentiated between samples that were blanks, not detected, or had detected measurements below MDL.

After processing the dataset as a whole, a second R script was implemented to parse and filter the dataset for individual watershed RSMK analysis. Each analysis focused on one nutrient for a single watershed. Minimum requirements were established that each dataset had to meet before the RSMK test could be performed. Only one datapoint for each monitoring date was allowed. If a site was sampled more than once on a given day, the average was calculated for those measurements. For the total N dataset, samples with inorganic N and TKN measurements were added together to make a calculated total N value. If more than one total N value existed for a given sampling event, those values were averaged together. A scatter plot was then generated of the sites and the sample counts to guide in selecting a time period for analysis (see Figure 8 for an example plot and Appendix D for all other watershed plots). Data points were grouped by year for each site and those groups were categorized as either "4 or more" samples or "less than 4" samples in that year. A label of "less than 4" indicated that all four seasons were not represented in a given year for a given site, which was useful when selecting the time period. If the data distribution supported more than one time period, the RSMK was performed multiple times. Once the time period was established, sites were removed based on the following criteria. For a site to be included in the RSMK test, it had to have at least four years of data spanning a minimum of ten years. Additionally, the site had to have at least one data point at the start year and one at the end year of the time period selected for analysis. Lastly, at least two sites must pass this filtering process for RSMK to be performed on the dataset.

A second scatter plot was then generated to visualize the final dataset for the RSMK test. Months in the final dataset were categorized into seasons for blocking purposes. Ideally, the final dataset should include samples from all four seasons, but RSMK was performed even when there were only three seasons. Winter was the only season occasionally not represented in the RSMK analyses. Seasons were defined as follows: spring included March, April, May; summer included

June, July, August; fall included September, October, November; and winter included December, January, and February. The RSMK test was then performed on the dataset. Results were exported as an excel file and summarized the number of sites and the seasons included in the analysis, the number of blocks analyzed and ignored, and the test statistic along with associated statistical information.

3. Results and Discussion

3.1 Challenges with Data Acquisition and Data Processing

During the data acquisition phase, most organizations were willing to share water quality data, and were helpful with follow-up questions as well as with providing additional information. The only two sources we were unsuccessful with acquiring data were USACOE Memphis District and the Tennessee Valley Authority (TVA). The USACOE Memphis office responded to our communications but were unable to locate a water quality representative for their region. The TVA did not respond to our data requests. It may be noted that TVA monitors 528 streams sites for water quality concentrations and flow. Therefore, obtaining data from them in the future and adding to the TN nutrient database may prove to be beneficial for trend analysis.

Accepting data from a variety of organizations presented some challenges. Since data was not received in a standardized format, some form of data wrangling and harmonization occurred at almost every step along the way. This was because discrepancies in the data continued to emerge and had to be resolved before uploading to the database or performing statistical analysis. Even the data that was exported from the WQP was not consistent. The WQP dataset did not have standardized characteristic names; nutrient speciation and fractions were sometimes left blank; samples were not always labeled with the type of water source they were collected from; and some samples were below MDL but were missing the MDL value. Furthermore, pertinent information was discovered in a column reserved for comments when the information should have been recorded in a more appropriate column. For example, the sample results of some data were described as “undetected” in a comment column and would have been better suited in the “result detection condition” column, where the “not detected” and “below MDL” labels are located.

The data cleaning, filtering, and formatting made up a large portion of this project. However, it was necessary not only for database upload, but also to more accurately identify data gaps. Moreover, without the cleaning, statistical analysis would have been impossible.

3.2 Tennessee Nutrient Database

The Tennessee Nutrient Database currently houses data spanning from 1946 to 2021 from 17 different organizations. This data represents 53 HUC-8 watersheds, 8 level 3 ecoregions, and 29 level 4 ecoregions. The database can be found in Appendix F.

3.3 Gaps Discovered in the Site and Water Quality Data

Housing all the site and sample data in a database allowed the watersheds to be examined collectively. For example, the count tables generated to assess the number of sites in each HUC-8 watershed and the organizations that operated them, illustrate how watersheds across Tennessee are not monitored equally (Tables 9a and 9b). The count table displaying the types of nutrients measured in each watershed and the number of samples collected, further supports that watersheds lack uniformity in regards to how they are monitored (Appendix C). Additionally, the database currently has data for 53 out of 55 watersheds in the state of Tennessee. Watershed East Fork Clarks River (HUC8 6040006) and Forked Deer River (HUC8 8010206) are currently not represented in the database.

Table 9a: Monitoring Site Count per Watershed

HUC 8 CODE	Watershed Name	Site count based on monitoring location name	Site count based on monitoring location GPS	Active Interval*
3150101	Conasauga	43	31	1957-2020
5110002	Barren	77	55	1985-2020
5130101	Upper Cumberland	38	34	1981-2020
5130103	Upper Cumberland - Lake Cumberland	4	3	1994-2020
5130104	South Fork Cumberland	81	64	1964-2020
5130105	Obey	61	55	1964-2020
5130106	Upper Cumberland -Cordell Hull Reservoir	50	43	1962-2021

HUC 8 CODE	Watershed Name	Site count based on monitoring location name	Site count based on monitoring location GPS	Active Interval*
5130107	Collins	69	58	1964-2019
5130108	Caney Fork**	174	145	1962-2021
5130201	Lower Cumberland - Old Hickory Lake	114	98	1962-2020
5130202	Lower Cumberland - Sycamore**	251	209	1964-2021
5130203	Stones**	204	173	1964-2020
5130204	Harpeth**	148	120	1966-2020
5130205	Lower Cumberland	74	60	1964-2021
5130206	Red**	144	108	1964-2020
6010101	North Fork Holston	4	3	1995-2020
6010102	South Fork Holston**	277	185	1963-2020
6010103	Watauga**	198	146	1967-2020
6010104	Holston**	153	108	1965-2021
6010105	Upper French Broad	25	17	1946-2021
6010106	Pigeon	48	41	1968-2021
6010107	Lower French Broad**	199	170	1967-2021
6010108	Nolichucky**	224	166	1966-2021
6010201	Watts Bar Lake**	348	280	1960-2021
6010204	Lower Little Tennessee**	126	104	1966-2019
6010205	Upper Clinch	96	71	1967-2021
6010206	Powell	59	43	1965-2020
6010207	Lower Clinch**	322	258	1961-2021
6010208	Emory**	167	135	1960-2019
6020001	Middle Tennessee - Chickamauga**	345	269	1957-2020
6020002	Hiwassee**	255	210	1966-2020
6020003	Ocoee	87	63	1965-2020
6020004	Sequatchie	119	94	1965-2021
6030001	Guntersville Lake	59	48	1967-2021
6030002	Wheeler Lake	21	21	1999-2021
6030003	Upper Elk	100	82	1966-2020
6030004	Lower Elk	51	39	1966-2020
6030005	Pickwick Lake	45	36	1967-2020
6040001	Lower Tennessee – Beech**	166	145	1961-2021
6040002	Upper Duck**	185	162	1962-2020

HUC 8 CODE	Watershed Name	Site count based on monitoring location name	Site count based on monitoring location GPS	Active Interval*
6040003	Lower Duck**	157	125	1964-2020
6040004	Buffalo	48	43	1963-2020
6040005	Kentucky Lake	112	95	1960-2021
8010100	Lower Mississippi - Memphis	50	37	1988-2021
8010202	Obion	113	85	1960-2021
8010203	South Fork Obion	77	62	1957-2021
8010204	North Fork Forked Deer	97	73	1961-2020
8010205	South Fork Forked Deer	76	59	1960-2020
8010207	Upper Hatchie	56	45	1960-2020
8010208	Lower Hatchie**	147	105	1960-2020
8010209	Loosahatchie	116	78	1960-2020
8010210	Wolf	112	75	1961-2020
8010211	Horn Lake - Nonconnah	73	49	1975-2021
	Total # of sites	6445	5083	

*Active interval does not indicate consecutive years of monitoring. The years stated are the first and last year that samples were collected from a watershed.

