DEVELOPMENT OF A WATER QUALITY INDEX (WQI) FOR THE SUSQUEHANNA RIVER BASIN

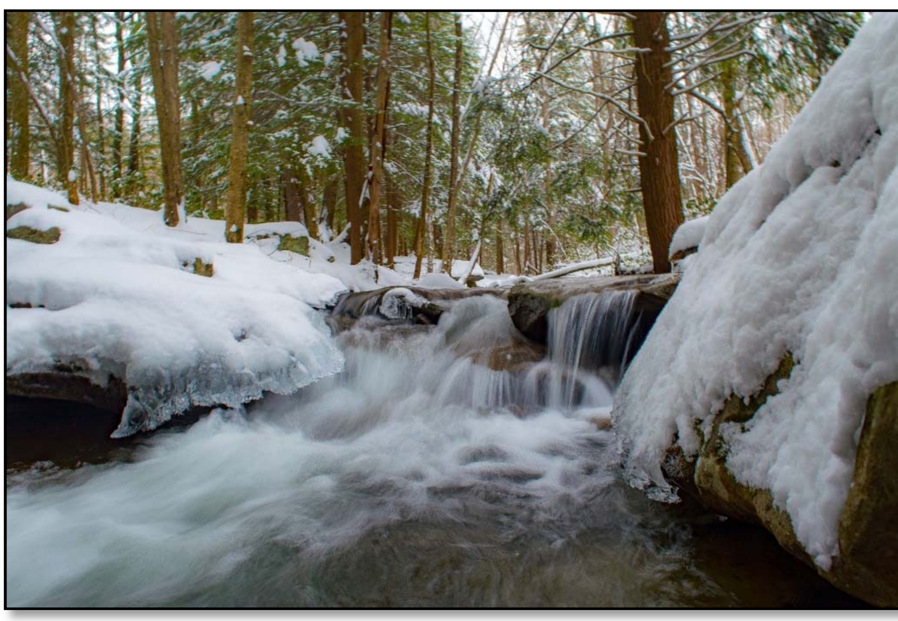
Publication No. 322

January 2020

Joanna L. Berry
Associate Environmental Scientist

Luanne Y. Steffy
Aquatic Biologist

Matthew K. Shank
Aquatic Biologist



SUSQUEHANNA RIVER BASIN COMMISSION
NY · PA · MD · USA

TABLE OF CONTENTS

FIGURES

cover photo: Moose Creek, Clearfield County, PA (courtesy of Matthew Shank, February 2018)

TABLES

APPENDICES

ABSTRACT

Water quality indices (WQIs) are tools used to evaluate water quality monitoring data and allow scientists to place discrete monitoring results into a larger context. This allows for meaningful interpretation of water chemistry results, particularly when concentrations are below water quality standards. Due to confounding environmental conditions, WQIs perform best at regional or local scales. A WQI was developed for the Susquehanna River Basin (SRB) to 1) assess water quality of streams during baseflow conditions and allow for comparisons between monitoring sites within the watershed, 2) allow water quality information to be easily understood and used by decision makers and the public, and 3) serve as the basis for a stressor gradient to allow for the evaluation of biological condition. Three tiers of the WQI were developed. The first score compared results across the whole SRB (WQI_{SRB}), the second compared only similarly sized streams to one another, and the third compared similarly sized streams within four main ecoregions (level III) in the SRB. The Susquehanna WQI converts raw concentrations of nine commonly monitored parameters into a unitless number between 0 and 100 where the greater the number, the better the water quality. The nine parameters are grouped into three categories: metals (aluminum, iron, and manganese); nutrient enrichment (nitrate, total phosphorus, and total organic carbon); and development (chloride, sodium, and sulfate). Each parameter score is determined based on percentile ranking and then averaged into a categorical score. The three categorical scores are then averaged to produce an aggregate score. Further analyses indicate that the WQI_{SRB} showed significant correlation with biological assemblage and land use data, resulting in a useful water quality assessment and biomonitoring tool for use within the SRB.

INTRODUCTION

Water quality is complex and multidimensional, varying throughout time and space, and is affected by anthropogenic and natural influences. The multitude of factors influencing water quality make it challenging to define. Moreover, the importance of the concentration of a parameter in a water sample below established water quality standards is difficult to conceptualize. When concentrations of parameters are not placed into context, answering questions about how "good" water quality is, how water quality has changed over time, or how to compare water quality among various bodies of water is difficult. Oftentimes, the most that can be said is that a parameter exceeds a water quality standard, and these questions remain unanswered.

The first water quality index (WQI) was created by Horton (1965), and WQIs have since become a popular tool in water resource management. WQIs condense selected parameters into a single unitless score, allowing for changes over time or comparisons between streams to be easily understood. WQIs perform best at regional or local scales; therefore, a unique set of watershed-specific parameters should be selected for individual river basins (Sutadian et al., 2016). WQIs are used to evaluate water quality monitoring data and allow scientists to place discrete monitoring results into a larger context. This allows for meaningful interpretation of water chemistry results, particularly when concentrations are below water quality standards. WQIs are also useful in presenting water quality information in an easily understandable way to the public.

The Susquehanna River Basin Commission (Commission) routinely monitors biological, chemical, and physical characteristics of streams across the Susquehanna River Basin (SRB). These monitoring efforts range from specific short-term projects focused on one type of impact (e.g., abandoned mine drainage (AMD), agriculture, or stormwater) to long-term basinwide monitoring efforts like the Remote Water Quality Monitoring Network and Subbasin Survey (https://www.srbc.net/our-work/programs/monitoring-protection/).

When analyzing water quality data, Commission scientists face an ongoing challenge of how to effectively evaluate the condition of the water quality of a stream, much less compare those conditions to nearby streams or across the SRB as a whole. The public often asks about the condition of local streams, and decision makers need to have ready answers to questions about a stream's status. As scientists, an ongoing dilemma exists with how to best translate technical data into something stakeholders and the public can readily understand and use. Commission scientists determined that an easily understandable WQI could meet a wide variety of needs and serve multiple purposes.

The Susquehanna WQI was developed to 1) assess water quality at a stream site over time and allow for comparisons between sites within the SRB during baseflow conditions, 2) allow water quality information to be easily understood and used by decision makers and the public, and 3) serve as the basis for a stressor gradient to allow for the evaluation of biological condition. The Susquehanna WQI will serve as an essential tool in the evaluation of streams and allow citizens and scientists to make more meaningful and quantitative assessments of spatial and temporal water quality conditions.

METHODS

Study Area

The Susquehanna River Basin (SRB) covers 27,510 square miles of New York, Pennsylvania, and Maryland and provides about 50 percent of the freshwater flowing into the Chesapeake Bay (Seitz, 1971). Land use is predominately forested (62.5 percent) followed by cultivated (27.5 percent) and developed lands (4.2 percent) (SRBC, 2013). The SRB includes parts of four major ecoregions: Central Appalachian Ridges and Valleys (RV, 34 percent), Northern Appalachian Plateau and Uplands (NAPU, 32 percent), North Central Appalachians (NCA, 22 percent), and Northern Piedmont (11 percent) (Figure 1; Omernik, 1987).

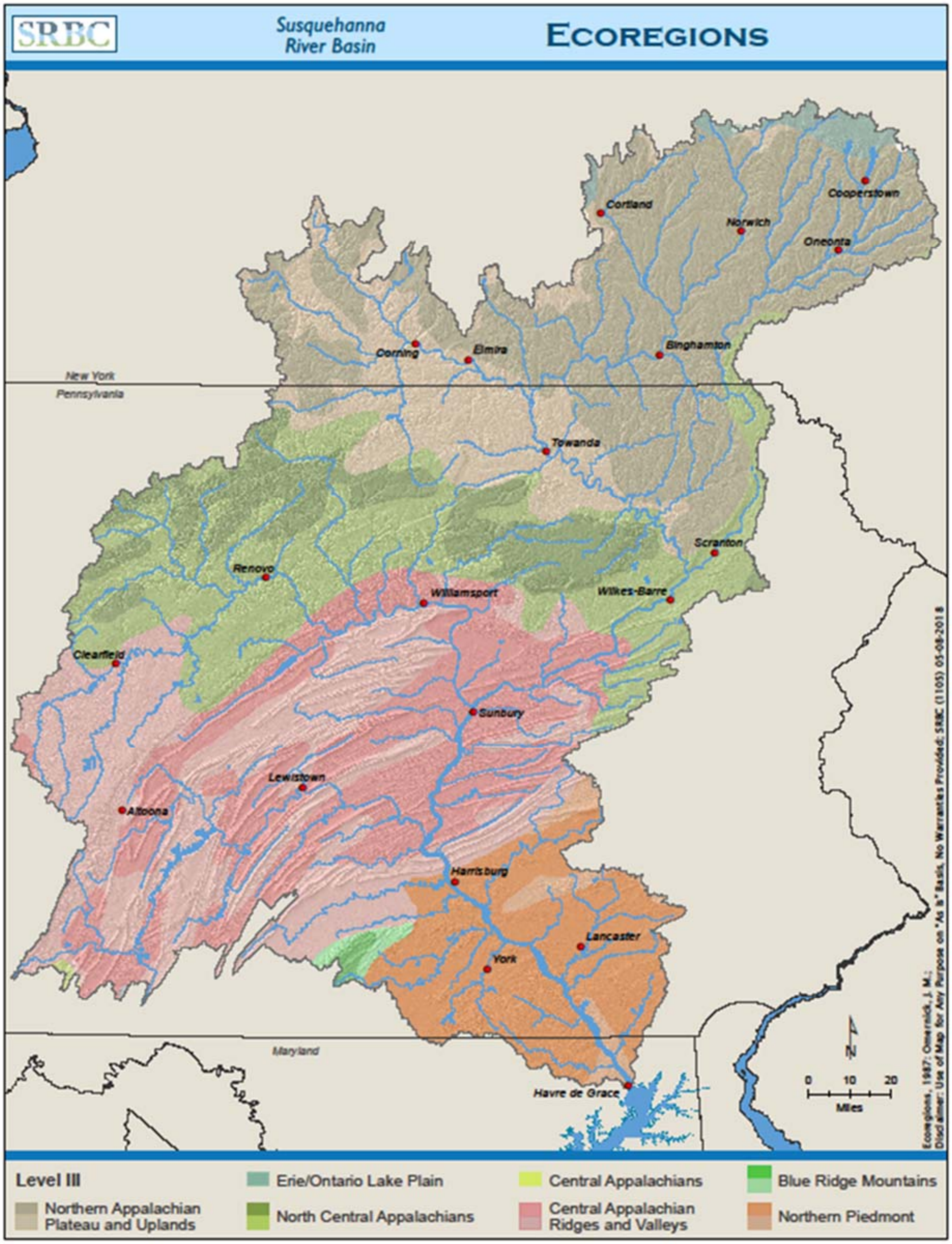


Figure 1. Ecoregions of the Susquehanna River Basin (Four major ecoregions underlay the SRB: Central Appalachian Ridges and Valleys (RV), Northern Appalachian Plateau and Uplands (NAPU), North Central Appalachians (NCA), and Northern Piedmont. Three minor ecoregions, Blue Ridge Mountains, Central Appalachians and Erie/Ontario Lake Plains, also fall within the watershed.)

The Ridge and Valley ecoregion is a highly diverse, long, and narrow ecoregion extending from New York to Alabama. Due to extreme folding and faulting, the region is comprised of roughly parallel ridges and valleys. Mountainous areas are highly forested, and overall forests cover about 50 percent of the region. Underlying geology includes limestone, dolomite, shale, siltstone, sandstone, chert, mudstone, and marble (USEPA, 2013).

The NAPU ecoregion is similar to the Ridges and Valleys ecoregion with low mountains and hills, open valleys, and Appalachian oak forests. Mixed hardwood forests are also present at higher elevations. Land use in the NAPU is a mix of agriculture (croplands and pasturelands) and forest. The region is underlain with glacial till from the Wisconsin age and has horizontally bedded shales, siltstones, and sandstones (USEPA, 2013).

The NCA is a highly forested ecoregion comprised of plateaus, high hills, and low mountains that were largely unaffected by glaciation present in the NAPU and other northern ecoregions. Land is characteristically less fertile and forests are more rugged in the NCA when compared to NAPU; agricultural land use is low while forestry and recreational uses are common. Some natural gas and coal activities are present in the region’s western reaches. The region is underlain with horizontally bedded sandstone, shale, siltstone, conglomerate, and coal (USEPA, 2013).

The Northern Piedmont ecoregion is a transitional region with small rolling hills, scattered plains, and open valleys. These valleys contain a larger proportion of agriculture use (croplands) than the Piedmont ecoregion to its south. Forests are predominantly Appalachian oak. The region is underlain by metamorphic, igneous, and sedimentary rocks (USEPA, 2013). Ecoregion and underlying geology influence natural concentrations of many water quality parameters that differ across the SRB.

Data Collection and Compilation

Water samples were collected using depth-integrating sampling across a transect perpendicular to streamflow. Only samples collected during baseflow conditions were included in the dataset to avoid misrepresenting water quality parameter concentrations during elevated flow or suspended sediment loads. Raw water samples were preserved in the field and all filtering was done by the analyzing laboratory; all laboratory results were reported as total fractions of the parameter. Concentrations of water quality parameters were determined by accredited laboratories using U.S. Environmental Protection Agency (USEPA)-approved methods.

