Evaluation of Sampling Frequency Impact on the Accuracy of Water Quality Status as Determined Considering Different Water Quality Monitoring Objectives

Camilo Torres · Margaret W. Gitau · Diego Paredes-Cuervo · Bernard Engel ·

Camilo Torres · Margaret W. Gitau* · Bernard Engel
Department of Agricultural and Biological Engineering, Purdue University, West Lafayette, IN, USA
*Corresponding author e-mail: mgitau@purdue.edu

Diego Paredes-Cuervo
Department of Basic Environmental Sciences, Universidad Tecnológica de Pereira, Pereira, Risaralda, Colombia

Camilo Torres ·
Department of Civil Engineering, Pontificia Universidad Javeriana, Bogotá D.C, Colombia

Abstract Water quality sampling is a key element in tracking water quality monitoring objectives. However, frequencies adapted by different agencies might not be sufficient to provide an accurate indication of water quality status. In this study, data from low and high-resolution water quality datasets were analyzed to determine the extent to which monitoring objectives could be achieved with different sampling frequencies, with a view to providing recommendations and best practices for water quality monitoring frequency in places with limited resources with which to implement a high-frequency monitoring plan. Water quality data from two watersheds (Maumee River and Raisin River) located in the Western Lake Erie Basin (WLEB) were used since these watersheds have consistent records over substantial periods of time, and the water quality data available have a high resolution (at least daily). The water quality constituents analyzed included suspended solids (SS), total phosphorus (TP), soluble reactive phosphorus (SRP), and nitrate+nitrite (NO_{2+3}). Sources of pollutants for watersheds located in the WLEB include contributions from point sources like discharges from sewage treatment plants and non-point sources such as agricultural and urban storm runoff. Weekly, bi-weekly, monthly and seasonal datasets were created from the original datasets, following different sampling rules based on the day of the week, week of the month, and month of the year. The resulting datasets were then compared to the original dataset to determine how the sampling frequency would affect the results obtained in a water quality assessment when different monitoring objectives are considered. Results indicated that constituents easily transported by water (such as sediments and nutrients) require more than 50 samples/year to provide a small error (<10%) with a confidence interval of 95%. Monthly and seasonal sampling were found appropriate to report a stream's prevailing water quality status and statistical

properties. However, these resolutions might not be sufficient to capture long-term trends, in which case bi-weekly samples would be preferable. Limitations of low-resolution sampling frequency could be overcome by including rainfall events and random sampling during specific time windows as part of the monitoring plan.

Keywords Water Quality · Sampling Frequency · Monitoring Objectives   Water Quality Indices

Declarations

Funding This work was funded in part by USDA National Institute of Food and Agriculture, Hatch Project IND00000752

Conflict of interest The authors declare no competing interests

Availability of data and material The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request

Code availability Not applicable

Acknowledgments The authors are thankful to Maria Camila Fernandez for her work testing and conducting part of the exploratory analysis with the water quality data. Mr. Torres is also grateful to the Fulbright Program and the Colombian Institute of Educational Credit and Technical Studies Abroad (ICETEX) for providing the funding for his doctoral studies under the “Fulbright-Pasaporte a la Ciencia – Foco Alimentos” fellowship program.

Introduction

Water quality monitoring protocols have been implemented as a way to understand water quality processes and to obtain water quality variables for planning, designing, and operating water resources and wastewater treatment systems (Harmancioğlu, 1999). Whitfield (1988) indicated that some of the reasons for conducting water quality monitoring are: 1) assessment of trends in variables of concern, 2) compliance with standards, 3) estimation of mass transport, 4) assessment of environmental impact, and 5) general surveillance. To incorporate a temporal basis for water quality monitoring, Sherwani & Moreau (1975) defined short-term, intermediate-term, and long-term goals. The primary purposes for short-term goals were to: monitor and investigate complaints; prevent water pollution emergencies (e.g., fish kills); set, amend, or repeal water quality standards; and, develop effluent standards. Intermediate and long-term goals focused on: evaluating the effectiveness of activities aimed to control water pollution; determining the nature and extent of pollution in areas of interest; long-range program and policy planning; evaluation of trends; and, prediction of water quality status.

Depending on monitoring goals, different sampling frequencies have been adopted by different entities. In the National Water Quality Assessment (NWQA) program developed by the U.S. Geological Survey (USGS), for example, the minimum and most common sampling frequency for basic fixed-sites is monthly for two years, while high-frequency sampling on intensive fixed-sites is typically weekly (or bi-weekly for large basins) and lasts between 3 to 9 months (Gilliom et al., 1995). In 2013, the U.S. Geological Survey National Water Quality Network (NWQN) was formed with the merger of the National Stream Quality Accounting Network (NASQAN) and the NWQA. Under this new program, the frequency of sample collection is determined based on historically observed variability in water concentrations and pollutant loads. Samples are collected through a seasonal weighted, fixed-interval regime, and the number of samples ranges between 12 to 18 samples collected in a year (Lee et al., 2017). In 2018, the USGS started a pilot of the Next Generation Water Observing System (NGWOS) in the Delaware River Basin. This program involves continuous monitoring of temperature and specific conductance is being conducted in selected streamflow stations, as well as the use of remote sensing technologies to monitor suspended-sediment concentration, channel erosion, and harmful algal blooms (USGS, 2021). The U.S. Environmental Protection Agency (USEPA), under the National Rivers and Streams Assessment (NRSA) program, collects data in a single site visit in the summer for two years to provide a national snapshot of a stream's water quality over time as captured during the sampling period. The NRSA is conducted every five years, with the last assessment completed in 2018-2019 (U.S. EPA, 2020).

The European Union's Water Framework Directive recommended monthly sampling should be completed for priority substances and every three months for other pollutants. However, the Directive allows that Member States define their own monitoring frequency based on the conditions and variability of their water bodies. The only requirement that Member States have to meet for selecting their own frequency is that this frequency provides a reliable assessment of the status of all water bodies (European Commission & Directorate-General for the Environment, 2003). In Australia and New Zealand, the sampling frequency depends on the water monitoring objectives and the jurisdiction where the study takes place. The Queensland's Monitoring and Sampling Manual indicates that monthly sampling is usually adequate for baseflow (dry water concentration) water quality sampling; however, the frequency can be adjusted based on the monitoring goals (DES, 2018).

Table 1 shows a summary of sampling frequencies used for different water quality monitoring objectives. Regulatory agencies and frameworks suggest sampling frequencies ranging from monthly to annually. However, recent research indicates that some of the frequencies indicated by regulatory agencies might not be sufficient to accomplish specific water quality monitoring objectives (Bowes et al., 2009; Halliday et al., 2015; Ross et al., 2015; Vilmin et al., 2018). Vilmin et al. (2018) concluded that it is not possible to specify one single optimal sampling frequency under the E.U. Water Directive Framework (WFD); instead, the optimal sampling should be defined for each variable and location. The authors disclosed that major urban effluents increase the variability in the receiving environment; therefore, weekly sampling may be needed to capture this variability. Babitsch et al. (2021), after comparing monthly samples with subsamples of 10, 8, 6, 4, 2, and 1 measurements, concluded that low sampling frequency reduces the reliability of temporally variable water quality data. Kotamäki et al. (2019) identified that more sampling is required beyond the recommendation provided by the WFD, especially for rivers with a class status of "High," "Bad," or "Poor. The researchers also indicated that the frequency and coverage of monitoring designs should be systematically and iteratively evaluated in relation to monitoring objectives for the water body.

According to the United Nations Environment Program (UNEP, 2016), sampling frequencies in the U.S. can vary from hourly to annually depending on the purpose of sampling. Conversely, the average monitoring frequency for Latin America was four samples per year from 1990 to 2010. The Colombian Andean Region follows a trend similar to that documented for the Latin American region. Water quality sampling is conducted three to four times a year, generally considering the annual wet and dry periods. Recent studies completed in Colombia indicated that monitoring with a higher temporal and spatial resolution is desirable (Díaz-Casallas et al., 2019; Holguin-Gonzalez et al., 2013; Rodríguez et al., 2013). These authors stated that there are periods with no information and the ecological water quality assessment may be limited.

Sampling frequency and recurrence are widely discussed in handbooks, official guidelines, and scientific papers, especially when optimization of water quality monitoring programs is discussed (Behmel et al. 2016). Several issues emerge from work and discussions on sampling frequencies: 1) Even though sampling frequencies are often based on water quality monitoring objectives, there does not seem to be any consensus as to the frequency to be adopted for any one objective. For example, (Naddeo et al., 2013) concluded that the assessment of the water environmental quality did not change with seasonal sampling in the Sele River Basin. In contrast, (Halliday et al., 2015) stated that sampling frequency and collection time had a significant impact on water quality status under the WFD. In the U.S., the EPA and the USGS have established monitoring programs with seasonal or annual sampling with the objective of assessing the streams' water quality status; however, at the state level, environmental agencies may set monthly sampling as the sampling frequency to accomplish the same monitoring objective (Florida Department of Environmental Protection, 2020; Indiana Department of Environmental Management (IDEM), 2017; Pennsylvania Department of Environmental Protection, 2019); 2) Water quality monitoring programs are better defined and established in North America and Europe. For the most part, the different planning and decision-making agencies can access relevant, accurate, and up-to-date information about water quality status. In many other parts of the world, there is less water quality information, and water quality monitoring may be less structured. These are also the regions in which resource constraints greatly impact water quality monitoring, hence the need for more definitive information on sampling frequencies; and, 3) The value of high-frequency sampling over lower-frequency sampling needs to be established (Jiang et al., 2020). Furthermore, there is the need to identify and eliminate redundancy (Guerreiro et al., 2020; Karamouz et al., 2009), this being the point beyond which higher frequency sampling does not result in a gain in information.

The purpose of this study is to provide recommendations and best practices for water quality monitoring frequency, particularly for places with limited resources to implement a high-frequency monitoring plan. Specifically, to: 1) Compare extent and applicability of water quality information obtained based on different sampling frequencies; 2) Assess the performance of the frequencies based on different monitoring objectives using long-term datasets with a high resolution; 3) Develop recommendations for water quality frequencies considering potential regional constraints.

Materials and Methods

Water quality data from the Western Lake Erie Basin (WLEB, Fig. 1) were used. This basin was selected because a substantial amount of data was readily available, with water quality parameters sampled at high resolutions, in some cases up to four times a day. An exploratory analysis of the data from the Maumee watershed led to the following critical question: how does monitoring frequency affect the accuracy of information obtained considering different objectives? From this question, five sub-questions were formulated to narrow the scope of the main question. The sub-questions are:

a) Is there an ideal sampling frequency to identify changes in water quality parameters?
b) What would be the adequate water quality sampling frequency to assess long-term trends?
c) Can monthly and seasonal sampling be improved in a way that can provide more accurate and reliable results for long-term trend monitoring?
d) How does the interpretation of water quality status change if sampling is conducted at different times during the year?
e) What are the impacts of different sampling frequencies on water quality status as reported using Water Quality Index (WQI)?

To provide answers to these questions, a combination of different analysis were completed, including the estimation of the number of samples required in a year using the water quality means methodology, a trend analysis using a Mann-Kendall trend test, and a comparison of the Water Quality Sub-Indexes proposed by Mijares et al. (2019). Data pre-processing and statistical analysis

Table 1 Summary of sampling frequencies based on different monitoring objectives

Agency/Reference	Purpose of Monitoring	Frequency	Notes
U.S. Environmental Protection Agency (EPA) National Rivers and Streams Assessment (NRSA) (U.S. EPA, 2020)	Establish a baseline of the condition of wadeable streams and extent of major stressors Assessment of long-term trends.	Single sample every 5 years	1853 sites sampled. Site selection based on a stratified random sampling design
Water Framework Directive – European Commission (EC) (2000/60/E.C.) (EC, 2003)	Establish the status of those bodies identified as being at risk of failing to meet their environmental objectives	Monthly for priority substances and every three months for other pollutants	45 pollutants are considered priority substances, including heavy metals, PAHs, and POPs. More frequent sampling may be necessary to detect long-term trends
U.S. Geological Services (USGS) National Water-Quality Assessment (NAWQA) Project (Gilliom et al., 1995)	Assessment of current status and long-term trends	Streams: bi-monthly and 8-18 seasonally weighted samples Agricultural and urban sites: monthly and 12 seasonally weighted samples	110 sites (streams and rivers) with consistent streamflow and water-quality information
Indiana Department of Environmental Management (IDEM) (IDEM, 2017)	Assessment of current status and long-term trends	Fixed stations: monthly Probabilistic selected sites: 3 seasonal samples (May-Oct) Targeted sites: At least 3 samples per year	165 fixed sites, 38-50 probabilistic selected sites. Targeted locations are selected based on a variety of factors depending on monitoring objectives, including known impairments, permitted dischargers, land use, etc. Sites change annually
Florida Department of Environmental Protection (FDEP) (FDEP, 2020)	Assessment of current status and long-term trends	Trend monitoring network: monthly Status monitoring network: 1 sample	- 78 fixed stations for trend monitoring - 240 sites are sampled for status monitoring (canals: 60, streams: 90, rivers: 90) Sampling occurred in Jan-Feb (canals), Apr-May (rivers), Jul-Sep (streams)
Pennsylvania Department of Environmental Protection (PADEP) (PADEP, 2019)	Assessment of current status and long-term trends	Standard stations: monthly (physical/chemical) and annually (biological)	CIM measure water temperature, specific conductance, pH, and dissolved oxygen and are completed for one year or less to capture time periods of specific interest. Several CIM

Agency/Reference	Purpose of Monitoring	Frequency	Notes
	Assessment of effluent limitations for the National Pollutant Discharge Elimination System (NPDES) permits	Chesapeake Bay stations: monthly (physical/chemical), every other year (biological), 8 times/year during storm events Reference stations: monthly (physical/chemical) and annually (biological) Continuous Instream Monitoring (CIM): every 15 min	deployments were maintained for multiple years to understand year-to-year differences and observe trends
New York City Department of Environmental Protection (NYCDEP) (NYCDEP, 2019)	Assessment of current status and long-term trends	Fixed frequency: monthly Automated stream monitoring (ASM): every 15min	100 fixed frequency sampling sites and 6 ASM stations. ASM stations monitoring for water temperature, specific conductivity, and turbidity
State of Queensland Department of Natural Resources, Mines, and Energy (DES, 2018)	Assessment of condition and trend of Queensland's freshwater aquatic ecosystem health	Manual samples were collected 1-4 times a year. Continuous time-series measurements of temperature and electrical conductivity at selected stations	161 stations (68% of all stations) have the equipment to continuously measure temperature and electrical conductivity. Manual water quality sampling is conducted in 229 stations
IIHR Hydroscience & Engineering (Weber et al., 2016)	Research	Samples were taken every 15 min	28 sites in Iowa monitoring for nitrate and nitrite, turbidity, temperature, specific conductance, pH, and dissolved oxygen.
Skeffington et al. (2015)	Research	Hourly (3 sites) and 2-4 times/day (1 site)	4 sites were assessed in the U.K. Parameters analyzed included pH, temperature, dissolved oxygen, and phosphorus (T.P., SRP, ortho-phosphorus)
The Heidelberg Tributary Loading Program (HTLP) (Roerdink, 2017)	Research	3 times/day (04:00, 12:00, 20:00)	18 sites in Ohio and Michigan, sampling for nutrients (especially phosphorus), sediments, and pesticides