**Watersheds with more than 100 sites based on location coordinates

Table 9b: Site Count per Watershed and Organization

HUC 8 CODE	Watershed Name	Organization Name	Site count based on monitoring location name	Site count based on monitoring location GPS	Active Interval*
3150101	Conasauga	Georgia DNR Env	2	1	2000-2013
		TDEC DWR	26	26	2011-2020
		TDEC	14	14	1999-2007
		USGS GA	1	1	1957-2001
5110002	Barren	TDEC DWR	28	28	2010-2020
		USGS TN	3	3	1985-1985
		TDEC	46	44	1999-2009
5130101	Upper Cumberland	TDEC DWR	17	17	2009-2020
		USGS TN	8	8	1981-1982
		TDEC	13	13	1999-2008
5130103	Upper Cumberland-Lake Cumberland	TDEC DWR	1	1	2009-2020
		USACOE Nashville Dist	1	1	1994-2019
		TDEC	2	2	1999-2009

HUC 8 CODE	Watershed Name	Organization Name	Site count based on monitoring location name	Site count based on monitoring location GPS	Active Interval*
5130104	South Fork Cumberland	NPS WRD	11	11	2009-2009
		TDEC DWR	24	24	2009-2020
		USGS TN	26	25	1964-1981
		TDEC	20	20	1999-2006
5130105	Obey	TDEC DWR	11	11	2009-2020
		USACOE Nashville Dist	27	27	1994-2019
		USGS TN	11	11	1964-1981
		TDEC	12	12	1999-2009
5130106	Upper Cumberland-Cordell Hull Reservoir	TDEC DWR	19	19	2009-2021
		USACOE Nashville Dist	12	12	1994-2019
		USGS TN	3	3	1962-1976
		TDEC	14	14	1999-2009
		City of Cookeville	2	2	2001-2019
5130107	Collins	TDEC DWR	31	31	2010-2019
		USACOE Nashville Dist	1	1	2002-2003
		USGS TN	9	9	1964-1997
		TDEC	28	26	1999-2013
5130108	Caney Fork	TDEC DWR	44	44	2009-2021
		USACOE Nashville Dist	41	41	1994-2019
		USGS TN	14	14	1962-1981
		TDEC	68	67	1999-2009
		City of Cookeville	6	6	1997-2019
		NC Dept Enviro	1	1	1972-1982
5130201	Lower Cumberland-Old Hickory Lake	TDEC DWR	57	57	2009-2020
		USACOE Nashville Dist	13	13	1994-2019
		USGS TN	3	3	1962-1994
		TDEC	41	41	1999-2009
5130202	Lower Cumberland-Sycamore	TDEC DWR	96	96	2009-2021
		USACOE Nashville Dist	14	14	1994-2019
		USGS TN	14	14	1964-1992
		TDEC	68	67	1999-2009
		Nashville Storm Water	59	59	1998-2020
5130203	Stones	NPS WRD	13	13	1997-2017
		TDEC DWR	68	67	2009-2020
		USACOE Nashville Dist	17	17	1994-2019
		USGS TN	22	22	1964-2001

HUC 8 CODE	Watershed Name	Organization Name	Site count based on monitoring location name	Site count based on monitoring location GPS	Active Interval*
		TDEC	61	59	1999-2009
		City of Murfreesboro	9	9	2014-2019
		Nashville Storm Water	14	14	2013-2019
5130204	Harpeth	NPS WRD	3	3	2007-2020
		TDEC DWR	73	73	2009-2020
		USACOE Nashville Dist	2	2	1994-2019
		USGS TN	5	5	1966-2017
		TDEC	55	55	1999-2012
		Nashville Storm Water	10	10	2013-2019
5130205	Lower Cumberland	NPS WRD	5	5	2003-2004
		TDEC DWR	33	33	2010-2021
		USACOE Nashville Dist	7	7	1994-2019
		USGS TN	1	1	1964-1965
		TDEC	28	27	2001-2008
5130206	Red	KY Division of Water	3	3	1999-2014
		TDEC DWR	59	58	2009-2020
		USACOE Nashville Dist	2	2	1994-2019
		USGS TN	8	8	1964-1997
		TDEC	72	68	1999-2009
6010101	North Fork Holston	TDEC DWR	2	2	2009-2020
		USGS TN	1	1	1995-1998
		TDEC	1	1	1999-2009
6010102	South Fork Holston	VA Dept Enviro	1	1	2001-2010
		TDEC DWR	103	103	2009-2020
		USGS TN	11	11	1963-1998
		TDEC	161	146	1998-2018
		USGS NY	1	1	2010-2011
6010103	Watauga	TDEC DWR	84	84	2010-2020
		USGS TN	11	11	1967-1998
		TDEC	98	94	1999-2009
		USGS NY	5	5	2010-2011
6010104	Holston	TDEC DWR	69	68	2009-2021
		USGS TN	10	10	1965-1998
		TDEC	74	71	1999-2014
6010105	Upper French Broad	TDEC DWR	11	11	2010-2021
		USGS TN	2	2	1946-1998

HUC 8 CODE	Watershed Name	Organization Name	Site count based on monitoring location name	Site count based on monitoring location GPS	Active Interval*
		TDEC	11	11	2001-2006
		USGS NY	1	1	2010-2011
6010106	Pigeon	NPS WRD	18	18	1993-2019
		TDEC DWR	9	9	2010-2021
		USGS TN	5	5	1968-1998
		TDEC	11	11	2001-2008
		NC Dept Enviro	1	1	1973-2016
		USGS NY	4	4	2011-2011
6010107	Lower French Broad	NPS WRD	91	90	1993-2019
		TDEC DWR	54	54	2009-2021
		USGS TN	14	14	1967-2009
		TDEC	39	39	2001-2009
		USGS NY	1	1	2011-2011
6010108	Nolichucky	TDEC DWR	95	95	2009-2021
		USGS TN	20	20	1966-2012
		TDEC	109	106	1996-2009
6010201	Watts Bar Lake	NPS WRD	63	63	1993-2019
		TDEC DWR	96	96	2009-2021
		USGS TN	55	54	1960-2021
		TDEC	134	134	1999-2009
6010204	Lower Little Tennessee	NPS WRD	44	44	1993-2019
		TDEC DWR	27	27	2010-2019
		USGS TN	17	17	1966-2009
		TDEC	37	36	1999-2013
		NC Dept Enviro	1	1	2004-2019
6010205	Upper Clinch	VA Dept Enviro	8	8	2005-2019
		TDEC DWR	36	36	2009-2021
		USGS TN	10	10	1967-2012
		TDEC	42	39	1999-2009
6010206	Powell	NPS WRD	5	5	1990-1997
		VA Dept Enviro	1	1	2009-2009
		TDEC DWR	25	25	2009-2020
		USGS TN	6	6	1965-1998
		TDEC	22	22	1998-2009
6010207	Lower Clinch	TDEC	148	144	1993-2019
		TDEC DWR	71	71	2009-2021