The WQI dataset was comprised of Commission data collected during the 10-year period from 2008-2017 throughout the SRB for all monitoring purposes. Data have been systematically stored in a Microsoft Access database and underwent a QA/QC check prior to analysis to ensure data quality. The QA/QC check involved general data clean-up and replacing non-detects or “present below quantification limit” results with the value of the detection limit. The final dataset included 8,119 records from 1,394 unique stream sites.

Water Quality Index Development

Building a WQI involves four main steps: the selection of parameters, creation of sub-indices, determining parameter weights, and the aggregation process to generate the final water quality score (Abbasi and Abbasi, 2012). The considerations and motives for how we handled each of the four steps are explained below.

I. Parameter Selection

Because the dataset was comprised of previously collected data, parameter selection was limited to parameters with an adequate sample size. Fourteen routinely sampled parameters were considered (Table 1). The nine parameters included in the Susquehanna WQI—aluminum (Al), chloride (Cl), iron (Fe), manganese (Mn), nitrate (NO_3), phosphorus (P), sodium (Na), sulfate (SO_4), and total organic carbon (TOC)—all had an adequate sample size in addition to being meaningful water quality constituents within the SRB (Figure 2). Each of the nine parameters were reported and analyzed as total concentrations. All of these naturally occurring parameters can be detrimental to stream biota when present at elevated concentrations. Alkalinity, calcium (Ca), magnesium (Mg), pH, and potassium (K) also influence stream biota but were left out of the index because each may vary considerably across the SRB due to natural variability in ecoregion and underlying geology.

Table 1. Number of Records and Ranges of Each Parameter in the WQI Dataset (Minimum and maximum values are reported as mg/l for all parameters except pH. Parameters selected for inclusion are in bold.)

Parameter	Alkalinity	**Al**	Ca	**Cl**	**Fe**	Mg	**Mn**	**NO_3**	pH	**P**	K	**Na**	**SO_4**	**TOC**
N	4453	**4061**	3516	**4220**	**3512**	3500	**3287**	**4536**	7895	**3774**	1935	**2921**	**4352**	**4122**
Minimum (mg/l)	0	**0.05**	0.529	**0.5**	**0.02**	0.29	**0.003**	**0.015**	2.4	**0.01**	0.25	**0.306**	**1**	**0.5**
Maximum (mg/l)	287	**21.7**	138	**243**	**44.4**	58.6	**45**	**13.8**	11.9	**2.619**	9.5	**128**	**1361**	**41.1**

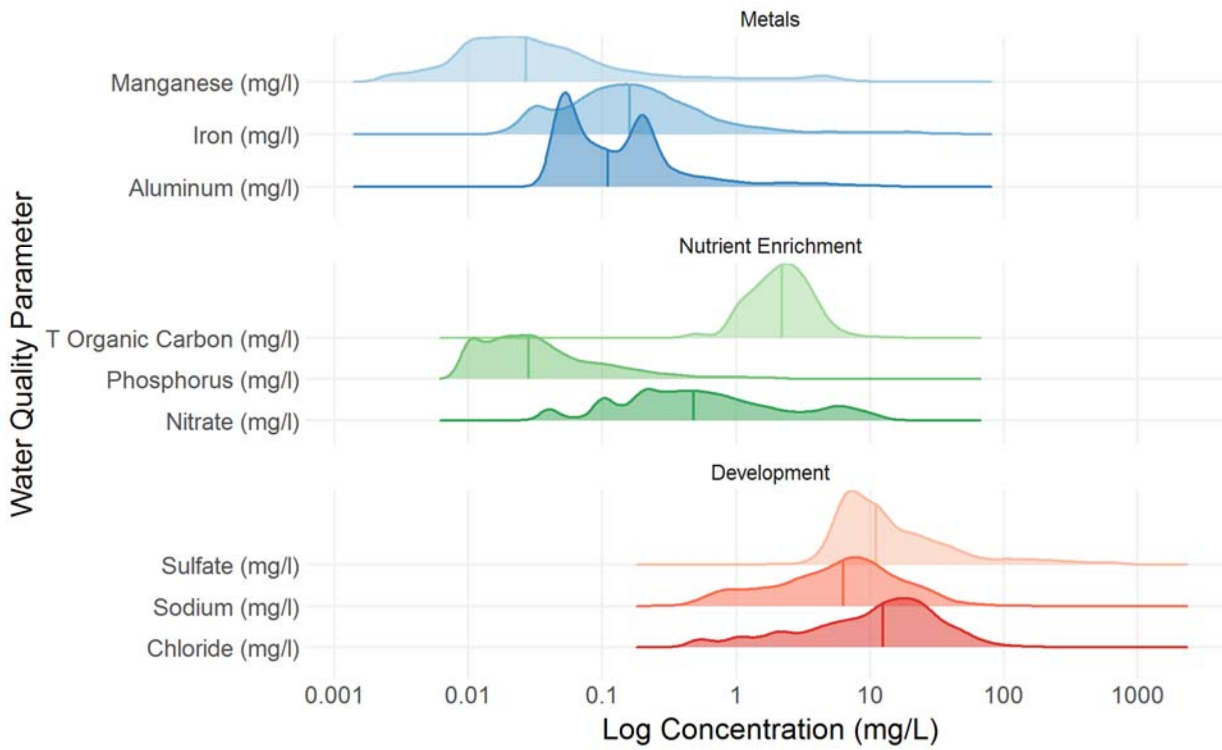


Figure 2. Proportional Frequency and Ranges of Water Quality Parameters Selected for Inclusion in the WQI Database (Vertical line indicates median value of parameter concentration.)

II. Category Creation

Many WQIs focus on wastewater management, and relatively few include metals (Abbasi and Abbasi, 2012). Although largely forested, the SRB is faced with three major threats to water quality: AMD, agriculture, and urban development (SRBC, 2013). These three broad threats to water quality reflect regionally specific land uses (past and present), emphasizing the need for a SRB-specific WQI. The nine parameters were sorted into three sub-indices or categories to reflect these threats to water quality: metals (Al, Fe, Mn), nutrient enrichment (NO_3, P, TOC), and development (Cl, Na, SO_4) (Figure 2).

A. Metals Category

The metals category includes aluminum, iron, and manganese and identifies streams that are affected by AMD resulting from historic coal mining. The Commission routinely monitors and works to restore AMD-impaired streams. Aluminum, iron, and manganese are the three indicators of impairment or improvement as they are the three metals most commonly associated with AMD in the SRB (Chaplin, 2005; Clark, 2013). When iron precipitates from pyrite, it coats the channel substrate and makes the stream unsuitable for aquatic life (Stoe, 1998). In humans, excess manganese can cause neurotoxicity and lead to a form of Parkinsonism (Perl and Olanow, 2007). Limited research has been conducted on its effects on freshwater aquatic life (Chaplin, 2005); however, manganese is a known threat to marine life (Pinsino et al., 2012). Aluminum is more toxic to aquatic life than iron or manganese; concentrations above 0.2 mg/l can impair life

(Baker and Schofield, 1982), while chronic levels greater than 0.5 mg/l can be lethal to fish and many macroinvertebrates (Earle and Callaghan, 1998).

B. Nutrient Enrichment Category

The nutrient enrichment category contains nitrate, phosphorus, and TOC, which are parameters that may be present due to anthropogenic influences including agriculture (e.g., fertilizer and animal waste entering water from cultivated fields and poor riparian buffers) and point source waste water treatment discharges (Gregory et al., 1991). An evaluation of 2,265 paired nitrate and total nitrogen samples within the SRB revealed that most of the nitrogen is comprised of nitrate (median value of 94 percent contribution, SRBC, unpublished data). High levels of nitrate are considered to be a human health risk, and PA has a drinking water standard of 10 mg/l (Zambrano and Stoner, 1998). Although they are not toxic, in excess, both nitrogen and phosphorus can be detrimental to water quality and aquatic life. Soluble forms are readily taken up by plants and algae, which can lead to algal blooms. Large amounts of algae can lower dissolved oxygen (DO) concentrations; lower DO concentrations in turn can kill fish and other aquatic species (USEPA, 2005). Total organic carbon (TOC) is a measurement of the total amount of carbon, including both dissolved and non-dissolved forms, in organic compounds in a system. TOC is a non-specific test generally used as an indicator of the level of organic contamination in a body of water, with riverine concentrations naturally occurring between one and 10 mg/l and municipal waste waters up to 1,000 mg/l (Whitehead, 2018).

C. Development Category

The development category includes chloride, sodium, and sulfate, which aligns with the threat of urban development. Elevated sodium and chloride are linked to transportation corridors (e.g., road salts), especially in snow affected areas (Stets et al., 2018; Minnesota PCA, 2017). High concentrations of chloride are harmful to aquatic life and can increase the mobility of metals (Novotny et al., 1998; Nelson et al., 2009; Duan and Kaushal, 2015), leading to some of the aforementioned metal threats and increased toxicity. Sulfate may result from industrial discharge and land disturbance (PADEP, 2009; WHO, 2004). High concentrations of sulfate are a human health risk and drinking water contaminant (PA Bulletin, 2017).

III. Determining Parameter Weights

We chose to assign equal weights to all parameters in each category in an attempt to remove some of the subjectivity over which parameters should be considered most important. The process of assigning and applying weights to parameters is not used in all WQIs (Abbasi and Abbasi, 2012). The relative importance of each parameter and its influence on water quality is subjective. When polled, experts often give different ranks to parameters (i.e., nitrate is more/ less important than phosphorus when determining water quality) or give different weights to parameters even if they have ranked them in the same order. Additionally, the overall score is in danger of being biased towards the most heavily weighted parameter if the score is weighted too high (Sutadian et al., 2016). While individual parameters are not actively weighted, the metals category is effectively weighted in the event that either iron or aluminum exceed aquatic life use standards through a “zeroing” out of that category score. See Section IV part C for more details.

IV. Aggregation Methods

A. Minimum Data Requirements

If data from a discrete sample are available for at least two of the three parameters in a category, the respective category score can be calculated. Further, if data are available for at least two of three parameters in all three categories, a WQI score can be calculated. If fewer than two of the three parameters are available for any category, a WQI score will not be calculated. However, category scores may be used independently from the WQI score. For example, in cases where the collection of all nine parameters is outside the scope of a study, there are benefits to scoring an individual category even when the WQI score cannot be calculated. Category scores can be compared across time and sites in the same manner as the WQI scores. Ideally, all three parameters in one category would be used to calculate the category score. Caution should be used when comparing sites that use a different combination of two parameters in the scoring (i.e., one site uses nitrate and TOC and one site uses nitrate and phosphorus).

B. Percentile Ranking

Each parameter in a category is percentile ranked based on all available data from the 2008-2017 WQI dataset. When concentrations of parameters were below minimum detection limits in the dataset, the detection limits were reported. As a result, non-detects represent the best water quality possible and fall at the 0^{th} percentile. If the parameter is greater than the maximum value in the dataset, the value for that parameter is assigned the 100^{th} percentile, representing the worst water quality possible.

C. Category Score

Once the percentile ranks are calculated for each parameter, the category score is calculated by subtracting the average of the parameter percentiles within that category from 100. This results in an intuitive scoring scale of 0 = worst and 100 = best.

Any parameter in high concentration may be detrimental to aquatic life and reduce the number of sensitive species, shifting assemblages towards metal, nutrient, or salt tolerant species (Tlili et al., 2016). Aquatic life standards have not been developed for all nine parameters included in the WQI. Aquatic life standards have been established for aluminum and iron, and we chose to use the most conservative values of the three SRB member states (which were from PA, Pennsylvania Bulletin, 2017) in the WQI. Because aluminum and iron are toxic to aquatic life at high concentrations, we chose to "zero" the metals category score if concentrations of either parameter exceeded the PA aquatic life standard, resulting in a lower WQI score. If the metals category has a parameter exceeding one or both aquatic life use standards, it will become zero regardless of individual parameters percent rank, effectively weighting the WQI score so that the WQI score strongly reflects the water quality violation (Table 1 within Figure 3). Other parameters in high concentrations lead to secondary effects and cascades, and should be considered when evaluating the WQI score, but are not lethal, so do not merit weighting the category score.

D. WQI_{SRB} Score

The three category scores are then averaged to yield a WQI_{SRB} score between 0-100 (Figure 3), with 0 indicating the worst water quality and 100 indicating the best water quality. Scores are then assigned classified as excellent, good, fair, poor, or very poor based on the entire SRB for assessment purposes.

Classification Thresholds Selection

Classification thresholds were established based on the relationship between WQI_{SRB} scores and macroinvertebrate index of biotic integrity (IBI) scores. Data were compiled from sites where macroinvertebrate samples were collected using riffle-run protocols simultaneously with water samples in the index period November–May (PADEP, 2013). WQI_{SRB} and the size-appropriate PADEP IBI score (small or large freestone IBI) were calculated for each paired sample. Water quality classifications are described in Table 2. For more information regarding classification threshold determination for the WQI_{SRB} using concurrently collected biological data, see Appendix A.