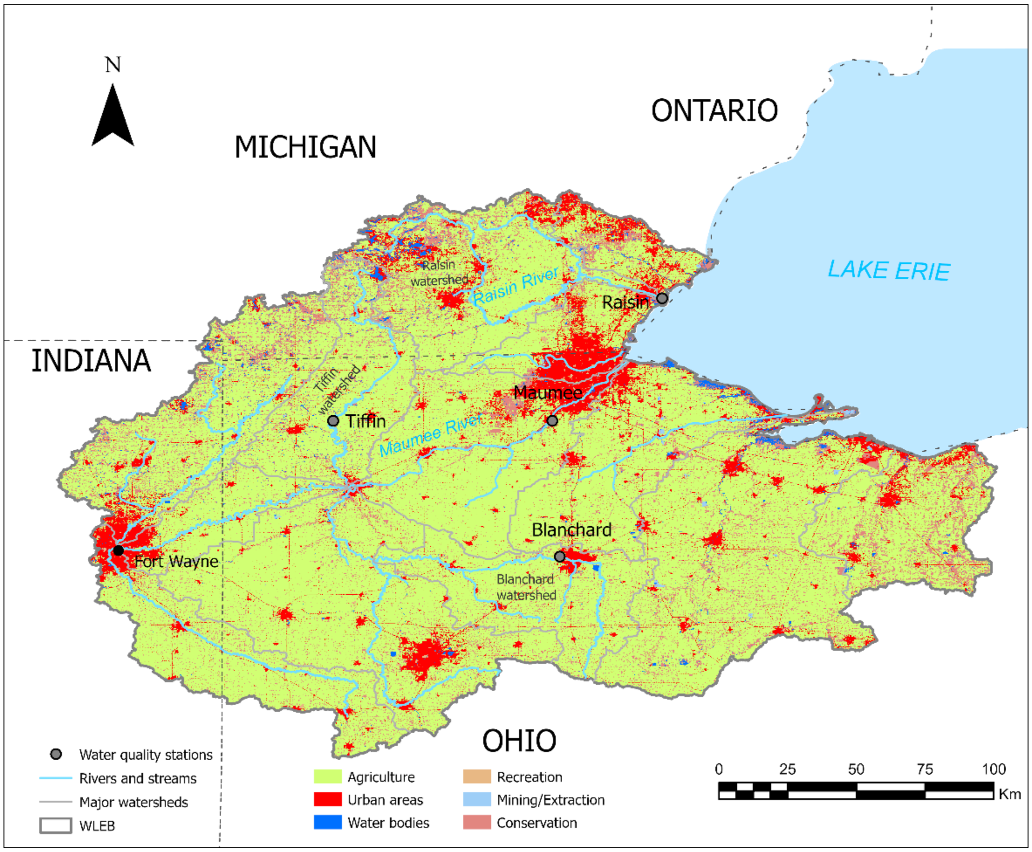

Fig. 1 Western Lake Erie Basin (WLEB) with the location of the water quality stations used in this study

were completed using R 4.0.0. Missing data were completed using embedded functions in R.

Data collection and preparation

Data from four water quality stations located in the Maumee (Maumee, Blanchard, and Tiffin stations) and Raisin (Raisin) Rivers watersheds (Fig. 1) were obtained from the Heidelberg's University's National Center for Water Quality Research (NCWRQ), part of the Heidelberg Tributary Loading Program (HTLP). Data for suspended solids (SS), total phosphorus (TP), soluble reactive phosphorus (SRP), and nitrates+nitrites (NO_{2+3}) were obtained for this study. Since water quality samples had been taken at varying intervals throughout the day, flow-weighted average daily values were computed and used for this analysis. These flow-weighted daily values constituted the baseline datasets for this study.

Data resampling

Similar to the method used by Tanos et al. (2015), a systematic approach was used to extract data from the original datasets and create datasets for use with the analysis. In particular, for: 1) weekly values, datasets were generated constituting samples taken on each day of the week (DOW) every week. This generated seven datasets with 52-54 samples/year/set; 2) bi-weekly values, datasets were generated constituting samples taken on each DOW every other week on even weeks and similarly for odd weeks. This generated 14 datasets with 26-27 samples/year/set; 3) monthly values, datasets were generated constituting samples taken on each DOW every month generating 31 datasets with 12 samples/year/set; and, 4) seasonal samples, datasets generated constituted samples taken on days 2, 6, 7, 10, 15, 16, 20, 21, 22, or 24 for March, June, September, and December; and a dataset for samples taken on Tuesday, Wednesday, or Thursday for week 1-4 generating 22 datasets with 4 samples/year/set. Four additional datasets were created by selecting random samples. These random datasets were created by selecting one sample from Tuesday, Wednesday, or Thursday from a randomly selected week 1-4, by setting various seed numbers and using R's replicate function.

Station Datasets

For the Maumee and Raisin stations, data were available at least daily from January 1986 to December 2015, and a total of 103 datasets were created for these stations. For the Tiffin and Blanchard stations, the data were primarily reported weekly from July 2007 to December 2015, and a total of 96 datasets were created. After creating the datasets, we conducted a comparison of statistical properties and essential characteristics among datasets.

Since a comprehensive comparison among all datasets within each category would be burdensome and might not necessarily provide additional information to address the study's objective, we analyzed datasets at each sampling frequency to see if there were appreciable differences among the datasets. Additionally, we conducted a Kruskal-Wallis test (Table S2) to check if the datasets for each sampling frequency had an identical distribution. Based on this analysis, subsets were extracted at random from the datasets at each sampling frequency and used for further analysis. To check if the selected datasets were from the same distribution as the original dataset, a Kolmogorov-Smirnov test was completed. The statistical properties evaluated in this analysis were mean, standard deviation, standard error, median, minimum, maximum, 90, 95, and 99 quantiles, skewness, and kurtosis. The essential characteristics were: number of samples over

90, 95, and 99 percentile; number of samples over thresholds; number of samples over the threshold in each season, and water quality sub-indices. This last characteristic was only estimated for monthly and seasonal datasets. A summary of the most relevant statistical properties is included in Table 2.

Ideal Sampling Frequency

As part of formulating a water quality monitoring program, agencies face the question of defining the minimum number of samples required to meet the water quality monitoring objectives. One of the most commonly used methods to estimate the number of samples is based on water quality means. This method has the purpose of defining a rational sampling frequency criterion based on the relationship between sampling frequency and the magnitude of half confidence interval of the annual mean variable concentration (Sanders & Adrian, 1978).

Table 2 Summary statistics for daily, monthly, and seasonal sampling frequencies (mg/L)

	Maumee			Raisin			Tiffin			Blanchard		
	Daily	Monthly	Seasonal	Daily	Monthly	Seasonal	Daily*	Monthly	Seasonal	Daily*	Monthly	Seasonal
Suspended Solids												
Mean	72.00	72.63	71.57	48.71	50.95	50.73	54.43	63.24	36.68	35.85	34.38	24.71
SD	100.08	94.02	90.56	93.05	138.73	87.51	70.15	94.55	25.89	66.58	60.61	36.30
Total Phosphorus												
Mean	0.224	0.228	0.230	0.120	0.116	0.123	0.179	0.186	0.144	0.280	0.291	0.275
SD	0.146	0.141	0.144	0.116	0.131	0.114	0.144	0.158	0.101	0.155	0.144	0.134
Soluble Reactive Phosphorus												
Mean	0.0548	0.0582	0.0576	0.0241	0.0263	0.0257	0.0542	0.0554	0.0497	0.1548	0.1706	0.1678
SD	0.0432	0.0436	0.0440	0.0248	0.0229	0.0227	0.0372	0.0346	0.0361	0.1032	0.1220	0.1155
Nitrate+nitrite												
Mean	4.40	4.53	4.69	2.90	2.93	2.69	3.16	2.91	2.79	5.70	6.30	6.01
SD	3.121	3.073	3.089	2.304	2.324	2.121	2.317	1.599	1.566	2.531	3.088	2.679

*Sampling frequency for Tiffin and Blanchard varies from daily to weekly sampling

These researchers used streamflow data instead of water quality parameters since high-frequency sample were not available. Additionally, using streamflow data is a logical assumption since some water quality constituents (sediments, major ions, and salts) are highly correlated to the river flow. Our study compares the number of samples required based on the streamflow, SS, TP, and SRP data obtained for the Maumee and Raisin Rivers. NO_{2+3} data were not included because it does not meet the condition of having an approximated normal distribution. This method established that the half-expected confidence interval of the mean (R) is a function of the standard deviation of the observed residuals (S), the constant from the Students' t-distribution (t), and the number of samples (n). The equation used to estimate the values of R is:

$$R = \frac{t_{\alpha/2}S}{\sqrt{n}} \tag{1}$$

Before applying the aforementioned method, the data must be modeled to isolate random, independent, and identically distributed residuals. To achieve this purpose, the methodology indicated by Sanders and Adrian (1978) was used, where the long-term trend component is removed from the time series. Then the time series is converted to natural logs, and a first-order autoregressive moving average model is created. Finally, the standard deviation is calculated and used in equation 1.

Sampling Frequency for Long-Term Trend Monitoring Goals

Water quality monitoring objectives often include the identification of long-term trends. Therefore, a trend analysis was conducted to identify if there was a difference in the results obtained from this analysis for the different sampling frequencies used in this study. Since daily, weekly, bi-weekly, and monthly datasets may be serial correlated, the seasons' means were used in the trend analysis. Once the means were estimated, we conducted a Mann-Kendall test to assess if there were trends in the water quality parameters during the observed period. The analysis was performed for each of the sampled seasons (Winter, Spring, Summer, and Fall).

Improving Monthly and Seasonal Sampling for Long-Term Trends

Since monthly and seasonal sampling has been reported as potentially inadequate to capture long term trends (László et al., 2007; Vilmin et al., 2018), the option to include samples corresponding to rainfall events was considered as an alternative to improve the accuracy and reliability of the results obtained from trend analysis. Therefore, an initial comparison between wet days and dry days was completed to check if there was a significant difference between both types of days. For the purpose of this study, a wet day was defined as a day that has a rainfall equal to or greater than 0.1 mm. The metric used for this analysis was the Water Quality Index as defined by Mijares et al. (2019), computed for each of the sampled seasons. Water Quality Indexes (WQIs) are simple methods by which to summarize the water quality status of a water body by grouping values of different water quality parameters into a single value. One of the issues associated with WQI is that there are different WQI models currently in use (Gitau et al., 2016).

Water quality sub-indexes (WQSI) are transformations that allow values of specific constituents to be expressed on a common scale (U.S. EPA, 2009). As defined by Mijares et al. (2019), the calculation of these sub-indexes is based on the respective constituent's threshold value. The estimated sub-index numerical value can range from 0-100; sub-index values below 40 indicate that the water quality falls below the constituent threshold. The ratings defined by Mijares et al. (2019) are as follows: 90-100: Pristine; 70-89: Good; 50-69: Fair; 40-49: At Risk; 30-39: Poor; 0-29: Unsuitable for all uses. For seasonal and annual computations, we used flow-adjusted concentrations to estimate the sub-indexes. Once the comparison was completed, monthly and seasonal datasets were complemented with additional observations corresponding to storm events. Storm samples were defined using streamflow data from each of the stations. A storm sample constituted data points where the daily streamflow exceeded the annual 90th percentile of the daily streamflow distribution.

Different combinations including storm samples (monthly+storm samples and seasonal+storm samples) were used to create additional datasets. The initial combination included monthly or seasonal samples plus eight storm samples, which follows the approach used by Zhang and Hirsch (2019). The remaining datasets included the monthly or seasonal samples plus 12, 16, 20, and all storm samples.

Variation in water quality status depending on the day of sampling

When assessing water quality status, a common question that arises is if the status varies depending on the day when sampling is conducted. To provide an answer to this question, a comparison between water quality sub-indexes was completed. The initial comparison was conducted for summer since a higher concentration of sediments and nutrients is expected during this season due to spring and summer showers and farming activities conducted during the summer months. The dates selected for the analysis were July 1st, July 30th, August 29th, and September 28th. This analysis was further expanded for the other seasons selecting four days, including the 1st, 30th, 60th, and 90th days of the season.

Analysis of Water Quality Status Reported as Water Quality Indexes (WQI) for different frequencies

Water quality sub-indexes were estimated for weekly, biweekly, monthly, and seasonal frequencies using the Mijares et al. (2019) method. Statistical properties for each frequency were then estimated and compared to identify if there was a significant difference in the water quality status as determined.

Results and Discussion

Data validation and statistical properties

Results from the Kolmogorov-Smirnov test validated that the datasets selected came from the same population ($p>0.05$). The statistical properties for the different sampling frequencies (Table S1) did not show a significant difference between the means except for seasonal samples, ranging from 7 to 15%. SS and TP were the constituents with a larger difference in the concentrations at the 99 percentile between daily and seasonal sampling.

Fig. 2 shows the distribution of water quality parameters for various sampling frequencies. Medians and interquartile range values (IQR) for all analyzed frequencies were similar with noticeable differences being primarily in the number of outliers. This was confirmed in the suspended solids time-series (Fig. 3) plot, where we observed that a reduction in the number of samples taken, from weekly to seasonal, represented a reduction in the peak concentration events observed. Moreover, the distributions varied among the different analyzed sampling frequencies. None of the analyzed frequencies was close in shape or magnitude to the daily distribution. The weekly, bi-weekly, monthly, and seasonal distributions were more evenly distributed, with shorter peaks, compared to the daily frequency.

These results were consistent with the findings reported by Ross et al. (2015). The reduced number of high concentration events captured in monthly and seasonal sampling may be troublesome since some of these events could exceed the allowable acute or chronic contaminant levels.