HUC 8 CODE	Watershed Name	Organization Name	Site count based on monitoring location name	Site count based on monitoring location GPS	Active Interval*
		USGS TN	93	92	1961-1993
		ORNL	10	10	2009-2020
6010208	Emory	NPS WRD	53	46	1979-2009
		TDEC DWR	34	34	2011-2019
		USGS TN	34	33	1960-2004
		TDEC	46	45	1999-2008
6020001	Middle Tennessee-Chickamauga	NPS WRD	2	2	2002-2004
		Georgia DNR Env	1	1	2001-2001
		TDEC DWR	184	183	2009-2020
		USGS TN	29	29	1957-2001
		TDEC	127	125	1999-2009
		USGS GA	2	2	1994-2001
6020002	Hiwassee	TDEC DWR	149	149	2009-2020
		USGS TN	13	13	1966-1994
		TDEC	93	92	1999-2009
6020003	Ocoee	TDEC DWR	46	46	2009-2020
		USGS TN	6	6	1965-1985
		TDEC	35	35	1999-2009
6020004	Sequatchie	TDEC DWR	63	63	2009-2021
		USGS TN	9	9	1965-1988
		TDEC	47	47	1999-2009
6030001	Guntersville Lake	AL Dept Enviro	1	1	2009-2020
		TDEC DWR	28	28	2010-2021
		USGS TN	5	5	1967-1986
		TDEC	25	20	1999-2006
6030002	Wheeler Lake	AL Dept Enviro	1	1	2009-2009
		TDEC DWR	17	17	2012-2021
		USGS TN	1	1	1999-1999
		TDEC	2	2	2003-2004
6030003	Upper Elk	TDEC DWR	47	47	2009-2020
		USGS TN	19	19	1966-2001
		TDEC	34	34	1999-2009
6030004	Lower Elk	AL Dept Enviro	1	1	2005-2020
		TDEC DWR	19	19	2009-2018
		USGS TN	10	10	1966-2001
		TDEC	21	21	2002-2009

HUC 8 CODE	Watershed Name	Organization Name	Site count based on monitoring location name	Site count based on monitoring location GPS	Active Interval*
6030005	Pickwick Lake	NPS WRD	4	4	2007-2020
		AL Dept Enviro	2	2	2015-2020
		TDEC DWR	15	15	2009-2020
		USGS TN	7	7	1967-2001
		TDEC	17	17	1999-2009
6040001	Lower Tennessee-Beech	NPS WRD	16	16	1999-2007
		TDEC DWR	36	36	2009-2021
		USGS TN	30	30	1961-2001
		TDEC	84	84	1999-2009
6040002	Upper Duck	EPA Reg 4 Athens GA	12	12	2017-2017
		TDEC DWR	48	48	2009-2020
		USGS TN	42	42	1962-2005
		TDEC	83	82	1999-2009
6040003	Lower Duck	NPS WRD	6	6	2007-2020
		EPA Reg 4 Athens GA	6	4	2017-2017
		TDEC DWR	54	54	2009-2020
		USGS TN	17	17	1964-2001
		TDEC	74	73	1999-2009
6040004	Buffalo	NPS WRD	3	3	2007-2020
		TDEC DWR	11	11	2009-2020
		USGS TN	19	19	1963-2001
		TDEC	15	15	1999-2009
6040005	Kentucky Lake	TDEC DWR	27	27	2009-2021
		USGS TN	19	19	1960-2001
		TDEC	66	64	1999-2009
8010100	Lower Mississippi-Memphis	TDEC DWR	31	30	2010-2021
		USGS TN	1	1	1988-2015
		TDEC	17	17	1999-2008
		USGS KY	1	1	2011-2011
8010202	Obion	TDEC DWR	53	53	2009-2021
		USGS TN	13	13	1960-1997
		TDEC	47	47	1999-2008
8010203	South Fork Obion	TDEC DWR	33	33	2009-2021
		USGS TN	6	6	1957-1965
		TDEC	38	38	1999-2008
8010204		TDEC DWR	37	37	2009-2020

HUC 8 CODE	Watershed Name	Organization Name	Site count based on monitoring location name	Site count based on monitoring location GPS	Active Interval*
	North Fork Forked Deer	USGS TN	3	3	1961-1965
		TDEC	57	57	1999-2009
8010205	South Fork Forked Deer	TDEC DWR	25	25	2009-2020
		USGS TN	5	5	1960-1968
		TDEC	46	46	1999-2009
8010207	Upper Hatchie	TDEC DWR	19	19	2009-2020
		USGS TN	3	3	1960-1962
		TDEC	34	34	1999-2008
8010208	Lower Hatchie	TDEC DWR	70	70	2009-2020
		USGS TN	2	2	1960-1995
		TDEC	75	73	1999-2009
8010209	Loosahatchie	TDEC DWR	47	47	2009-2020
		USGS TN	14	14	1960-1995
		TDEC	55	55	1999-2009
8010210	Wolf	TDEC DWR	45	45	2009-2020
		USGS TN	5	5	1961-2004
		TDEC	62	60	1999-2014
8010211	Horn Lake-Nonconnah	TDEC DWR	29	29	2009-2021
		USGS TN	2	2	1975-1975
		TDEC	42	42	1999-2009
		Total # of sites	6445	6362	

*Active interval does not indicate consecutive years of monitoring. The years stated are the first and last year that samples were collected from a watershed.

-Note: Per TDEC discussion, the two divisions of TDEC listed above, were not combined

Site distribution for many of the watersheds appeared sparse or sometimes uneven (Figure 7a; Appendix A). In the northern areas of TN, the watersheds North Fork Holston and Lake Cumberland do not comprise a large amount of area in TN, therefore, it is not surprising that each one only had a few sites (Tables 9a and 9b; Appendix A). However, other watersheds in that area, such as South Fork Cumberland and Cordell Hull, are larger and have more monitoring locations, but site coverage across those watersheds seemed lacking considering both are homes to popular outdoor recreational areas in TN. In comparison, metropolitan areas, such as Nashville, Chattanooga, and the Tri-Cities, were located in watersheds that had a relatively high number of monitoring sites, which included Hiwassee, Lower Cumberland – Sycamore, Stones, and South

Fork Holston watersheds. Watts Bar, Chickamauga, and Lower Clinch watersheds had the most monitoring locations, with over 250 sites each (Table 9a). A high number of monitoring sites could be because each of these watersheds are homes to nuclear reactors, which might require a higher level of ecosystem and water quality monitoring. Furthermore, according to the maps, Chickamauga appears to have even coverage in terms of site distribution, whereas Lower Clinch and Watts Bar have a more clustering-type distribution, again this could be due to locations of specialized facilities as well as the variety of organizations monitoring those areas (Appendix A). The Lower French Broad watershed also exhibited strong site clustering, but that appears to be due to the National Park Service operating a higher number of sites in the Great Smoky Mountains National Park (Table 9b; Appendix A). As can be seen in Tables 7 and 9b, the vast majority of sites across the state of TN were operated by TDEC, followed by USGS TN and National Park Services. Additionally, it should be noted that using site coordinates when counting monitoring locations gives a more accurate count of how many sampling locations exist for a given watershed. This is demonstrated in Table 9a where the number of sites drops for all but one watershed when counting sites based on GPS coordinates instead of the site name.