Table 2. Classifications for WQI_{SRB} Scores

Classification	WQI Score Range	Description of How Range was Determined
Excellent	≥ 85	> 80th percentile of WQI_{SRB} scores with "Good" IBI scores
Good	62 - 85	> 20th percentile of WQI_{SRB} scores with "Good" IBI scores
Fair	43 - 62	> 20th percentile of WQI_{SRB} scores with "Fair" IBI scores
Poor	31 - 43	> 20th percentile of WQI_{SRB} scores with "Poor" IBI scores
Very Poor	< 31	< 20th percentile of WQI_{SRB} scores with "Poor" IBI scores

Potential Impacts of High Flow and Total Suspended Sediment

The WQI was intended for use in evaluating baseflow (non-storm impacted) water quality conditions. High flow conditions could potentially misrepresent water quality results by contributing higher than typical amounts of surface runoff in relation to baseflow, which could result in atypical parameter concentrations. A high flow warning was added to the WQI tool to address this, where any sample taken during hydrological events exceeding 3.25 cubic feet per second per square mile (cfm) will be flagged and the user will be instructed to interpret results with caution. The 3.25 cfm threshold represents a percent streamflow exceedance value of 10 (P_{10}) at all unregulated USGS stream gages across the SRB, which is considered to be a high flow event (SRBC, 2012; DePhilip and Moberg, 2010). More details on the methods and validation for determining this flow threshold are described in Appendix B.

Similarly, turbid instream conditions following intense precipitation events or other stream disturbances may misrepresent water quality due to the affinity of water quality constituents to bind to suspended sediment particles. An additional warning was implemented due to the fact that suspended sediment concentrations between streams can vary greatly despite similar hydrologic conditions (Steffy and Shank, 2018) and by different flow conditions within the same stream (Beck and Birch, 2011). The application of a single value threshold for total suspended solids (TSS) to the state or basin level by either concentration or duration is challenging (Rowe et al., 2003).

With this in mind, an 80 mg/l TSS threshold was set based on an internal dataset of 21,653 samples covering 20 sites and including a wide range of TSS concentrations and WQI scores. However, this TSS warning concentration of 80 mg/l should be used with caution by end users, with the realization that this concentration may not represent the lowest impact threshold in all streams. For additional information on how both flow and TSS warning thresholds were validated, see Appendix B.

Susquehanna River Basin Water Quality Index (WQI) Flow Chart

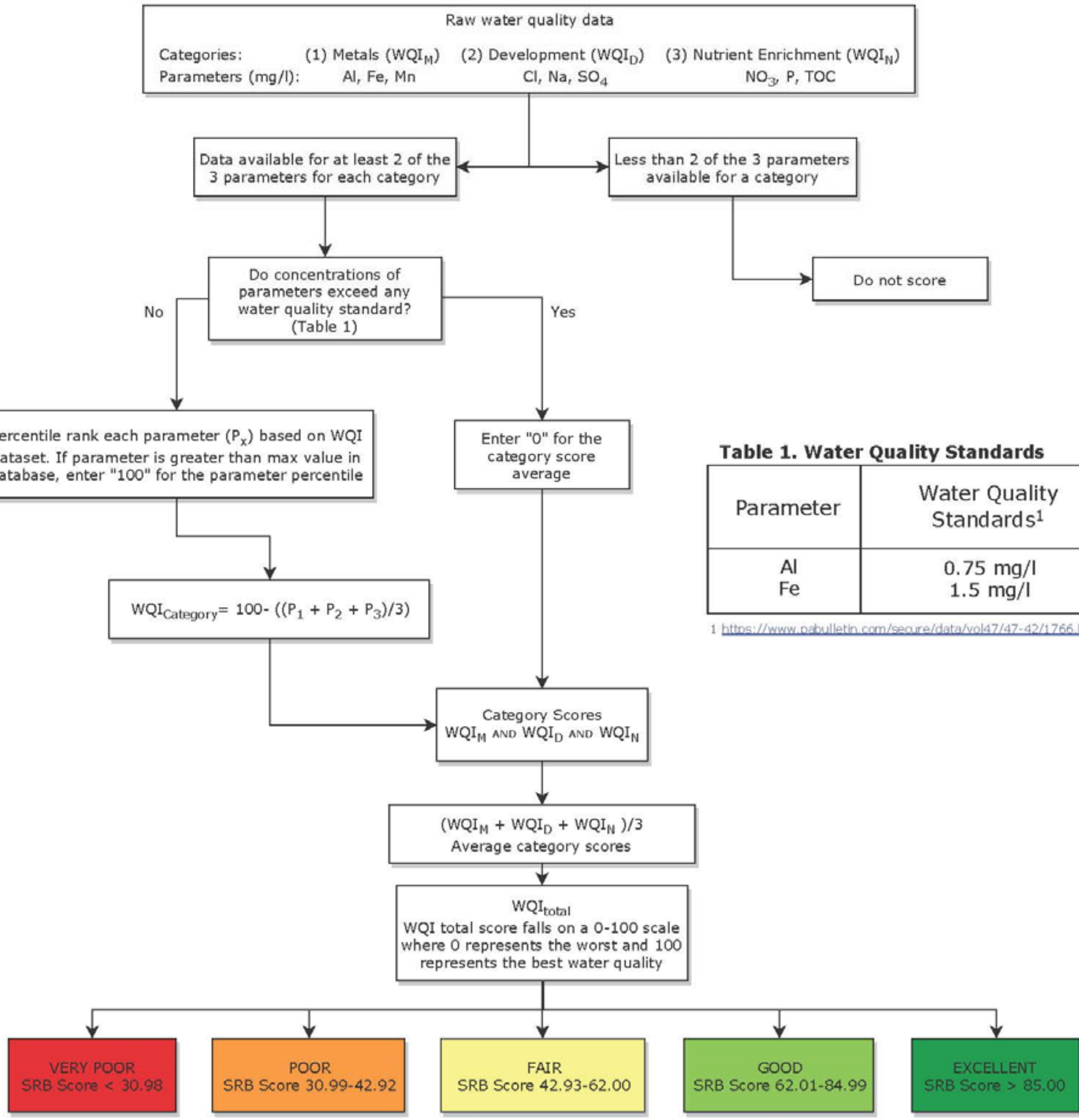


Table 1. Water Quality Standards

Parameter	Water Quality Standards[1]
Al	0.75 mg/l
Fe	1.5 mg/l

1 https://www.pabulletin.com/secure/data/vol47/47-42/1766.html

Figure 3. Susquehanna River Basin Water Quality Index Flow Chart (Classifications are only assigned to WQI_{SRB} scores, not categorical (WQI_m, WQI_D, and WQI_N) scores.)

Scoring Resolutions

The Susquehanna WQI follows a three-tiered approach, allowing the user to generate three different scores that can be used independently or in concert depending on the type of analysis of interest and project objectives. Additional spatial or temporal scoring resolutions above and beyond the ones described herein are possible and may be explored in the future as needs arise.

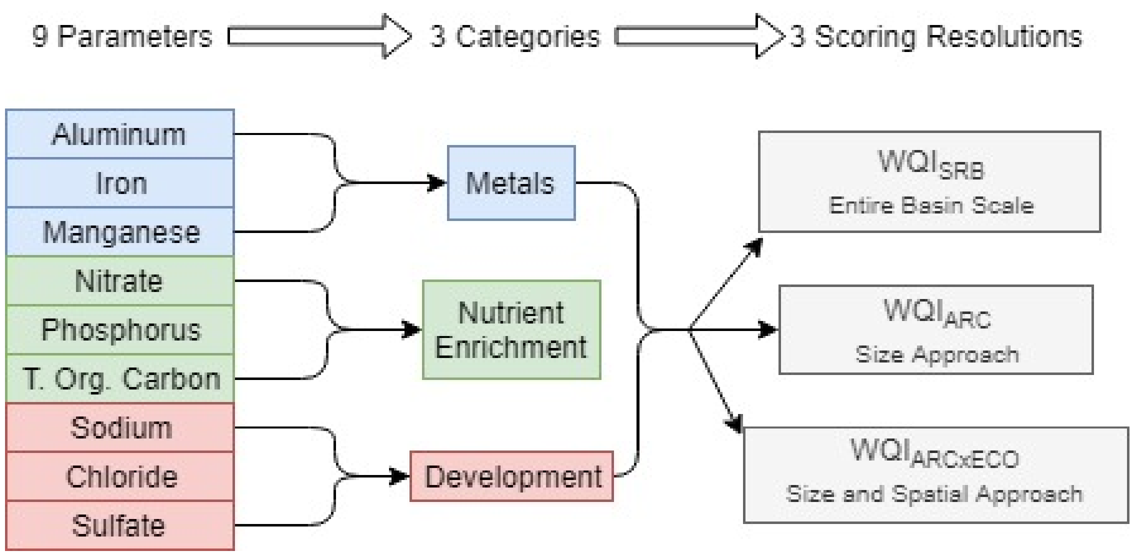


Figure 4. Flow Chart of Scoring Resolutions for the Susquehanna River Basin Water Quality Index (The nine parameters are condensed into three categories that are aggregated into a total WQI score using a three-tiered approach. The WQI$_{SRB}$ score compares all sites across the SRB. WQI$_{ARC}$ compares the site of interest only to similarly sized streams. WQI$_{ARCxECO}$ compares the site to similarly sized streams within its ecoregion.)

I. Basinwide Approach

As previously outlined, WQI$_{SRB}$ compares water quality conditions at a site to conditions at sites across the entire SRB by percentile ranking parameters against all records in the WQI dataset (Figure 4).

II. Stream Size Approach

WQI$_{ARC}$ generates a WQI score by comparing water quality only at similarly sized sites by percentile ranking parameter concentrations based on all data within the same aquatic resource class (ARC) groupings, a classification based on watershed area (SRBC, 2012). Headwaters and creeks with a drainage area of <50 mi^2 are included in the smallest grouping (ARC 1&2); small and medium tributary rivers are included in the intermediate grouping (ARC 3&4; 50-1000 mi^2); and medium mainstem and large rivers are included in the largest grouping (ARC 5&6; >1000 mi^2) (Table 3). Grouping sites in this manner accounts for river continuum controls (Vannote et al., 1980; DePhilip and Moberg, 2010). These groupings also align with Pennsylvania Department of Environmental Protection (PADEP) Benthic Index of Biotic Integrity Scores, as small and large freestone scoring align with ARC 1&2 and ARC 3&4 groups, respectively (Shull and Pulket, 2018), while ARC 5&6 sites correspond to PADEP's semi-wadeable multi metric index (Shull and Lookenbill, 2017).

Table 3. Drainage Area Ranges for Aquatic Resource Classes (SRBC, 2012)

ARC	Drainage Area Range (mi^2)
1	0-10
2	10-50
3	50-200
4	200-1000
5	1000-5000
6	>5000

III. Stream Size and Spatial Approach

$WQI_{ARCxECO}$ includes an even finer scale resolution, where a WQI score is generated by comparing a site to only those of a similar size within the same ecoregion. This tier percentile ranks a parameter within both ARC size groupings and level 3 ecoregion (Figure 4). There are three minor ecoregions represented in the SRB in addition to the main four. Due to the small sample sizes and limited coverage in contributing watershed area, these were grouped with adjacent ecoregions. Sites in the Blue Ridge Mountains (n=5) were grouped into the neighboring Northern Piedmont while sites in the Central Appalachians (n=3) were grouped into the Central Appalachian Ridges and Valleys. Sites in the Erie/Ontario Lake Plain ecoregion (n=3) were grouped into the NAPU (Figure 1).

The ecoregion that comprises the largest percentage of drainage area for a site determines the ecoregion for the site, as opposed to the ecoregion in which the site is located. Thus, there are no ARC 5&6 sites classified as Northern Piedmont because large rivers, such as the mainstem of the Susquehanna River, may drain entire ecoregions (Figures 1 and 4). Distribution of sites within $WQI_{ARCxECO}$ in this manner generally resulted in a sufficient sample size across drainage categories and parameters. Smaller sample sizes were available for ARC 5&6 for all parameters and ecoregions, which is representative of the distribution of streams in each ecoregion of the SRB (Figure 5).

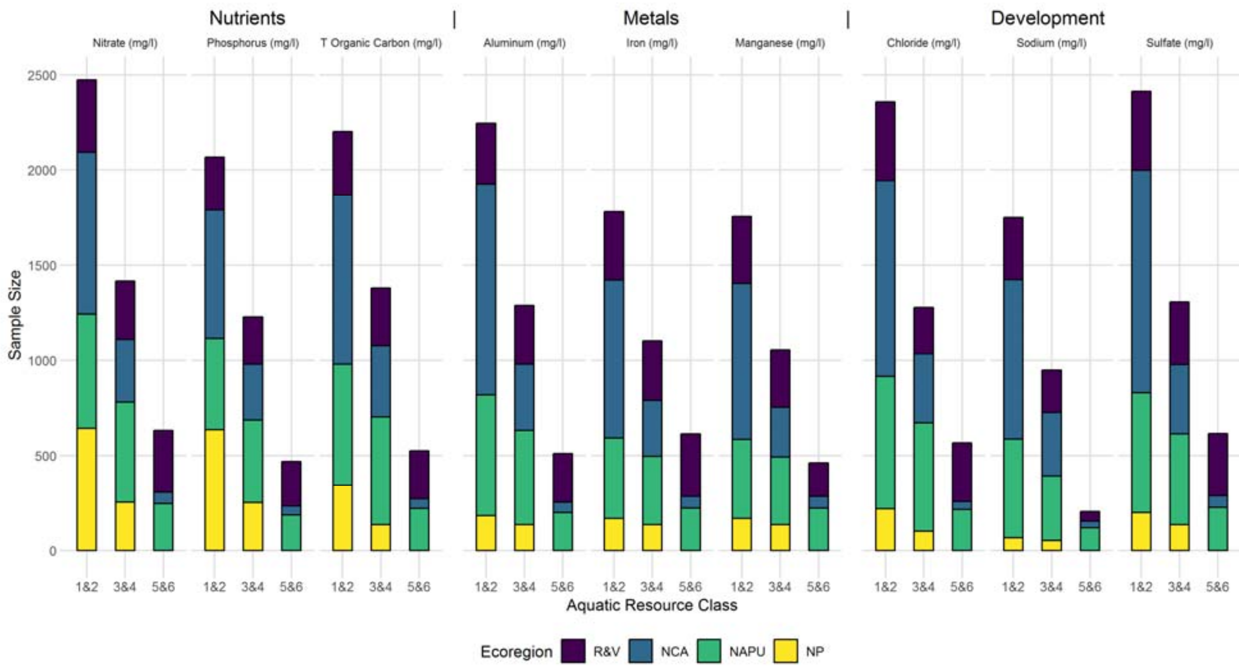


Figure 5. Records from the WQI Dataset Sorted into the Four Major Ecoregions and Aquatic Resource Class Groupings (These groupings were used to calculate percentile ranks for the WQI_{ARC} and $WQI_{ARCxECO}$ scores.)

