Reassessing Criteria for Impaired Urban Waters and Stream Restorations

Report to the U. S. Environmental Protection Agency in Fulfillment of EPA Wetlands Program Development Grant - CD-00D51116-0

North Carolina Department of Environmental Quality Division of Water Resources

December 18, 2018

Prepared by: Lawrence Eaton

Table of Contents

Introduction

In the 1980s NC Division of Water Quality, now the Division of Water Resources (NCDWR), developed macroinvertebrate biological criteria based on a multi-habitat sampling method called a Full Scale collection with criteria assigned based on Total Taxa Richness and EPT Taxa Richness (Lenat 1988). In the late 1980s, it was discovered that Total Taxa Richness was too variable due to scour after rain events and was therefore dropped from the criteria. In the early 1990s, a rapid method was developed and the rapid method criteria were correlated with the Full Scale criteria (Eaton and Lenat 1991). In 1993, Tolerance Values were developed for most taxa in North Carolina and a Biotic Index was added as one of the metrics used to assign bioclassifications (Lenat 1993). Since that time, other methods have been developed by NC DWQ (e.g. Qual4 and Swamps), but EPT Taxa Richness and Biotic Index have remained the two metrics upon which most biocriteria are assigned (https://ncdenr.s3.amazonaws.com/s3fs-public/Water%20Quality/Environmental%20Sciences/BAU/SOPBenthosSept_2012.pdf). No other metric has been seriously evaluated for flowing streams since.

For years, the stream restoration community has struggled to document biological uplift, or improvement, for their projects (Adams et. al. 2002, Penrose 2004, Sudduth et. al. 2011; Tullos et. al., 2009; Violin et. al 2011). This is a problem because one of the stated outcomes of nearly every stream restoration is some version of “improve water quality,” and the current metrics, that are based on macroinvertebrate biological criteria, only infrequently change enough to “document improvement”. As documented by a previous EPA grant (CD 95415709-01), one effect of this difficulty documenting improvement is that mitigation for impacts in urban areas are being provided in rural areas, often far from the impacts. The inability to document ecological uplift in stream restorations is problematic because of the EPA policy of no net loss of wetlands or streams. An integral part of adhering to this policy is utilizing stream restorations to return previously lost ecological functions, when approval is given to impact streams as part of development. Without successful restorations returning ecological function, the regulatory agencies would have to choose between ignoring their no net loss mandate or allowing only very small development projects.

For almost 20 years, biologists and regulators have struggled to find a way to measure biological improvement (Penrose 2002). North Carolina first extended the monitoring period to five years post restoration before awarding all credits, then other biological metrics were proposed (e.g. Dominants in Common metric, and Keystone species) in lieu of the standard water quality ratings (e.g. bioclassifications of Fair or Poor). These proposed metrics have not proved to be consistent indicators, in 2015 the Wilmington District – U.S. Army Corps of Engineers increased the window to document improvement in stream mitigations, using standard biological metrics, to seven years, in the hope that more recovery time would be what was needed to document biological improvement.

The metrics used to measure water quality have not changed in North Carolina in nearly 25 years. During that time, biological uplift of stream restorations has been documented very infrequently. There are now proposals in the restoration community to move away from biological improvement as the goal for earning mitigation credit for stream restorations and more towards a societal good (e.g. building a greenway beside an urban stream could count as successful restoration deserving of a credit award) (Smith, et. al. 2016). Perhaps what is needed is to find some new metrics or criteria that are sensitive to smaller levels of ecological uplift in impaired sites instead of giving up on stream restorations being able to improve the quality of a stream. This effort would also be of interest to municipalities who for years have forced developers in urban areas to install expensive best management practices (BMPs), but have been unable to document any water quality changes because of them (Dave Phlegar, NC League of Municipalities, personal communication).

Recently, Will Harman (Stream Mechanics) developed a multi-metric model to predict which streams would be most likely to show biological uplift and, alternatively, calculate uplift on a restored stream (Harman and Jones 2016). His Quantification Tool (metrics listed in Appendix 1) makes biological improvement the last, most difficult, way to earn mitigation credits. While this tool recognizes that pollutant removal is important, even if it doesn't lead to measurable biological improvements, it would be better if there were a way to document that pollutant reductions led to improved biological function. In this model, he proposes a previously untested metric, % Shredders, that is designed to be a surrogate for trees growing along the bank, providing shade, bank stability and nutrients (leaf packs) to a stream. He also proposes other metrics that could be tested with current DWR sampling methods, which include both water chemistry and a habitat assessment form: specific conductance, temperature, bank erosion, canopy cover, and riparian area quality. Another metric that would be worth testing is % Tolerance Value < 6.0 (TV<6). This metric was designed to measure the shift in the chironomid community as urbanization increases (Gresens et. al. 2007). It was developed specifically to assist in a high school science project where it was found to outperform the rapid NC Stream Assessment Method (http://www.saw.usace.army.mil/Portals/59/docs/regulatory/publicnotices/2013/NCSAM_Draft_User_Manual_130318.pdf) in correctly separating small urban streams from suburban streams and also from least impaired streams. This grant looks at the feasibility of using any of these metrics, and others if proposed, as additional ways to measure biological function in impaired systems. These metrics could provide a way to measure water quality in areas where the inability to do so is becoming problematic in regard to EPA's goal of maintaining water quality.

Because in recent years most stream restorations and urban stormwater BMPs have been built in small watersheds, it was decided to concentrate efforts on finding new metrics for these small streams. In 2009, NCDWR concluded a study that decided that small stream (<3 mi^2) monitoring would be done using the Qual4 method and assessed using biocriteria where Biotic Index was the only metric.

New Bioclassification

One of the listed deliverables for this grant was to look at the possibility of developing additional bioclassifications for two of North Carolina's level 3 ecoregions (Mountains and Piedmont) in addition to the five bioclasses DWR has used for over 30 years (Poor, Fair, Good-Fair, Good and Excellent). Figure 1 is a scatter plot of the Piedmont sites rated Poor, Fair or Good-Fair by NCDWR. The X axis of the graph is the EPT Taxa Richness (EPTS) at each site and the vertical criteria lines are based on EPTS criteria from the EPT method. The Y axis is Biotic Index and the horizontal criteria lines are derived from DWR Small Stream biocriteria. Both a linear and a polynomial line of best fit were drawn through the data. The polynomial curve had a slightly better correlation than the linear regression at R^2=0.65, which shows that there is a significant amount of correlation between the two metrics, but not enough to consider the two metrics auto-correlated.

The graph also shows the effects of how the biocriteria were developed differently for each metric. According to Dave Lenat, the developer of both biocriteria, EPTS biocriteria cut offs were developed by finding the number of taxa that generally represented unimpaired conditions. Samples with that many or more taxa were considered Excellent. The remaining range below that number was divided equally into four bioclasses. The Biotic Index criteria were done slightly differently. Not only was the unimpaired (Excellent) range identified through best professional judgement, but the Poor range was also identified as streams that were significantly impaired and the remaining values between unimpaired and significantly impaired were equally subdivided into Good, Good-Fair and Fair bioclassifications. As a result, the BI and EPTS biocriteria lines intersect to form equal sized rectangles in

Fair, Good-Fair and Good (if it had been included) bioclasses. The exception is the Poor Biotic Index range, which is much larger than the others. This raises the possibility of splitting the Poor BI biocriteria and forming a new biocriteria called Very Poor.

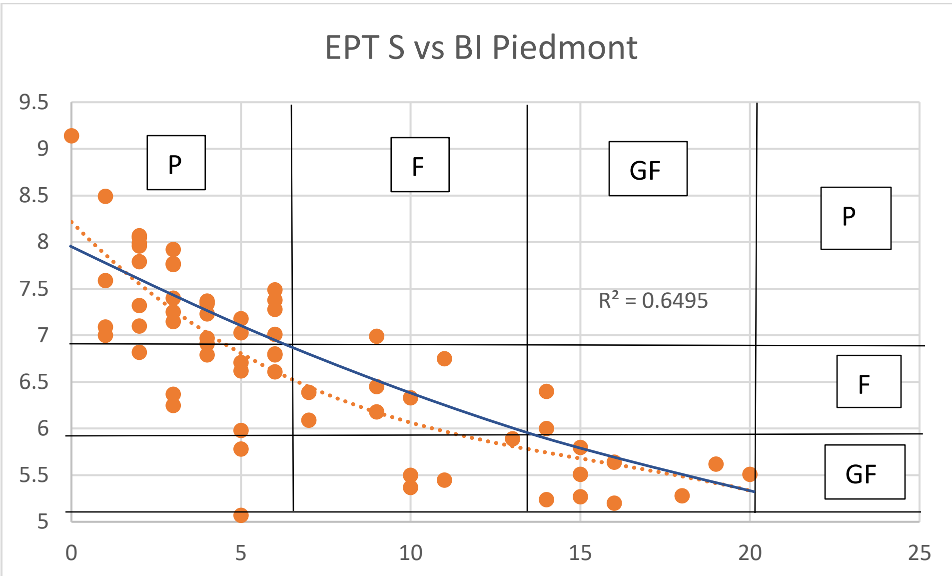

Figure 1. DWR Data, EPTS vs BI in the Piedmont – Poor, Fair and Good-Fair bioclasses.

Figure 2 shows the same data as figure 1, only with a new bioclass, Very Poor. Since the Good, Good-Fair and Fair bioclasses have a range of about 1.1, adding that number to the cut off for the Poor range, 6.9, yields a cut off for Very Poor in the Piedmont as >8.0. Figure 3 shows the same situation for the Mountain ecoregion. In that case the width of each Biotic index Bioclass is about 0.9, so a Very Poor bioclassification for the Mountains would be >7.4.

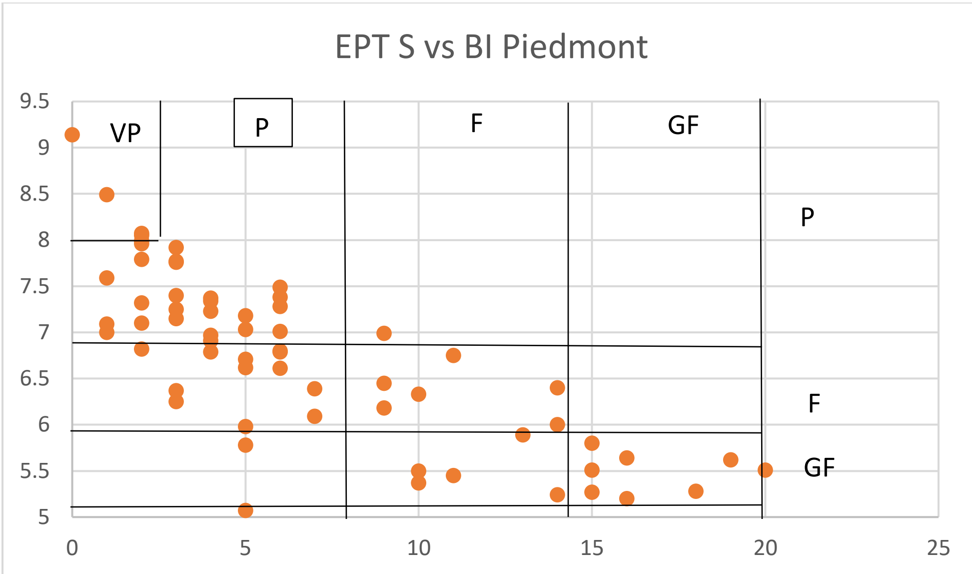

Figure 2. DWR Data, EPTS vs BI in the Piedmont – V. Poor, Poor, Fair and Good-Fair bioclasses

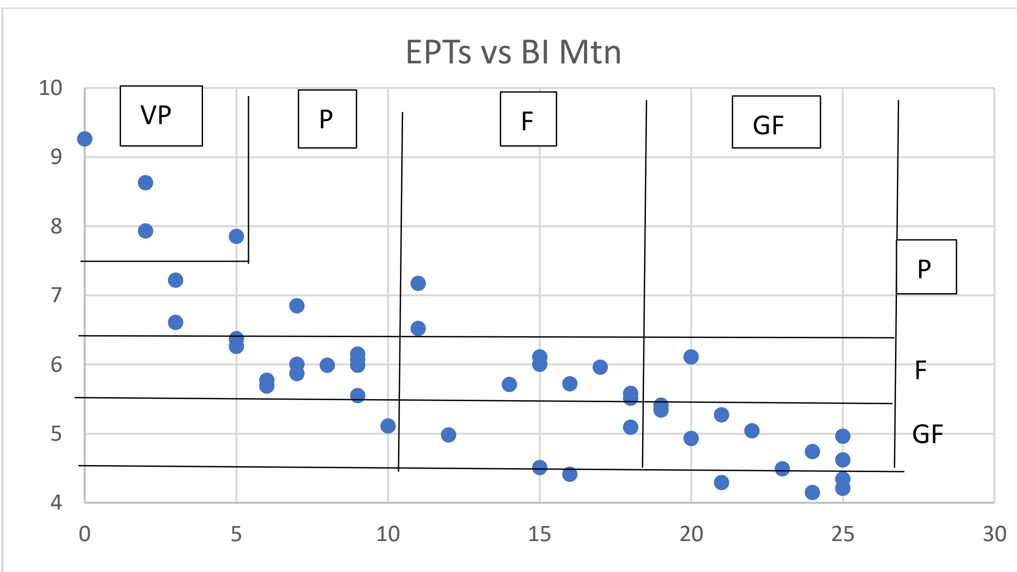

Figure 3. DWR Data, EPTS vs BI in the Mountains – V. Poor, Poor, Fair and Good-Fair bioclasses

While there are not very many observations in the proposed Very Poor classification, it still serves a function. In a past grant that looked at magnitude and extent of stream impacts below a dam (CD-00D01312), 12 of the 13 dams sampled showed a water quality decline of at least one bioclassification immediately below the dam. The one stream that did not decline in water quality was a dam in an urban area that was rated Poor before flowing into the impoundment. If the Very Poor bioclassification was used, the measured impairment below even that urban impoundment (BI=8.5) would have been enough to cause a decline in bioclass to Very Poor below the dam. This would have supported a firmer statement that all impoundments cause significant declines in stream water quality from above the pool to immediately below the dam.