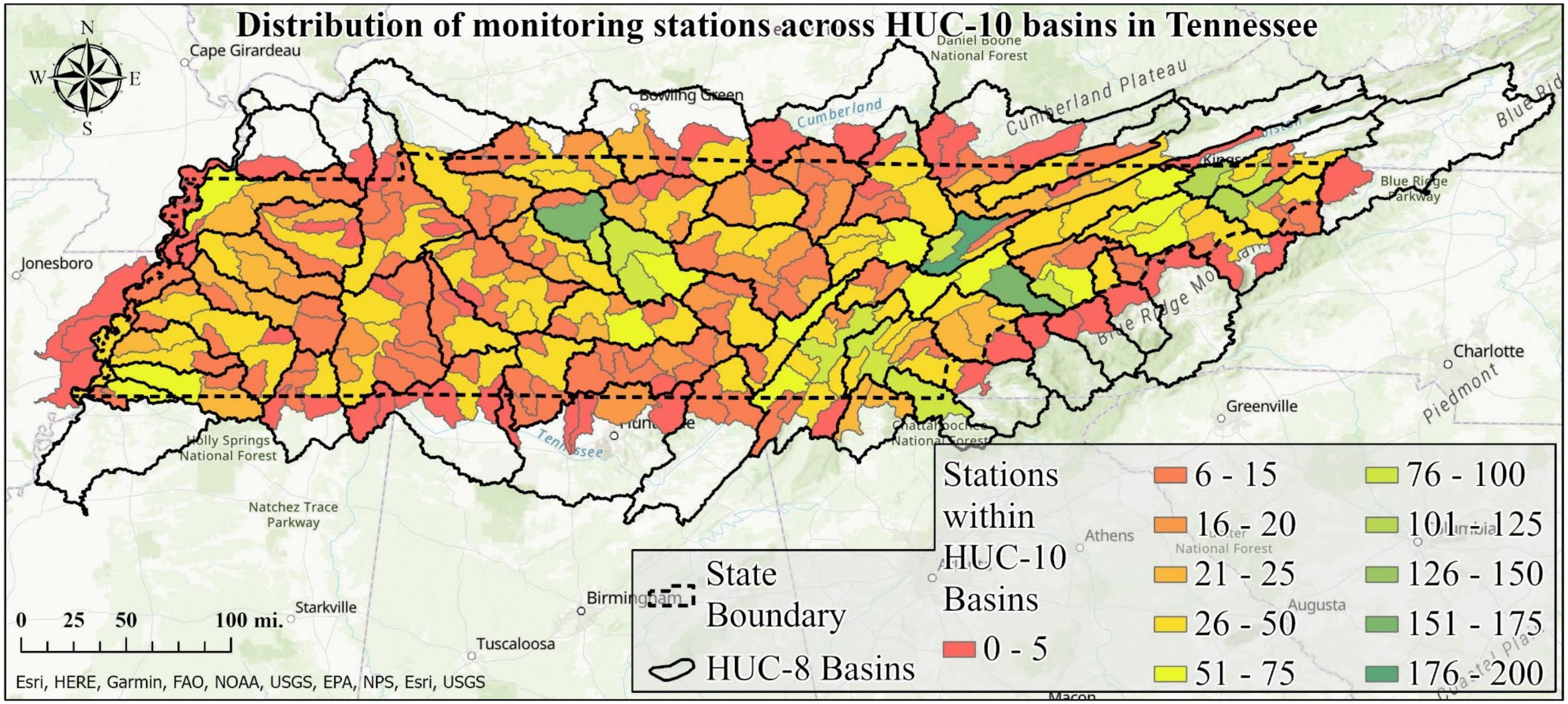


Figure 7a: Heatmap illustrating the number and distribution of monitoring sites in each HUC-8 watershed in Tennessee. The light grey lines represent the HUC-10 watershed boundaries.