Only the WQI_{SRB} scores are given water quality classifications because WQI_{SRB} scores had the strongest linear relationship with IBI score. Scores may vary depending on the WQI tier used. All three scores can be used collectively to create a comprehensive picture of the water quality of a stream in a variety of contexts. The utility and benefits of including multiple scoring resolutions within the Susquehanna WQI is emphasized in the examples included in Appendix C.

Correlation with Land Use

The correlation between WQI_{SRB} scores and land use were examined using principal components analysis (PCA). Land use was calculated for the entire watershed upstream of each sampling point using the 2006 Chesapeake Bay Watershed Land Cover Data Series (Irani and Claggett, 2010). To reduce the number of land use categories, the sum of urban (low, medium, and high intensity development), forested (deciduous, coniferous, mixed, and shrub/scrub), and agricultural (hay/pasture and cultivated crops) land uses were calculated and used as eigenvectors along with water, developed open space, wetland, and barren land uses. PCA analysis was conducted in R (R Core Team, 2018) using the prcomp function and visualized using the ggbiplot function (Vincent, 2011).

Validation with Biology

We also examined the correlation between biology and WQI_{SRB} score. This was an additional validation exercise, since water quality is a known stressor to biology (Metcalfe, 1989; Cairns and Pratt, 1993; USEPA, 2000, 2005; NC DEQ, 2007; Cooper et al., 2009; Abbasi and

Abbasi, 2012; MDE, 2018). To examine the relationship, we compiled data at sites where macroinvertebrates were collected simultaneously with water quality in the index period November-May. We calculated the appropriate PADEP macroinvertebrate Index of Biotic Integrity score (small or large freestone; PADEP, 2013) and then constructed an ordinary least squares linear regression model with the paired samples using WQI score to predict IBI score (283 samples at 255 unique sites). Model diagnostics were examined to ensure that constant variance and normal distribution of residuals were obtained. Significance was assessed at $\alpha = 0.05$ and all analyses were conducted in R (R Core Team, 2018).

RESULTS AND DISCUSSION

WQI_{SRB} Scores Correlation with Land Use

WQI_{SRB} scores correlated well with land use patterns across the basin (Figure 6). The West Branch Susquehanna River subbasin in Northcentral PA contains the largest contiguous tract of mature forest in the SRB and was found to have the highest WQI_{SRB} scores in the entire basin. However, there are streams in the western section of the West Branch subbasin affected by AMD – these sites have lower WQI_{SRB} scores than the rest of the subbasin. Likewise, the anthracite coal region of the Middle subbasin also had low WQI_{SRB} scores, reflecting a legacy of coal mining which results in widespread AMD influences today.

Similarly, the Lower Susquehanna subbasin in the Northern Piedmont has high agricultural and developed land uses and resulted in depressed WQI_{SRB} scores. The Cohocton River in the Chemung subbasin had low scores in all three categories. Many reaches of the mainstem have agricultural lands and/or roads with poor riparian buffers neighboring the stream. Better riparian buffers could protect the stream from salts and nutrients that may be making their way into the stream during snowmelt or heavy rain events. Additionally, the NAPU ecoregion where the Cohocton River is located has a higher proportion of reactive minerals due to the glacial till deposits (Rogers, 1989). Spatially, the WQI_{SRB} scores reflected knowledge of general land use issues in the basin.

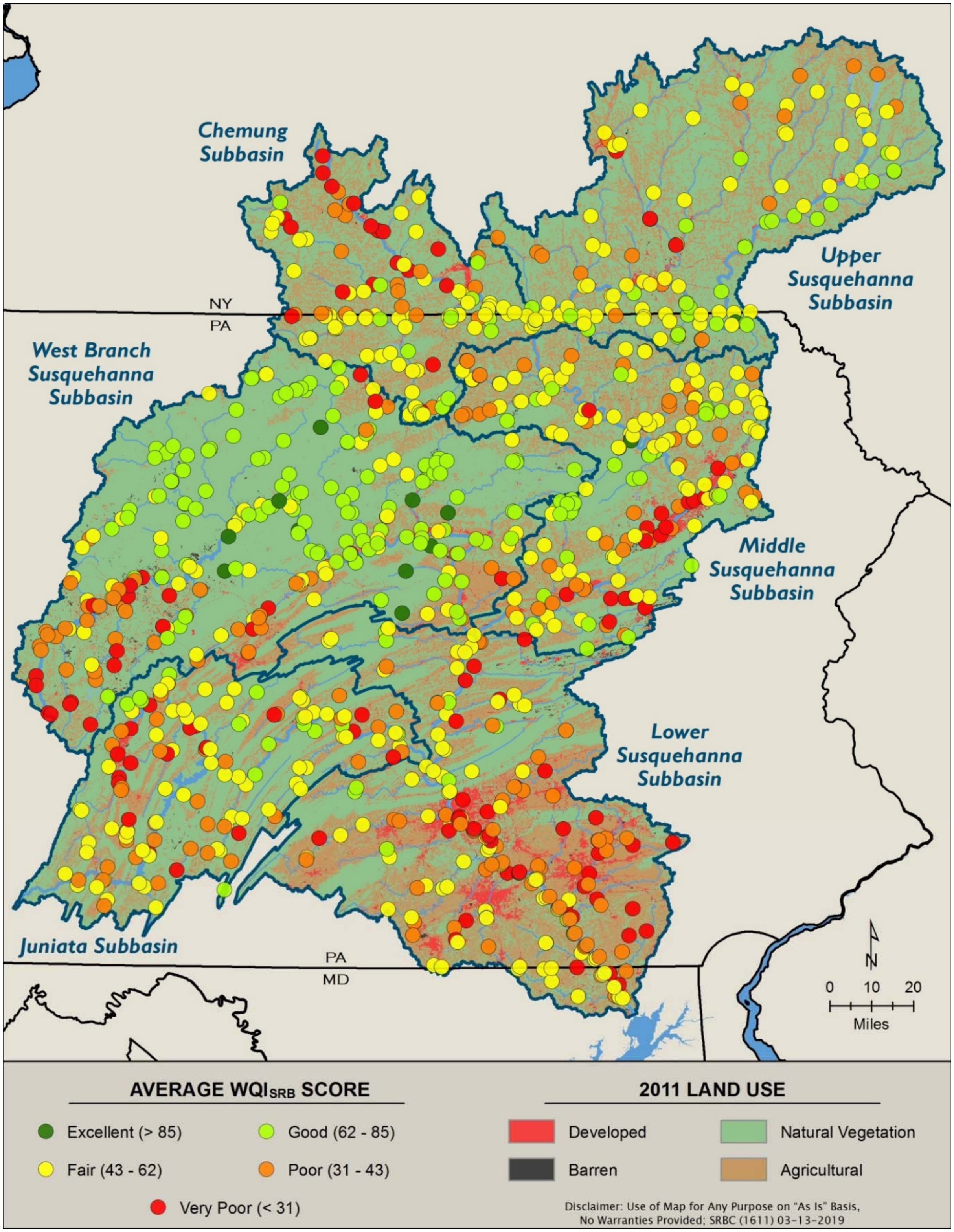


Figure 6. Average WQI$_{SRB}$ Scores at 1,016 Unique Sites in the SRB (2,750 records were scored and then averaged by site. Subbasins are outlined in blue and labeled.)

These relationships are reinforced through a PCA of land use and WQI_{SRB} scores. The PCA revealed a majority of variation can be described by agricultural and forested land uses in watersheds (horizontal axis of Figure 7). A smaller but meaningful amount of variation is described by the amount of developed, urban, and/or barren land in watersheds (vertical axis of Figure 7). Further, separation exists between the grouping of our sites. As expected, forest is associated with WQI_{SRB} scores >75 while agriculture and urban lands are associated with WQI_{SRB} scores <50 (Figure 7).

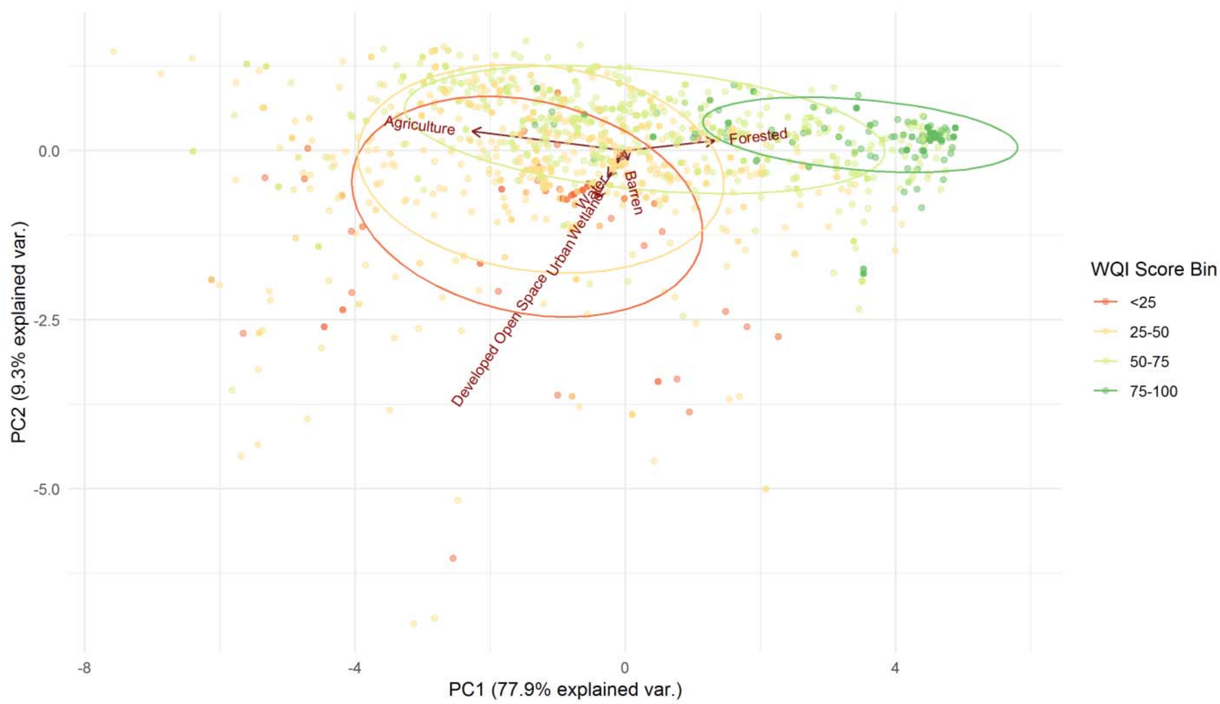


Figure 7. Principal Components Analysis (PCA) Plot of Aggregated Land Use at 1015 Sites (n=2750 observations), Symbolized by WQISRB Score Bin (Correlation circles show normal probability groupings of sites within each WQI_{SRB} score bin.)

High Flow and TSS Results

All but one parameter had significant relationships with both flow and TSS, but responses were stronger with one or the other depending on their soluble phase aqueous chemistry (Table 4). Nitrate was the only parameter that did not have a significant relationship with TSS. All relationship directions (positive or negative) were consistent between the two variables.

Table 4. Summary of Parameter Models with Flow and TSS (All parameters were log transformed with the exception of nitrate (*) which was inverse square root transformed. All flow values were log transformed and all TSS concentrations were inverse square root transformed. The stronger relationship for each parameter is shown in bold.)

Parameter	Response to Flow			Response to TSS		
	R^2 (adj) %	p-value	relationship	R^2 (adj) %	p-value	relationship
Aluminum	41.28	<0.001	+	**61.38**	<0.001	+
Iron	52.49	<0.001	+	**73.17**	<0.001	+
Manganese	20.36	<0.001	+	**29.80**	<0.001	+
Nitrate*	**0.24**	0.025	-	0.01	0.266	-
Phosphorus	23.01	<0.001	+	**48.95**	<0.001	+
TOC	20.54	<0.001	+	**51.26**	<0.001	+
Chloride	**15.88**	<0.001	-	2.27	<0.001	-
Sodium	**19.33**	<0.001	-	04.76	<0.001	-
Sulfate	**16.17**	<0.001	-	11.46	<0.001	-

The Susquehanna WQI was created to contextualize and classify water quality during baseflow (non-storm impacted) conditions. Generating WQI scores using water samples collected during high flow or high TSS conditions has the potential to misrepresent a stream's water quality as the nine parameters respond differently to non-baseflow conditions (see Appendix B). Given the soluble phase properties of the three metals, phosphorus, and TOC, data integrity of samples with TSS exceeding 80 mg/L may be compromised if samples were not filtered prior to preservation. Scores for WQI_{SRB}, WQI_{ARC}, and $WQI_{ARCxECO}$ are artificially raised or lowered in response to high flow or TSS. Therefore, WQI scores of samples collected during high flow or TSS events should be interpreted with caution. Both a high flow warning and a TSS caution were added to the WQI scoring tool to alert the user when flows are greater than 3.25 CFM or TSS concentrations are greater than 80 mg/l.