New Metrics

Additional data

While 450 Qual 4 sites were collected by NCDWR from 2000-2017, only 112 of those were rated Poor, Fair and Good-Fair. Since that is the range of water quality most associated with urban streams and stream restorations, it would be good to have additional data points from the lower water quality ranges. Additional data was available from people who have retired from NCDWR, yet continued working with macroinvertebrates. These people are:

David R. Lenat – Developed Full Scale, EPT and Qual 4 metrics and modified the midwestern-developed Biotic Index to work in the southeast by assigning tolerance values to hundreds of invertebrates

David R Penrose – Led push to use macroinvertebrates to document ecological uplift in stream restorations and developed Taxa with TV<4 metric

Lawrence E Eaton – Developed Small Mountain Stream Criteria, Biocriteria for meso and poly-haline estuarine systems and developed taxa lists to separate intermittent from perennial streams.

Data collected from these three individuals, post retirement, add another 169 samples to the assessment.

The NC Division of Mitigation Services (DMS) commissioned a study looking at the effects of power lines going over stream restorations. The US Army Corps of Engineers (USACE) has claimed that all power line crossings harm the ecological integrity of restoration projects and has refused to award mitigation credit for these sections. This has amounted to several million dollars of lost credit to DMS. The study, which looked at 26 sites in the Piedmont and Mountains of North Carolina, showed that significant impairment, defined as a decline of at least one bioclassification compared to a site upstream, occurred at 9 of the 26 sites (35%). DMS plans to use this data to negotiate a different credit ratio for power line crossings with USACE. This data set, compiled by Penrose and Eaton, allowed for a comparison of how several metrics react to a known stressor (power lines) and not just a correlation with water quality in general.

Table 1 is a summary of the results of this study and comparison of the Biotic index with many of the other metrics. Of the 10 sites in the mountains, five sites showed a decline of one bioclassification or more and five streams did not. As predicted, most metrics that moved the same direction as the Biotic Index did so whether the impairments were large or small. Metrics such as EPT Taxa Richness (EPTS), EPT Abundance (EPTN) and Taxa<TV 6 tracked well regardless of impairment size. The Piedmont results, however were more problematic. When the effect of the power line was large (a decline of one bioclass), all metrics moved with the BI at least half the time, but when the differences were smaller (no change in bioclass), most metric movements correlated with the Biotic Index less than half the time. The only outliers were EPT N and Shredder abundance which tracked half the time. Also of note, in four of the 12 Piedmont sites that had impairment less than one bioclass, all of the tested metrics went in the opposite direction of the BI.

DMS data

Table 1. Summary of DMS data used to test metrics tracking with Biotic Index.

	Piedmont			Mountain	
	Same Bioclass	Diff Bioclass		Same Bioclass	Diff Bioclass
# pairs	12	4	# pairs	5	5
Number moved w/ BI			Number moved w/ BI		
Total Taxa Richness	5	2	Total Taxa Richness	2	3
EPT Taxa Richness	3	2	EPT Taxa Richness	3	3
EPT Abundance	6	2	EPT Abundance	3	3
Taxa < BI=4	4	2	Taxa < BI=4	2	2
%BI=4/tot	4	3	%BI=4/tot	2	4
Taxa < BI=6	4	3	Taxa < BI=6	3	4
%BI=6/tot	4	3	%BI=6/tot	3	4
Taxa Richness Shredders	5	2	Taxa Richness Shredder	4	2
Abundance Shredders	6	3	Abundance Shredders	2	1
Biotic Index	12	4	Biotic Index	5	5
Color Code:	>50% agree	50% agree			

Shredder metrics

Shredders are a group of animals that tear up rotting leaves and eat the bacteria and fungus that are performing the leaf decomposition. Poff et. al. (2006) lists a very small number of taxa that shred, most of which are very intolerant to pollution: Stoneflies (except Perlidae, Perlodidae and Chloroperlidae), the caddisflies *Lepidostoma* and *Pycnopsyche*, and the cranefly *Tipula*. Merritt et. al. (2008) also lists the caddisfly genera *Ironoquia* and *Triaenodes,* as well as the midge *Brillia* and the mayfly *Eurylophella funeralis,* as shredders. Not included in this list were taxa that are more omnivorous, where shredding was one of several ways an animal can derive nutrition (*Peltodytes*, *Helichus*, *Cricotopus, Hydrobaenus, Rheocricotopus, Chironomus, Endochironomus, Glyptotendipes, Polypedilum, Stenochironomus*). This fairly long list of taxa that may or may not be shredding, including the very speciose genera of *Cricotopus* and *Polypedilum*, adds a great deal of uncertainty to just how much shredding is going on in a stream segment. Also making it difficult to assess how much shredding is going on in a stream are the aspects of behavior changes over time (i.e. plasticity) and animal size. Some shredders, like *Pycnopsyche*, shred almost entirely when they are small, however as they get closer to their final molt, they will attach their case to a rock or log and scrape algae. The number of shredders that can be supported by any given stream segment is also dependent on the size of the animal. A mature *Tipula* is roughly 20 times larger than a mature *Leuctra* so an equal number of *Tipula* and *Leuctra* at a site represents vastly different rates of shredding. Any assessment of the usefulness of shredder metrics needs to bear in mind the uncertainty of just who is shredding.

The DMS data set provided the best opportunity to investigate the usefulness of shredder metrics (Abundance and Taxa Richness), since other data sets have samples from multiple seasons that would add more uncertainty to the analysis. Data for the DMS study was all collected in February 2017, so there was little variation in invertebrate life stage over the course of the study. As shown in Table 1, Shredder Abundance was one of only two metrics tested on this data set to track with the Biotic Index even half the time in the Piedmont, both when major impairments occurred and also when impairments were smaller. Shredder Taxa Richness also consistently tracked with the Biotic Index in the mountains when impairment levels were small.

The abundance of shredders in the Piedmont and Mountains, shown in Figures 4 and 5, mostly depict a large amount of variability between sites with Excellent to Good-Fair water quality. It also shows that most shredders are intolerant to pollution (80% of Shredders had TV<5). This is an issue because as the water quality in a stream segment declines, there are fewer obligate shredder taxa that can live in the stream no matter how much food is available. In the Piedmont, this point appears to be around a Biotic Index value of 6.5, which is in the middle of the Fair range, while in the Mountains this decline seems to be around a Biotic Index of 5.8, at the top of the Fair range. A line of best fit between Shredders and Biotic Index in both ecoregions finds a correlation (R^2) of only 23%

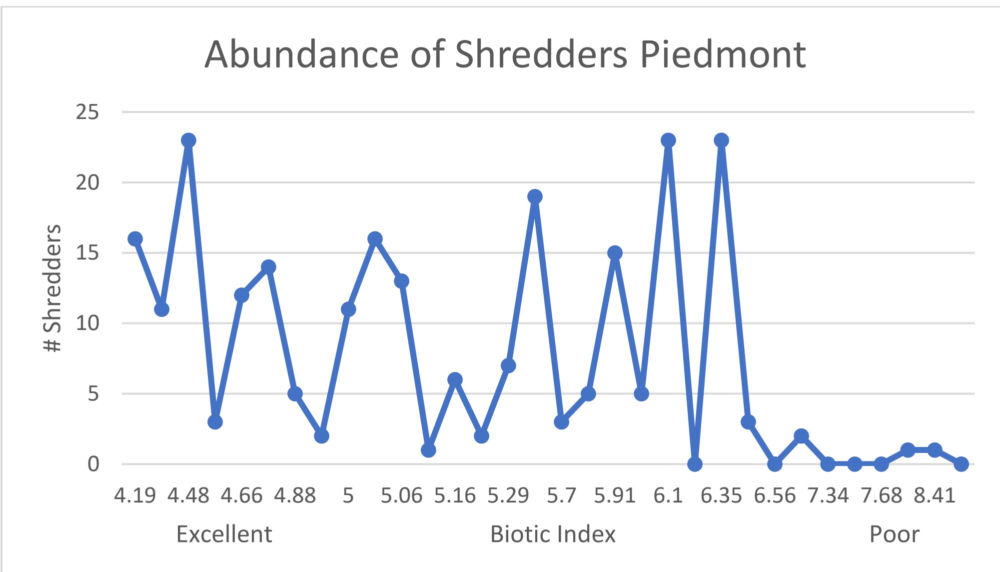

Figure 4. Abundance of Shredders in the Piedmont

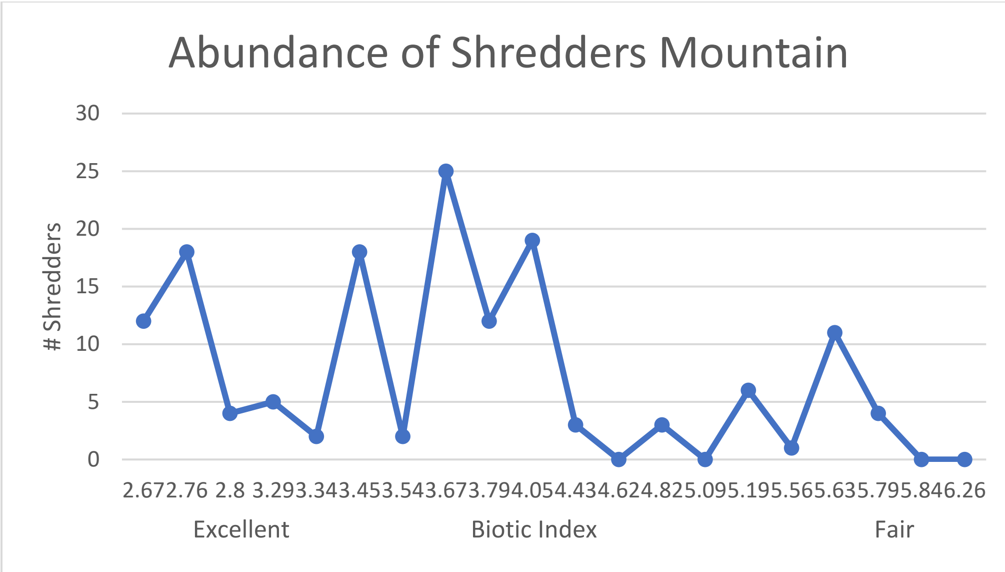

Figure 5. Abundance of Shredders at Mountain sites.

Shredder Taxa Richness in the Piedmont and Mountains are shown in Figures 6 and 7. They show that while Taxa Richness doesn't have the abrupt drop off in the lower quality streams, there is still a large amount of between-site variability. The lowest Biotic Index with no shredders was 6.2 in the Piedmont (Fair), but 4.6 in the Mountains (Good). The correlation (R^2) between Shredder Taxa Richness and Biotic index was 48% in the Piedmont and 41% in the Mountains. Even though the correlation between Shredder Taxa and BI is nearly twice as much as Shredder Abundance and BI, there is still too much variability within bioclasses to develop useful biocriteria.