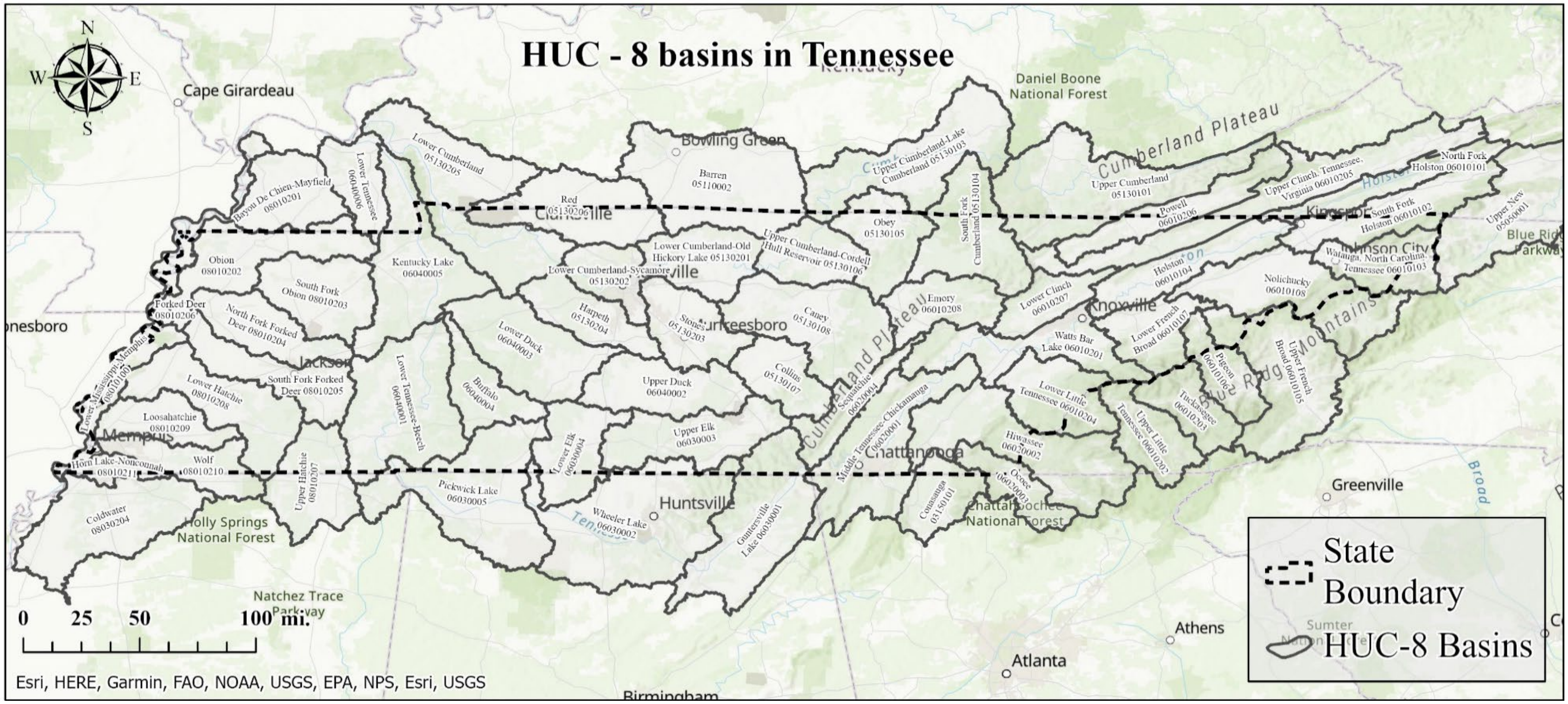


Figure 7b: HUC-8 watersheds in Tennessee.

The water quality data also exhibited gaps, including gaps detected in the frequency that samples were collected and in the different types of nutrients that were measured. Appendix C shows the sample count for each nutrient characteristic, how they were distributed among watersheds, and the years they were collected. The table shows that the majority of data was collected between the early 1990s and 2021, the last year currently in the database. Additionally, all the watersheds have been monitored somewhat recently; the most recent samples in the database were collected between 2019 and 2021. Some watersheds, such as Barren, Collins, Emory, and Upper French Broad, were not continuously monitored for more than a few years at a time between the early 1990s and 2021. As seen in Appendix C, samples were collected more sporadically for these watersheds. In some cases, Upper Cumberland and Wheeler Lake watersheds for example, there were timespans of three or more years when the watershed was not monitored at all. The table also shows that the nutrient data is not consistent among watersheds. The sample count for the combination of total N and total N mixed forms was less than 20,000, and almost half of the watersheds were barely monitored for total N directly. Whereas inorganic N and TKN had sample counts over 70,000 each and were measured in every watershed, suggesting that most organizations likely measure these characteristics in order to calculate total N in place of directly measuring total N. Total ammonia and total P were also measured in every watershed and monitored more frequently than total N. Interestingly, even though orthophosphate was measured in all but one watershed and all orthophosphate fractions were combined for statistical analysis, only six watersheds were measured frequently enough to qualify for the RSMK test. Additionally, the different fractions of nutrients posed challenges for data analysis. The database consisted of 16 different nutrient characteristic names and when including the different fractions, there were 51 different nutrient and fraction combinations. Many sample data points were omitted from statistical analysis because not all nutrient characteristics could be combined due to the different fractions. If nutrient testing was more standardized, it would improve the data pool for statistical analysis.

3.4 RSMK Analysis Challenges and Results

Data gaps were revealed during the statistical analysis process, in addition to those identified from the maps and count tables. Part of the statistical process involved visualizing datasets to guide selection of time periods to be analyzed (Appendix D). During this step, gaps were found in the sample and site data, and were more or less evident depending on which nutrient was being analyzed. One observed gap was that many watersheds lacked consistent, yearly sampling that included all four seasons. This gap, specifically, did not always disqualify a watershed from the RSMK test, because the RSMK test can accommodate missing data, but having yearly data collected during all four seasons would make the analysis more robust and conclusions could be drawn more confidently when examining the outcomes. For each nutrient analyzed, there were watersheds that did not pass the filtering process due to having less than four datapoints, a sampling period of less than 10 years, or a lack of sites sampled in the same time frame. A few times the RSMK test was discarded because the minimum requirement for blocking was not met, usually because samples were collected from only one or two seasons. Another issue that was noticed during scatter plot development was that some of the site coordinates were very similar (Figure 8 and Appendix D) and should probably have been grouped together for the RSMK analysis. Using tools such as GIS, a strategy could be developed to group geographically-similar sites together, however that was outside the scope of this project.

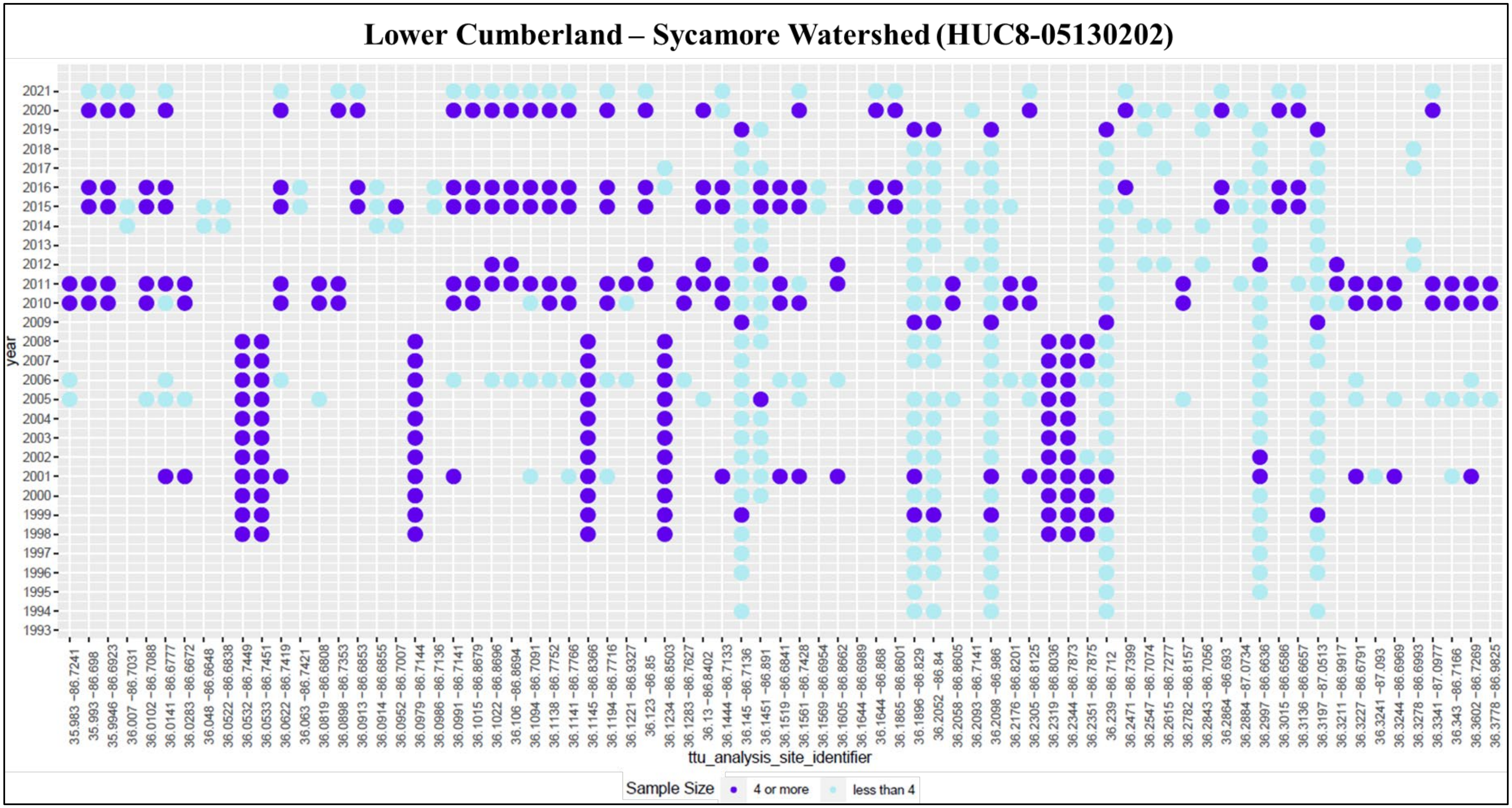


Figure 8: Sample data distribution for sites in the Lower Cumberland - Sycamore watershed for total nitrogen.