Correlation with Biology

The relationship of WQI_{SRB} vs. IBI score is highly significant ($p< 0.001$) in ARC groups 1&2 and 3&4 and describes a good deal of variation (49.4 percent and 47.7 percent, respectively) (Figure 8; Table 5). Larger rivers categorized in ARC 5&6 (>1000 mi^2) are outside of the recommended range for the PADEP Freestone IBI and a greater number of samples applicable to the Semi-Wadeable Macro Metric Index (SWMMI; Shull and Lookenbill, 2017) are needed to establish meaningful relationships.

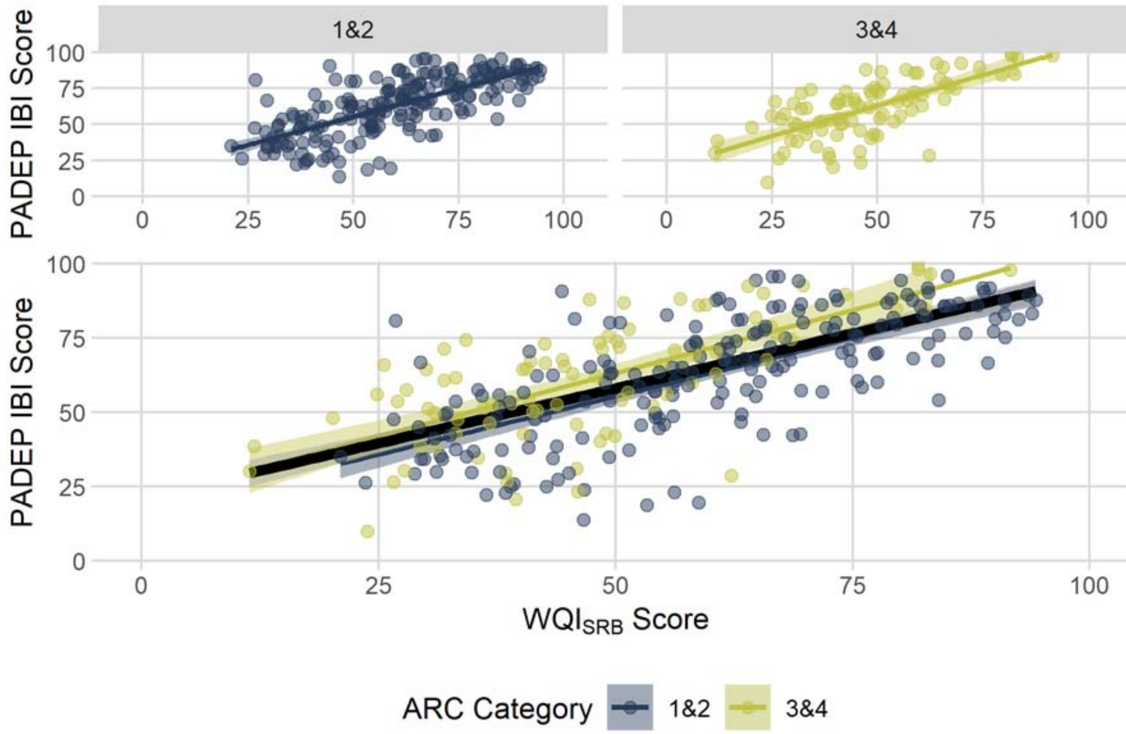


Figure 8. Scatterplot of WQISRB vs. IBI Scores, Shown for ARC Groupings 1&2 and 3&4 and All Combined (This dataset only includes samples collected during November-May index period (n= 283).)

Table 5. Regression Metrics of WQISRB vs IBI by ARC Groups (This dataset includes only samples collected during November-May index period. Note: *p<0.1; **p<0.05; ***p<0.01.)

	ARC 1&2	ARC 3&4	All ARCs
Intercept	16.227*** (3.565)	20.802*** (4.807)	21.106*** (2.818)
Slope	0.782*** (0.057)	0.848*** (0.096)	0.738*** (0.048)
Observations	195	88	283
R2	49.4%	47.7%	45.9%
Adjusted R2	49.2%	47.1%	45.7%
P	<0.001	<0.001	<0.001
Residual Std. Error	14.339 (df = 193)	14.720 (df = 86)	14.832 (df = 281)
F Statistic	188.613*** (df = 1; 193)	78.346*** (df = 1; 86)	238.772*** (df = 1; 281)

Despite the exclusion of ARC 5 and 6 sites, we have a highly significant model that describes 45.7 percent of variation in IBI score (Table 5). This is a promising result and indicates the WQI_{SRB} score may be used as an effective stressor gradient for biological analyses.

Unexplained IBI vs WQI Scores

Over 97 percent of paired WQI_{SRB} and IBI scores fell within expected margins. Less than three percent of sites fell outside of the expected ranges for WQI and IBI scores (Figure 9). For the two percent of sites that appear to have good water quality but poor macroinvertebrates, habitat could be a potential explanatory variable responsible for the poor macroinvertebrate IBI score.

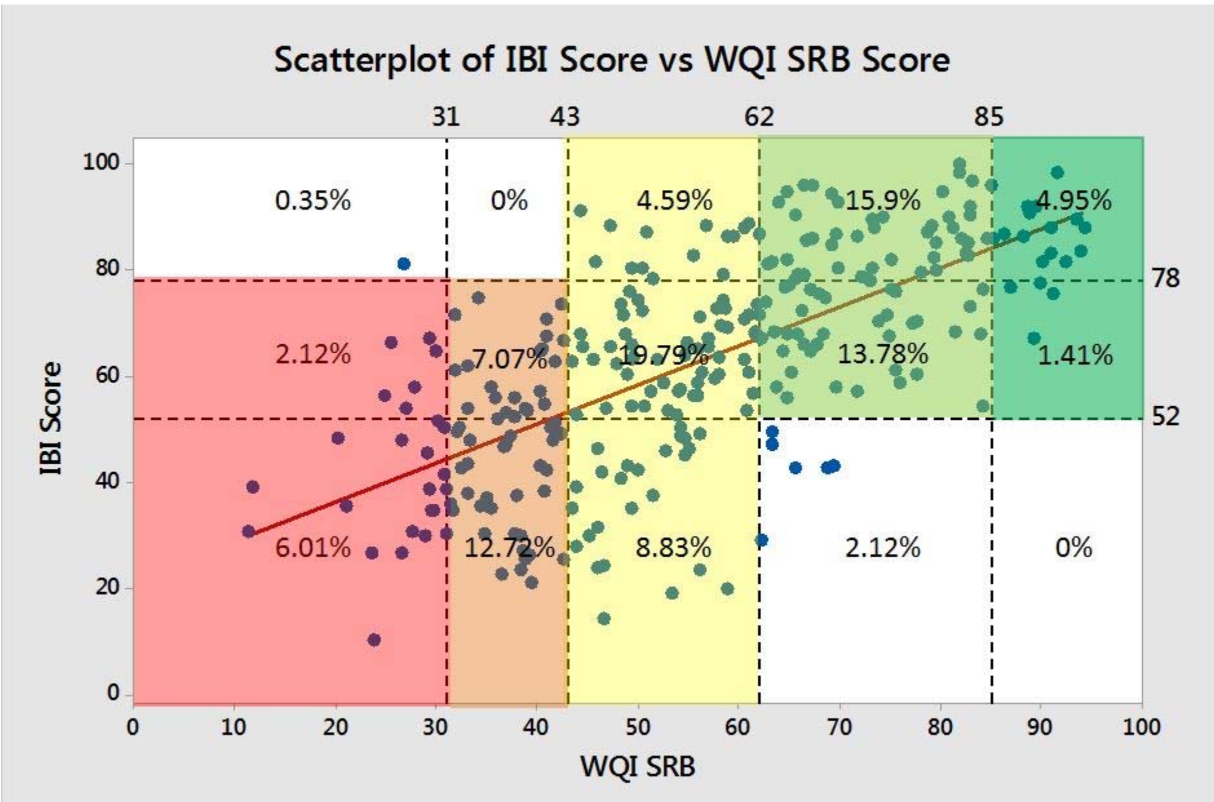


Figure 9. Scatterplot of Macroinvertebrate IBI Score vs. WQISRB Score with Water Quality Thresholds (n= 283) (X-axis reference lines indicate WQI_{SRB} classification thresholds; y-axis reference lines indicate generally considered good IBI scores of > 78 and poor IBI scores of < 52 (Dustin Shull, personal communication). Over 97 percent of sites fell within expected margins (colored sections); less than 3 percent fell outside.)

Limitations of WQIs

There are hundreds of additional parameters that could be tested for in a water sample; a WQI only reflects the parameters included. Any index has the potential to miss something simply because a particular parameter has not been included in lab analysis. There is no way to be fully objective in the parameter selection process (Abbasi and Abbasi, 2012). Data availability and best professional judgement were considered when determining which parameters to include based on knowledge of the land use/land history of the SRB in addition to its geophysical components.

Addressing Common Issues with Aggregation Methods

WQIs created using aggregation methods are known to be subject to ambiguity, eclipsing, compensation, and rigidity (Abbasi and Abbasi, 2012). Each of these four potential weaknesses and how to address them within the Susquehanna WQI were considered carefully.

Ambiguity occurs when the final index score exceeds a critical value without any of the individual categories exceeding the critical value. Because we averaged at both the category and final score levels, ambiguity is not an issue for the Susquehanna WQI.

Eclipsing occurs when at least one of the category scores exceeds a critical value but the final score does not. In the metals category of the Susquehanna WQI, if aluminum or iron exceed their respective aquatic life standard, the entire category will be scored as a "0." This solves the potential eclipsing issue at the category level. Eclipsing could still be an issue at the WQI_{SRB} level, because there are no aquatic life use standards for parameters in the nutrient enrichment or development categories. One category score of zero will greatly reduce the overall WQI_{SRB} score, but it may remain in the "fair" range if the nutrient enrichment or development categories score well. Eclipsing is not an issue at the WQI_{SRB} level if the nutrient enrichment or development categories are also low.

Compensation is considered good when "not biased towards extremes (i.e., highest or lowest [category] value)" (Abbasi and Abbasi, 2012). The issue of some eclipsing in the WQI is balanced by having good compensation. Assigning equal weights, averaging parameter percentiles to generate a category score, and averaging category scores to generate a final score all prevent any one parameter from exerting undue influence on the final result. Therefore, we can conclude compensation in the Susquehanna WQI is adequate.

The final common issue of aggregated WQIs is rigidity versus flexibility, or how easily additional variables or categories can be added to the index. Often a regional level index is not sufficient to address more localized environmental conditions within the larger basin, creating a need for more fine scale resolution variations of the regional scale model (Abbasi and Abbasi, 2012). The Susquehanna WQI addresses this by having three tiers; nested WQI_{ARC} and $WQI_{ARCxECO}$ scores are calculated concurrently with the WQI_{SRB} score.

Additionally, the requirement for two of three variables per category to obtain a score builds in flexibility. If additional parameters were of interest, they could be added to a category or a new category could be created and averaged into the existing Susquehanna WQI. However, comparisons would only be equivalent across that category score or that set of WQI scores; scores should only be compared when the same parameters are contributing in the index. Currently, no provision exists for adding a new parameter that was not previously sampled for to the previously scored dataset. In the event a new pollutant or parameter of interest arises, caution should be used to only compare scores that were generated with the new parameter in the index. The simplicity of our aggregation method (percentile ranking followed by subsequent averaging within each category) strikes a good balance between rigidity and flexibility. The Susquehanna WQI is also flexible in that the design allows for updates and expansions of the WQI dataset in the future. However, scores will retroactively change when data are added to the WQI dataset as percentile ranks may shift slightly. Due to the large range of included data (Table 1) and spatial coverage across the SRB (Figure 8), we do not anticipate shifts greater than a few points if additional data were added to the dataset.

CONCLUSIONS AND RECOMMENDATIONS

A water quality index was created for the Susquehanna River Basin to assess ambient stream water quality under baseflow conditions within and among monitoring sites throughout time. The Susquehanna WQI is a powerful tool that contextualizes data across a 27,510-square-mile watershed and allows for enhanced water quality monitoring and analysis and wide-ranging applications. The user-friendliness of the tool and an intuitive 0-100 scoring scale are two of its greatest strengths. Additionally, the benefits of being able to easily track changes in water quality across spatial and temporal scales cannot be overstated. The clarity and simplicity of its output coupled with its ease of use make this tool applicable for the technical scientific user as well as water resource managers, policy makers, and non-scientists alike.

Further analyses and steps could include 1) adding a tier to the calculator that examines ecoregion independently of stream size group (ARC); 2) evaluating relationships between each parameter and/or category and IBI or WQI score; and 3) pursuing a dashboard interface to speed up the scoring process and to allow users to place their own data into the context of the WQI dataset.