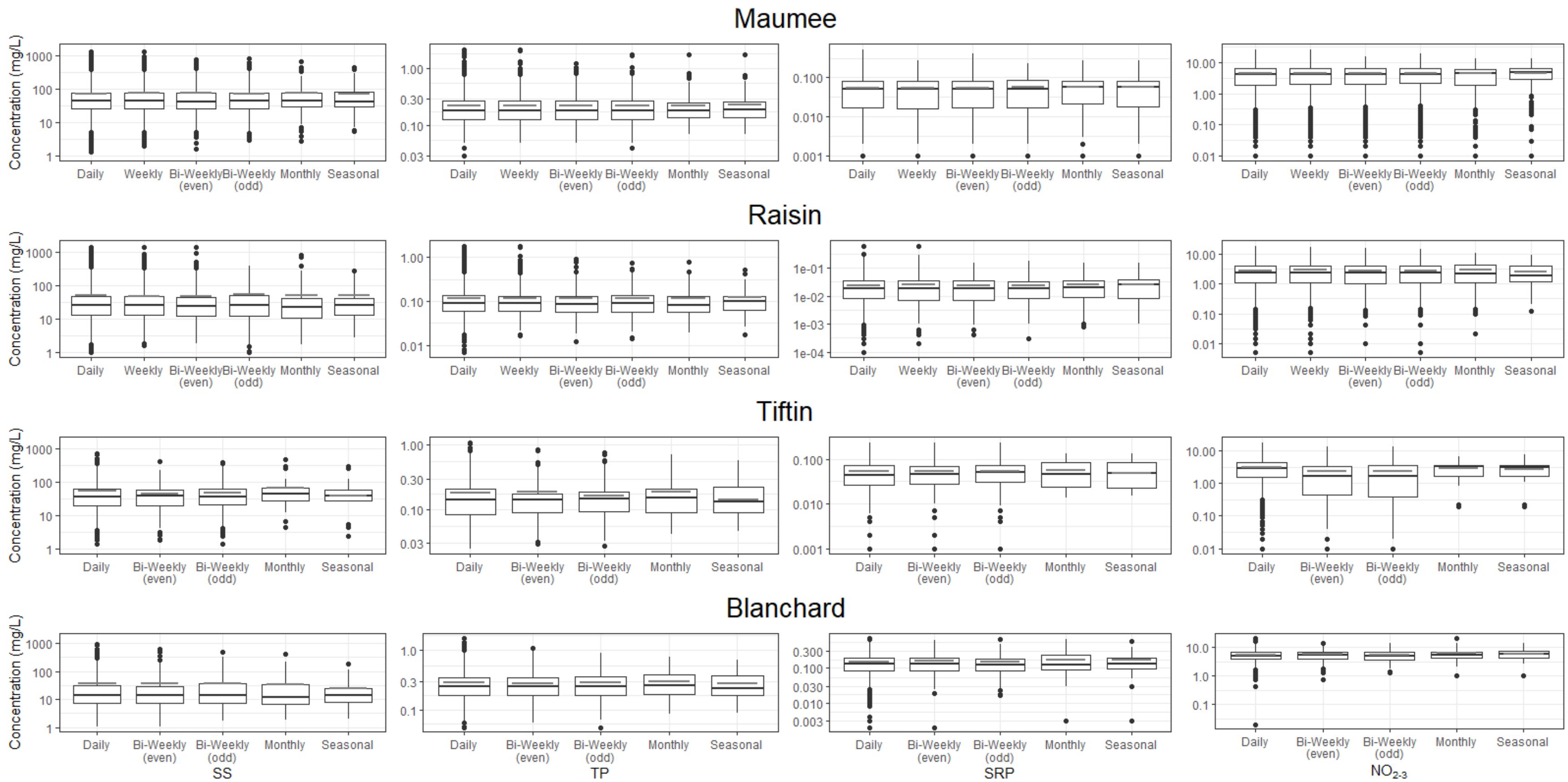

Fig. 2 Summary of the distribution for analyzed constituents for each sampling frequency in selected stations in the Maumee and Raisin watersheds.

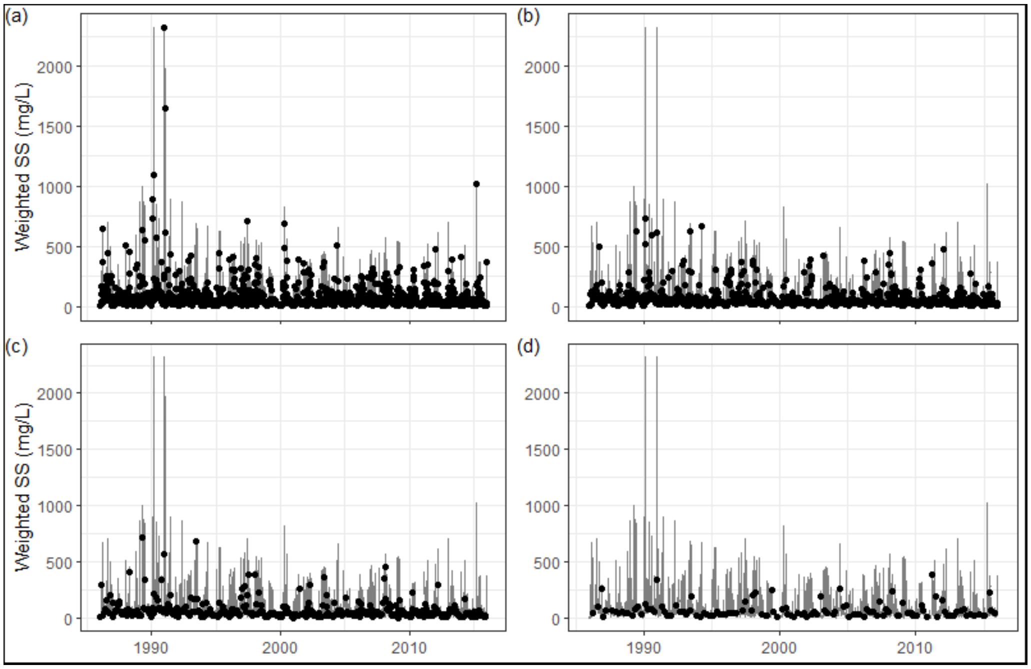

Fig. 3 Suspended solids time series plots comparing daily frequency sampling (grey line) with generated datasets simulating samples collected on a weekly (a), bi-weekly (b), monthly (c), and seasonal (d) basis.

<u>Estimation of the number of samples</u>

The datasets' resolution was sufficient to allow the use of the water quality means method to estimate the number of samples required for the sampled constituents at the selected stations in the Maumee and Raisin Rivers. Fig. 4 shows the magnitude of the confidence interval's half-width of the means' random component (R) for streamflow, SS, TP, and SRP as a function of the number of samples. The value of R rapidly decreases as the number of samples increases, indicating that a high number of samples (>100) reduces the values of R to 0.25 or less. Streamflow, SS, and SRP require more samples to reach an error equal to or below 10% (R=10), while the same error can be reached for TP with 50-100 samples, depending on the station. Even though mixed results have been found regarding the relationship between streamflow and water quality trends, Murphy and Sprague (2019) indicated that streamflow regimes more commonly influenced trends in major ions, salinity, and sediments, which coincide with the behavior observed between streamflow, SS, and SRP. In contrast, TP seemed to be independent of streamflow, which conforms with findings reported in the literature (Murphy & Sprague, 2019). Table 3 shows the values of R at the Maumee Station for the different sampling frequencies analyzed in this study. In relation to seasonal samples, it was observable that a 56% reduction in the value of R occurred when seasonal and monthly sampling were compared, and an additional 15% reduction was observed when values were compared with those from bi-weekly sampling. Additionally, the results suggest SRP is the constituent that should be sampled more frequently since the rate at which R decreases is slower compared to that for the other two constituents analyzed. A bi-weekly sampling frequency (26 samples) would give an R value of 0.411, which can be achieved with 7 and 11 samples for TP and SS, respectively.

Table 3 Summary of the magnitude of the confidence interval's half-width of the means' random component (R) for different sampling frequencies (Maumee Station)

Parameter	Frequency									
	Daily		Weekly		Biweekly		Monthly		Seasonal	
	n	R	n	R	n	R	n	R	n	R
Streamflow	365	0.115	52	0.311	26	0.448	12	0.693	4	1.573
Suspended Solids	365	0.083	52	0.223	26	0.321	12	0.497	4	1.128
Total Phosphorus	365	0.045	52	0.121	26	0.175	12	0.271	4	0.614
Soluble Reactive Phosphorus	365	0.106	52	0.285	26	0.411	12	0.636	4	1.442

* The table shows the values of R as a function of the number of samples. A low value of R (<0.10) indicates that the water quality data's variance is small, providing higher confidence that the data is a good representation of the streams' water quality. A large value of R (>0.50) indicates high variability in water quality data, which reduces the confidence that the data is representative for the streams' water quality.

Long-Term Trend Analysis

Fig. 5 shows the results for the trend analysis for suspended solids during winter, where the slope for the decreasing and increasing trend for seasonal sampling was greater compared to the other frequencies. Table S3 to S6 show a summary of the results obtained from the Mann-Kendall trend analysis. Overall, the tau and p-values obtained were similar for the daily, weekly, and bi-weekly samples. However, the results for monthly and seasonal sampling varied among them, making them less reliable as a tool for identifying trends. This finding is consistent with what was reported by Raimonet et al. (2015), who indicated that monthly sampling is not sufficient to assess water quality for highly variable constituents like ammonia and nitrite. The aforementioned authors also concluded that other constituents' fluxes like for nitrate could be captured with monthly sampling. This finding could not be verified for the Maumee River since nitrogen concentrations were reported as nitrate+nitrite (NO_{2+3}). However, the nitrite variability

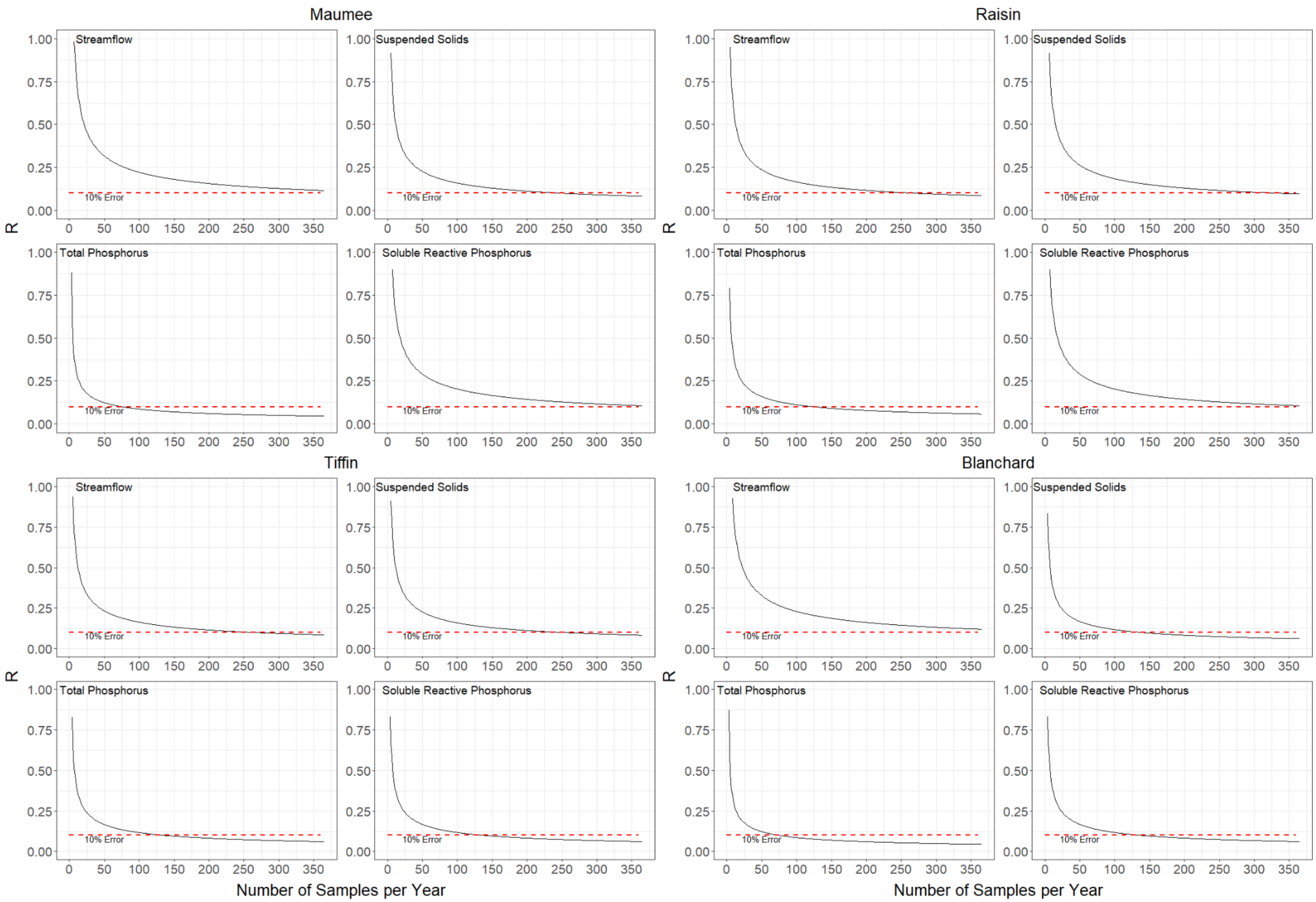

Fig. 4 Magnitude of the confidence interval's half-width of the means' random component (R) for streamflow, suspended solids (SS), total phosphorus (TP), and soluble reactive phosphorus (SRP) versus the number of samples per year for four water quality stations located in the Maumee and Raisin watersheds. The red line represents the threshold to achieve a 10% error.