As mentioned before, five nutrient characteristics were analyzed with the RSMK test, total N, total ammonia, inorganic N, total P, and orthophosphate. These characteristics were selected because they are common parameters investigated when evaluating nutrient pollution in surface water (USEPA, 2000), they were measured in almost all of the watersheds, and each had over 60,000 samples. Orthophosphate was the exception and had closer to 20,000 samples.

The results of the RSMK showed a mix of significant trends in different watersheds across the state. Significance was reported for p-values < 0.1 and < 0.5. Watersheds with significant RSMK results can be found in Tables 10a – 10d; RSMK results for all watersheds can be found in Appendix E. A positive trend indicates an increase in nutrient concentrations over time, while a negative trend indicates a decrease in nutrient concentrations. Total N, total ammonia, and inorganic N all exhibited both positive and negative trends among various watersheds. Total P, however, mostly exhibited positive trends with Powell as the only watershed that showed a decrease in total P concentrations and it was over a span of 21 years. Total N had significant trends in nine watersheds, total ammonia had significant trends in 39 watersheds, inorganic N had significant trends in 16 watersheds, and total P had significance in 12 watersheds. Most of the watersheds lacked sufficient orthophosphate data for the RSMK analysis and the test was performed on only six watersheds, in which no significant trends were detected. Significant trends in total ammonia concentrations were detected in four times more watersheds than total N. This could be due to a difference in sampling patterns or perhaps biological events, such as nitrification, were occurring causing ammonia concentrations to shift while keeping total N concentrations stable. Interestingly, some watersheds, Lower Cumberland - Sycamore, Cordell Hull, and Guntersville Lake for example, had significant trends, but the trends were in different directions depending upon time period. Lower Cumberland - Sycamore exhibited an increase in total N from 1996 to 2019, but exhibited a decrease from 2001 to 2020; Cordell Hull exhibited a decrease in total ammonia from 1995 to 2019, but exhibited an increase from 2000 to 2020; and Guntersville Lake exhibited an increase in total ammonia from 2005 to 2020, but exhibited a decrease from 2011 to 2020. Each of these watersheds had sites with similar GPS coordinates, which can be observed for the Lower Cumberland – Sycamore watershed in Figure 8. Therefore, grouping sites that are close together in each watershed could possibly resolve this observed trend contradiction. Furthermore, since not all sites were sampled during the same years (Figure 8 and Appendix D), the sites included in an RSMK test would change depending on the time period selected for analysis, which could have also contributed to conflicting RSMK results.

Table 10a: Significant Results from the RSMK analysis of Total Nitrogen

HUC 8	**Watershed Name**	**Start Year**	**End Year**	**# of Years**	**Mann Kendall** ^ p.value < 0.1 * p.value < 0.05	**Kendall Tau**	**Direction of Trend**
5130106	Upper Cumberland - Cordell Hull Reservoir	2005	2019	15	^	-0.1820673	Decreasing
5130108	Caney	2007	2018	12	*	-0.2492484	Decreasing
5130202	Lower Cumberland - Sycamore	1996	2019	24	^	0.1555827	Increasing
5130202	Lower Cumberland - Sycamore	2001	2020	20	^	-0.2819383	Decreasing
5130205	Lower Cumberland	1994	2019	26	^	0.1497637	Increasing
6010102	South Fork Holston	2000	2019	20	*	0.1855104	Increasing
6010109	Nolichucky	2005	2020	16	*	-0.1981873	Decreasing
6040004	Buffalo	2000	2019	20	*	0.2184211	Increasing
8010202	Obion	2010	2020	11	^	0.320122	Increasing
8010203	South Fork Obion	2010	2020	11	^	0.2690217	Increasing

Table 10b: Significant Results from the RSMK analysis of Total Ammonia

HUC 8	**Watershed Name**	**Start Year**	**End Year**	**# of Years**	**Mann Kendall** ^ p.value < 0.1 * p.value < 0.05	**Kendall Tau**	**Direction of Trend**
5110002	Barren	2000	2020	21	^	0.2580645	Increasing
5130104	South Fork Cumberland	2000	2020	21	*	0.56917	Increasing
5130105	Obey	2000	2010	11	*	-0.2511292	Decreasing
5130105	Obey	2000	2019	20	*	-0.1928946	Decreasing
5130106	Upper Cumberland - Cordell Hull Reservoir	1995	2019	25	*	-0.4264373	Decreasing
5130106	Upper Cumberland - Cordell Hull Reservoir	2000	2020	21	^	0.5151515	Increasing
5130108	Caney	1996	2018	23	*	-0.1846343	Decreasing
5130202	Lower Cumberland - Sycamore	1994	2019	26	*	-0.3011123	Decreasing
5130203	Stones	1994	2019	26	*	-0.192437	Decreasing