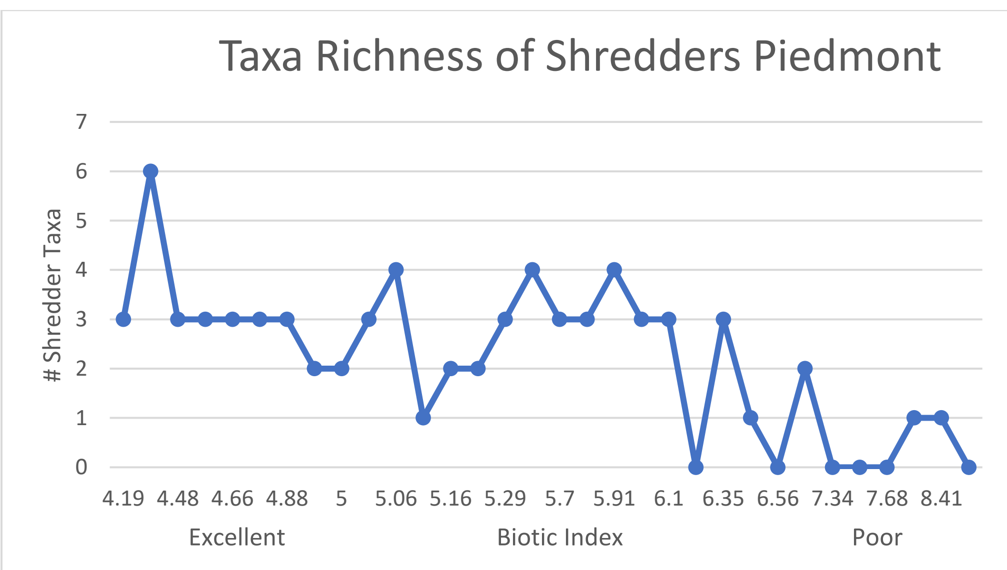

Figure 6. Taxa Richness of Shredders in the Piedmont

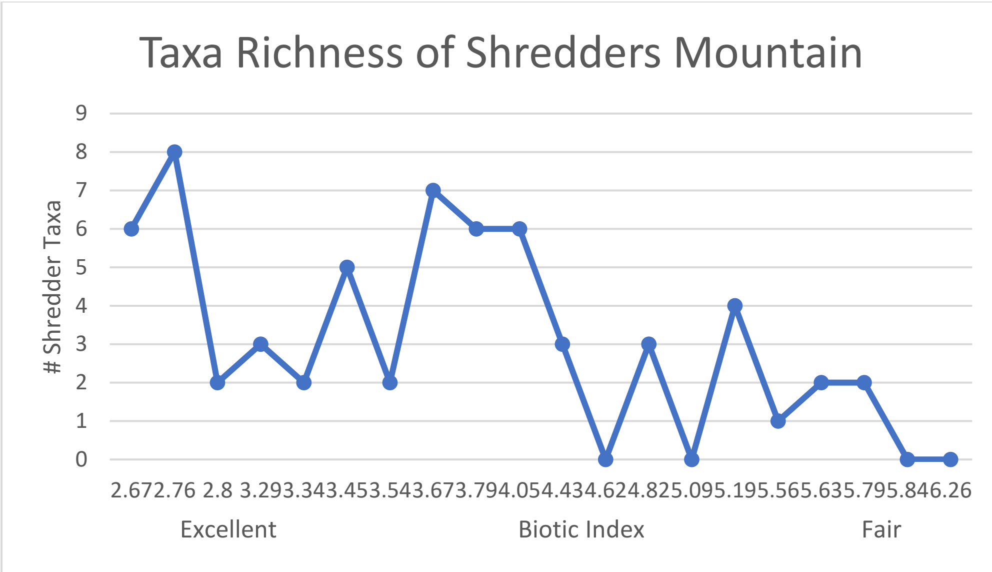

Figure 7. Taxa Richness of shredders at mountain sites.

NCDWR Data

Data in the next two sections will be graphically presented using Box and Whisker plots. The general layout will be Poor sites will be on the left of the graph in the green box, Fair sites are in the blue box in the center and Good-Fair sites are in the reddish box on the right. Generally speaking, if there is space between the top of one box and the bottom of the adjacent one, a biocriteria cut off can be made where there should be relatively few misclassifications.

% Impervious Surface

Figures 8 and 9 show the % impervious surface of the 112 sites sampled by NCDWR rated Fair, Poor and Good-Fair in the Piedmont and the Mountains.

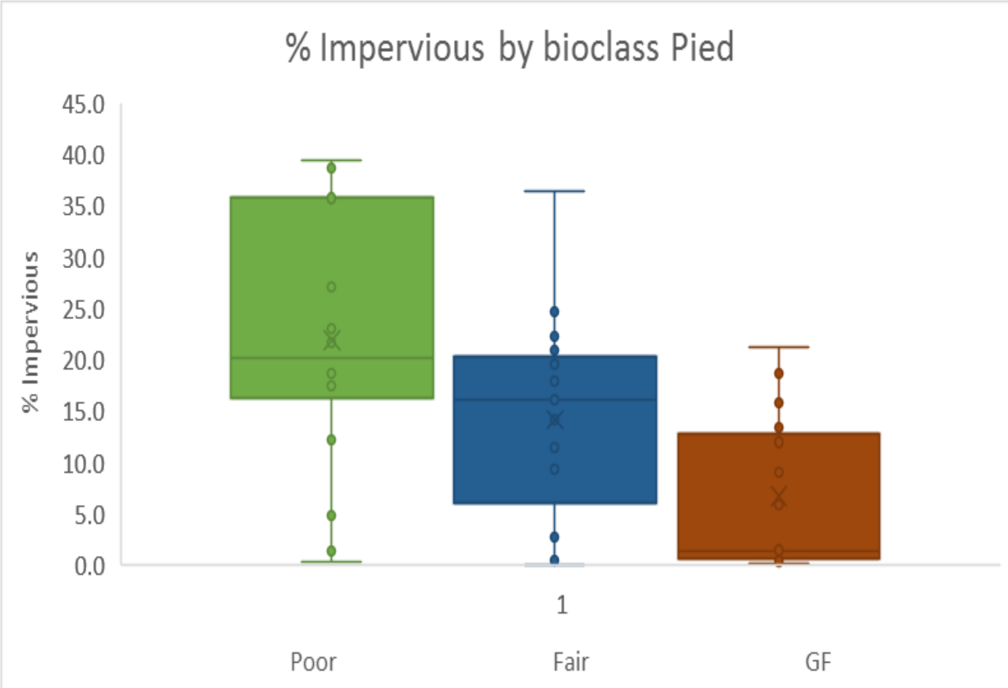

Figure 8. NCDWR % Impervious Surface – Piedmont

The Piedmont sites (Figure 8) show approximately what one would expect from this metric – improving water quality with declining Impervious surface in the watershed. The issue comes from trying to develop criteria for this metric. The median value for sites rated Poor is approximately 21%, while the median value for sites rated Fair is about 18%. Choosing a cut off between these two numbers would, by necessity, lead to a large number of misclassifications. The problem of misclassifications is the same, although to a lesser degree between Fair and Good-Fair, where a criteria line between 8 and 10% would probably be optimal.

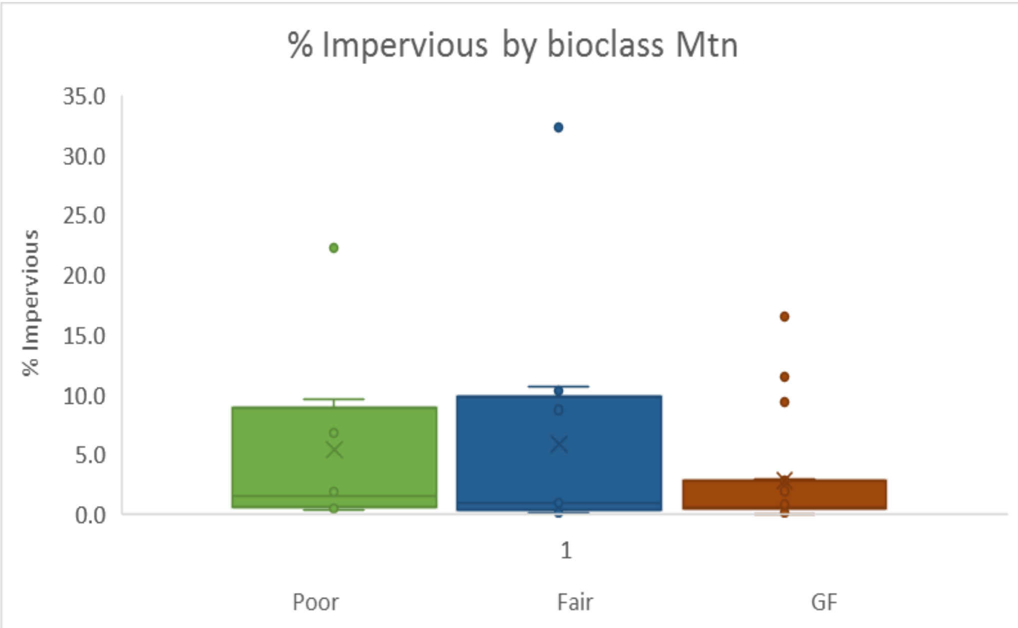

Figure 9. NCDWR % Impervious Surface – Mountain

The NCDWR dataset for the Mountains (Figure 9) appears to be unsuited to testing the % Impervious metric. Upon examination, most mountain samples in this data set were collected from rural areas and looked at agricultural issues (e.g. trout farm impacts, pesticide and herbicide runoff, and sedimentation from land disturbing activity) where impairments had nothing to do with urban runoff. It should be noted that Eric Fleek (NCDWQ 2009) found that in terms of land use in all NC ecoregions, % Forested had a much better correlation with water quality than did any other land use type, including % impervious.

Specific Conductance

Specific Conductance is a measure of ions in the water. While most pollutants do, in fact, elevate the number of ions in a stream, the background ionic concentration in streams can be variable, especially when only a single measurement is taken at a location. While there is general agreement that impairment is frequently observed in Piedmont streams when the Specific Conductance is >70 μ/S, this is not always the case. The author has found Piedmont headwater streams with watersheds entirely within the boundary of State Parks with Specific Conductance of >100 μ/S. Another example is Wilson Creek in Chapel Hill, where the upstream portions of the stream have been regularly rated Good or Excellent and the Specific Conductance has always been in the 130-160 μ/S range. Figures 10 and 11 are the NCDWR values for Specific Conductance for Piedmont and Mountain sites.

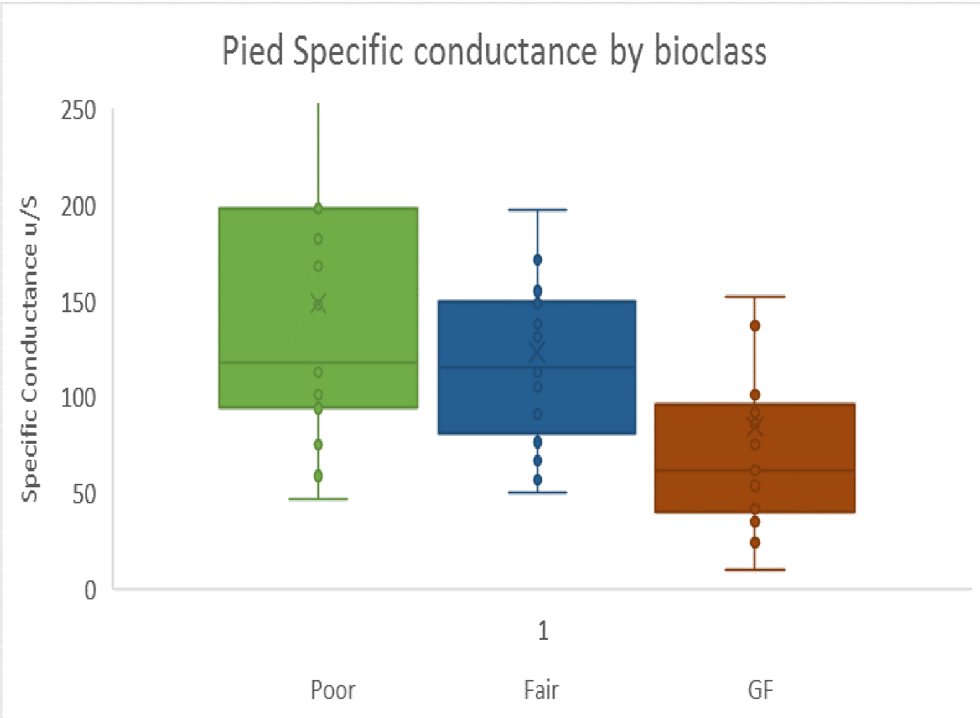

Figure 10. NCDWR Specific Conductance – Piedmont

Specific Conductance in the Piedmont (Figure 10) agrees with previous observations– while there is a breakpoint between impaired and unimpaired sites around 70-80 µ/S, there is too much ecoregion wide variability to assign additional levels of biocriteria with any degree of confidence.

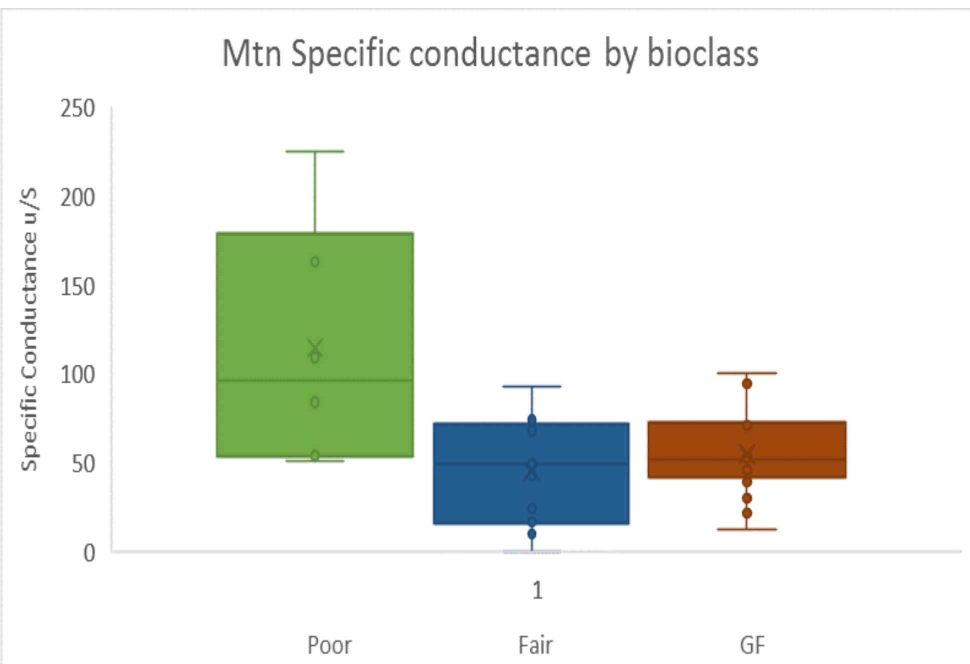

Figure 11. NCDWR Specific Conductance – Mountains

Specific Conductance in the Mountains (Figure 11) is even more problematic. The variability of the Fair bioclass range totally encompasses that of the Good-Fair range, with the Poor range being somewhat elevated above them both. While the Specific Conductance of approximately 70 μ/S appears to be a break point, in this case that break is between Poor sites and everything else. While Specific Conductance seems to be limited as a metric across ecoregions, it may still prove to be a powerful tool at a watershed level, such as a stream restoration. Currently Dave Penrose is building a model using Specific Conductance to predict stream water quality. In its current form, this model can ascribe 70% of between site changes in specific conductance to changes in water quality.