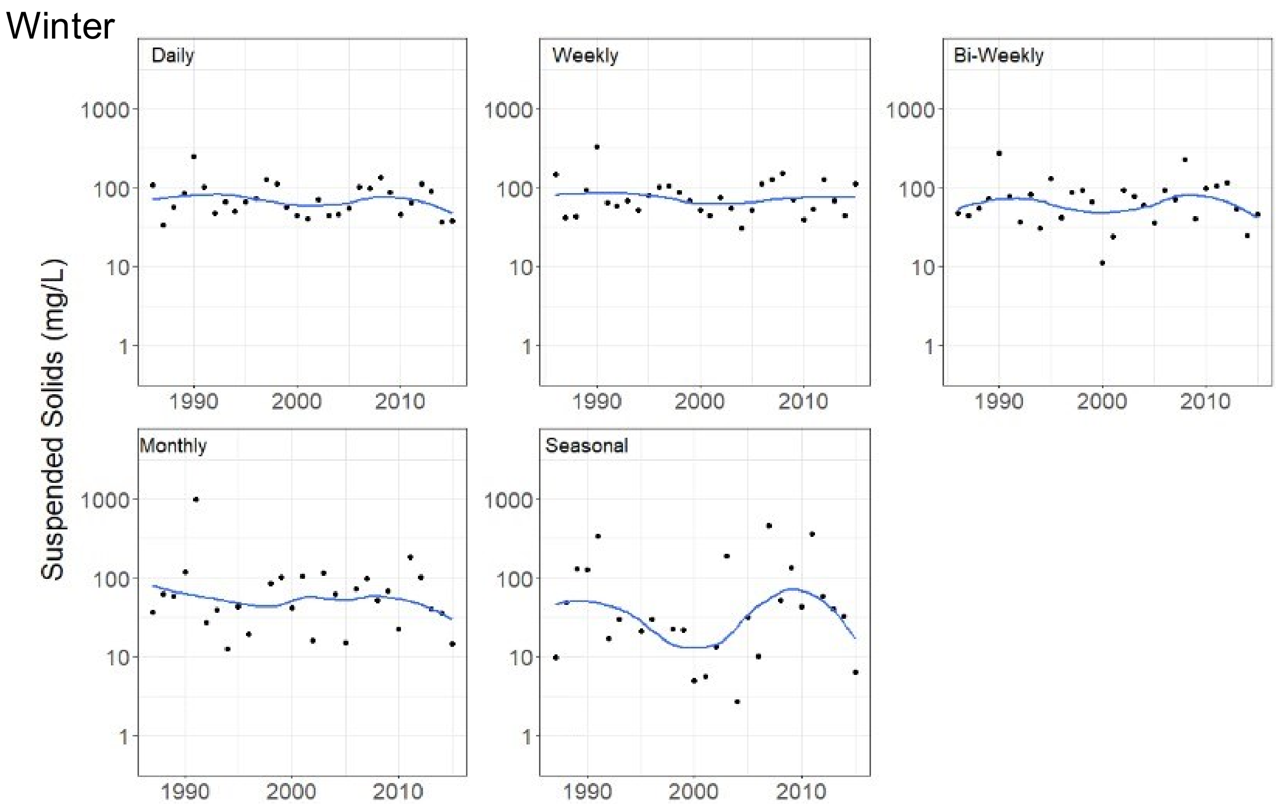

Fig. 5 Trend results for different sampling frequencies (daily, weekly, bi-weekly, monthly, and seasonal) for Maumee Station.

tended to be more predominant since the trend analysis results for nitrate+nitrite had the highest variability for the constituents sampled for the Maumee River.

<u>Monthly and Seasonal Sampling for Long-Term Trends</u>

Fig. 6 and Fig. 7 show a comparison for water quality parameters among monthly, seasonal, and monthly+storm events sampling. The monthly+storm events sampling captured more high concentration events for all constituents except for NO_{2+3} (Fig. 6). The inability to capture high concentration events for NO_{2+3} may be attributable to the dilution of nitrate during the early stages of the storm events (Blaen et al., 2017; Outram et al., 2014). Fig. 7 shows the distributions for the different sampling schemes and it was evident that the monthly+storm events sampling median were higher by between 17% to over 100% compared to those obtained from monthly and seasonal sampling. Discrepancies were highest for suspended solids (102% - 114%), followed by TP (50% - 58%), SRP (22% - 29%), and NO_{2+3} (17%) for the Maumee station. These findings are consistent with the results reported by (Chanat et al., 2016), who indicated that phosphorus and suspended solids had the more pronounced relative difference in medians when monthly+storm event samples

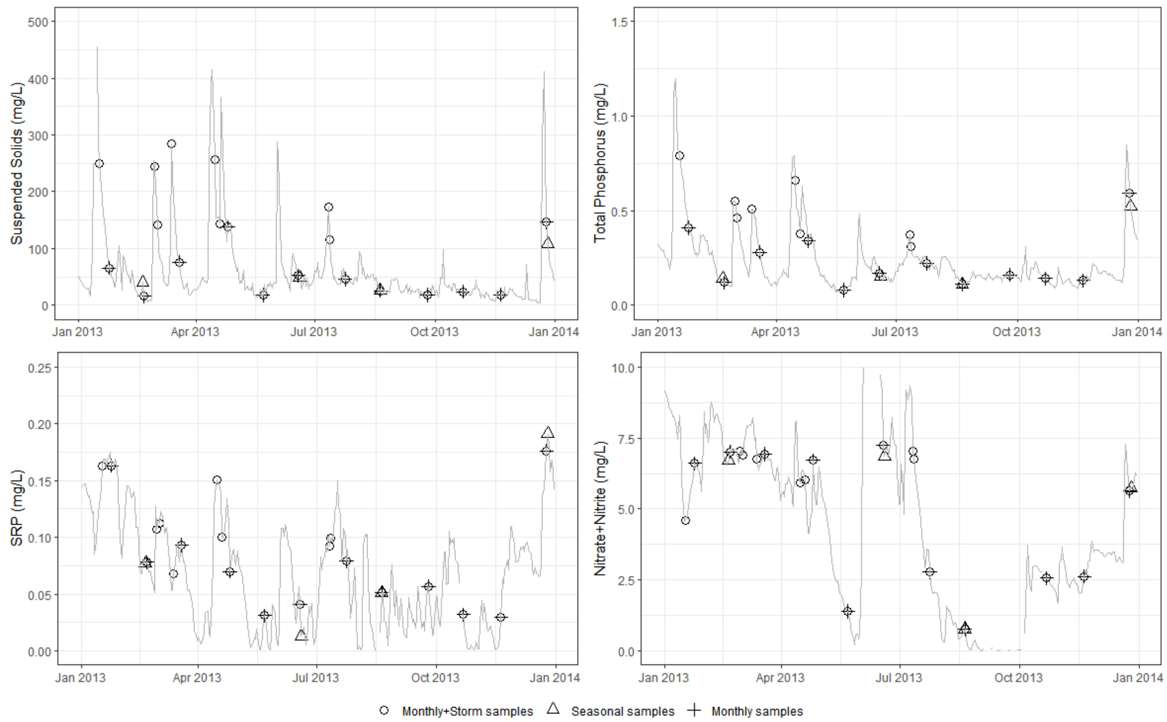

Fig. 6 Variation in concentrations in 2013 for the different constituents sampled at the Maumee station. The solid gray line corresponds to daily concentrations

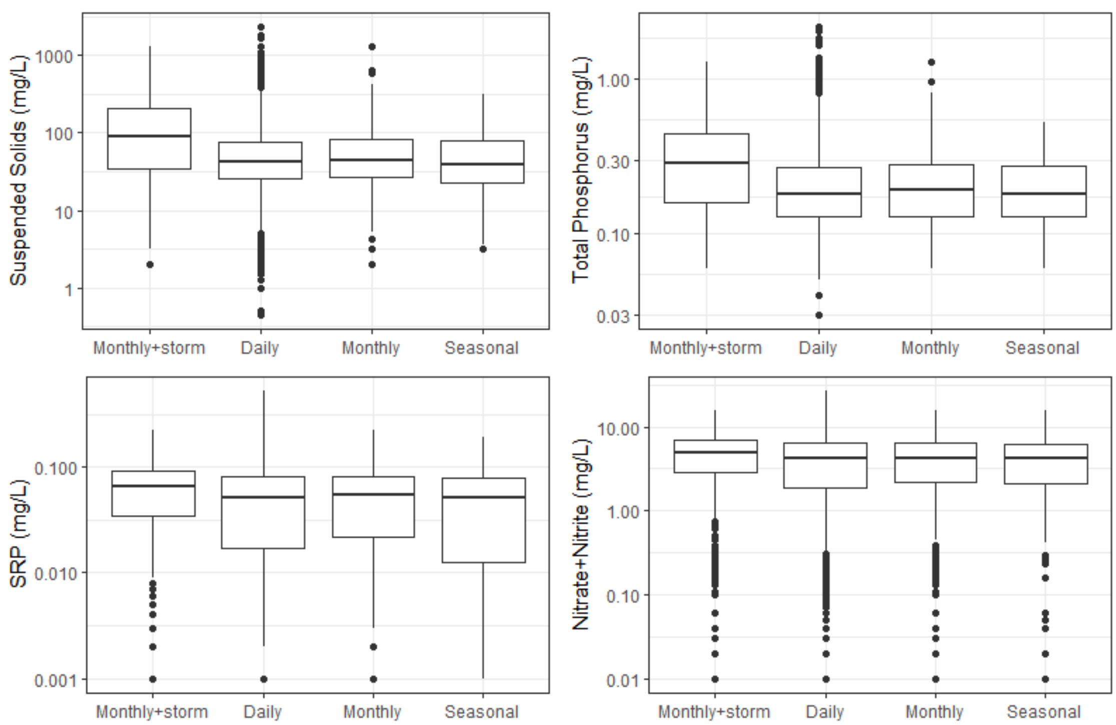

Fig. 7 Comparison of the distribution for different sampling frequencies (daily, monthly, seasonal, and monthly+storm datasets) for the Maumee station.

were compared to a baseline dataset made of all observations. The discrepancies in the median for the analyzed sampling frequencies might be attributable to the concentration-discharge relation for the analyzed constituents. SS, TP, and SRP are more easily mobilized through sub-surface than NO_{2+3}; therefore, any storm event generates a spike in their concentrations.

Regarding the identification of long-term trends, the monthly+storm event datasets did not seem to improve the results previously obtained with monthly and seasonal sampling datasets. Based on the results, the slope direction and relationship strength differed among sampling frequencies for the four parameters analyzed. The daily dataset exhibited a negative slope for SS, TP, and NO_{2+3}, while the monthly+storm event sample dataset had a positive slope. For SRP, both datasets had a positive slope, but the relationship strength for the daily set was more than twice the corresponding relationship strength for the monthly+storm event dataset. This difference could be associated with the hysteresis of SRP, which makes concentrations increase or decrease lag behind an increase or decrease in streamflow (Moosmann et al., 2005). Therefore, trends would be more sensitive to daily variations than storm events. These findings differed from what was reported by Chanat et al. (2016), who indicated that there was little evidence that the trend found using a design guideline subsample (monthly+storm event) would lead to different findings regarding trend direction or shape when compared to baseline data (daily).

Analysis of Water Quality Status Reported as Water Quality Indexes (WQI) for different frequencies

Table 4 shows the mean, minimum, and maximum values for WQSI by season. The results showed that WQSI values tended to increase when sampling frequency decreased. For SS, WQSIs ranged from 18 to 78 (unsuitable to good) for weekly sampling, 11 to 83 (unsuitable to good) for bi-weekly sampling, 12 to 88 (unsuitable to good) for monthly sampling, and 29 to 86 (poor to good) for seasonal sampling. For SRP, WQSI values tended to increase only for maximum values up to monthly sampling. Seasonal samples for spring and summer (SRP) had a higher maximum WQSI compared to weekly samples but lower than bi-weekly and monthly samples. The mean WQSI values changed slightly within seasons, ranging from 35 to 38, which did not affect the overall classification of the stream. In contrast, WQSIs for TP ranged from 16 to 59 for weekly sampling, 10 to 59 for bi-weekly sampling, 7 to 80 for monthly sampling, and 23 to 83 for seasonal sampling. In this case, the differences were substantial and changed the stream water quality classification from unsuitable to fair (weekly and bi-weekly) to unsuitable to good for monthly and seasonal sampling. For NO_{2+3} values ranged from 33 to 85 (poor to good) for weekly sampling, 33 to 86 (poor to good) for bi-weekly sampling, 26 to 89 (unsuitable to good) for monthly sampling, and 26 to 89 (unsuitable to good) for seasonal sampling. Additionally, the highest WQSI values occurred during the Summer, when the stream reached the "Good" category for this constituent.

Based on the results obtained, variation in WQSIs tended to increase as the sampling frequency decreased from weekly to seasonal sampling. This variation was considered acceptable at a weekly sampling frequency because the WQSIs generally fluctuated within two adjacent categories. However, the variation would be critical where the WQSIs spanned the range from "unsuitable" to "good,"—as was the case with some seasonal samples—, because this could lead to a misleading assessment of the streams' water quality status. Thus, if WQSIs are estimated using low-resolution water quality data (monthly or seasonal sampling), the results obtained should be used as a reference value, rather than a tool to assess the effectiveness of a water quality management strategy or to make decisions about the stream.

<u>Variation in water quality status depending on the day of sampling</u>

Fig. 8 shows the seasonal distribution of water quality sub-index values for SS in the Maumee River. The days in the Fair and At Risk categories occurred primarily between in the Spring and Summer, reaching the lowest median value on July 1st. These are the seasons in which there is higher precipitation and more agricultural activities in the region. Results for the remaining constituents sampled at the Maumee River station are presented as supplementary material (Fig. S1-Fig**. S3**). Changes in the water quality status for TP and SRP, did not vary during the selected dates and the median subindexes values always remained below the target threshold. This is consistent with work by Mijares (2017), which did not show significant changes in these values regardless of the season. High phosphorus concentrations are a persistent problem in the Maumee River Watershed. SRP concentrations have increased since the mid-1990s due to an increase in annual discharges into the river, perpetuating the prevalence of harmful algal blooms in Lake Erie. Hence results are consistent with what has been reported for the watershed. For NO_{2+3}, the median subindex values mainly fell in the *Fair* category, with five dates falling in the *Good* category in Summer (3) and Fall (2). In general, it was observed that for NO_{2+3}, as with SS, coarse sampling (monthly and seasonal) tended to present a better overall outcome than when high-frequency sampling was used. This was of special concern for NO_{2+3} during the summer since monthly and seasonal sampling would not capture a potential water quality impairment due to the agricultural activities conducted at the study site.

Based on these findings, we can infer that water quality status as reported based on water quality subindexes did not change drastically at different times throughout the season. Even though there may be a risk of misclassification, some water quality parameters were more stable than others and sampling at any date during the season would not change the reported water quality status. These results were consistent with what was reported by Skeffington et al. (2015), who indicated that pH, temperature, and phosphorus could be assigned an unambiguous category with monthly samples.