HUC 8	Watershed Name	Start Year	End Year	# of Years	Mann Kendall ^ p.value < 0.1 * p.value < 0.05	Kendall Tau	Direction of Trend
5130205	Lower Cumberland	1994	2019	26	*	-0.3276956	Decreasing
5130206	Red	2000	2020	21	*	0.5889968	Increasing
5130206	Red	2010	2020	11	^	0.3520408	Increasing
6010102	South Fork Holston	1999	2018	20	*	0.4098684	Increasing
6010103	Watauga	1999	2017	19	*	0.6936937	Increasing
6010104	Holston	2000	2020	21	*	0.3792325	Increasing
6010105	Upper French Broad	2001	2020	20	*	0.4315068	Increasing
6010107	Lower French Broad	2005	2020	16	*	0.3979239	Increasing
6010108	Nolichucky	2000	2020	21	*	0.5418001	Increasing
6010108	Nolichucky	2005	2020	16	*	0.4317921	Increasing
6010201	Watts Bar Lake	2003	2017	15	*	0.5622951	Increasing
6010201	Watts Bar Lake	2006	2017	12	*	0.5141388	Increasing
6010204	Lower Little Tennessee	1999	2019	21	*	0.6436782	Increasing
6010205	Upper Clinch	2000	2020	21	*	0.5335868	Increasing
6010206	Powell	2000	2020	21	*	0.6105528	Increasing
6010207	Lower Clinch	2008	2019	12	^	0.1717949	Increasing
6020001	Middle Tennessee - Chickamauga	1999	2020	22	*	0.3706395	Increasing
6020001	Middle Tennessee - Chickamauga	2010	2020	11	^	0.1740891	Increasing
6020002	Hiwassee	1999	2018	20	*	0.2740741	Increasing
6020003	Ocoee	2006	2017	12	*	0.572238	Increasing
6020004	Sequatchie	2005	2020	16	*	0.5306513	Increasing
6030001	Guntersville Lake	2005	2020	16	^	0.3381295	Increasing
6030001	Guntersville Lake	2011	2020	10	*	-0.2469636	Decreasing
6030003	Upper Elk	2002	2018	17	*	0.2430769	Increasing
6030004	Lower Elk	2008	2018	11	*	-0.5660377	Decreasing
6030005	Pickwick Lake	2002	2017	16	*	0.5719844	Increasing
6040002	Upper Duck	2000	2020	21	*	0.362069	Increasing
6040003	Lower Duck	1999	2019	21	*	0.5617886	Increasing
6040004	Buffalo	1999	2019	21	*	0.6625	Increasing
6040005	Kentucky Lake	1999	2019	21	*	0.4292649	Increasing
8010202	Obion	2001	2020	20	*	0.4012121	Increasing
8010204	North Fork Forked Deer	2002	2018	17	*	0.2852174	Increasing
8010205	South Fork Forked Deer	2001	2017	17	^	0.187602	Increasing

HUC 8	Watershed Name	Start Year	End Year	# of Years	Mann Kendall ^ p.value < 0.1 * p.value < 0.05	Kendall Tau	Direction of Trend
8010208	Lower Hatchie	2000	2020	21	*	0.3285714	Increasing
8010209	Loosahatchie	2002	2018	17	*	0.4349975	Increasing
8010210	Wolf	2003	2019	17	*	0.4627171	Increasing
8010211	Horn Lake - Nonconnah	2004	2020	17	*	0.2321429	Increasing

Table 10c: Significant Results from the RSMK analysis of Inorganic Nitrogen

HUC 8	Watershed Name	Start Year	End Year	# of Years	Mann Kendall ^ p.value < 0.1 * p.value < 0.05	Kendall Tau	Direction of Trend
5130108	Caney	2007	2018	12	*	-0.196758	Decreasing
5130203	Stones	1994	2019	26	*	-0.1945831	Decreasing
5130203	Stones	2001	2018	18	*	-0.1999608	Decreasing
5130204	Harpeth	2001	2017	17	^	-0.1656516	Decreasing
6010102	South Fork Holston	1999	2019	21	*	0.2398325	Increasing
6010102	South Fork Holston	2007	2018	12	*	0.2958199	Increasing
6010104	Holston	2000	2020	21	*	0.1912568	Increasing
6020001	Middle Tennessee - Chickamauga	1999	2020	22	*	0.1949846	Increasing
6020001	Middle Tennessee - Chickamauga	2005	2020	16	*	0.2599049	Increasing
6020001	Middle Tennessee - Chickamauga	2010	2020	11	*	0.2701031	Increasing
6020002	Hiwassee	1999	2019	21	*	0.1406026	Increasing
6020002	Hiwassee	2007	2018	12	^	0.2307692	Increasing
6020004	Sequatchie	2001	2020	20	*	0.3604061	Increasing
6020004	Sequatchie	2005	2020	16	*	0.2251521	Increasing
6030003	Upper Elk	2002	2018	17	^	0.1958525	Increasing
6030003	Upper Elk	2006	2018	13	*	0.2271605	Increasing
6040003	Lower Duck	2003	2019	17	*	0.1765893	Increasing
6040004	Buffalo	1999	2020	22	^	0.1384217	Increasing
8010202	Obion	2001	2020	20	^	0.1304348	Increasing
8010203	South Fork Obion	2001	2020	20	^	0.2477876	Increasing
8010203	South Fork Obion	2005	2020	16	*	0.3069054	Increasing
8010203	South Fork Obion	2010	2020	11	*	0.3117978	Increasing

HUC 8	Watershed Name	Start Year	End Year	# of Years	Mann Kendall ^ p.value < 0.1 * p.value < 0.05	Kendall Tau	Direction of Trend
8010204	North Fork Forked Deer	2002	2018	17	^	0.209901	Increasing
8010210	Wolf	1999	2019	21	^	-0.1375661	Decreasing
8010211	Horn Lake - Nonconnah	2005	2020	16	^	-0.2358722	Decreasing
8010211	Horn Lake - Nonconnah	2001	2017	17	*	-0.2744565	Decreasing

Table 10d: Significant Results from the RSMK analysis of Total Phosphorus

HUC 8	Watershed Name	Start Year	End Year	# of Years	Mann Kendall ^ p.value < 0.1 * p.value < 0.05	Kendall Tau	Direction of Trend
3150101	Conasauga	2006	2017	12	^	0.381295	Increasing
5130108	Caney	1996	2019	24	^	0.1426535	Increasing
5130203	Stones	1994	2019	26	^	0.1005581	Increasing
6010107	Lower French Broad	2005	2020	16	*	0.3611111	Increasing
6010206	Powell	2000	2020	21	^	-0.1795775	Decreasing
6020001	Middle Tennessee - Chickamauga	2000	2020	21	*	0.2234818	Increasing
6030003	Upper Elk	2006	2018	13	*	0.4061033	Increasing
6030005	Pickwick Lake	2008	2017	10	*	0.390625	Increasing
6040002	Upper Duck	2000	2020	21	*	0.2959831	Increasing
6040003	Lower Duck	2008	2019	12	^	0.1333333	Increasing
8010208	Lower Hatchie	2000	2020	21	^	0.129085	Increasing
8010211	Horn Lake - Nonconnah	2005	2020	16	^	0.1699346	Increasing
8010211	Horn Lake - Nonconnah	2006	2017	12	*	0.2222222	Increasing

4. Conclusions and Future Recommendations

This study successfully compiled nutrient water quality data spanning from 1946 to 2021 from 17 different agencies. Data were wrangled and harmonized for storage in the Tennessee

Nutrient Database, which was developed during this study. This database allowed for multi-agency nutrient data to be examined simultaneously and to be queried and partitioned for analysis in a variety of ways, such as per watershed, ecoregion, nutrient characteristic, or sampling period. The ability to examine data holistically is crucial for understanding the trends in nutrient concentrations in the surface waters of Tennessee.

Some challenges were encountered during the study. For example, some data lacked necessary information or had errors in documentation that prevented the data from being analyzed. To address this problem, a standardized reporting format for site and sampling data should be developed and made available to relevant organizations across the state. The database developed from this project is structured and able to house such water quality data and should be utilized as such since stakeholders in the Tennessee Nutrient Reduction Taskforce are able to administer how it is used, unlike larger databases, such as EPA WQP. The standardized reporting format should limit the terminology and water quality parameters that can be uploaded to the database. Specifically, an improved reporting format should include requiring and standardizing information such as the characteristic names, units, species, and fractions. Additionally, sample measurements should be properly labelled as "not detected" or "below MDL", if applicable, and if measurements are below MDL, the MDL value should be reported. Furthermore, participating organizations should be encouraged to collect samples more frequently. If possible, samples should be collected yearly and during each season. Monitoring more frequently would allow a more robust statistical trend analysis to be performed on the data. Furthermore, if stream flow was measured along with water quality sample collection, a loads analysis could be implemented.