The Susquehanna River Basin WQI will be most useful when used in conjunction with biological monitoring. The WQI is a form of stressor-based monitoring, meaning that it attempts to link chemical stressors (water quality parameters) to potential biological responses. However, this link is only correlation, not causation. Response-based monitoring allows for a more comprehensive view of community structure, species richness, diversity, tolerance, as well as overall trophic structure (Abbasi and Abbasi, 2012). This tool could eventually be incorporated into Commission environmental reviews in combination with biological surveys to assess potential regulatory and/or water resource management actions. By integrating use of the Susquehanna WQI, biological monitoring–primarily fish and macroinvertebrate–and habitat assessments, an extensive, more complete picture of water quality and stream health will appear, allowing for improved management of the waters of the Susquehanna River Basin.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the Susquehanna River Basin Commission for funding the Remote Water Quality Monitoring Network (www.srbc.net/), from which came a large portion of records used in the WQI dataset. We also acknowledge D. Hintz for her continual commitment to database improvement and maintenance, which allows for easy and consistent access to such a large dataset. We thank G. Markowitz for assistance in determining the flow per drainage area cutoff used to create the high flow warning tool within the WQI scorer. J. Shallenberger, B. Hutchison, E. Campbell, H. Hollier, and T. Shenk provided comments and feedback to improve this document.

REFERENCES

Abbasi, T. and S.A. Abbasi. 2012. Water Quality Indices. Oxford (UK): Elsevier.

Baker, J.P. and C.L. Schofield. 1982. Aluminum toxicity to fish in acidic waters. Water, Air, and Soil Pollution, 18:289-309.

Beck, H.J. and G.F. Birch. 2011. Spatial and temporal variance of metal and suspended solids relationships in urban stormwater-Implications for monitoring. Water, Air, and Soil Pollution, 223(3):1005-1015.

Cairns, Jr., J. and J.R. Pratt. 1993. A history of biological monitoring using benthic macroinvertebrates. In: D.M. Rosenberg and V.H. Resh, editors. Freshwater Biomonitoring and Benthic Macroinvertebrates. New York, NY. Springer-Verlag (US): 10-27.

Chaplin, J.J. 2005. Effects of historical coal mining and drainage from abandoned mines on streamflow and water quality in Bear Creek, Dauphin County, Pennsylvania-March 1999-December 2002. U.S. Geological Survey Scientific Investigations Report 2005-5042.

Clark, T.J. 2013. Bear Run Watershed Renaissance Mine Drainage Restoration Project. Susquehanna River Basin Commission, Pub. No. 290. Harrisburg, PA.

Cooper, T.F., J.P. Gilmour, and K.E. Fabricius. 2009. Bioindicators of changes in water quality on coral reefs: review and recommendations for monitoring programmes. Coral Reefs, 28(3):589-606.

Correll, D.L. 1998. The role of phosphorus in the eutrophication of receiving waters: a review. Journal of Environmental Quality, 27(2):261-266.

DePhilip, M. and T. Moberg. 2010. Ecosystem flow recommendations for the Susquehanna River basin. The Nature Conservancy. Harrisburg, PA.

Duan, S. and S.S. Kushal. 2015. Salinization alters fluxes of bioreactive elements from stream ecosystems across land use. Biogeosciences, 12(23):7331-7347.

Earle, J. and T. Callaghan. 1998. Impacts of mine drainage on aquatic life, water uses, and man-made structures. Chapter 4 in Coal Mine Drainage Prediction and Pollution Prevention in Pennsylvania. Edited by: K.B.C. Brady, M.W. Smith, and J. Schueck. Department of Environmental Protection, Harrisburg, PA.

Gregory, S.V., F.J. Swanson, W.A. McKee, and K.W. Cummins. 1991. An ecosystem perspective of riparian zones. BioScience, 41(8):540-551.

Hintz, D.R. and L.Y. Steffy. 2015. Remote Water Quality Network: Data Report of Baseline Conditions for 2010-2013. Susquehanna River Basin Commission, Pub. No. 297. Harrisburg, PA.

Horton, R.K. 1965. An index number system for rating water quality, Journal of Water Pollution Control Federation, 37(3), pp 300–306.

Irani, F.M. and P.R. Claggett. 2010. Chesapeake Bay Watershed Land Cover Change Data Series: U.S. Geological Data Series 505. U.S. Geological Survey, Reston, VA.

Maryland Department of the Environment. 2018. Frequently asked questions: Chesapeake Bay water quality standards-Designated uses in the Chesapeake Bay. https://mde.maryland.gov/programs/Water/TMDL/WaterQualityStandards/Pages/faqs.aspx.

Metcalfe, J.L. 1989. Biological water quality assessment of running waters based on macroinvertebrate communities: History and present status in Europe. Environmental Pollution, 60(1-2):101-139.

Minnesota Pollution Control Agency. 2017. Salty issue: What to do about chloride in wastewater. https://www.pca.state.mn.us/featured/salty-issue-what-do-about-chloride-wastewater.

Nagorski, S.A., J.N. Moore, T.E. McKinnon, and D.B. Smith. 2003. Geochemical response to variable streamflow conditions in contaminated and uncontaminated streams. Water Resources Research, 39(2;1044): 1-14.

Nelson, S.S., D.R. Yonge, and M.E. Barber. 2009. Effects of road salts on heavy metal mobility in two Eastern Washington soils. Journal of Environmental Engineering, 135(7):505-510.

Nelson, W. and M. Frazier. 2014. Sediment total organic carbon: is this a useful indicator of sediment condition for pacific northwest estuaries? Presented at Pacific Estuarine Research Society, Newport, OR, April 03-05, 2014.

North Carolina Department of Environmental Quality. 2007. Water quality stressors and sources of impairment in the White Oak River basin. https://files.nc.gov/ncdeq/Water%20Quality/Planning/BPU/BPU/White%20Oak/White%20Oak%20Plans/2007%20Plan/Chapter%207.pdf.

Novotny, V., D. Muehring, D.H. Zitomer, D.W. Smith, and R. Facey. 1998. Cyanide and metal pollution by urban snowmelt: impact of deicing compounds. Water Science and Technology, 38(10): 223-230.

Omernik, J.M. 1987. Aquatic Ecoregions of the Conterminous United States. U.S. Geological Survey, Reston, VA.

Paul, J.F., S.M. Cormier, W.J. Berry, P.R. Kaufmann, R.L. Spehar, D.J. Norton, R.E. Cantilli, R. Stevens, W.F. Swietlik, and B.K. Jessup. 2008. Developing water quality criteria for suspended and bedded sediments. U.S. Environmental Protection Agency, Water Practice, Water Environment Federation, Alexandria, Virginia. 1(2):1-17.

Pennsylvania Bulletin. 2017. Table 5: Water quality criteria for toxic substances. https://www.pabulletin.com/secure/data/vol47/47-42/1766.html.

Pennsylvania Department of Environmental Protection (PADEP). 2013. A benthic macroinvertebrate index of biotic integrity for wadeable freestone riffle-run streams in Pennsylvania. Pennsylvania Department of Environmental Protection. Division of Water Quality Standards.

PADEP. 2009. Permitting strategy for high total dissolved solids (TDS) wastewater discharges. http://files.dep.state.pa.us/Water/Wastewater%20Management/WastewaterPortalFiles/MarcellusShaleWastewaterPartnership/high_tds_wastewater_strategy_041109.pdf.

Perl, D.P. and C.W. Olanow. 2007. The neuropathology of manganese-induced Parkinsonism. Journal of Neuropathology and Experimental Neurology, 66(8):675-682.

Pinsino, A., V. Matranga, and M.C. Roccheri. 2012. Manganese: A new emerging contaminant in the environment. InTech, Rijeka, 17-36. http://cdn.intechopen.com/pdfs/29309/InTech-Manganese_a_new_emerging_contaminant_in_the_environment.pdf.

R Core Team. 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Rogers, R.J. 1989. Geochemical comparison of ground water in areas of New England, New York, and Pennsylvania. Ground Water, 27(5):690-712.

Rowe, M., D. Essig, and B. Jessup. 2003. Guide to selection of sediment targets for use in Idaho TMDLs. Idaho Department of Environmental Quality. https://www.epa.gov/sites/production/files/2015-10/documents/sediment-appendix5.pdf.

Seely, O. 2018. “Solubility Rules.” http://www5.csudh.edu/oliver/chemdata/solrules.htm.

Seitz, R.C. 1971. Drainage area statistics for the Chesapeake Bay freshwater drainage basin. Chesapeake Bay Institute, Baltimore, Maryland.

Shull, D.R. and M.J. Lookenbill. 2017. A benthic macroinvertebrate multimetric index for large semi-wadeable rivers: Technical report. Pennsylvania Department of Environmental Protection, Office of Water Programs, Harrisburg, PA.

Shull, D.R. and M. Pulket. 2018. Assessment methodology for rivers and streams. Pennsylvania Department of Environmental Protection Office of Water Programs, Harrisburg, PA. http://files.dep.state.pa.us/Water/Drinking%20Water%20and%20Facility%20Regulation/WaterQualityPortalFiles/Methodology/2015%20Methodology/Assessment_Book.pdf.

Steffy, L.Y. and M.K. Shank. 2018. Considerations for using turbidity as a surrogate for suspended sediment in small, ungaged streams: Time-series selection, streamflow estimation, and regional transferability. River Research and Applications, 1-11.

Stets, E.G., C.J. Lee, D.A. Lytle, and M.R. Schock. 2018. Increasing chloride in rivers of the conterminous U.S. and linkages to potentially corrosivity and lead action level exceedances in drinking water. Science of the Total Environment, 613-614:1498-1509.

Stoe, T.W. 1998. Water Quality and Biological Assessment of the Wiconisco Creek Watershed. Susquehanna River Basin Commission, Pub. No. 193. Harrisburg, PA.

Susquehanna River Basin Commission (SRBC). 2013. Comprehensive Plan for the Water Resources of the Susquehanna River Basin. https://www.srbc.net/our-work/programs/planning-operations/ docs/comp-plan-no-appendices.pdf

SRBC. 2012. Technical Guidance for Low Flow Protection Related to Withdrawal Approvals. FileSite #170,477. Harrisburg, PA.

Sutadian, A.D., N. Muttil, A. Yilmaz, and C. Perera. 2016. Development of river water quality indices-a review. Environmental Monitoring and Assessment, 188:58.

Tlili, A., A. Berard, H. Blanck, A. Bouchez, F. Cássio, K.M. Eriksson, S. Morin, B. Montuelle, E. Navarro, C. Pascoal, S. Pesce, M. Schmitt-Jansen, and R. Behra. 2016. Pollution-induced community tolerance (PICT): towards an ecologically relevant risk assessment of chemicals in aquatic systems. Freshwater Biology, 61(12):2141-2151.

United States Environmental Protection Agency (USEPA). 2013. Primary distinguishing characteristics of level III ecoregions of the continental US. https://www.epa.gov/eco-research/level-iii-and-iv-ecoregions-continental-united-states.

USEPA. 2005. National estuary program-challenges facing our estuaries. Key management issues: Nutrient overloading. https://www.epa.gov/nep/how-national-estuary-programs-address-environmental-issues.

USEPA. 2000. Stressor identification guidance document. https://nepis.epa.gov/Exe/ZyPDF.cgi/20003F6L.PDF?Dockey=20003F6L. PDF.

Vannote, R.L., G.W. Minshall, K.W. Cummins, J.R. Sedell, and C.E. Cushing. 1980. The river continuum concept. Canadian Journal of Fisheries and Aquatic Sciences, 37(1):130-137.

Vincent, Q. Vu. 2011. ggbiplot: A ggplot2 based biplot. R package version 0.55. http://github.com/vqv/ggbiplot.

Whitehead, P. 2018. Total organic carbon (TOC) and its measurement. https://www.elgalabwater.com/blog/total-organic-carbon-toc.

World Health Organization. 2004. Sulfate in drinking-water. Background document for development of WHO Guidelines for Drinking-water Quality. https://www.who.int/water_sanitation_health/dwq/chemicals/sulfate.pdf.

Zambrano, J. and S. Stoner. Reissued 1998. Ambient water quality standards and guidance values. New York State Department of Environmental Conservation, Division of Water. https://www.dec.ny.gov/docs/water_pdf/togs111.pdf.

APPENDIX A

Water Quality Classification Threshold Determination

Classification thresholds were established based on the relationship between WQI_{SRB} scores and macroinvertebrate index of biotic integrity (IBI) scores. Data were compiled from sites where macroinvertebrate samples were collected using riffle-run protocols simultaneously with water samples in the index period November-May (PADEP, 2013). WQI_{SRB} and the appropriate PADEP biotic index (small or large freestone IBI) were calculated for each sample.

The macroinvertebrate samples were first categorized as Good (IBI >78), Fair (52-78), or Poor (<52) (Dustin Shull, personal communication). After biological categorization, the corresponding WQI_{SRB} scores for these sites were used to assign categorical WQI cutoffs. The scoring threshold for Excellent water quality represents the 80th percentile of WQI scores of sites that had an IBI score >78. The Good WQI score threshold represents the 20th percentile of sites that had an IBI score >78. The Fair WQI score threshold represents the 20th percentile of WQI scores that fell between the IBI scores of 52 and 78. The Poor WQI score threshold represents the 20th percentile of sites with IBI scores <52. Sites that had WQI scores below 31 were classified as Very Poor. Overall, the WQI_{SRB} water quality classifications are: 85-100, Excellent; 62-85, Good; 43-62, Fair; 31-43, Poor; and 0-31, Very Poor (Figure A1, Table 3 from text). Only the WQI_{SRB} scores are given water quality classifications because WQI_{SRB} scores had the best linear relationship with IBI score. For simplicity and ease of use, WQI_{ARC} and $WQI_{ARCxECO}$ were not given classifications because each ARC within each ecoregion would need its own set of cutoffs.