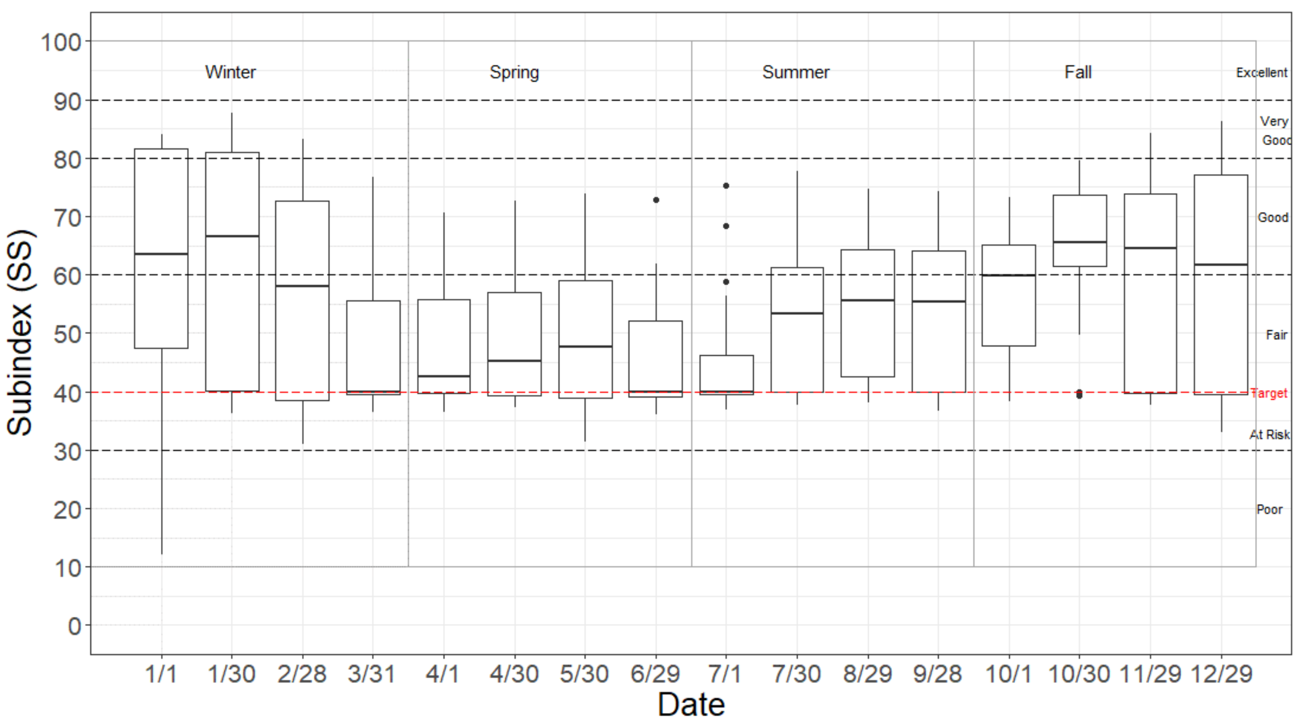

Fig. 8 Seasonal subindex distribution for suspended solids at the Maumee River water quality station

Table 4 Summary statistics for water quality sub-indexes for suspended solids, total phosphorus, soluble reactive phosphorus, and nitrate+nitrite (mg/L)*

Statistics	Fall				Spring				Summer				Winter			
	Weekly	Bi-weekly	Mthly	Seasr	Weekly	Bi-weekly	Mthly	Seasr	Weekly	Bi-weekly	Mthly	Seasr	Weekly	Bi-weekly	Mthly	Seasr
SS																
Mean	48	51	57	60	39	40	43	46	46	48	50	58	40	42	49	48
Min	18	11	32	31	31	31	29	29	33	32	30	35	23	19	12	30
Max	78	81	88	86	62	63	86	72	71	70	79	77	73	83	86	84
Interpretation	At Risk	Fair	Fair	Fair	Poor	At Risk	At Risk	At Risk	At Risk	At Risk	Fair	Fair	At Risk	At Risk	At Risk	At Risk
TP																
Mean	36	37	39	41	35	35	36	37	37	37	38	38	35	35	36	37
Min	16	10	28	26	28	27	24	24	29	29	25	30	20	15	7	23
Max	59	59	77	83	39	40	57	50	39	40	71	50	39	40	80	80
Interpretation	Poor	Poor	Poor	At Risk	Poor	Poor	Poor	Poor	Poor	Poor	Poor	Poor	Poor	Poor	Poor	Poor
SRP																
Mean	35	35	36	35	36	36	37	36	36	36	37	38	35	35	35	37
Min	29	27	19	19	32	31	30	25	30	30	28	30	29	29	25	26
Max	49	65	77	68	58	71	74	68	41	71	80	68	40	39	50	68
Interpretation	Poor	Poor	Poor	Poor	Poor	Poor	Poor	Poor	Poor	Poor	Poor	Poor	Poor	Poor	Poor	Poor
NO_{23}																
Mean	59	60	62	58	53	53	54	54	69	70	71	78	56	57	57	57
Min	33	33	30	27	38	37	34	26	42	37	37	57	38	38	26	32
Max	78	79	89	88	75	76	81	89	85	86	89	89	69	69	75	74
Interpretation	Fair	Fair	Fair	Fair	Fair	Fair	Fair	Fair	Fair	Good	Good	Good	Fair	Fair	Fair	Fair

Ratings: 90-100 Pristine; 70-89 Good; 50-69 Fair; 40-49 At Risk; 30-39 Poor; 0-29 Unsuitable for all uses. The interpretation was reported for the mean values.

*The goal is not to compare water quality across seasons, rather to determine whether sampling frequencies make a difference in water quality status determination within a season

General Discussion

Ideally, sampling frequencies are subject to diverse factors like stream characteristics, nature of the constituents of concern, and climate (Halliday et al., 2015; László et al., 2007; Vilmin et al., 2018). From our analysis, the number of samples required to obtain a 10% error with a 95% confidence ranged from 50 to 350 samples in a year for the stations analyzed in this study. The constituent that required the least amount of samples was TP (50 to 100 samples/year), followed by SS (100 to 250 samples/year) and SRP (100 to 350 samples/year), translating to at least weekly for TP, and twice a week for SS and SRP. These findings were consistent with results reported by Khalil et al. (2014), who pointed out that TP and total suspended solids (TSS) would require more than 12 samples per year due to the high variation in these constituents and recommended that these constituents should ideally be sampled on a biweekly basis. If the contaminants came from urban sources, additional aspects needed to be considered. Vilmin et al. (2018) indicated that optimal sampling depended on the location of the sampling site relative to the major anthropogenic effluents since major effluents increase the variability in the receiving environment. In the case of urban pollutant sources, Raimonet et al. (2015) and Vilmin et al. (2018) noted that at least weekly sampling was required to account for the high-variability of ammonia and nitrite.

Regarding the estimation of long-term trends, the results obtained in this study showed that there were differences in the trends as determined from the different sampling frequencies analyzed. Results from the Mann-Kendall test showed that daily, weekly, and bi-weekly sampling gave consistent results, capturing the direction and magnitude of the potential trend. In contrast, there was not a consistent pattern for the trends for monthly and seasonal sampling. The trend's direction and strength vary when compared to the different datasets at the same frequency level (monthly or seasonal). Moreover, in some cases, the trends (positive or negative) identified in seasonal sampling can be amplified due to the high variability of the analyzed constituents (Fig. 6). Vilmin et al. (2018) and Raimonet et al. (2015) indicated that monthly sampling does not capture the high variability of constituents originating from urban effluents. As a result, monthly or seasonal sampling may not be adequate to estimate long-term trends. Similar findings were reported by Bowes et al. (2009), who indicated that the temporal resolution of the datasets should be increased from monthly sampling if the monitoring goal is to assess water quality improvement due to the implementation of management practices. Halliday et al. (2012) concluded that weekly sampling could be used to obtain general water quality characteristics, long-term trends, and seasonality changes for nitrate, sulfate, chloride, dissolved organic carbon, iron, and aluminum.

Additionally, the use of a high-frequency sampling strategy may reduce the length of the record required to identify long-term trends. Liu et al. (2020) determined that good trend detection probability was achievable with 5-years of continuous (15-min sampling) records for electrical conductivity and turbidity. Likewise, Moosmann et al. (2005) showed the relationship between observed trends, the length of study, and the number of samples required to identify the desired trend in nutrient loads. If the purpose was to identify small trends (< 3%/yr), the number of samples or the length of study had to be increased. The researchers indicated that for TP, 150 samples in a 3-year study or 30 samples in a 5-year study would be necessary to identify a 3%/year trend. In contrast, Naddeo et al. (2013) were able to identify trends using water quality samples ranging from one to six months. Hunt et al. (2008) found that there was redundancy for DO and chlorophyll samples collected in the Boston Harbor and Massachusetts Bay, and a frequency sampling reduction from 17 to 4 per year could be accomplished without affecting the quality of the data. The reduction in sampling frequency can only be achieved after a robust statistical analysis indicates that the low-frequency data would not be significantly different from the high-frequency data.

Concerning the improvement of monthly and seasonal sampling to assess long-term trends, the inclusion of 8 randomly selected storm samples helped to capture some of the peak concentration events. However, the median and IQR were higher for SS and TP. The difference in those values was consistent with what was reported by Thompson et al. (2021), who found that the estimated error for TSS loads was 32% for the median and 106% for the IQR when storm samples were included in a weekly dataset. This type of error could lead to misleading conclusions when the effectiveness of water quality improvement practices is assessed.

Overall, our results indicate that using the same sampling frequency for all constituents may not be suitable to meet all water quality monitoring objectives, especially when high-frequency sampling is not available. Unfortunately, setting clear objectives may not be easily accomplished since there is no consensus among existing definitions monitoring objectives due to external constraints like social, legal, economic, and administrative factors (Harmancioğlu, 1999). While agencies at the local, regional, and national levels might have similar monitoring objectives; often different sampling frequencies are used to assess the objectives. Due to the uncertainty in defining an optimal sampling frequency, future research needs to focus on how high-frequency sampling can be used to define the appropriate sampling frequency tailored to the site-specific needs and aligned to the water quality monitoring objectives. Likewise, the lack of extensive data makes it difficult to define optimal sampling frequencies. Based on the GEMStat database, the density of water quality stations ranges from 1.5 to 4 stations per 10,000 km^2 in the U.S. and Europe, while in Latin America there are 0.3 stations per 10,000 km^2, in Asia is 0.08 stations per 10,000 km^2, and in Africa is 0.02 stations per 10,000 km^2 (UNEP, 2016).

Sampling frequency also depends on the constituent analyzed and the source of water pollution. Water quality parameters with high variability, like nutrients and sediments, require at least weekly sampling. Additionally, areas near urban area effluents or another known source of pollutants require a high-resolution sampling frequency (Raimonet et al., 2015; Vilmin et al., 2018). Regarding long-term trends, our study showed differences depending on when the samples were taken in the monthly and seasonal datasets. Therefore, assessments made based only on this information might be biased. In relation to water quality status, we observed that WQI could be different depending on the sampling frequency. Weekly and bi-weekly sampling seemed to capture the constituents' high variability and the changes associated with seasonality, while monthly and seasonal sampling could lead to misclassification in the water quality status thereby misrepresenting a stream's condition.

A limitation of this study is associated with the number of constituents analyzed. Further research could evaluate the behavior of water quality parameters related to urban pollution like BOD_5, heavy metals, ammonia, and contaminants of emerging concern. Additionally, the confidence interval method to estimate the number of samples per constituent was the only method used in this study since it was the more frequently used method, as reported by Nguyen et al. (2019). A comparison among methods to estimate the number of samples would be of interest to identify differences and potential shortcomings for each method. Additionally, since the Western Lake Erie Basin is located in the Great Lakes area, conducting similar studies in other regions and landscapes could provide additional information to formulate new recommendations.

Case Application

The Colombian Andean Region has an area of 283,000 km^2, located in the country's central part along the East, Central, and Western Andes mountain ranges. The elevation ranges from 500-5000 meters above mean sea level, making it a region with numerous valleys, canyons, and plateaus. The main rivers are the Magdalena and Cauca Rivers, forming the Magdalena-Cauca Hydrological Area. This region also has most of the country's water resources and the most productive soils (Universidad del Rosario, 2015). According to the 2018 Census, around 26 million people live in the region (57% of the country's population), mostly concentrated in urban areas (82% live in urban centers) (Baena Salazar et al., 2020). The region also represents 65% of the GDP, followed by manufacturing activities (13%), real estate (10%), and administrative activities (DANE, 2020). Sub-areas of the Andean region have been identified for having water-related issues. Eight out of the eleven region's Departments have a high erosion potential; the deficit in precipitation events is more recurrent across the region; and the water bodies are highly impacted by contaminant loads (IDEAM, 2018).

Due to the region's economic, social, and environmental significance, it is a priority to set strategies, practices, and tools to guarantee its water resources' sustainable use. The Otun River, a major tributary of the Cauca River, has been identified as a river with issues associated with erosion, water quality, and water scarcity. Additionally, this watershed is located in Colombia's coffee axis, an area where coffee production boosted the national economy during most of the 20th century and brought the emergence of cities like Pereira/Dosquebradas and Manizales. These two cities became relevant urban areas since they are intermediate points between Bogotá, the capital city and the country's major economic center, and the Pacific Ocean. Pereira and Dosquebradas are the largest urban centers and made up to 99% of the total population (Consorcio Ordenamiento Cuenca del Río Otún, 2017). According to the 2008 Census, Pereira and Dosquebradas had 467,269 and 217,178 inhabitants, where 83% and 92% of the total population live in the urban area, respectively.

The Otun River Watershed's (ORW) water management falls under the Regional Environmental Authority (CARDER) jurisdiction. CARDER is Risaralda's environmental authority and its responsibilities include classifying surface water, indicating the water's intended use, establishing short, mid, and long-term water quality objectives, and setting regulations to preserve water quality (Ministerio de Ambiente, Vivienda y Desarrollo Territorial, 2015). The main urban center in the ORW is the Pereira/Dosquebradas metropolitan area. In 2015, CARDER approved the Otun River and Dosquebradas Creek Water Ordinance Plan. Under this ordinance, the short (5 years), mid (10 years), and long-term (20 years) objectives were set for both streams. Each of the streams is subdivided into stretches and objectives are tailored for each of them, based on the intended water use. The water objectives for both streams can be found as supplementary material (Table S7 and Table S8). The ordinance plan also indicates that the plan's revision and adjustment must be completed during its extent (20 years), based on the results obtained from the monitoring plan. Under this plan, sampling is conducted three times a year, two during the wet period (January-June and September-December) and one during the dryer period (July-August). The number of sampling points in the Otun River ranges from 15 to 17, and in the Dosquebradas Creek ranges from 8 to 17 sampling points. Samples are analyzed for 25 physical-chemical parameters, including temperature, pH, turbidity, dissolved oxygen, fecal coliforms, total solids, biochemical oxygen demand (BOD_5), nitrates, and phosphates.

The previously listed parameters are used to calculate a WQI, which is the metric used to report the streams' water quality status. This WQI is estimated using the methodology proposed by the National Sanitation Foundation (Brown et al., 1970). In this methodology, each parameter has a weighted factor as follows: dissolved oxygen (0.17), fecal coliform (0.16), pH (0.11), BOD_5 (0.11), change in temperature (0.10), phosphate (0.10), nitrates (0.10), turbidity (0.08), and total solids (0.07). In the 2017 Risaralda's water quality report, the Otun River's water quality was classified as good in eight stations (57%), medium in five stations (36%), and bad in one station (7%). Meanwhile, Dosquebradas Creek's water quality was classified as good in five stations (29%), medium in ten stations (59%), and bad in two stations (12%). The water quality decline for the aforementioned streams is associated with agricultural activities occurring in the ORW midsection and urban raw sewage discharges from Pereira and Dosquebradas.

Results of our analysis indicate that water quality status could be worse than indicated due to sampling frequency. Depending on the water concerns in the area of study, the monitoring objectives may be different. The most common objectives are to estimate the trends and assess the status. Based on the results obtained in this study, it is necessary to have a minimum of bi-weekly sampling frequency to capture the water quality trends with a high degree of confidence. The sampling frequency may be reduced for relatively stable constituents, like TP, or it needs to be increased for easily transported constituents or highly variable like NO_{2+3}, SS, and SRP.