Total Taxa Richness (TotS)

In the 1980s and 90s, Total Taxa Richness (TotS) was used, as well as EPT Taxa Richness and EPT Abundance, as a biocriteria metric for Full Scale samples. The metric was unsatisfying because it was documented that following rain events, natural scour and drift would decrease the Total Taxa Richness at a site. This effect was especially pronounced for sandy sites, but even unimpaired mountain streams would see post-precipitation declines due to scour. When the Biotic Index was developed for the southeast, Total Taxa Richness was dropped in favor of Biotic Index**.** Figures 12 and 13 show the variability of the metric between the bioclassifications.

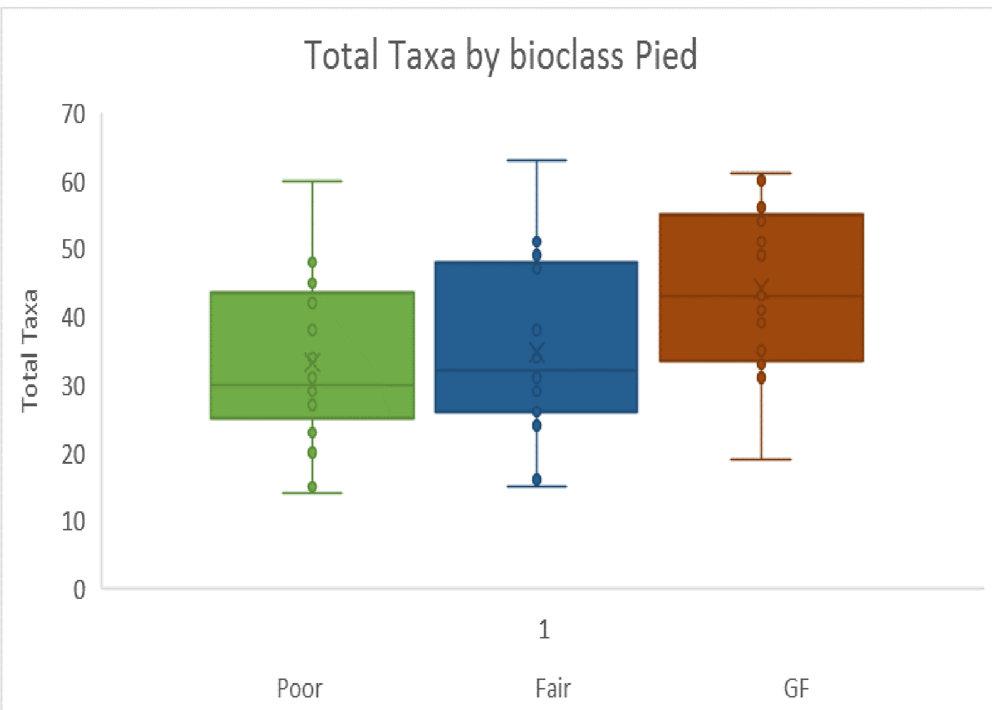

Figure 12. NCDWR Total Taxa Richness - Piedmont

For TotS in the Piedmont (Figure 12), there is surprisingly little difference between the ranges of Total Taxa within each bioclassification. The Poor range of TotS is entirely a subset of the Fair range. The Good-Fair range is slightly higher than Fair/Poor, median of 40 taxa compared to 32; however, the within bioclass variability leads to a large amount of overlap.

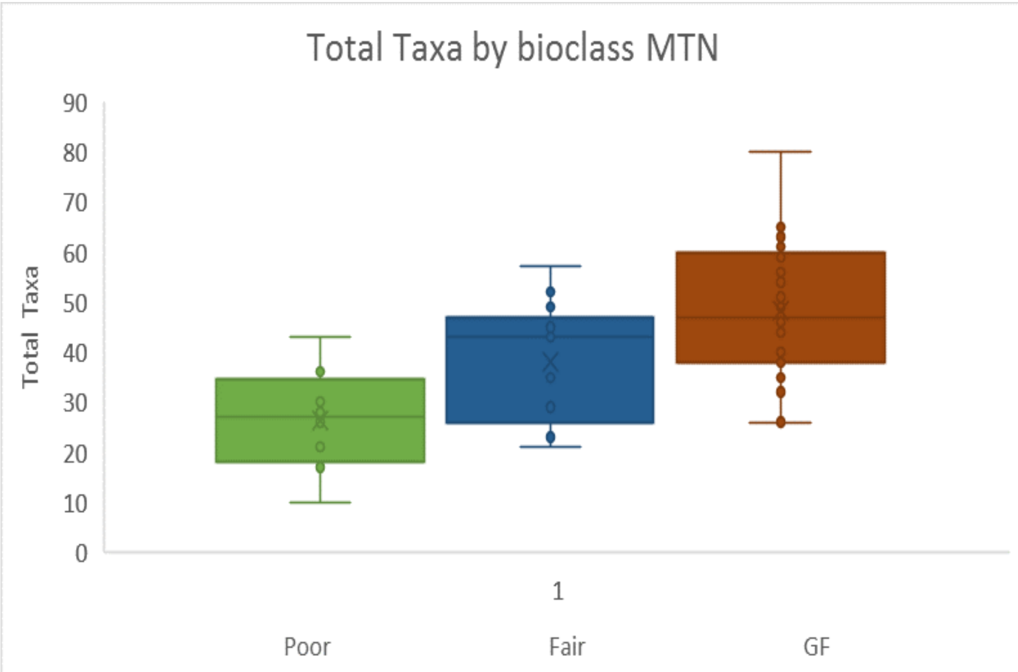

Figure 13. NCDWR Total Taxa Richness- Mountains

Total Taxa Richness in the Mountains (Figure 13) had lower variability than TotS in the Piedmont, however there is still a large amount of overlap between bioclasses. With the median TotS value for Poor sites around 28 and Fair sites around 42, one could propose a Poor/Fair biocriteria cut off around 33 taxa, however there would still be a significant number of misclassifications with that line. A proposed Fair/Good-Fair cutoff would be much more difficult. With the Fair range having a median value of 42 and the Good-Fair median being 45, nearly half of each bioclass would likely be misclassified.

EPT Taxa Richness (EPTS)

The EPT Taxa Richness (EPTS) metric was invented in the 1980s. EPT stands for the orders Ephemeroptera (mayflies), Plecoptera (stoneflies) and Trichoptera (caddisflies). These orders are generally considered to be the aquatic insect groups most intolerant to pollution. Therefore, more species of these intolerant orders present in a sample would indicate better water quality. The EPTS metric is used by NCDWR in its Full Scale, EPT and Swamp biocriteria. It was tested to be an additional metric for the Qual 4 method, but despite it working better than every other metric besides the Biotic Index, it was decided that it was too variable to use as a metric (NCDWQ 2009). Figures 14 and 15 show the EPTS values for the Piedmont and Mountains, respectively.

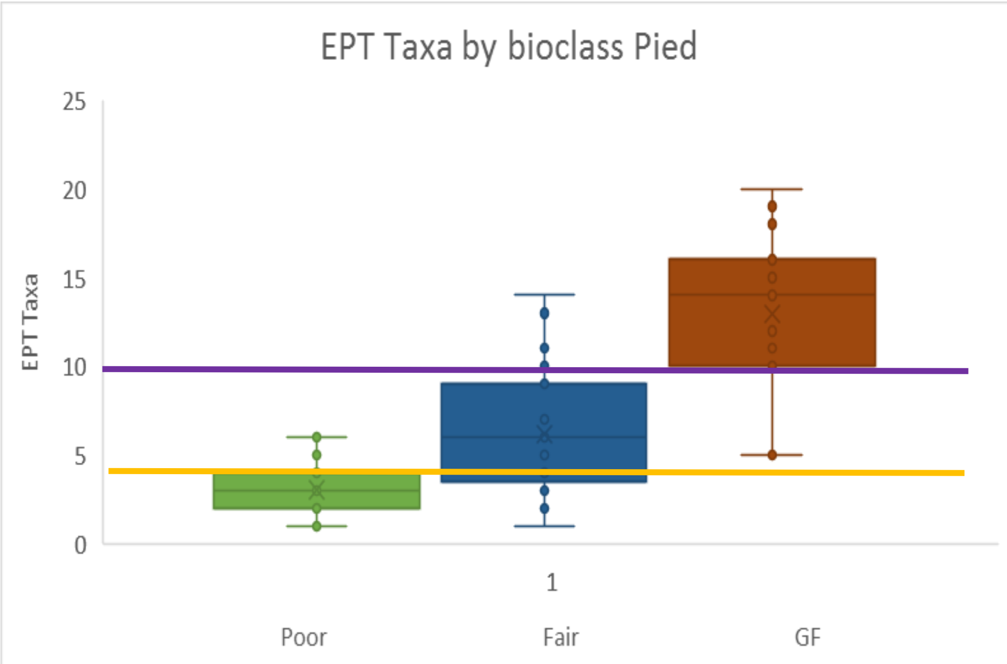

Figure 14. NCDWR EPTS Piedmont

Of all of the metrics tested as part of this study, EPTS has the least amount of overlap between bioclasses for the Poor/Fair and Fair/Good-Fair ranges, suggesting that EPTS could be a viable metric for these lower levels of water quality. Using this DWR data, a good case could be made to draw the Poor/Fair biocriteria cut off line at 4 EPT taxa in the Piedmont and 7 taxa in the Mountains and the Fair/Good-Fair line at 10 taxa in the Piedmont and 15 EPT in the Mountains.

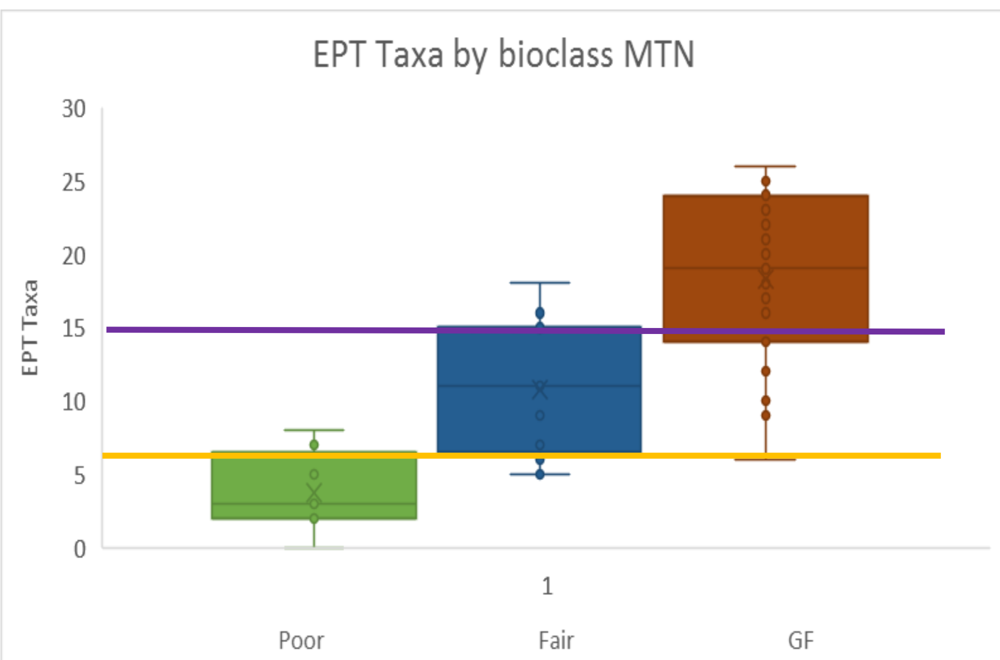

Figure 15. NCDWR EPTS Mountains

EPT Abundance (EPTN)

EPT abundance (EPTN) is a count of the number of individuals in the groups Ephemeroptera, Plecoptera and Trichoptera. This is an ancillary metric to EPTS and was introduced to try to separate sites where there was a high number of EPT taxa, but low abundance, from sites with both high taxa richness and abundance, as one would expect in an unimpaired stream. Examples of when one might find high EPTS, but low EPTN, would be a sandy stream with embedded riffles and little habitat or a site just downstream of a point source discharge where intolerant bugs from above the discharge were drifting into a much more impaired stream segment below it. In his small stream criteria work, Eric Fleek (NCDWQ 2009) found this metric to work third best for small streams, behind BI and EPTS. Figures 16 and 17 show relatively decent separation between most bioclasses in both the Piedmont and the Mountains.