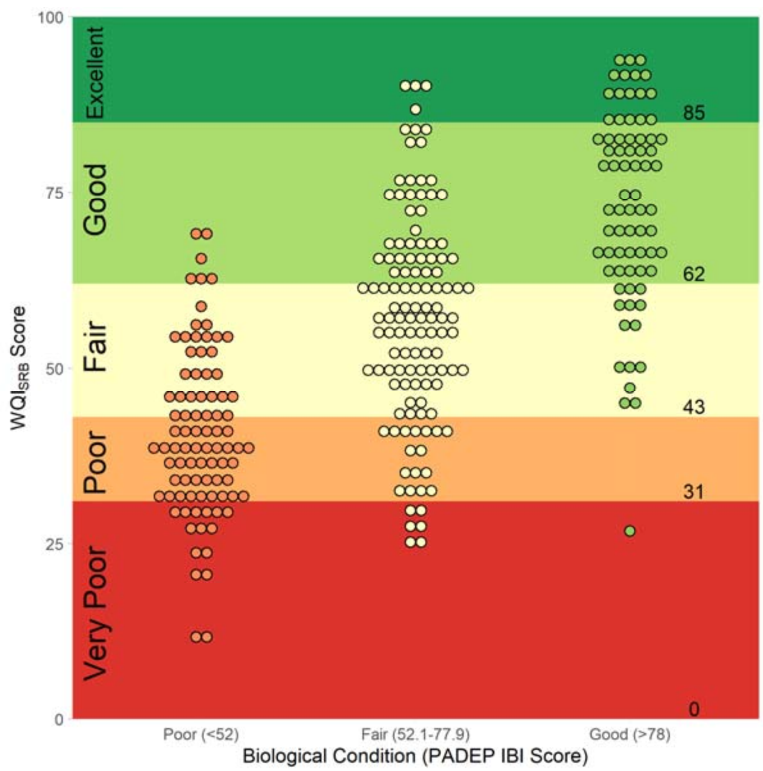


Figure A1. Water Quality Classification Thresholds for the WQI_{SRB} Scores (Thresholds were determined based on 283 paired WQI and IBI samples from the November-May index period.)

APPENDIX B

High Flow and Total Suspended Sediment Threshold Determination

High streamflow conditions could potentially misrepresent water quality results by contributing higher than typical proportions of runoff in relation to baseflow, which could result in unrepresentative parameter concentrations. Additionally, suspended sediment concentrations between streams can vary greatly despite similar hydrologic conditions (Steffy and Shank, 2018) and by different flow conditions within the same stream (Beck and Birch, 2011). For these reasons, we sought to determine thresholds of streamflow and total suspended solids (TSS) above which Water Quality Index (WQI) scores may not be representative of baseflow conditions at individual sites.

High Flow Threshold Determination

A streamflow event with a percent exceedance value of 10 (P_{10}) is considered to be a high flow event in the SRB (SRBC, 2012; DePhilip and Moberg, 2010). We obtained average daily flow (ADF) time series from 143 hydrologically unregulated U.S. Geological Survey (USGS) stream gages in the SRB with an adequate period of record to generate flow statistics (>10 years). The median streamflow yield that represented annual P_{10} conditions at unregulated gages was 3.25 cubic feet per second per square mile (CFM; Figure B1). A high flow warning was added to the WQI tool to address this, where any sample taken during hydrological events exceeding 3.25 CFM will be flagged and the user will be instructed to interpret results with caution.

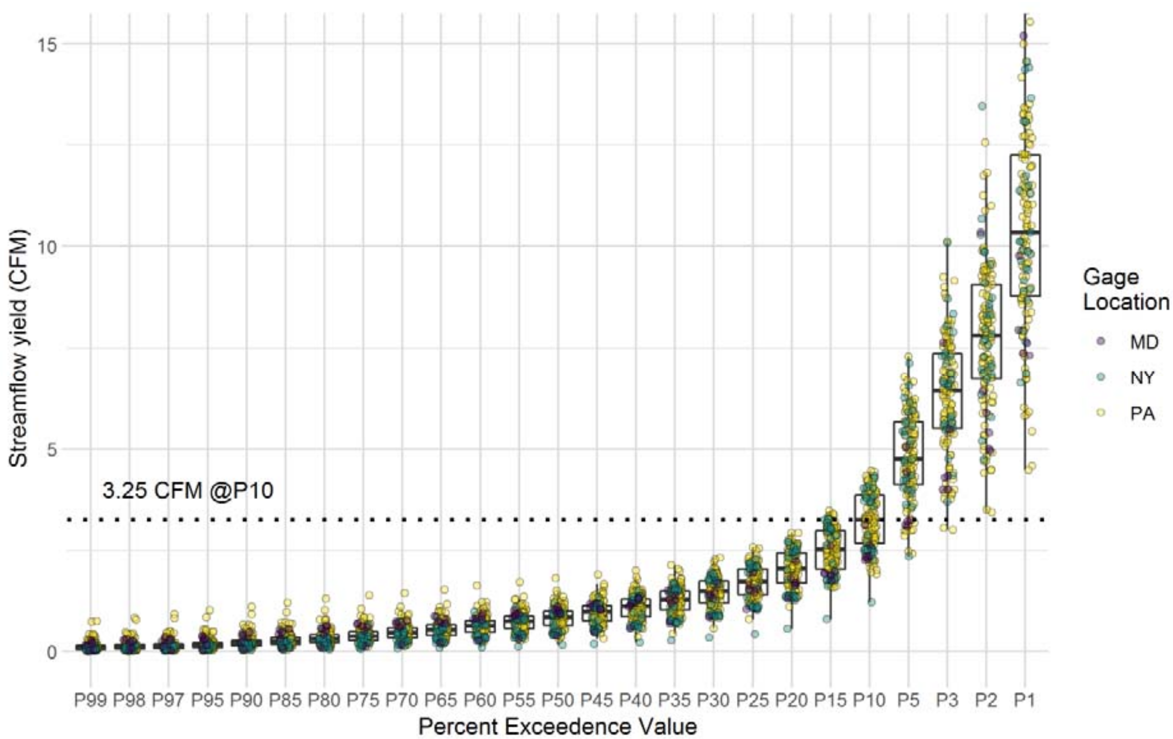


Figure B1. Relationship between Percent Exceedance Value and Streamflow Yield in Cubic Feet Per Second Per Square Mile (CFM) at All Unregulated USGS Stream Gages in the Susquehanna River Basin (n=143) (Median streamflow yield equivalent to P_{10} conditions is 3.25 CFM.)

Total Suspended Solids (TSS) Threshold Determination

Applying a limit for TSS to the state or basin level is challenging due to complexities of quantifying concentration and duration of exposure. Many states have tried to establish thresholds, but no consistent concentration exists, and, as a result, only a few states use suspended solids as a water quality criterion (Paul et al., 2008). A study in Idaho showed that fish communities are negatively impacted by long-term exposure to TSS concentrations above 80 mg/L (Rowe et al., 2003).

In order to identify a practical, regionally applicable TSS threshold to use as warning within the SRB WQI, a large internal dataset of 21,653 samples was compiled. These samples covered 20 sites within the SRB and included a wide range of TSS concentrations and WQI_{SRB} scores. Sample number per site ranged from 127 to 1,806 observations. Total and dissolved fractions of each parameter were plotted against TSS when available. The divergence of dissolved and total phosphorus was consistently between 60-100 mg/L at all 20 sites. As a result, the TSS threshold was set at 80 mg/L within the SRB WQI to alert users that soluble phase properties of some parameters may impact WQI score above this concentration of TSS.

Streamflow and TSS Threshold Validation Methods

In order to account for potential impacts of high flow and/or high TSS on WQI_{SRB} score, a validation exercise was completed to determine if the streamflow and TSS thresholds of 3.25 CFM and 80 mg/l, respectively, resulted in significantly different WQI_{SRB} scores within sites. The WQI dataset is comprised primarily of samples collected during baseflow conditions. In order to examine the impacts of high flow and/ or elevated TSS, a separate dataset was used to evaluate the effect of elevated flow and TSS on WQI_{SRB} score. This dataset was pulled from the Sediment and Nutrient Assessment Program (SNAP; https://www.srbc.net/portals/water-quality-projects/sediment-nutrient-assessment/) and included data collected under a variety of flow conditions (n= 1,656 samples from 26 sites; min n= 37, max n=100; 2008-2013).

First, we determined the strength of individual relationships between WQI parameters and streamflow and TSS. Bivariate ordinary least squares (OLS) linear regression models of streamflow and TSS (predictors) and each individual WQI parameter (responses) were created to examine linear relationships. Box-Cox analysis indicated that flow (CFM) and TSS (mg/l) needed to be $\log_{10}$ and inverse square root transformed, respectively, to fit model assumption of normality. Each WQI parameter was then $\log_{10}$ transformed except nitrate, which was inverse square root transformed, to meet model assumptions. The adjusted r-squared value of each model was evaluated to determine whether flow or TSS was more strongly correlated with each parameter.

Second, we determined the overall response of WQI_{SRB} scores to streamflow and TSS thresholds using two-way ANOVAs. Flow and TSS "bins" were created by categorizing whether the sample was collected during conditions above or below the thresholds of 3.25 CFM and 80 mg/L TSS, respectively. Two-way ANOVAs were then used to determine if WQI_{SRB} scores were significantly different when flow and TSS were below or above thresholds, by site.

Streamflow and TSS Threshold Validation Results

All but one parameter had significant relationships with both flow and TSS, but responses were stronger with one or the other (Table 4). Nitrate did not have a significant relationship with TSS. All relationship directions (positive or negative) were consistent between the two variables. All parameters were reported and analyzed as total concentrations. Metals parameters (aluminum, iron, and manganese) were more strongly correlated with TSS than flow (Table 4). Metals are more likely to be chemically bound to suspended sediment than in dissolved form (Beck and Birch, 2011). Therefore, the stronger relationship with TSS compared to flow was not surprising. Our finding is consistent with Nagorski et al. (2003), who found strong correlations ($r \geq 0.88$, $p < 0.01$) between TSS and aluminum, iron, and manganese. The positive direction of this relationship (Table 4 from main text) indicates that as TSS and/or flow increases, the concentration of metals will increase, resulting in a decreased metals category score.

Reprinted from main text:
Table 4. Summary of Parameter Models with Flow and TSS (All parameters were log transformed with the exception of nitrate (*) which was inverse square root transformed. All flow values were log transformed and all TSS concentrations were inverse square root transformed. The stronger relationship for each parameter is shown in bold.)

Parameter	Response to Flow			Response to TSS		
	R^2 (adj) %	p-value	relationship	R^2 (adj) %	p-value	relationship
Aluminum	41.28	<0.001	+	**61.38**	<0.001	+
Iron	52.49	<0.001	+	**73.17**	<0.001	+
Manganese	20.36	<0.001	+	**29.80**	<0.001	+
Nitrate*	**0.24**	0.025	-	0.01	0.266	-
Phosphorus	23.01	<0.001	+	**48.95**	<0.001	+
TOC	20.54	<0.001	+	**51.26**	<0.001	+
Chloride	**15.88**	<0.001	-	2.27	<0.001	-
Sodium	**19.33**	<0.001	-	04.76	<0.001	-
Sulfate	**16.17**	<0.001	-	11.46	<0.001	-

Chloride, sodium, and sulfate are soluble and present as dissolved ions in surface waters (Seely, 2018) and were more strongly correlated with flow than TSS (Table 4). The relationship was negative, indicating that as flow increases, concentrations of the three urban parameters are diluted, which could result in an increased category score. Nitrate is also soluble (Seely, 2018) and had the same response pattern as the urban parameters, decreasing in concentration as flow increased (Table 4).

An USEPA National Coastal Assessment found TOC to be generally positively correlated with sediment (Nelson and Frazier, 2014). Total phosphorus is present in streams in a variety of forms; dissolved species exist in the water column and various forms of phosphorus are bound to sediment (Correll, 1998). These parameters were both more highly correlated with TSS compared to streamflow (Table 4).

The results of the two-way ANOVA of flow bin (above or below 3.25 CFM) on WQI_{SRB} score revealed a significant main effect of flow on WQI_{SRB} score, $F(1,1)= 9.96$, $p= 0.002$. WQI_{SRB} scores were higher below the flow caution threshold compared to above. Similarly, the results of the two-way ANOVA of TSS bin (above or below 80 mg/l) revealed a significant main effect of TSS concentration, $F(1,1)= 109.02$, $p <0.001$. When TSS was below 80 mg/l, WQI_{SRB} scores were higher than when TSS was above 80 mg/l, which yielded the TSS caution threshold. The flow and TSS interaction was significant, $F(1,1)= 9.75$, $p= 0.002$, indicating that flow and TSS are not independent of each other. Dissolved and total phosphorus data from 20 sites were plotted against TSS concentrations to provide insight into where along the TSS gradient an inflection point becomes evident (Figure B2). This analysis provided an additional line of evidence indicating a TSS caution threshold of 80 mg/l was appropriate.

Both TSS and flow were found to influence WQI_{SRB} scores (Figure B3). Of the 20 sites in the dataset, eight had significantly different mean WQI_{SRB} scores by flow bin (above or below the threshold of 3.25 CFM), while seven of the 20 had significantly different mean WQI_{SRB} scores by TSS bin (above or below 80mg/l). Six sites had significantly different mean WQI_{SRB} scores by both flow bin and TSS bin, indicating that strong correlation exists between TSS and streamflow.