Another important consideration that can be made when defining the sampling frequency is the watershed characteristics and its potential source of pollutants. In the case of the ORW, the upper section does not have many sources of pollutants (the only major source of pollutants is the Pezfresco trout farming and processing factory, located in the borderline between the watershed's upper and mid sections), and since part of this section corresponds to a protected environmental zone, the sampling frequency may be limited to seasonal sampling or even reduced to an annual basis, with the purpose being to assess the Otun River and its tributaries' water quality status. This is consistent with findings from Levine et al. (2014), who indicated that sampling at mixed frequencies might decrease the number of times that each site is sampled. On the other hand, the ORW mid-section has been identified as the area contributing most of the pollutant loads into the Otun River and Dosquebradas Creek (Consorcio Ordenamiento Cuenca del Río Otún, 2017). For that reason, it is highly recommended that the monitoring efforts should be focused on this part of the watershed. As high-frequency water quality monitoring plans are expensive to implement, an initial focus on increasing the sampling frequency from seasonal to monthly in this area would provide benefits in capturing pollutant trends. The monthly sampling may, however, not capture all the high concentration events, but the values for median concentration and trend analysis would be closer to the values obtained with a daily sampling frequency. Over the long term, environmental authorities should consider the use of emerging monitoring technologies. Automatic high-frequency monitoring is becoming more common and affordable to implement. Additionally, passive and active remote sensing can be used as complementary data sources. Citizen science—a commonly used resource in environmental monitoring—could potentially provide improved statistical power of datasets and facilitate the observation of difficult to observe phenomena (as reported by Jollymore et al., 2017).

Regarding water quality status, it is important to select an appropriate metric to report the water bodies' status. The water quality status in the ORW is reported using the categories established by the National Sanitation Foundation (NSF) (Brown et al., 1970). In contrast, the Colombian Institute of Hydrology, Meteorology, and Environmental Studies (IDEAM) uses the Universal Water Quality Index (UWQI) methodology. Gitau et al. (2016) concluded that the use of more objective and less rigid formulations would provide a better way of assessing water quality status. Mijares et al. (2019) developed subindex formulations that were then built into the Unweighted Multiplicative Water Quality Index (UMWQI). The advantage of using these subindex formulations is that they incorporate water quality thresholds. Additionally, an unweighted quality index is not restrictive with the water quality parameters to be used and can be tailored to a specific location. Regardless of the methodology used to assess and report the water quality status, it is essential to consider the pollutants of concern in the area. For the ORW, the pollutants of concern come from agricultural activities and the Pereira/Dosquebradas raw sewer discharges. As a result, constituents like dissolved oxygen, BOD, nitrogen, phosphorus, pH, suspended solids, and ammonia should be included in the metric used for the water quality status assessment. Moreover, since one of the primary uses for the Otun River's water is for human consumption, environmental authorities may want to consider a metric that includes constituents of interest for human health.

Conclusions

Water quality sampling frequency has typically been selected according to the monitoring objective to be addressed. However, the sampling frequency selected varies within sampling objectives with monitoring plans comprising weekly, monthly, seasonal, and/or annual sampling frequencies. Water quality sampling is expensive, thus the need to determine suitable sampling frequencies capturing multiple objectives. In this study, data representing weekly, bi-weekly, monthly, and seasonal sampling frequencies were compared against daily data to determine their suitability in capturing long-term trends and water quality status. Long-term trends determined from monthly and seasonal sampling were highly variable and, thus, these frequencies might not be appropriate for use in this type of analysis. Accuracy at these sampling frequencies could be improved by including rainfall events as part of the monitoring plan. Regarding water quality status, differences were found in the water quality sub-indexes from data sampled 30-days apart, suggesting higher frequency sampling would be more appropriate. Our results suggest that weekly sampling accounts for the high variability of some constituents, like sediments and nutrients, and changes in the constituents' concentrations associated with seasonal phenomena. If sampling at such resolution is not feasible, bi-weekly sampling would still provide reasonably accurate data. Bi-weekly sampling is suggested as an alternative sampling frequency since it is able to capture long-term trends for relatively stable constituents, and the water quality status as determined based on this frequency would generally be same as the one reported with weekly samples. For pristine or protected environments in which water quality is not at risk, different sampling frequencies can be considered as alternatives to sampling on the same day every year, for example, random sampling during a specific time window.

References

Babitsch, D., Berger, E., & Sundermann, A. (2021). Linking environmental with biological data: Low sampling frequencies of chemical pollutants and nutrients in rivers reduce the reliability of model results. *Science of The Total Environment*, *772*, 145498. https://doi.org/10.1016/j.scitotenv.2021.145498

Baena Salazar, D., Fuentes Hernández, J. S., Pino Reyes, L. T., Marín Durán, S., Horta Pérez, S. V., & Fonseca González, W. C. (2020). *Contexto Regional Andina*. http://hdl.handle.net/1992/47782

Behmel, S., Damour, M., Ludwig, R., & Rodriguez, M. J. (2016). Water quality monitoring strategies—A review and future perspectives. *Science of The Total Environment*, *571*, 1312–1329. https://doi.org/10.1016/j.scitotenv.2016.06.235

Blaen, P. J., Khamis, K., Lloyd, C., Comer-Warner, S., Ciocca, F., Thomas, R. M., MacKenzie, A. R., & Krause, S. (2017). High-frequency monitoring of catchment nutrient exports reveals highly variable storm event responses and dynamic source zone activation: High-Frequency Storm Event Monitoring. *Journal of Geophysical Research: Biogeosciences*, *122*(9), 2265–2281. https://doi.org/10.1002/2017JG003904

Bowes, M. J., Smith, J. T., & Neal, C. (2009). The value of high-resolution nutrient monitoring: A case study of the River Frome, Dorset, UK. *Journal of Hydrology*, *378*(1–2), 82–96. https://doi.org/10.1016/j.jhydrol.2009.09.015

Brown, R., McClelland, N., Deininger, R., & Tazer, R. (1970). A Water Quality Index—Do We Dare. *Water & Sewage Works*, *117*(10).

Chanat, J. G., Moyer, D. L., Blomquist, J. D., Hyer, K. E., & Langland, M. J. (2016). *Application of a Weighted Regression Model for Reporting Nutrient and Sediment Concentrations, Fluxes, and Trends in Concentration and Flux for the Chesapeake Bay Nontidal Water-*

Quality Monitoring Network, Results Through Water Year 2012 (Scientific Investigations Report Report 2015–5133; Scientific Investigations Report). U.S. Geological Survey.

Consorcio Ordenamiento Cuenca del Río Otún. (2017). *Ajuste Plan de Ordenación y Manejo de la Cuenca del Río Otún*.

Departamento Administrativo Nacional de Estadística (DANE). (2020, September 25). *Cuentas nacionales departamentales: PIB por departamento*. https://www.dane.gov.co/index.php/estadisticas-por-tema/cuentas-nacionales/cuentas-nacionales-departamentales#pib-base-2005

DES. (2018). *Monitoring and Sampling Manual: Environmental Protection (Water) Policy* (p. 285). Department of Environment and Science Government.

Díaz-Casallas, D. M., Castro-Fernández, M. F., Bocos, E., Montenegro-Marin, C. E., & González Crespo, R. (2019). 2008–2017 Bogota River Water Quality Assessment based on the Water Quality Index. *Sustainability*, *11*(6), 1668. https://doi.org/10.3390/su11061668

European Commission & Directorate-General for the Environment. (2003). *Monitoring under the water framework directive.* http://bookshop.europa.eu/uri?target=EUB:NOTICE:KH5103213:EN:HTML

Florida Department of Environmental Protection. (2020). *Florida Watershed Monitoring Status and Trend Program Design Document*.

Gilliom, R. J., Alley, W. M., & Gurtz, M. E. (1995). *Design of the National Water-Quality Assessment Program: Occurrence and Distribution of Water-Quality Conditions* (Circular No. 1112; Circular). United States Geological Survey.

Gitau, M. W., Chen, J., & Ma, Z. (2016). Water Quality Indices as Tools for Decision Making and Management. *Water Resources Management*, *30*(8), 2591–2610. https://doi.org/10.1007/s11269-016-1311-0

Guerreiro, M. S., Abreu, I. M., Monteiro, Á., Jesus, T., & Fonseca, A. (2020). Considerations on the monitoring of water quality in urban streams: A case study in Portugal. *Environmental Monitoring and Assessment*, *192*(6), 347. https://doi.org/10.1007/s10661-020-8245-y

Halliday, S. J., Skeffington, R. A., Wade, A. J., Bowes, M. J., Gozzard, E., Newman, J. R., Loewenthal, M., Palmer-Felgate, E. J., & Jarvie, H. P. (2015). High-frequency water quality monitoring in an urban catchment: Hydrochemical dynamics, primary production and implications for the Water Framework Directive. *Hydrological Processes*, *29*(15), 3388–3407. https://doi.org/10.1002/hyp.10453

Halliday, S. J., Wade, A. J., Skeffington, R. A., Neal, C., Reynolds, B., Rowland, P., Neal, M., & Norris, D. (2012). An analysis of long-term trends, seasonality and short-term dynamics in water quality data from Plynlimon, Wales. *Science of The Total Environment*, *434*, 186–200. https://doi.org/10.1016/j.scitotenv.2011.10.052

Harmancioğlu, N. (1999). *Water quality monitoring network design*. Kluwer Academic Publishers.

Holguin-Gonzalez, J. E., Everaert, G., Boets, P., Galvis, A., & Goethals, P. L. M. (2013). Development and application of an integrated ecological modelling framework to analyze the impact of wastewater discharges on the ecological water quality of rivers. *Environmental Modelling & Software*, *48*, 27–36. https://doi.org/10.1016/j.envsoft.2013.06.004

Hunt, C. D., Rust, S. W., & Sinnott, L. (2008). Application of statistical modeling to optimize a coastal water quality monitoring program. *Environmental Monitoring and Assessment*, *137*(1–3), 505–522. https://doi.org/10.1007/s10661-007-9785-0

IDEAM. (2018). *Estudio Nacional del Agua 2018* (p. 438).

Indiana Department of Environmental Management (IDEM). (2017). *Water Quality Monitoring Strategy 2017-2021* (p. 95). IDEM.

Jiang, J., Tang, S., Han, D., Fu, G., Solomatine, D., & Zheng, Y. (2020). A comprehensive review on the design and optimization of surface water quality monitoring networks. *Environmental Modelling & Software*, *132*, 104792. https://doi.org/10.1016/j.envsoft.2020.104792

Jollymore, A., Haines, M. J., Satterfield, T., & Johnson, M. S. (2017). Citizen science for water quality monitoring: Data implications of citizen perspectives. *Journal of Environmental Management*, *200*, 456–467. https://doi.org/10.1016/j.jenvman.2017.05.083

Karamouz, M., Nokhandan, A. K., Kerachian, R., & Maksimovic, Č. (2009). Design of on-line river water quality monitoring systems using the entropy theory: A case study. *Environmental Monitoring and Assessment*, *155*(1–4), 63–81. https://doi.org/10.1007/s10661-008-0418-z

Khalil, B., Ou, C., Proulx-McInnis, S., St-Hilaire, A., & Zanacic, E. (2014). Statistical Assessment of the Surface Water Quality Monitoring Network in Saskatchewan. *Water, Air, & Soil Pollution*, *225*(10), 2128. https://doi.org/10.1007/s11270-014-2128-1

Kotamäki, N., Järvinen, M., Kauppila, P., Korpinen, S., Lensu, A., Malve, O., Mitikka, S., Silander, J., & Kettunen, J. (2019). A practical approach to improve the statistical performance of surface water monitoring networks. *Environmental Monitoring and Assessment*, *191*(6), 318. https://doi.org/10.1007/s10661-019-7475-3

László, B., Szilágyi, F., Szilágyi, E., Heltai, G., & Licskó, I. (2007). Implementation of the EU Water Framework Directive in monitoring of small water bodies in Hungary, I. Establishment of surveillance monitoring system for physical and chemical characteristics for small mountain watercourses. *Microchemical Journal*, *85*(1), 65–71. https://doi.org/10.1016/j.microc.2006.06.007

Lee, C. J., Murphy, J. C., Crawford, C. G., & Deacon, J. R. (2017). *Methods for Computing Water-Quality Loads at Sites in the U.S. Geological Survey National Water Quality Network* (Open-File Report No. 2017–1120; Open-File Report). U.S. Geological Survey.

Levine, C. R., Yanai, R. D., Lampman, G. G., Burns, D. A., Driscoll, C. T., Lawrence, G. B., Lynch, J. A., & Schoch, N. (2014). Evaluating the efficiency of environmental monitoring programs. *Ecological Indicators*, *39*, 94–101. https://doi.org/10.1016/j.ecolind.2013.12.010

Liu, S., Guo, D., Webb, J. A., Wilson, P. J., & Western, A. W. (2020). A simulation-based approach to assess the power of trend detection in high- and low-frequency water quality records. *Environmental Monitoring and Assessment*, *192*(10), 628. https://doi.org/10.1007/s10661-020-08592-9

Mijares, V., Gitau, M., & Johnson, D. R. (2019). A Method for Assessing and Predicting Water Quality Status for Improved Decision-Making and Management. *Water Resources Management*, *33*(2), 509–522. https://doi.org/10.1007/s11269-018-2113-3

Ministerio de Ambiente, Vivienda y Desarrollo Territorial. (2015). *Decreto 1076* (Decreto 1076). https://www.minambiente.gov.co/wp-content/uploads/2021/06/Decreto-1076-de-2015.pdf

Moosmann, L., Müller, B., Gächter, R., Wüest, A., Butscher, E., & Herzog, P. (2005). Trend-oriented sampling strategy and estimation of soluble reactive phosphorus loads in streams: TREND-ORIENTED ESTIMATION OF PHOSPHORUS. *Water Resources Research*, *41*(1). https://doi.org/10.1029/2004WR003539

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. https://doi.org/10.1016/j.scitotenv.2018.11.255

Naddeo, V., Scannapieco, D., Zarra, T., & Belgiorno, V. (2013). River water quality assessment: Implementation of non-parametric tests for sampling frequency optimization. *Land Use Policy*, *30*(1), 197–205. https://doi.org/10.1016/j.landusepol.2012.03.013

Nguyen, T. H., Helm, B., Hettiarachchi, H., Caucci, S., & Krebs, P. (2019). The selection of design methods for river water quality monitoring networks: A review. *Environmental Earth Sciences*, *78*(3), 96. https://doi.org/10.1007/s12665-019-8110-x

NYC Department of Environmental Protection. (2019). *2018 Watershed Water Quality Annual Report* (p. 233). NYC Department of Environmental Protection.