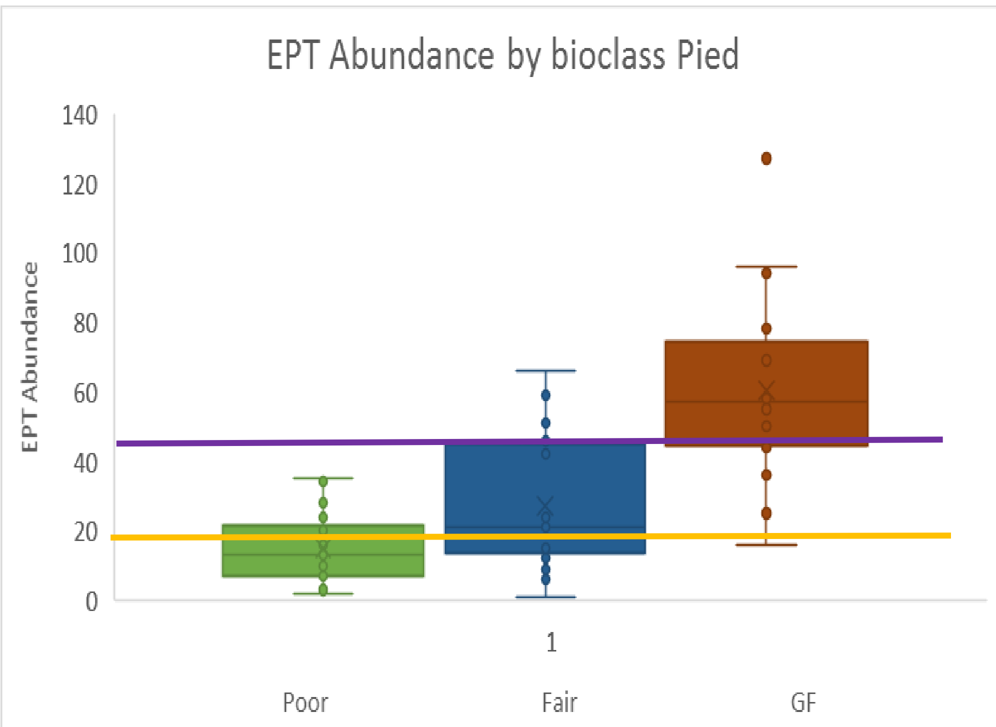

Figure 16. NCDWR EPTN Piedmont

For the Piedmont, Figure 16 shows an effective separation between Fair sites and Good-Fair sites at around 45 individual EPT. The Poor/Fair cut off is less defined, but still better than the separation of Total Taxa Richness bioclasses. A cut off value around 20 EPTN will lead to a few misclassifications; however, this metric should be subjected to further testing since it otherwise seems to perform well.

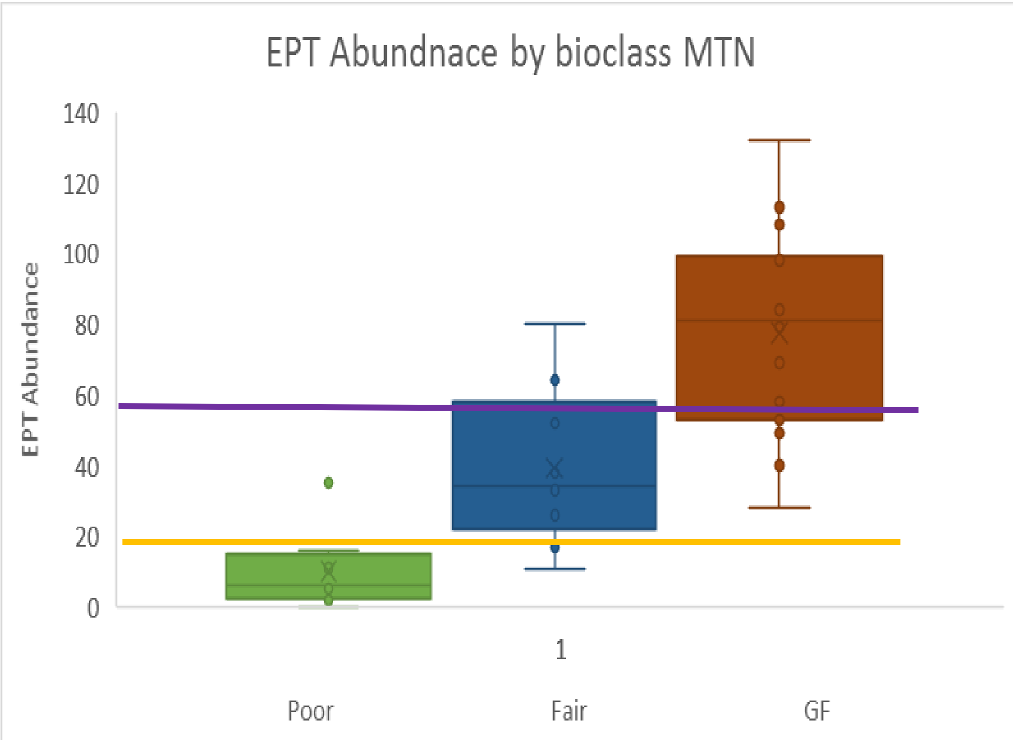

Figure 17. NCDWR EPTN Mountain

Figure 17 shows the EPTN for data from the Mountain ecoregion where the separation between Poor and Fair is much better than in the Piedmont. A Poor/Fair cut off value of 20 will cause almost no misclassifications between the two bioclasses and a Fair/Good-Fair cut off around 59-60 EPTN would only lead to minimal misclassifications.

TV<4

This metric has been used only recently and only by Dave Penrose, a long time State Biologist and contemporary of Dave Lenat. It is a count of the number of taxa in a sample that have a tolerance value <4, which are generally agreed to be intolerant taxa. This metric is a refinement of the EPT Taxa Richness metric, in that while there are mayflies, stoneflies and caddisflies that are tolerant to pollution, such as the hydropsychid caddisflies *Hydropsyche betteni* and *Cheaumatopsyche* spp, there are also intolerant taxa in other, non EPT, groups, such as the beetle *Psephenus herricki* and the snail *Elimia* spp. While Penrose has been using this metric (and TV<2.5) for several years, he has done no testing of these metrics to see if they work any better than EPTS. Figures 18 and 19 show a generally good separation between bioclasses in DWR data for both Piedmont and Mountain ecoregions. Since this is a new metric for DWR testing, all Qual 4 data was utilized, including 350 Excellent and Good rated sites in addition to the 112 Poor, Fair and Good-Fair samples.

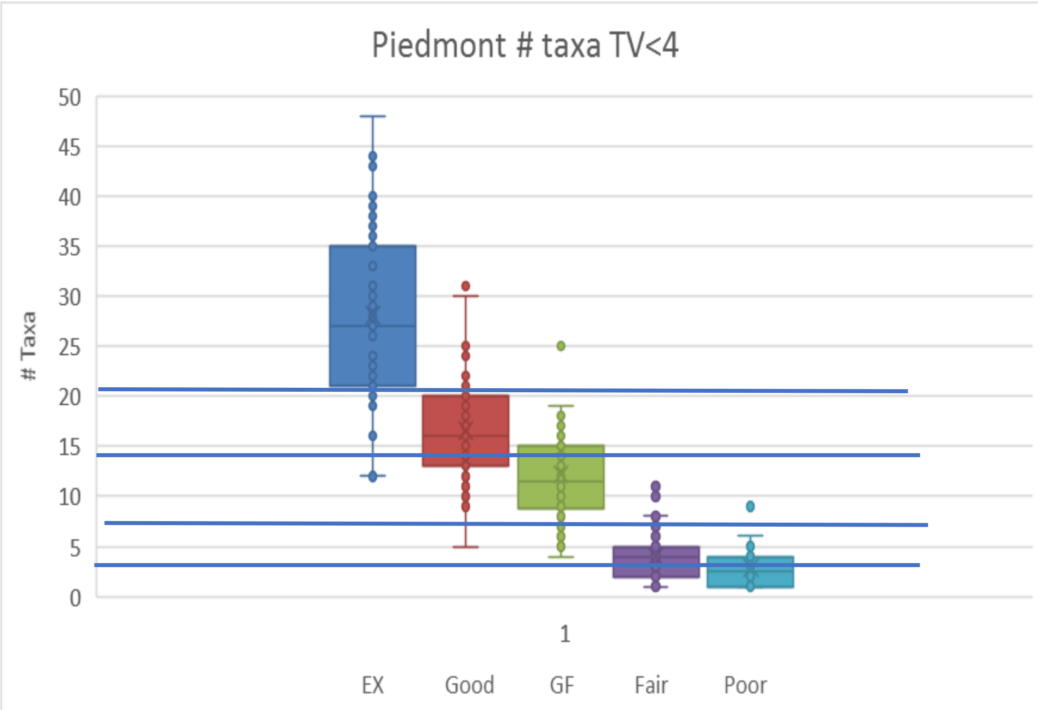

Figure 18. NCDWR TV<4 Piedmont

A generally successful job of drawing criteria cut off values can be made for the Piedmont TV<4 data (Figure 18). The Excellent/Good cut off at 20 taxa only includes mild outliers in each bioclass, while the Good/Good-Fair line at 14 has slightly more overlap, and thus a few more potential misclassifications. Good separation exists between Good-Fair and Fair bioclasses, so a cut off around 7 or 8 taxa would allow for minimal misclassifications. However, there is a large amount of overlap in the distributions of the Fair and Poor bioclasses and a cut off of 3 or 4 will likely have a significant number of misclassifications.

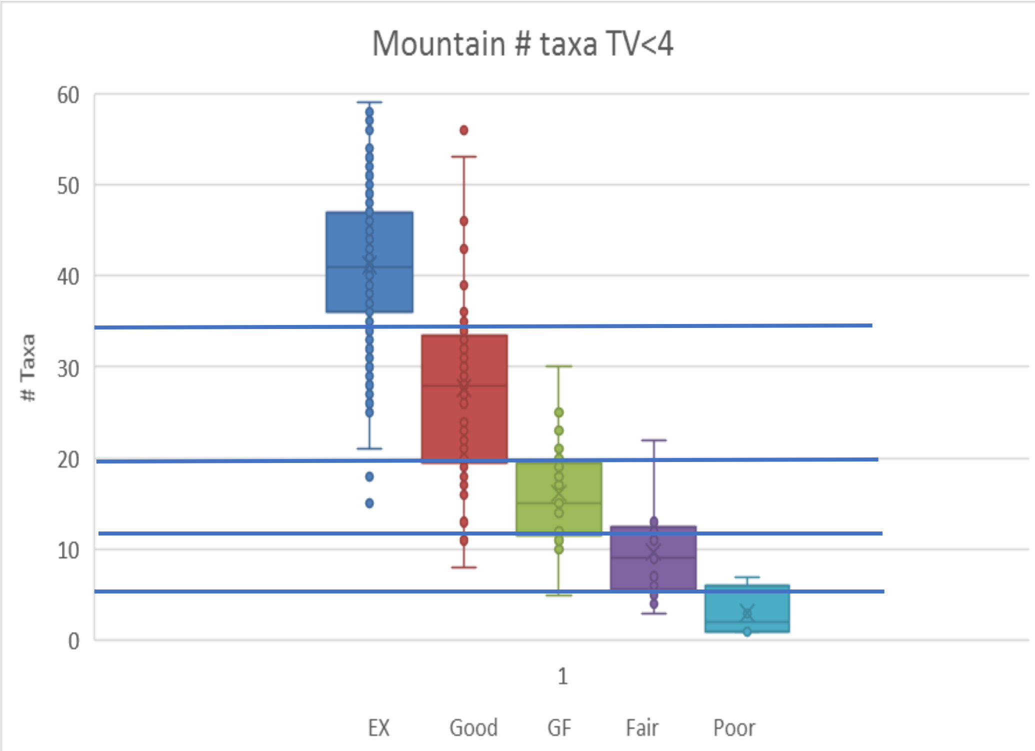

Figure 19. NCDWR TV<4 Mountains.

The cut off lines separating the five bioclasses in the Mountain ecoregion (Figure 19) are much cleaner than those in the Piedmont and will likely lead to a minimal number of misclassifications. Apparent cut off points are: Poor/Fair - 5; Fair/Good-Fair – 11; Good-Fair/Good – 20 and Good/Excellent – 34. The largest problem with the data is the wide variability in the Good and Excellent bioclassifications, which is almost certainly due to there being almost three times as much data in these two bioclasses than with Good-Fair, Fair and Poor bioclasses.

Chapel Hill/Carrboro/Durham Data

The main source of data beyond the DWR database is from the towns of Chapel Hill, Carrboro and Durham (CHCD). While Dave Lenat has been doing monitoring for the municipalities since 2005, I was only able to obtain taxa lists from 2012-2018, which were, unfortunately, not associated with water chemistry or land use data. This amounted to 86 samples from 2012-2017 plus an additional 26 sites collected in 2018 that were used as part of the validation database. Data from Raleigh and Greensboro were also acquired for this study, however, neither data sets were used. Raleigh data was volunteer collected and identified, and thus of unknown quality, while Greensboro data turned out to be entirely collected using the Full Scale technique and thus was not comparable to Qual4 results.