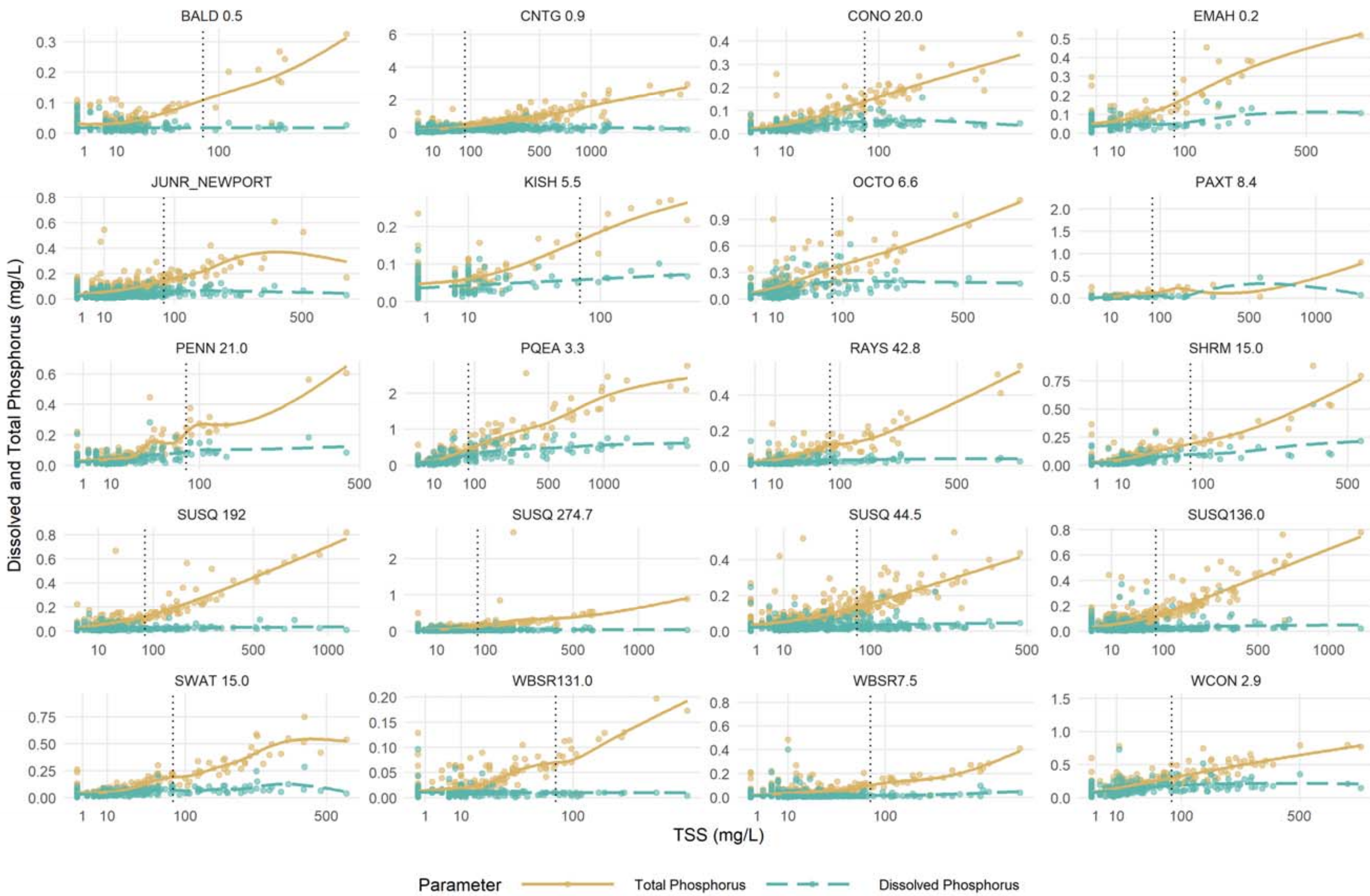


Figure B2. Scatterplots of Dissolved and Total Phosphorus Concentrations for 21,653 Samples from 20 Sites (Total phosphorus concentrations are in yellow and dissolved concentrations are in green. A vertical reference line representing the assigned cautionary threshold for TSS of 80mg/L was added to each plot.)

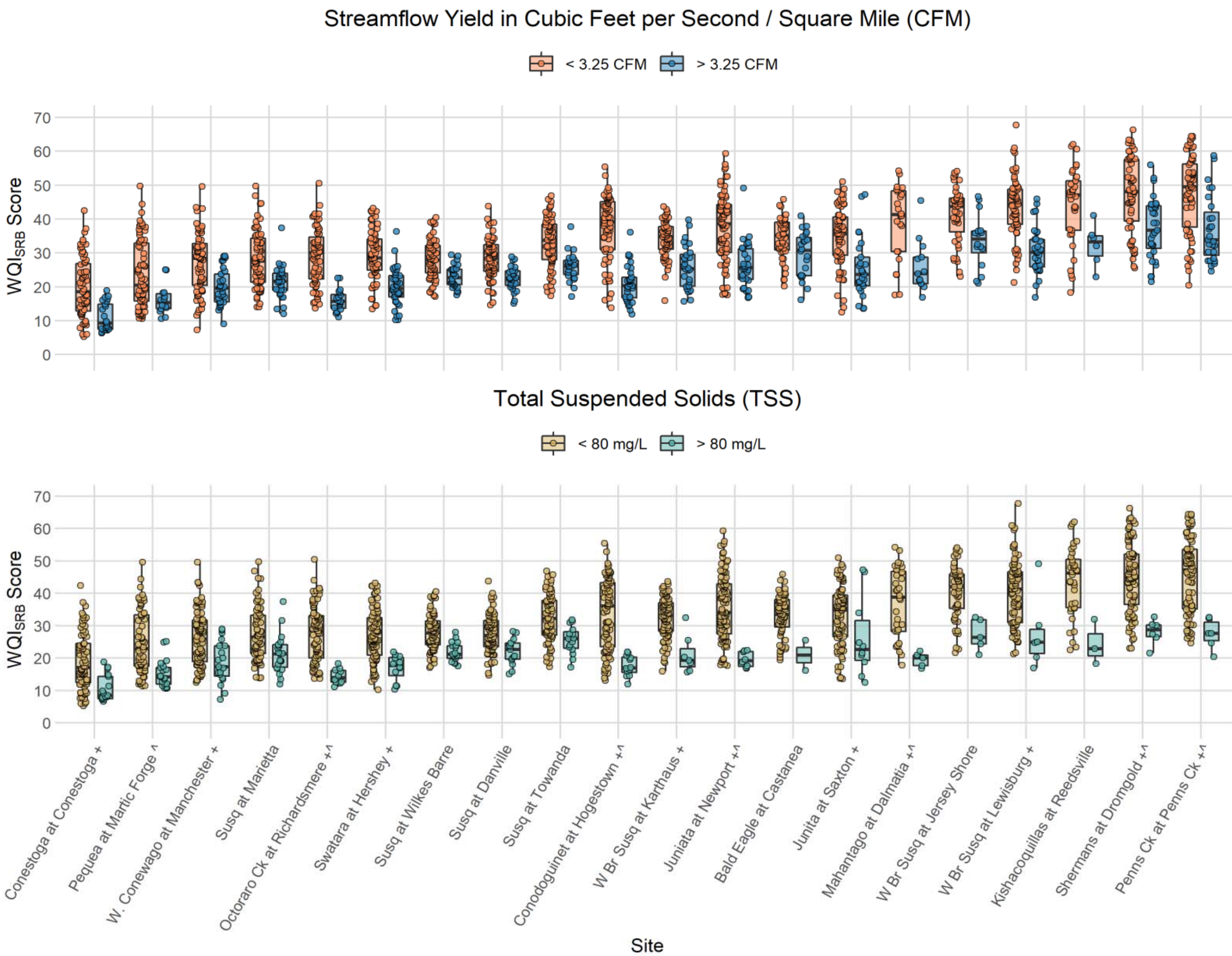


Figure B3. Boxplots of WQI_{SRB} Scores for 20 Sites Above and Below Flow and Total Suspended Solids (TSS) Threshold Cautions (Dots represent one sampling event. Significantly different mean WQI_{SRB} scores are denoted by site on the x-axis where '+' indicates significant difference in flow bin, and '^' indicates significant difference in TSS bin. Top: Orange boxes represent flows below the 3.25 CFM threshold; blue boxes represent flows above the 3.25 CFM caution. Bottom: Yellow boxes represent the range of WQI_{SRB} scores with TSS concentrations less than 80mg/l; green boxes represent scores associated with TSS concentrations greater than 80mg/l.)

Appendix C

The Utility of Scoring at Multiple Spatial Scales

A site in a highly disturbed ecoregion may score poorly when compared to the entire Susquehanna River Basin (low WQI_{SRB} score), but the site could score much higher when compared to proximal streams facing similar pressures and influences. The opposite scenario is also possible; a site may receive a good WQI_{SRB} score, but be in an ecoregion with many high quality streams of the same size and receive a lower score for the $WQI_{ARCxECO}$ value. The utility of the two nested scoring versions is clear in the cases of Octoraro Creek and Young Womans Creek which both have shifting scores (Figure C1). Octoraro Creek is an example of score improving with scoring resolution ($WQI_{ARCxECO}$ score is higher than WQI_{SRB} score) while Young Womans Creek does the opposite ($WQI_{ARCxECO}$ score is lower than WQI_{SRB} score). Dunning Creek's scores, however, remain consistent across scoring tiers. All three scores work collectively to create a comprehensive picture of the water quality of a stream.

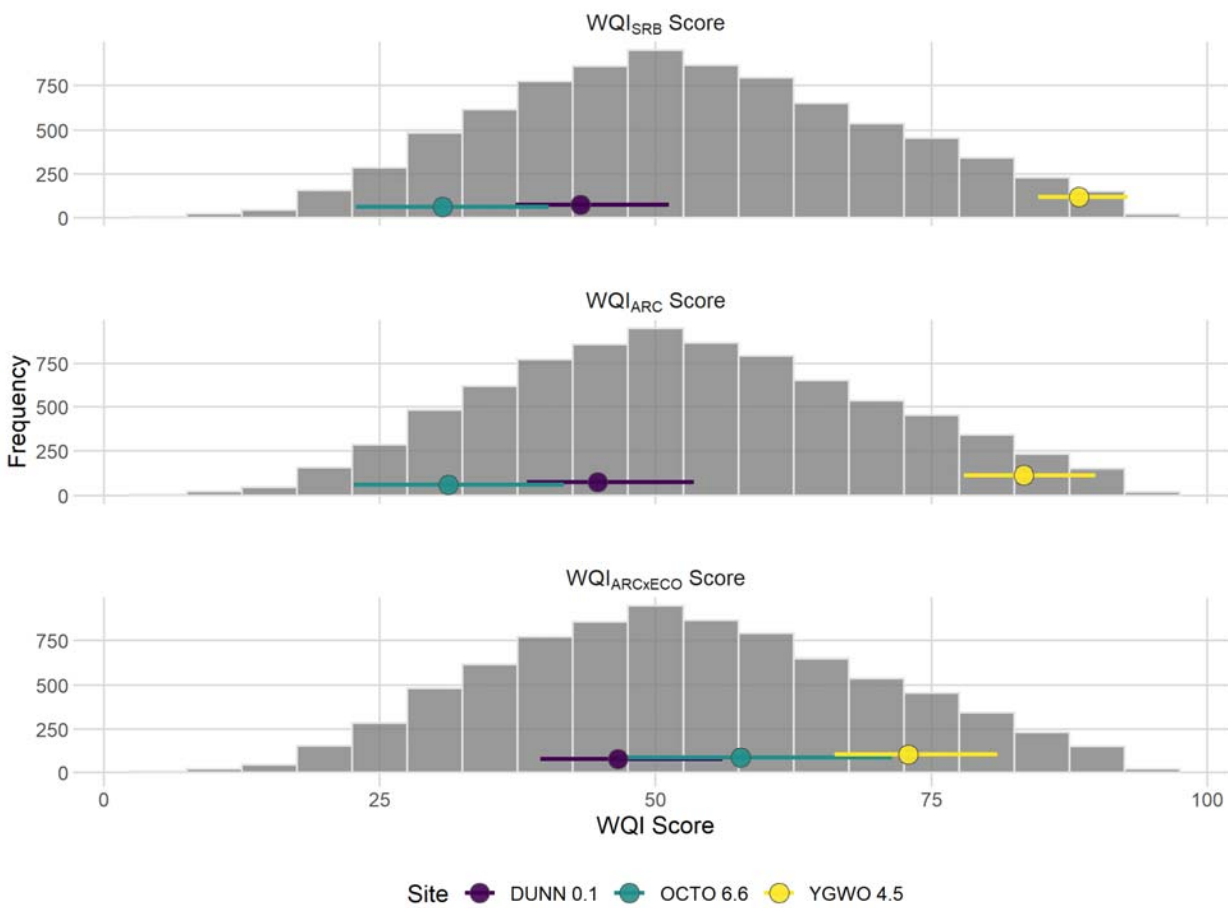


Figure C1. Mean WQI_{SRB}, WQI_{ARC}, and $WQI_{ARCxECO}$ I Scores for Three Stream Sites: Dunning Creek (DUNN 0.1, n=8), Octoraro Creek (OCTO 6.6, n=21), and Young Womans Creek (YGWO 4.5, n=23) (Error bars represent ± 1 SD. Gray bars indicate the frequency distribution of WQI scores within each Tier.)