In regards to statistical analysis, developing a procedure using GIS to methodically group sample data together based on geographical distance could greatly benefit the RSMK analysis and outcomes. Conducting the RSMK test again using re-grouped datasets could strengthen the results from this project and instill greater confidence in the trends observed.

Other studies that could complement the current study are recommended below:

1. Develop a web interface for the database that can be accessed via TDEC's website. This would allow more users to easily retrieve nutrient data from the Tennessee Nutrient Database.
2. Map/visualize nutrient data availability across watersheds or streams.

3. Evaluate streams in Tennessee for nutrient load trend analysis.
4. Conduct a nutrient trend analysis along with land-use change to evaluate factors that may be influencing the increasing or decreasing trends.

References

Alexander, R. B., Smith, R. A., Schwarz, G. E., Boyer, E. W., Nolan, J. V., & Brakebill, J. W. (2008). Differences in phosphorus and nitrogen delivery to the Gulf of Mexico from the Mississippi River Basin. *Environmental Science & Technology*, *42*(3), 822-830.

Conservation Biology Institute. Data Basin. (2022). https://databasin.org/maps/df7fa3b1a0cc4ee997a677a29b6e9523/active/#. Accessed 01/16/2022.

Dodds, Walter K., Wes W. Bouska, Jeffrey L. Eitzmann, Tyler J. Pilger, Kristen L. Pitts, Alyssa J. Riley, Joshua T. Schloesser, and Darren J. Thornbrugh. (2009). Eutrophication of US freshwaters: analysis of potential economic damages. *Environmental Science & Technology* *43*(1), 12-19.

Dowle M, Srinivasan A. (2021). _data.table: Extension of `data.frame`_. R package version 1.14.2, https://CRAN.R-project.org/package=data.table

Esri ArcGIS. (2022). TDEC HUC8 and EPA Ecoregions Level III and IV. https://www.arcgis.com/home/webmap/viewer.html?url=https%3A%2F%2Ftdeconline.tn.gov%2Farcgis%2Frest%2Fservices%2FHUC8%2FMapServer&source=sd. Accessed 01/16/2022.

Grolemund, G. and Wickham, H. (2011). Dates and Times Made Easy with lubridate. Journal of Statistical Software, 40(3), 1-25. https://www.jstatsoft.org/v40/i03/.

Heisler, John, Patricia M. Glibert, JoAnn M. Burkholder, Donald M. Anderson, William Cochlan, William C. Dennison, Quay Dortch et al. (2008). Eutrophication and harmful algal blooms: a scientific consensus. *Harmful Algae* 8, no. 1, 3-13.

Helsel, D. (2005). Nondetects and data analysis: *Statistics for censored environmental data*.

Helsel, D. R., & Frans, L. M. (2006). Regional Kendall test for trend. *Environmental Science & Technology*, 40(13), 4066-4073.

Henry L, Wickham H. (2020). _purrr: Functional Programming Tools_. R package version 0.3.4, https://CRAN.R-project.org/package=purrr.

Howarth, R. W., Sharpley, A., & Walker, D. (2002). Sources of nutrient pollution to coastal waters in the United States: Implications for achieving coastal water quality goals. *Estuaries*, *25*(4), 656-676.

Marchetto, A. (2021). _rkt: Mann-Kendall Test, Seasonal and Regional Kendall Tests_. R package version 1.6, https://CRAN.R-project.org/package=rkt.

McIsaac, G. F., David, M. B., & Gertner, G. Z. (2016). Illinois River nitrate-nitrogen concentrations and loads: Long-term variation and association with watershed nitrogen inputs. *Journal of environmental quality*, *45*(4), 1268-1275.

Murphy, J., & Sprague, L. (2019). Water-quality trends in US rivers: Exploring effects from streamflow trends and changes in watershed management. Science of the Total Environment, 656, 645-658.

Oelsner, G. P., & Stets, E. G. (2019). Recent trends in nutrient and sediment loading to coastal areas of the conterminous US: Insights and global context. *Science of the Total Environment*, *654*, 1225-1240.

R Core Team. (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

RStudio Team. (2022). RStudio: Integrated Development Environment for R. RStudio, PBC, Boston, MA. URL http://www.rstudio.com/.

Schuenemeyer, J. H., & Drew, L. J. (2011). Statistics for Earth and Environmental Scientists.

Smith, V. H. (2003). Eutrophication of freshwater and coastal marine ecosystems a global problem. *Environmental Science and Pollution Research*, *10*(2), 126-139.

Sprague, L. A., & Lorenz, D. L. (2009). Regional nutrient trends in streams and rivers of the United States, 1993− 2003. Environmental science & technology, 43(10), 3430-3435.

TDEC 303(d) List. https://www.tn.gov/environment/program-areas/wr-water-resources/water-quality/water-quality-reports---publications.html.

TDEC. (2015). Tennessee Nutrient Reduction Framework. https://www.tn.gov/content/dam/tn/environment/water/tmdl-program/wr-ws_tennessee-draft-nutrient-reduction-framework_030315.pdf

TDEC, n.d. Increasing Stakeholder Awareness and Participation in Tennessee's Nutrient Reduction Strategy.

TDEC: Laster, K. & Moore, N. (2021). "Re: MDL/Not Detected Data." Received by Grace McClellan and Tania Datta, 20 December. 2021. Email.

US Environmental Protection Agency. (2000). Nutrient Criteria Technical Guidance Manual: Rivers and Streams. (EPA-822-B-00-002). Office of Water. https://www.epa.gov/sites/default/files/2018-10/documents/nutrient-criteria-manual-rivers-streams.pdf

Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, https://ggplot2.tidyverse.org.

Wickham, H. (2019). _stringr: Simple, Consistent Wrappers for Common String Operations_. R package version 1.4.0, https://CRAN.R-project.org/package=stringr

Wickham, H., François, R., Henry, L., Müller, K. (2022). _dplyr: A Grammar of Data Manipulation_. R package version 1.0.9, https://CRAN.R-project.org/package=dplyr

Wickham, H., Girlich M. (2022). _tidyr: Tidy Messy Data_. R package version 1.2.0, https://CRAN.R-project.org/package=tidyr

Woodward, G., Gessner, M. O., Giller, P. S., Gulis, V., Hladyz, S., Lecerf, A., Malmqvist, B., McKie, B.G., Tiegs, S. D., Cariss, H., Dobson, M., Elosegi, A.,Ferreira, V., Graca, M.A.S., Fleituch, T., Lacoursiere, J.O., Nistorescu, M., Pozo, J., Risnoveanu, G., Schindler, M., Vadineanu, A., Vought, L.B.M. Chauvet, E. (2012). Continental-scale effects of nutrient pollution on stream ecosystem functioning. *Science*, *336*(6087), 1438-1440.