While DWR samples came from all over the State, the Towns of Chapel Hill and Carrboro and the City of Durham are within 20 miles of each other in two adjacent counties in the Piedmont ecoregion. All towns share a mix of level 4 ecoregions Triassic Basins and Slate Belt, both of whom are characterized by flashy flows and poor connection to groundwater which causes small streams to dry from drought conditions sooner than other areas in the Piedmont. Small Stream drying is a large enough issue with DWR that it recommends that small Triassic Basin streams not be assigned bioclassifications. This admonition does not help the towns, so they use DWR biocriteria anyway for their monitoring purposes, however for this reason this data set was analyzed separately from the DWR data, since frequent drying could shift the results.

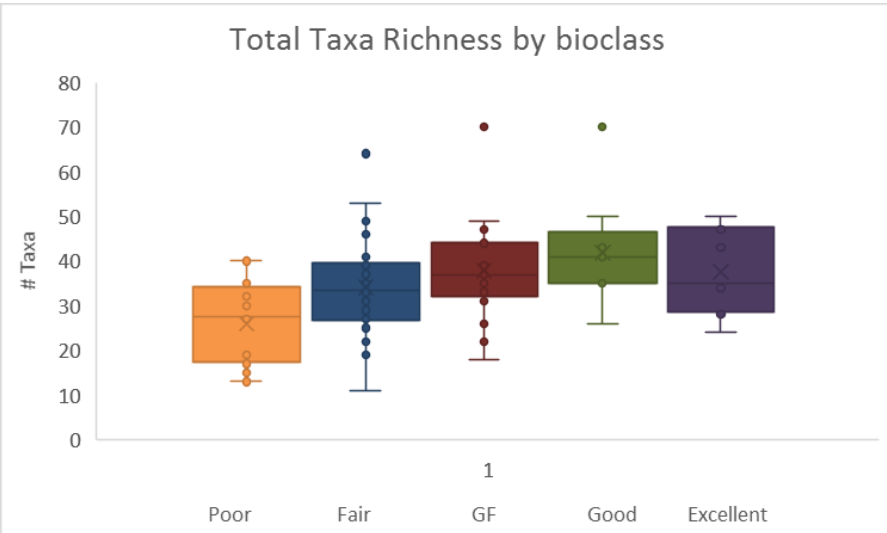

Figure 20. CHCD Total Taxa Richness.

Figure 13, Total Taxa Richness for DWR data, showed no real differences between bioclasses, and Figure 20, shows much the same thing with CHCD data. Not only is there a large amount of overlap between Poor, Fair and Good-Fair bioclasses, the Total Taxa Richness is indistinguishable between Good-fair, Good and Excellent.

EPT Taxa Richness (EPTS)

Figure 21 is the EPT Taxa Richness for the CHCD data. The criteria lines for Poor/Fair and Fair/Good-Fair were the 4 and 10 cut offs derived from the DWR data. The Poor/Fair cut off works well for both data sets, however there is more overlap between Fair and Good-Fair sites with the CHCD data. Like Total Taxa, the Good and Excellent ranges for EPTS are indistinguishable.

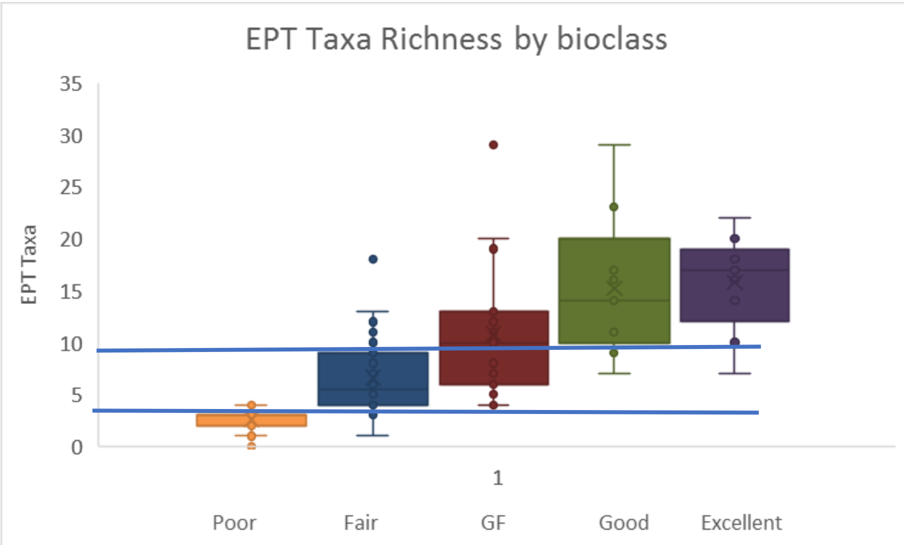

Figure 21. CHCD EPT Taxa Richness

TV<4, TV<6

In addition to the metric TV<4 discussed earlier, there are two other tolerance value cut off points used for different types of systems. Dave Penrose has refined his metric to very intolerant taxa (TV<2.5) for his work in the Mountain ecoregion and Larry Eaton has been using a more facultative TV<6 for intermittent streams and wetlands. While initial testing rejected TV<2.5 as too low to differentiate sites in less than pristine Piedmont streams, TV<6 was tested further to see if it would be able to more subtly parse differences between impaired Piedmont sites.

Figure 22 is the Town data for the TV<4 metric. While there does appear to be good separation between Poor/Fair, Fair/Good-Fair and Good-Fair/ Good-Excellent, the breaks in this data set are not quite in the same places as criteria lines from the DWR data, which are shown on the graph. This may be due to the lack of habitat in many Triassic Basin streams, a primary reason that DWR's Biological Assessment Unit SOP recommends against assigning bioclassifications to these streams using their metrics (DWR 2012).

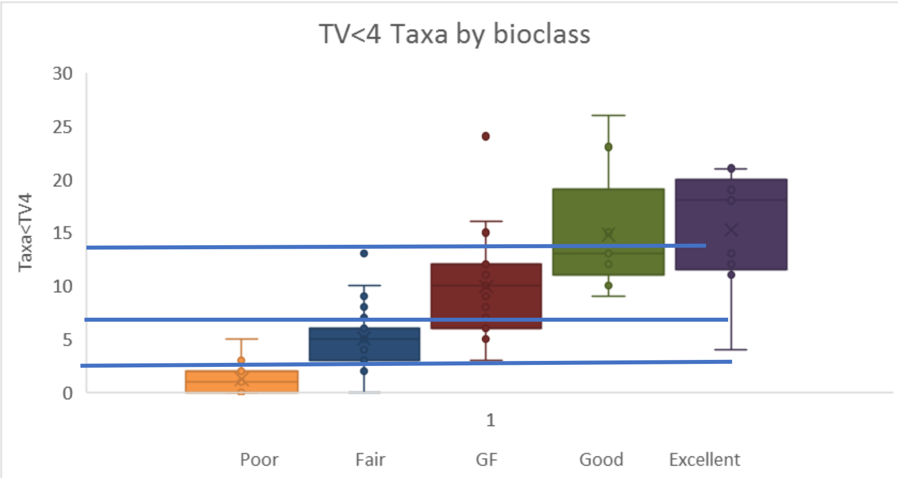

Figure 22. CHCD Taxa<TV4

Figure 23 is the town data for Taxa with TV<6. Despite hopes that this metric would do a better job than TV<4 at separating Poor from Fair Piedmont sites, this does not appear to be the case. While criteria can be drawn, around 6 and 14, there are more misclassified sites using TV<6 than TV<4.

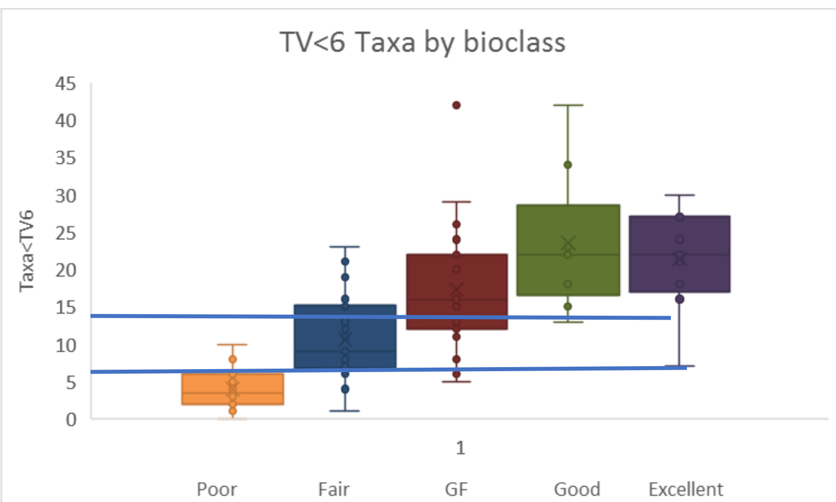

Figure 23. CHCD Taxa<TV6

Figures 24 and 25 evaluate the performance of the Abundance of taxa TV<4 and TV<6 to discriminate between bioclassifications. TV<4 Abundance was effective at separating the bioclasses, with the Poor/Fair cut off around 4, Fair/Good-Fair around 30 and Good-Fair/Good-Excellent around 49. Unfortunately, the TV<6 metric underperformed and could only discriminate between Poor and Fair without significant overlap and misclassifications.

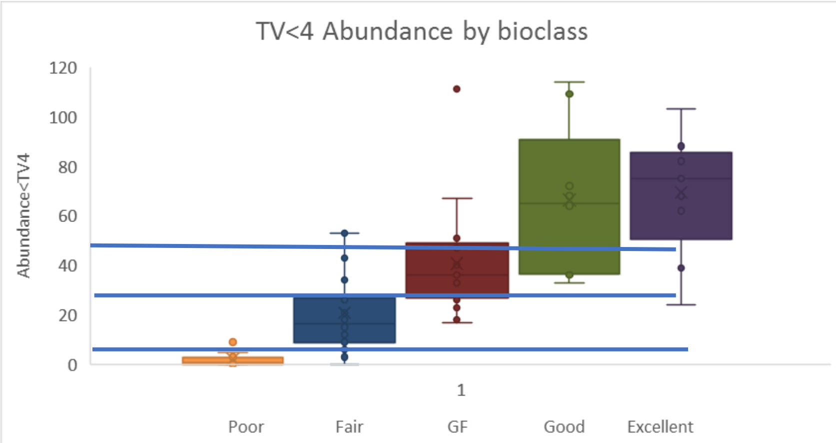

Figure 24. CHCD TV<4 Abundance

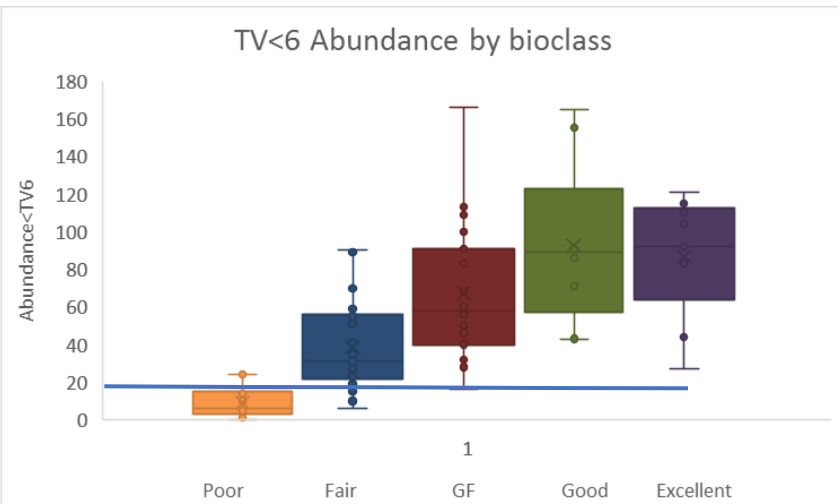

Figure 25. TV<6 Abundance

It appears that the TV<6 metric regularly underperforms the TV<4 metric in separating bioclasses for both taxa richness and abundance. While this metric may have some utility in wetlands or intermittent streams, where there is a dearth of intolerant taxa, it is not as useful as TV<4 and so will not be pursued.

Effect of Stream size

Throughout the above discussion, it was noted that while many taxa richness and abundance metrics did a decent job of separating sites with some level of impairment (Poor, Fair and Good-Fair), in most cases there was no way to distinguish between Good sites and Excellent sites (Figures 20-25). Figure 26, which was developed as part of a previous small streams biocriteria grant, shows that there is a maximum number of taxa and individuals that can fit into a small stream and that this maximum gets larger with increasing stream size.