Outram, F. N., Lloyd, C. E. M., Jonczyk, J., Benskin, C. McW. H., Grant, F., Perks, M. T., Deasy, C., Burke, S. P., Collins, A. L., Freer, J., Haygarth, P. M., Hiscock, K. M., Johnes, P. J., & Lovett, A. L. (2014). High-frequency monitoring of nitrogen and phosphorus response in three rural catchments to the end of the 2011–2012 drought in England. *Hydrology and Earth System Sciences*, *18*(9), 3429–3448. https://doi.org/10.5194/hess-18-3429-2014

Pennsylvania Department of Environmental Protection. (2019). *Water Quality Network Objectives*. Pennsylvania Department of Environmental Protection.

Raimonet, M., Vilmin, L., Flipo, N., Rocher, V., & Laverman, A. M. (2015). Modelling the fate of nitrite in an urbanized river using experimentally obtained nitrifier growth parameters. *Water Research*, *73*, 373–387. https://doi.org/10.1016/j.watres.2015.01.026

Rodríguez, J. P., McIntyre, N., Díaz-Granados, M., Quijano, J. P., & Maksimović, Č. (2013). Monitoring and modelling to support wastewater system management in developing mega-cities. *Science of The Total Environment*, *445–446*, 79–93. https://doi.org/10.1016/j.scitotenv.2012.12.022

Roerdink, A. (2017). *Water Quality in Ohio Rivers and Streams*. National Center for Water Quality Research, Heidelberg University.

Ross, C., Petzold, H., Penner, A., & Ali, G. (2015). Comparison of sampling strategies for monitoring water quality in mesoscale Canadian Prairie watersheds. *Environmental Monitoring and Assessment*, *187*(7), 395. https://doi.org/10.1007/s10661-015-4637-9

Sanders, T. G., & Adrian, D. D. (1978). Sampling frequency for river quality monitoring. *Water Resources Research*, *14*(4), 569–576. https://doi.org/10.1029/WR014i004p00569

Sherwani, J. K., & Moreau, David. H. (1975). *Strategies for Water Quality Monitoring* (No. 107; p. 152). Water Resources Research Institute - University of North Carolina.

Skeffington, R. A., Halliday, S. J., Wade, A. J., Bowes, M. J., & Loewenthal, M. (2015). Using high-frequency water quality data to assess sampling strategies for the EU Water Framework Directive. *Hydrology and Earth System Sciences*, *19*(5), 2491–2504. https://doi.org/10.5194/hess-19-2491-2015

Tanos, P., Kovács, J., Kovács, S., Anda, A., & Hatvani, I. G. (2015). Optimization of the monitoring network on the River Tisza (Central Europe, Hungary) using combined cluster and discriminant analysis, taking seasonality into account. *Environmental Monitoring and Assessment*, *187*(9), 575. https://doi.org/10.1007/s10661-015-4777-y

Thompson, J., Pelc, C. E., & Jordan, T. E. (2021). Water quality sampling methods may bias evaluations of watershed management practices. *Science of The Total Environment*, *765*, 142739. https://doi.org/10.1016/j.scitotenv.2020.142739

UNEP. (2016). *A Snapshot of the World's Water Quality: Towards a global assessment*. United Nations Environment Programme (UNEP).

Universidad del Rosario. (2015, February 16). *¿Cómo vamos en las regiones? - Universidad del Rosario*. https://www.urosario.edu.co/Home/Principal/boletines/Ediciones-OPIP-Regionales/Edicion01-Regiones/Como-vamos-en-las-regiones/

U.S. EPA. (2009). *Environmental Impact and Benefits Assessment for Final Effluent Guidelines and Standards for the Construction and Development Category—November 2009* (EPA-821-R-09-012; p. 374). U.S. Environmental Protection Agency.

U.S. EPA. (2020). *National Rivers and Streams Assessment 2013–2014: A Collaborative Survey* (EPA 841-R-19-001). U.S. Environmental Protection Agency. http://www.epa.gov/national-aquatic-resource-surveys/data-national-aquatic-resource-surveys

USGS. (2021, April 26). *Next Generation Water Observing System: Delaware River Basin | U.S. Geological Survey*. https://www.usgs.gov/mission-areas/water-resources/science/next-generation-water-observing-system-delaware-river-basin?qt-science_center_objects=0#overview

Vilmin, L., Flipo, N., Escoffier, N., & Groleau, A. (2018). Estimation of the water quality of a large urbanized river as defined by the European WFD: What is the optimal sampling frequency? *Environmental Science and Pollution Research*, *25*(24), 23485–23501. https://doi.org/10.1007/s11356-016-7109-z

Weber, L., Jones, C., & Davis, C. (2016). *IIHR 2015 Water Monitoring Report*. IIHR - Hydroscience and Engineering.

Zhang, Q., & Hirsch, R. M. (2019). River Water-Quality Concentration and Flux Estimation Can be Improved by Accounting for Serial Correlation Through an Autoregressive Model. *Water Resources Research*, *55*(11), 9705–9723. https://doi.org/10.1029/2019WR025338

Table S1. Summary of statistical properties for different sampling frequency datasets (Maumee Station)

	Daily	**Weekly**	**Bi-weekly (even)**	**Bi-weekly (odd)**	**Monthly**	**Seasonal**
SS						
Mean (mg/L)	72.0	71.8	71.7	71.9	73.7	78.9
SD (mg/L)	100.1	92.4	90.2	94.2	99.0	87.0
Min (mg/L)	0.5	2.1	2.9	2.2	3.3	5.4
Max (mg/L)	2325.1	1351.3	1046.4	1112.1	865.4	516.4
Q90 (mg/L)	156.9	158.8	155.8	159.5	157.7	176.3
Q99 (mg/L)	465.0	458.6	449.7	460.0	460.1	411.6
TP						
Mean (mg/L)	0.22	0.22	0.22	0.22	0.22	0.24
SD (mg/L)	0.15	0.14	0.14	0.14	0.15	0.14
Min (mg/L)	0.03	0.043	0.05	0.04	0.06	0.06
Max (mg/L)	2.17	1.45	1.23	1.32	1.09	0.80
Q90 (mg/L)	0.40	0.39	0.39	0.39	0.40	0.41
Q99 (mg/L)	0.78	0.78	0.79	0.77	0.75	0.68
SRP						
Mean (mg/L)	0.055	0.056	0.056	0.056	0.056	0.059
SD (mg/L)	0.043	0.043	0.043	0.042	0.043	0.043
Min (mg/L)	0.001	0.001	0.001	0.001	0.001	0.001
Max (mg/L)	0.525	0.351	0.334	0.265	0.244	0.202
Q90 (mg/L)	0.110	0.111	0.111	0.111	0.110	0.111
Q99 (mg/L)	0.176	0.170	0.173	0.165	0.163	0.172
NO_{23}						
Mean (mg/L)	4.40	4.43	4.43	4.43	4.45	4.71
SD (mg/L)	3.12	3.13	3.11	3.16	3.12	3.28
Min (mg/L)	0.01	0.01	0.01	0.01	0.012	0.02
Max (mg/L)	26.72	20.6	15.02	20.58	15.66	13.8
Q90 (mg/L)	8.47	8.49	8.41	8.55	8.47	8.72
Q99 (mg/L)	12.87	13.06	13.04	12.97	12.58	12.79

Table S2. Dataset comparison made with the Kruskal-Wallis test

Type of dataset	# of datasets	p-value			
		SS	TP	SRP	NO_{23}
Weekly					
DOW+random	10	0.992	0.999	0.331	0.996
Bi-weekly					
Even+random	10	0.967	0.996	0.860	0.999
Odd+random	10	0.999	0.999	0.647	0.999
Even+Odd+random	20	0.999	1.000	0.887	1.000
Monthly					
DOM+random	35	0.995	0.999	0.784	0.999
SEL+random	16	0.981	0.956	0.869	0.898
DOM+SEL+random	47	0.940	0.983	0.107	
Seasonal					
SELD+random	14	0.308	0.218	0.835	0.394
TWT+random	16	0.506	0.173	0.938	0.376
SELD+TWT+random	26	0.707	0.386	0.972	0.626

Table S3. Results from the Mann-Kendall test for suspended solids concentrations during the observed period (1986-2015) for Maumee Station

Dataset	**Suspended Solids**							
	Winter		Spring		Summer		Fall	
	tau	p-value	tau	p-value	tau	p-value	tau	p-value
Daily	-0.085	0.521	-0.306	< 0.05	-0.434	< 0.05	-0.237	0.069
Weekly-random 1	-0.002	> 0.999	-0.287	< 0.05	-0.448	< 0.05	-0.209	0.108
Weekly-random 2	-0.002	> 0.999	-0.287	< 0.05	-0.448	< 0.05	-0.209	0.108
Weekly-random 3	-0.002	> 0.999	-0.287	< 0.05	-0.448	< 0.05	-0.209	0.108
Bi-weekly even-random 1	0.053	0.695	-0.389	< 0.05	-0.370	< 0.05	-0.297	0.022
Bi-weekly even-random 2	-0.076	0.568	-0.237	0.069	-0.411	< 0.05	-0.338	< 0.05
Bi-weekly even-random 3	-0.094	0.475	-0.375	< 0.05	-0.343	< 0.05	-0.274	0.035
Bi-weekly odd-random 1	0.044	0.748	-0.251	0.054	-0.407	< 0.05	-0.186	0.153
Bi-weekly odd-random 2	-0.030	0.830	-0.168	0.199	-0.434	< 0.05	-0.149	0.254
Bi-weekly odd-random 3	-0.053	0.695	-0.301	< 0.05	-0.379	< 0.05	-0.218	0.094
Monthly-random 1	-0.048	0.737	-0.085	0.521	-0.292	< 0.05	-0.212	0.104
Monthly-random 2	-0.062	0.643	-0.191	0.143	-0.398	< 0.05	-0.255	< 0.05
Monthly-random 3	0.154	0.239	-0.131	0.318	-0.324	< 0.05	-0.237	0.069
Monthly-random 4	-0.223	0.087	-0.191	0.143	-0.389	< 0.05	-0.237	0.069
Seasonal_random 1	0.043	0.770	-0.163	0.212	-0.281	< 0.05	-0.197	0.156
Seasonal_random 2	-0.182	0.201	-0.057	0.669	-0.108	0.412	-0.076	0.574
Seasonal_random 3	-0.021	0.890	-0.099	0.464	-0.340	< 0.05	-0.074	0.594
Seasonal_random 4	-0.368	0.008	-0.168	0.199	-0.345	< 0.05	-0.164	0.228

Table S4. Results from the Mann-Kendall test for total phosphorus concentrations during the observed period (1986-2015) for Maumee Station

Dataset	Total Phosphorus							
	Winter		Spring		Summer		Fall	
	tau	p-value	tau	p-value	tau	p-value	tau	p-value
Daily	0.177	0.175	-0.136	0.301	-0.315	< 0.05	-0.186	0.153
Weekly-random 1	0.186	0.153	-0.147	0.261	-0.366	< 0.05	-0.113	0.392
Weekly-random 2	0.186	0.153	-0.147	0.261	-0.366	< 0.05	-0.113	0.392
Weekly-random 3	0.186	0.153	-0.147	0.261	-0.366	< 0.05	-0.113	0.392
Bi-weekly even-random 1	0.164	0.212	-0.242	0.063	-0.260	< 0.05	-0.166	0.205
Bi-weekly even-random 2	0.083	0.532	-0.071	0.592	-0.350	< 0.05	-0.175	0.181
Bi-weekly even-random 3	0.090	0.498	-0.113	0.392	-0.269	< 0.05	-0.168	0.199
Bi-weekly odd-random 1	0.168	0.199	-0.102	0.443	-0.357	< 0.05	-0.126	0.335
Bi-weekly odd-random 2	0.189	0.148	-0.111	0.402	-0.356	< 0.05	-0.094	0.475
Bi-weekly odd-random 3	0.173	0.187	-0.136	0.301	-0.240	0.066	-0.164	0.212
Monthly-random 1	0.247	0.069	0.044	0.748	-0.236	0.071	-0.065	0.630
Monthly-random 2	0.118	0.372	-0.134	0.309	-0.268	< 0.05	-0.245	0.063
Monthly-random 3	0.297	< 0.05	-0.063	0.642	-0.227	0.083	-0.317	< 0.05
Monthly-random 4	0.021	0.886	-0.058	0.668	-0.257	< 0.05	-0.137	0.300
Seasonal_random 1	0.219	0.117	0.021	0.886	-0.170	0.208	-0.078	0.579
Seasonal_random 2	-0.053	0.724	-0.007	0.971	-0.054	0.694	-0.063	0.652
Seasonal_random 3	0.174	0.205	-0.060	0.666	-0.165	0.221	-0.231	0.095
Seasonal_random 4	-0.210	0.137	-0.162	0.224	-0.343	< 0.05	-0.100	0.475

Table S5. Results from the Mann-Kendall test for soluble reactive phosphorus concentrations during the observed period (1986-2015) for Maumee Station

Dataset	**Soluble Reactive Phosphorus**							
	Winter		Spring		Summer		Fall	
	tau	p-value	tau	p-value	tau	p-value	tau	p-value
Daily	0.545	< 0.05	0.301	< 0.05	0.002	1.000	0.154	0.239
Weekly-random 1	0.526	< 0.05	0.241	0.064	0.030	0.830	0.159	0.225
Weekly-random 2	0.526	< 0.05	0.241	0.064	0.030	0.830	0.159	0.225
Weekly-random 3	0.526	< 0.05	0.241	0.064	0.030	0.830	0.159	0.225
Bi-weekly even-random 1	0.529	< 0.05	0.269	< 0.05	0.048	0.721	0.145	0.269
Bi-weekly even-random 2	0.548	< 0.05	0.408	< 0.05	0.030	0.830	0.154	0.239
Bi-weekly even-random 3	0.554	< 0.05	0.290	< 0.05	0.103	0.432	0.205	0.116
Bi-weekly odd-random 1	0.531	< 0.05	0.191	0.143	-0.057	0.669	0.177	0.175
Bi-weekly odd-random 2	0.485	< 0.05	0.336	< 0.05	0.025	0.858	0.205	0.116
Bi-weekly odd-random 3	0.559	< 0.05	0.209	0.108	0.011	0.943	0.237	0.069
Monthly-random 1	0.508	< 0.05	0.322	< 0.05	0.083	0.559	0.091	0.499
Monthly-random 2	0.529	< 0.05	0.023	0.872	0.111	0.409	0.069	0.605
Monthly-random 3	0.566	< 0.05	0.334	< 0.05	0.030	0.830	0.143	0.276
Monthly-random 4	0.442	< 0.05	0.118	0.372	0.046	0.734	0.005	0.986
Seasonal_random 1	0.595	< 0.05	0.355	< 0.05	0.080	0.591	0.049	0.738
Seasonal_random 2	0.284	0.050	0.049	0.738	0.148	0.268	0.195	0.162
Seasonal_random 3	0.430	< 0.05	0.157	0.270	-0.007	0.970	0.095	0.504
Seasonal_random 4	0.412	< 0.05	-0.035	0.807	-0.042	0.767	-0.041	0.791