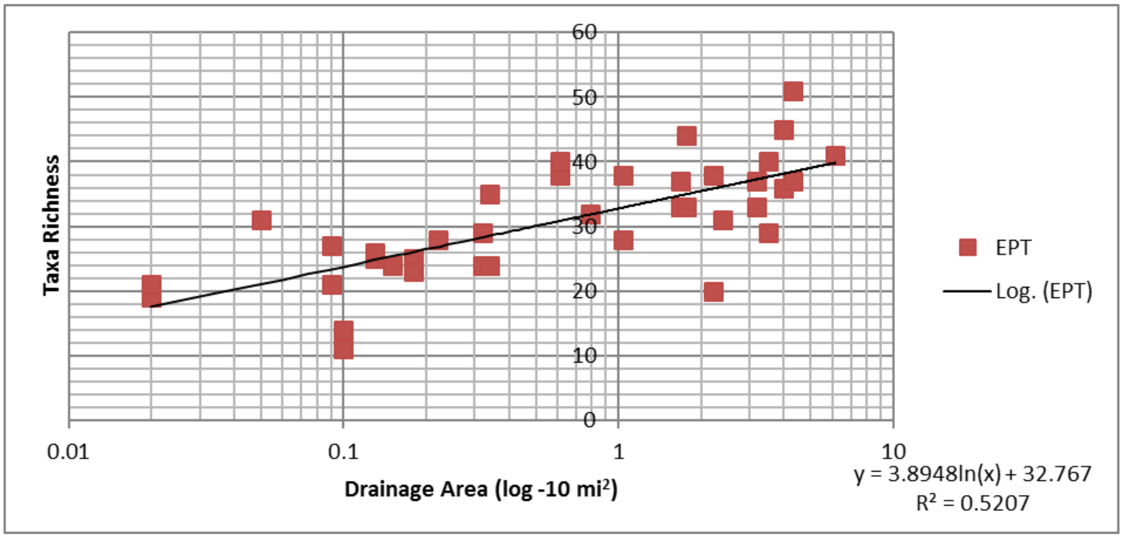

Figure 26. EPT Taxa Richness in Unimpaired Small Mountain Streams.

The results of this maximum can be seen in the following graph. Figure 27 is a graph of EPTS in small streams and their bioclasses overlain by the EPT bioclasses for streams > 3 m. Good-Fair small stream EPTS values fit in the large stream Fair range and Good and Excellent small stream EPTS values fall neatly in the Good-Fair category for large streams. Very few small stream observations reaching into the Good or Excellent large stream bioclasses. This is very likely the reason why metric analysis for small stream biocriteria in 2008 found that EPTS and EPTN did not work well enough in small streams to warrant inclusion as metrics.

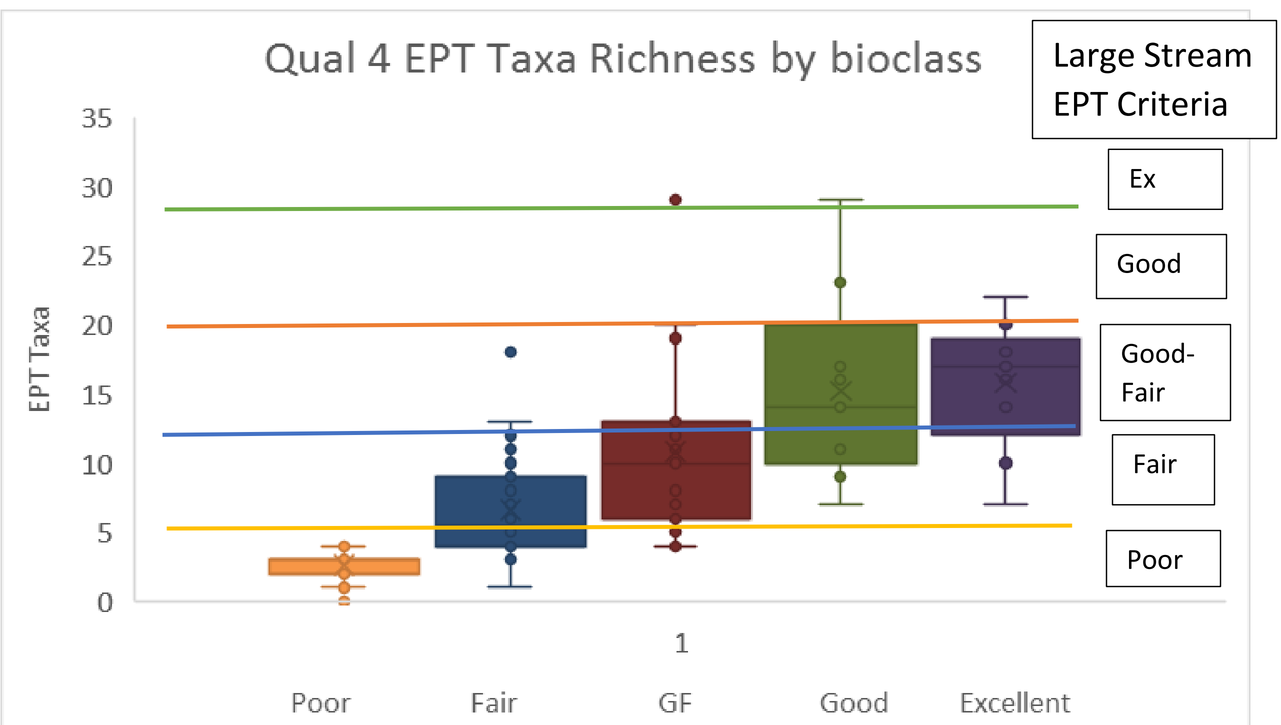

Figure 27. Small Stream Taxa Richness by Bioclass and EPT S Criteria for larger streams.

There are at least three ways to deal with this difference between metric reactions in large and small streams. First, one could develop site specific biocriteria based on the watershed size of the site and a maximum taxa line such as Figure 26. The largest problem with this approach is developing such a line for the Piedmont, where some level 4 ecoregions, such as the Slate Belt and Triassic Basins, have much greater lengths of intermittent streams than other ecoregions. Care would have to be taken that only perennial sites were sampled. Another problem would come in if one tried to compare sites of different sizes where a metric with a particular value may have different interpretations in a tiny stream versus a larger stream.

The second way to approach this is to accept the fact that in some instances, Excellent, Good and sometimes Good-Fair are inseparable and thus biocriteria should be developed only for Poor, Fair and Unimpaired waters. The third approach would be to try to normalize the metric to remove as many stream size differences as possible.

This third method is demonstrated in figures 28 and 29. Figure 28 is a ranking of the metric TV<4 for data from the towns of Chapel Hill, Carrboro and Durham by Biotic Index. Two sites with the highest TV<4 values, at Biotic Index values of 4.33 and 5.7, were from a stream rated Good or Excellent in different years with an 8 mi^2 watershed. The two sites with the lowest Biotic index values on the graph, were from an Excellent stream with a 0.2 mi^2 watershed. A line of best fit through this data showed an R^2 of 66%.

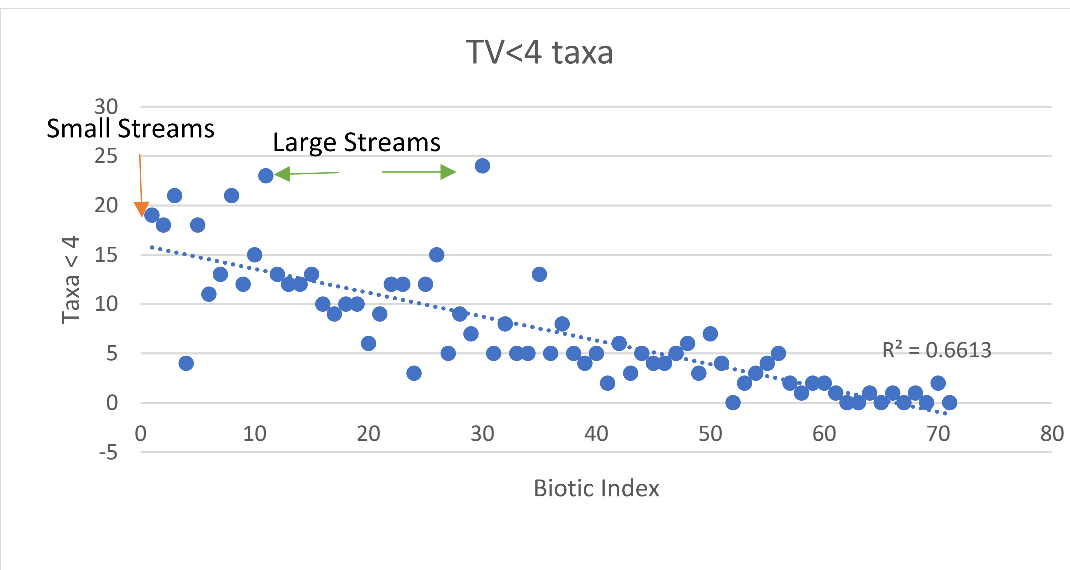

Figure 28. Taxa with TV<4 for streams in Piedmont towns.

Figure 29 are the same sites only the Taxa < TV4 metric has been divided by total taxa richness at the site to normalize for stream size. The result is a graph with much less variability, the R^2 is up to 74%, and thus an improved chance to develop criteria that are statistically significant. With an additional refinement, multiplying the normalized value by 100, the metric now becomes % Intolerant Taxa, which is a concept much more readily understood by non-scientists.

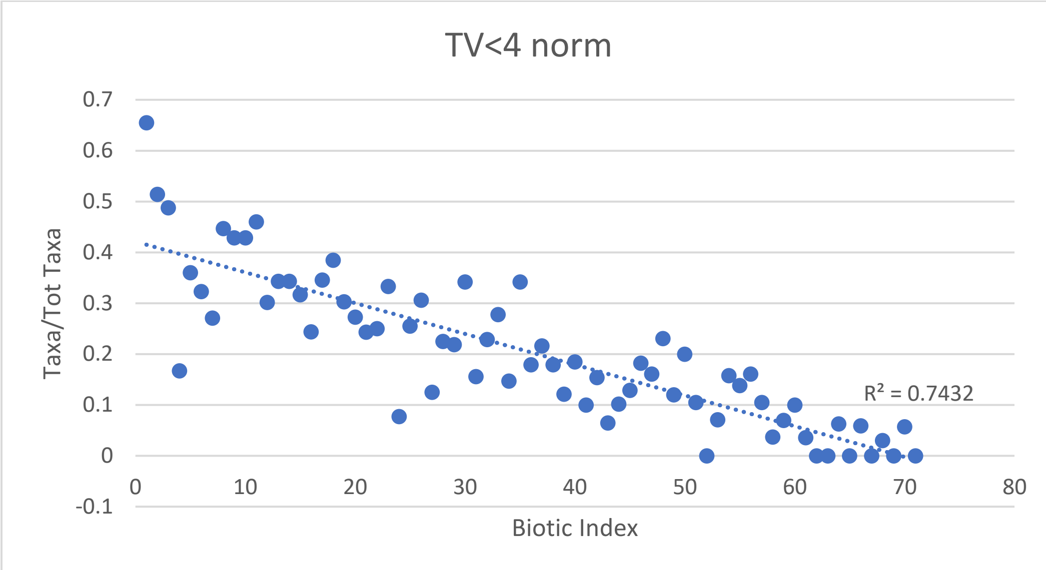

Figure 29. Taxa TV<4/Tot Taxa for streams in Piedmont towns.

Figures 30 and 31 are the % Intolerant Taxa metrics for the DWR data. In both the Mountains and the Piedmont ecoregions, criteria lines can be drawn between the bioclassifications at all five levels of biocriteria with mostly minimal misclassifications, though the Fair bioclass seems to deviate from the pattern in both the Piedmont and Mountain data.

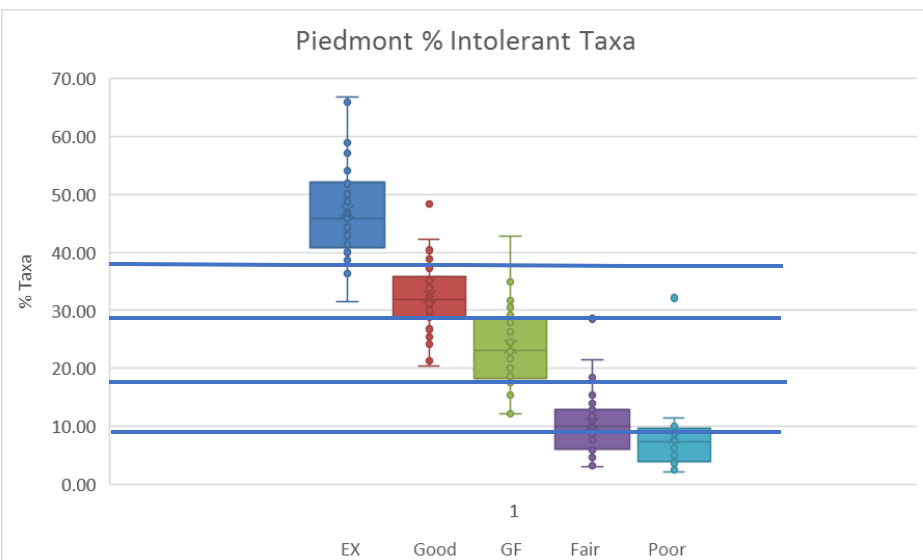

Figure 30. DWR Piedmont % Intolerant taxa by bioclass.

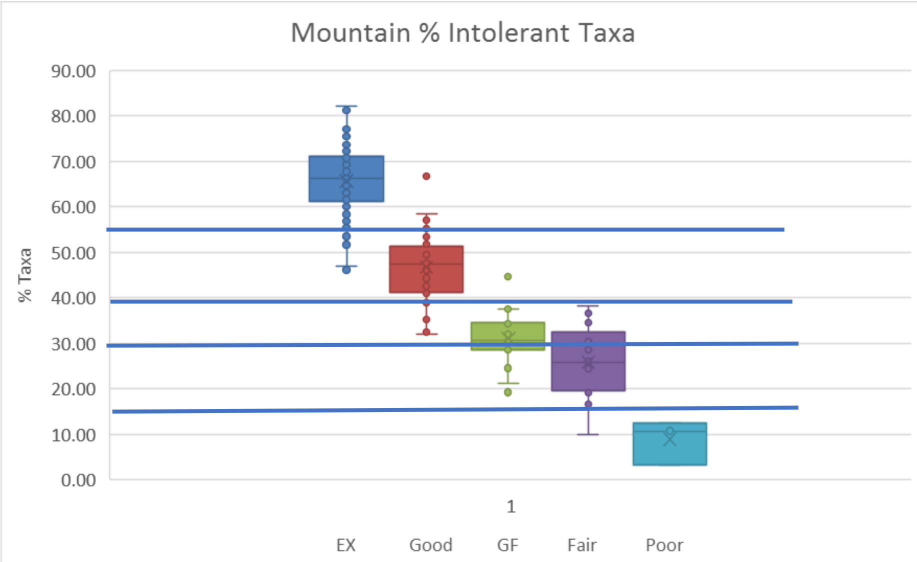

Figure 31. DWR Mountain % Intolerant taxa by bioclass.

Figures 32 and 33 are the CHCD data normalized for stream size for both TV4 and TV 6. Normalizing for stream size appeared to help in the comparison of cut off values between DWR data and CHCD data. Whereas the differences between TV<4 cutoffs increased as the water quality improved, all proposed cut offs of the % Intolerant Taxa metric were within a single taxon for all bioclassifications, except CHCD Good/Excellent, which still cannot be separated even after normalizing for stream size. One possible reason for this is the small number of observations of Excellent streams (9) is resulting in nearly twice the variability within this bioclass than any other in this data.

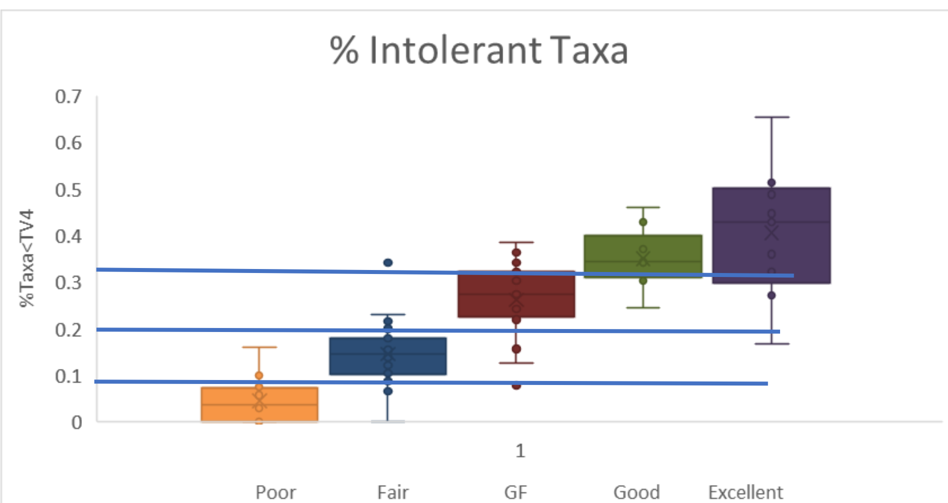

Figure 32. CHCD % Intolerant Taxa Piedmont

The Taxa with TV<6 metric was helped substantially by normalizing the data. Instead of a single clear cut off at Poor/Fair with the TV<6 metric, the normalized data yields a Poor/Fair cut off at 22%, Fair/Good-Fair at 35% and Good-Fair/Good-Excellent at 51%. The Excellent category still displays a large amount of variability that almost entirely includes the ranges of both Good and Good-Fair criteria, however.

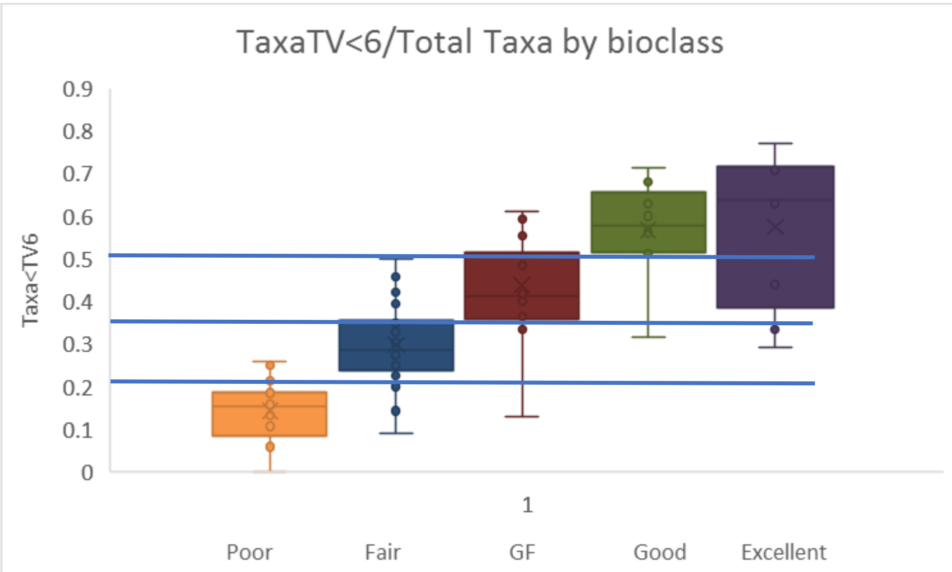

Figure 33. CHCD % Taxa <TV6

Proposed Biocriteria

Table 2, shows the proposed biocriteria for the metrics that were determined to have the best discriminatory power. While normalizing for stream size allowed five levels of biocriteria to be identified for % Intolerant Taxa, this normalization failed to separate Good from Excellent sites in the count metrics, EPT Taxa Richness, EPT Abundance and TV<4 Abundance, so criteria for these metrics were confined to Poor, Fair and Unimpaired (Good-Fair, Good and Excellent).

Table 2. Proposed biocriteria for most promising metrics for Mountain and Piedmont ecoregions.

	% Intolerant Taxa	
	MTN	Pied
Excellent	55+	40+
Good	41-55	30-40
Good-Fair	30-40	19-29
Fair	16-29	10-18
Poor	0-15	0-9

	EPT Abundance	
	MTN	Pied
Unimpaired	60+	45+
Fair	21-59	20-45
Poor	0-20	0-19

	EPT Taxa Richness	
	MTN	Pied
Unimpaired	15+	10+
Fair	8-15	5-9
Poor	0-7	0-4

	TV<4 Abundance	
	MTN	Pied
Unimpaired		26-45
Fair		6-25
Poor		0-5

Validation

To test proposed metrics, one needs a data set that was not part of the dataset used to develop the metrics. In this case, that validation dataset was 2018 data collected both from DWR, as part of the grant, and from local towns. The 19 Mountain and 7 Piedmont samples (26 total) collected by DWR were primarily resampling from stream restorations that were sampled five years ago and were part of a set of restorations with at least 10 years of post-restoration data. An additional 27 Piedmont samples were obtained from the towns of Chapel Hill, Carrboro and Durham for a total of 53 samples.

Samples were identified and assigned a bioclassification based solely on the Biotic Index, as per DWR Small Streams Criteria. Samples were then reassigned a bioclassification based on equally weighting the metrics Biotic Index and % Intolerant taxa with a tie breaker of EPT Abundance. EPT Abundance was selected over EPT Taxa Richness or TV<4 Abundance because 1) this metric is used as the tie breaker for Full Scale biocriteria and 2) it can capture cases where intolerant animals are drifting in from above a pollution source, such as an outfall, and skewing the Biotic Index downward.

Of the 26 Qual4 samples collected by DWR in 2018, these additional metrics did not change any bioclassifications. Of the 27 samples from the town data, five of the samples would have had their bioclassificatons downgraded. Overall, then, 90% of bioclassifications remained unchanged with the added metrics, so there would be few cases where streams would be going on or coming off the 303(d) list. All five downgrades in bioclassification were in streams with fairly low numbers of taxa with tolerance values (23, 27, 31, 32 and 35 taxa), which can often lead to increased variability of the Biotic Index, thus it would be difficult to declare the bioclassification shift as a misclassification on the part of the new metrics.

References

Adams S. M., W. R. Hill, M. J. Peterson, M. G. Ryon, J. G. Smith and A. J. Stewart. 2002. Assessing Recovery in a Stream Ecosystem: Applying Multiple Chemical and Biological Endpoints. Ecological Applications 12(5): 1510-1527.

Eaton, L.E. and D.R. Lenat. 1991. Comparison of a rapid bioassessment method with North Carolina's qualitative invertebrate collection method. J. N American Benthological Society 10(3): 335-338.

Harman, W. and C. Jones. 2016. Draft Functional Lift Quantification Tool for Stream Restoration Projects in North Carolina Spreadsheet User Manual. Environmental Defense Fund. Raleigh, NC. 40pp.

Gresens, S.E., K.T. Belt, J.A. Tang, D.C. Gwinn, P.A. Banks. 2007. Temporal and spatial responses of Chironomidae (Diptera) and other benthic invertebrates to urban stormwater runoff. Hydrobiologia 575:173-190.

Lenat, D.R. 1988. Water quality assessment of streams using a qualitative collection method for benthic macroinvertebrates. J. N American Benthological Society 7(3): 222-233.

Lenat, D.R. 1993. A biotic Index for the Southeastern United States: derivation and a list of tolerance values, with criteria for assigning water quality ratings. J. N American Benthological Society 12(3): 279-290.

Merritt, R.W, Cummings K.W. and Berg M.B. 2008. An Introduction to the Aquatic Insects of north America 4th edition. Kendall/Hunt Publishing Company. Dubuque, Iowa. 1158pp.

NC Division of Water Resources SOP https://ncdenr.s3.amazonaws.com/s3fs-public/Water%20Quality/Environmental%20Sciences/BAU/SOPBenthosSept_2012.pdf

NCDWQ. 2009. Biocriteria for the Small Streams of the North Carolina Mountains and Piedmont: Memorandum. NC Dept. of Environment and Natural Resources, Division of Water Quality. May 29, 2009.

Penrose, D.L. 2002. Benthic Macroinvertebrate Monitoring Protocols for Compensatory Stream Restoration Projects Version 3.0. North Carolina Division of Water Quality 401/Wetlands Unit. 46pp.

Penrose, D.L. 2004. Final Addendum. Ecological Functions of Restored Stream Systems Benthic Macroinvertebrates. North Carolina Division of Water Quality 401/Wetlands Unit. 25 pp.

Poff, N.L., Olden, J.D., Vieira, N.K.M., Finn, D.S., Simmons, M.P., and Kondratieff, B.C. 2006. Functional Trait Niches of North American lotic insects: Traits-based Ecological Applications in light of Phylogenetic Relationships. J. N. Am. Benthological Soc. 25(4):730-755.

Smith, R.F., R.J. Hawley, M.W. Neale, G.J. Vietz, E. Diaz-Pascacio, J. Hermann, A.C. Lovell, C. Prescott, B. Rios-Touma, B. Smith and R.M. Utz. 2016. Urban stream renovation: incorporation societal objectives to achieve ecological improvements. Freshwater Science 35(1): 364-379.

Sudduth, E.B., B.A. Hassett, P. Cada and E.S. Bernhardt. 2011. Testing the Field of Dreams Hypothesis: Functional Responses to Urbanization and Restoration in Stream Ecosystems. Ecological Applications 21: 1972-1988.

Tullos, D.D., D.L. Penrose, G.D. Jennings, W.G. Cope. 2009. Analysis of Functional Traits in Reconfigured Channels: Implications for the Bioassessment and Disturbance of River Restoration. J. N. American Benthological Society 28(1): 80-92.

Violin, C.R., P. Cada, E.B. Sudduth, B.A. Hassett, D.L. Penrose, E.S. Bernhardt. 2011. Effects of urbanization and urban stream restoration on the physical and biological structure of stream ecosystems. Ecological Applications 21(6): 1932-49.

http://www.saw.usace.army.mil/Portals/59/docs/regulatory/publicnotices/2013/NCSAM_Draft_User_Manual_130318.pdf

Appendix 1. Harman Quantification Metrics.

Functional Category	Function-Based Parameters	Measurement Method
Hydrology	Runoff	Storm EZ
	Flow Duration	NATHAT-DHRAM
	Catchment Hydrology	Catchment Assessment
Hydraulics	Floodplain Connectivity	Bank Height Ratio
		Entrenchment Ratio
Geomorphology	Large Woody Debris	LWD Index
	Lateral Stability	Erosion Rate (ft/yr)
		Dominant BEHI/NBS
		Percent Streambank Erosion (%)
	Riparian Vegetation	Canopy Coverage
		Basal Area
		Width
		Density
	Bed Material Characterization	Size Class Pebble Count Analyzer
	Bedform Diversity	Pool-Pool Spacing Ratio
		Pool Depth Ratio
		Percent Riffle
	Plan Form	Sinuosity
Physicochemical	Temperature	Upstream/Dwn Monitoring
	Nitrogen Loading	Falls Lake Nutrient Tool
	Phosphorus Loading	Falls Lake Nutrient Tool
	Specific Conductance	
	Bacteria Loading	Upstream/Dwn Monitoring
	Organic Matter	Leaf-Litter Processing Rate
	Stream Metabolism	Gross Primary Production
Biology	Macroinvertebrates	Biotic Index
	Fish	North Carolina Index of Biotic Integrity