Table S6. Results from the Mann-Kendall test for nitrate + nitrite concentrations during the observed period (1986-2015) for Maumee Station

Dataset	Nitrate + Nitrite							
	Winter		Spring		Summer		Fall	
	tau	p-value	tau	p-value	tau	p-value	tau	p-value
Daily	-0.241	0.064	-0.191	0.143	-0.278	< 0.05	-0.154	0.239
Weekly-random 1	-0.241	0.064	-0.241	0.064	-0.306	< 0.05	-0.136	0.301
Weekly-random 2	-0.241	0.064	-0.241	0.064	-0.306	< 0.05	-0.136	0.301
Weekly-random 3	-0.241	0.064	-0.241	0.064	-0.306	< 0.05	-0.136	0.301
Bi-weekly even-random 1	-0.246	0.059	-0.175	0.181	-0.264	< 0.05	-0.163	0.212
Bi-weekly even-random 2	-0.186	0.153	-0.214	0.101	-0.283	< 0.05	-0.154	0.239
Bi-weekly even-random 3	-0.223	0.087	-0.186	0.153	-0.274	< 0.05	-0.149	0.254
Bi-weekly odd-random 1	-0.241	0.064	-0.172	0.187	-0.283	< 0.05	-0.122	0.354
Bi-weekly odd-random 2	-0.223	0.087	-0.186	0.153	-0.186	0.153	-0.067	0.617
Bi-weekly odd-random 3	-0.223	0.087	-0.163	0.212	-0.283	< 0.05	-0.159	0.225
Monthly-random 1	-0.061	0.664	-0.099	0.454	-0.128	0.339	-0.067	0.617
Monthly-random 2	-0.191	0.143	-0.195	0.134	-0.193	0.139	-0.149	0.254
Monthly-random 3	-0.297	0.022	-0.103	0.432	-0.200	0.125	-0.214	0.101
Monthly-random 4	-0.241	0.064	-0.205	0.116	-0.202	0.129	-0.030	0.830
Seasonal_random 1	-0.077	0.588	-0.051	0.708	0.036	0.823	-0.267	0.065
Seasonal_random 2	-0.133	0.362	-0.117	0.372	-0.148	0.300	0.066	0.646
Seasonal_random 3	-0.323	0.017	0.054	0.694	-0.124	0.400	-0.016	0.921
Seasonal_random 4	-0.177	0.203	-0.049	0.722	0.004	> 0.999	-0.074	0.602

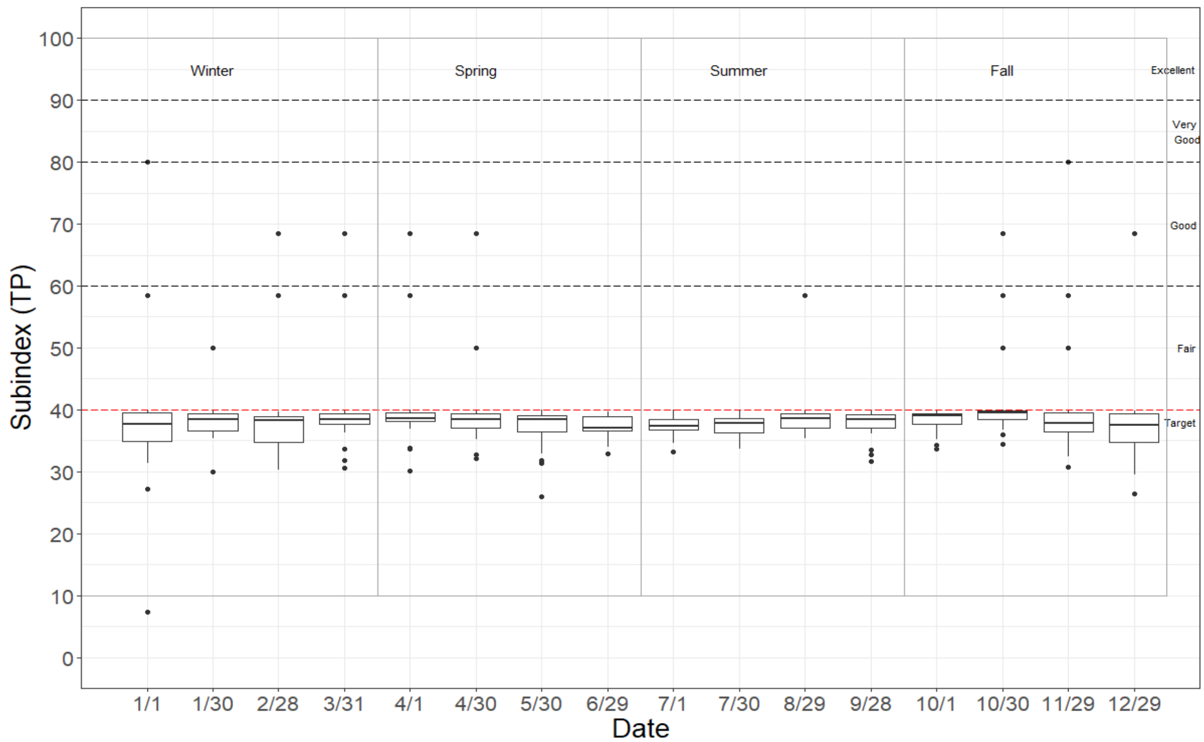

Fig. S1 Annual subindex distribution for total phosphorus at the Maumee River water quality station

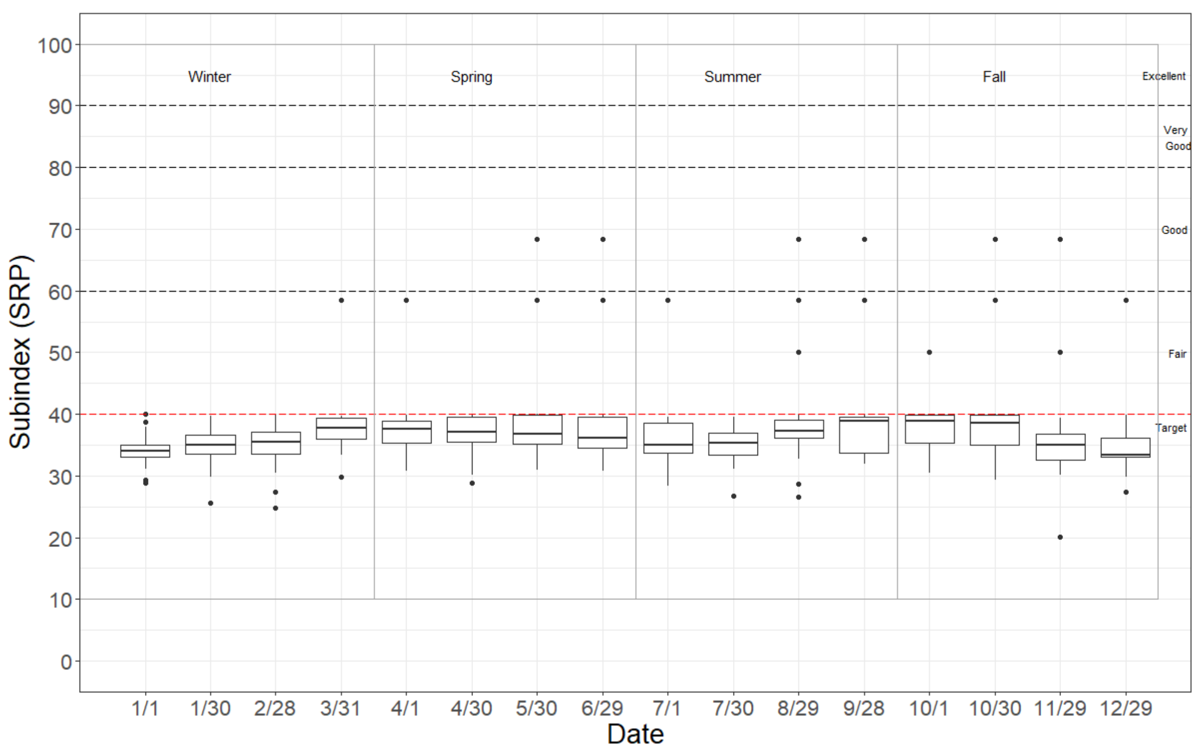

Fig. S2 Annual subindex distribution for soluble reactive phosphorus at the Maumee River water quality station

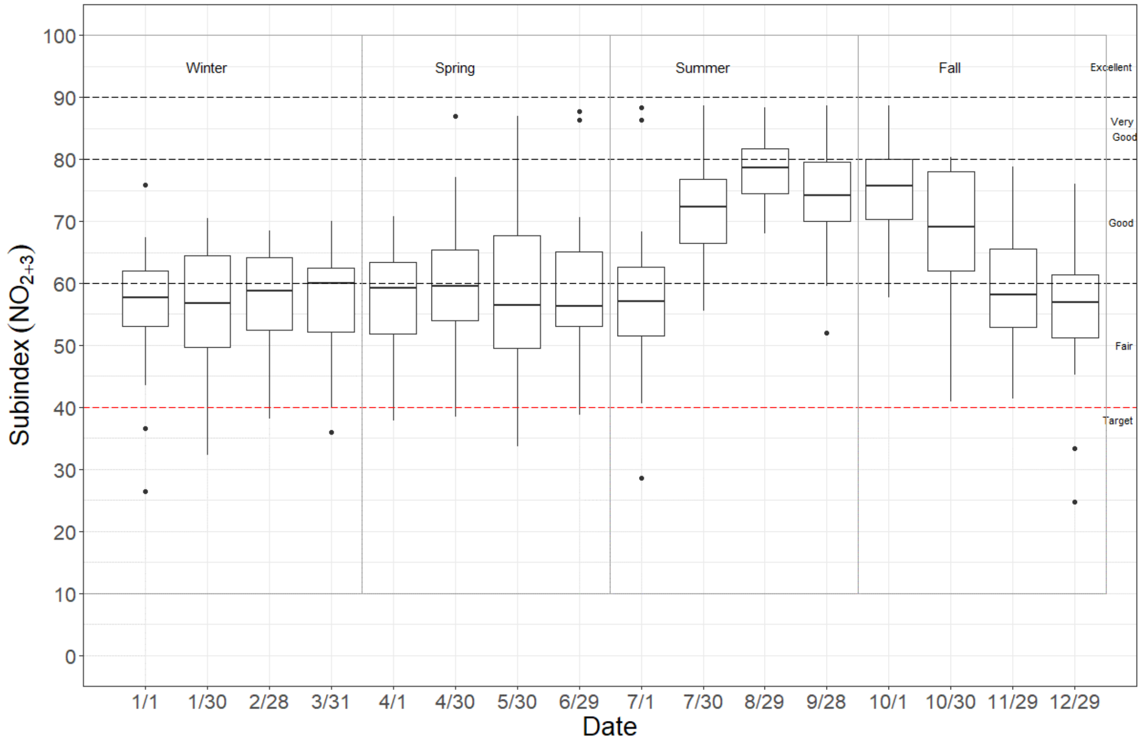

Fig. S3 Annual subindex distribution for nitrate+nitrite at the Maumee River water quality station

Table S7. Water Quality Objectives for the Otun River as indicated in the 2015 Water Ordinance Plan

Water Quality Parameter	**Water Quality Objectives (WQO)**														
	Stretch														
	1. Source to Barbo River			2. Barbo River to Nuevo Libaré intake			3. Nuevo Libaré intake to Gaitán Bridge			4. Gaitán Bridge to Belmonte power house			5. Belmonte power house to discharge		
	5 yr	10 yr	20 yr	5 yr	10 yr	20 yr	5 yr	10 yr	20 yr	5 yr	10 yr	20 yr	5 yr	10 yr	20 yr
BOD_5 (mg/L)	3	3	3	4	4	3	3	3	3	20	30	10	20	30	10
COD (mg/L)	5	6	5	5	6	5	5	6	5	60	60	40	60	60	40
pH	6.5–9.0	6.5–8.5	6.5–8.5	6.5–9.0	6.5–9.0	6.5–9.0	6.5–9.0	6.5–8.5	6.5–8.5	6.5–9.0	6.5–9.0	6.5–9.0	6.5–8.5	6.5–8.5	6.5–8.5
Dissolved Oxygen (mg/L)	> 6	> 6.5	> 6.5	> 6	> 6.5	> 6.5	> 6	> 6	> 6	> 4.5	> 4.5	> 6	> 5	> 5	> 6
Conductivity (µS/cm)	700	750	500	750	750	500	700	750	500	750	750	500	750	750	750
Total Coliforms (MPN/ 100 mL)	<20,000	<20,000	<1,000	<20,000	<20,000	<5,000	<20,000	< 20,000	< 5,000	1 X 10^6	1 X 10^6	50,000	1 X 10^6	1 X 10^6	50,000
Fecal Coliforms (MPN/ 100 mL)	< 2,000	< 2,000	< 200	< 2,000	< 2,000	< 2,000	< 2,000	< 2,000	< 2,000	10,000	10,000	10,000	10,000	10,000	10,000
Total Suspended Solids (mg/L)	6	5	4	6	5	4	6	5	4	40	40	30	40	40	30
Oil and Grease (mg/L)	Non present	Non present	Non present	Non present	Non present	Non present	Non present	Non present	Non present	Non present	Non present	Non present	Non present	Non present	Non present
Total Phosphorus (mg P-PO_4/L)	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
Total Nitrogen (mg N/L)	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5

Table S8. Water Quality Objectives for the Otun River as indicated in the 2015 Water Ordinance Plan

Water Quality Parameter	Water Quality Objectives (WQO)					
	Stretch					
	1. Aguazul Creek source to ACUASEO intake			2. ACUASEO intake to discharge		
	5 yr	10 yr	20 yr	5 yr	10 yr	20 yr
BOD_5 (mg/L)	3	3	3	20	30	10
COD (mg/L)	5	6	5	60	60	40
pH	6.5 – 9.0	6.5 – 8.5	6.5 – 8.5	6.5 – 9.0	6.5 – 8.5	6.5 – 8.5
Dissolved Oxygen (mg/L)	> 6	> 6.5	> 6.5	> 5	> 5	> 6
Conductivity (µS/cm)	700	750	500	750	750	750
Total Coliforms (MPN/ 100 mL)	< 20,000	< 20,000	< 1,000	1×10^6	1×10^6	50,000
Fecal Coliforms (MPN/ 100 mL)	< 2,000	< 2,000	< 200	10,000	10,000	10,000
Total Suspended Solids (mg/L)	6	5	4	40	40	Non reported
Oil and Grease (mg/L)	Non present	Non present	Non present	Non present	Non present	Non present
Total Phosphorus (mg $P\text{-}PO_4$/L)	2	2	2	2	2	2
Total Nitrogen (mg N/L)	0.5	0.5	0.5	0.5	0.5	0.5