Workshop to Develop a 2008 Baseline for the CBP Stream Health Outcome Indicator
Cacapon Resort State Park, WV
5th – 6th April 2018

Workshop Report

Executive Summary

Workshop participants recommended the family-level bioregion version of the Chessie BIBI index for Chesapeake Bay Program reporting purposes. The bioregion index is expected to be more attuned to natural landscape differences than the larger-scale regional index, and therefore may be able to detect improvement faster. A 6-year timeframe from 2006 to 2011 is recommended as the "2008 baseline." This timeframe overlaps the sampling designs of the monitoring programs represented at the workshop. Future time periods being compared to the 2006 – 2011 baseline should also have 6-year time spans for consistency. Workshop participants recommended using the Proportional Watershed method to communicate percentage of stream miles in the each of the five ratings (Excellent, Good, Fair, Poor, Very Poor). The method first weights each of the rating in smaller watersheds by the percentage of stations in the watershed with that rating, then apportions the percentage of each rating to the watershed's total stream miles and sums the results for the entire Chesapeake watershed. The workshop participants also recommended performing the closely related Bootstrap method to obtain error estimates around the rating percentages. Output from a predictive model (Maloney et al., in press) will be used in small watersheds with sparse or no stream biological data. The model uses twelve landscape variables to predict Chessie BIBI ratings with a high degree of accuracy. The mix of observed BIBI results supplemented with predicted results to estimate ratings everywhere in the Chesapeake watershed is not ideal, but it is currently the best solution for filling data gaps. Percentages of each Chessie BIBI rating for the baseline period are being calculated per the workshop recommendations. A follow-up conference call will be scheduled to review the results when they are available. Confidence in Chesapeake-wide trends may be low until a substantial dataset of many years can confirm short-term trends. Several additional recommendations were made: build consensus on what qualifies as 10% improvement; verify sample type (whether stations are randomly or systemically located versus targeted); refine poorly performing Chessie BIBI bioregion indices; perform power analyses; establish more sentinel sites; use CBP data protocols and the CEDR database; and develop other indicators of stream health.

Challenges

Higher quality streams are a desired outcome of the Chesapeake Bay's nutrient and sediment TMDL and many restoration and preservation efforts that can potential improve streams are underway in the watershed. A goal in the 2014 Chesapeake Bay Agreement is: *"Continually improve stream health and function throughout the watershed. Improve health and function of ten percent of stream miles above the 2008 baseline for the watershed."* The Chesapeake Bay Program (CBP) selected the "Chessie BIBI," or Chesapeake Basin-wide Index of Biotic Integrity, as its biological indicator of stream health. It will use the index to track and report progress towards accomplishing the stream health goal (Stream Health Management Strategy 2015). A "2008" baseline for the Chessie BIBI still needed to be established.

Several technical challenges had to be resolved first because the Chessie BIBI draws from a post hoc amalgamation of data from federal, state, county and citizen monitoring programs. The objectives, sampling frequencies, and coverage of these programs are not consistent across the Chesapeake watershed or across time. Several programs do not sample their entire sampling area in one year, and therefore have sampling designs that require multiple years to complete (e.g., state). Some sampling sites are randomly picked while others are revisited periodically. To manage these inconsistencies, the "2008" baseline must represent multiple years and have sufficient coverage to ensure statistical confidence in the results. Technical decisions had to be made on how to express results from individual monitoring sites in terms of stream miles and how to minimize geographic bias across the Chesapeake Bay watershed. Furthermore, there was no agreed-upon process or schedule for providing data to the CBP Data Center for Chesapeake-wide reporting. Finally, the question of what other stream health indicators could be developed—and whether their progress should be measured using the same baseline period—needed discussion.

A workshop was convened by ICPRB on April 5th-6th, 2018 at Cacapon State Park in Berkeley Springs, WV with the overarching goal of establishing a 2008 baseline. The workshop was attended by representatives of federal, state, and county government agencies, citizen monitors, and consultants. Many had contributed data to the BIBI database or analyzed the data for state-specific purposes. The first day, participants discussed the most appropriate method for reporting Chessie BIBI results across the Chesapeake Bay watershed in terms of stream miles. Six possible methods for doing this were explored. Five methods apply different statistical approaches to observed (monitoring) data and one method uses a predictive model that estimates the likely BIBI rating. Each method has strengths and weaknesses. The second day, participants discussed how CBP's Chesapeake Environmental Data Repository and its connections to the national Water Quality Exchange could facilitate efforts to maintain and update a shared stream monitoring database. Participants also explored other possible indicators of stream health in a discussion led by the CBP Stream Health Workgroup co-chairs.

Recommendations

The National River and Streams Assessment (NRSA) has implemented a national monitoring program and can provide spatially unbiased estimates of stream conditions in the Chesapeake watershed based on about 100 random samples collected every five years. This program does not address all the monitoring objectives of Chesapeake water resource agencies, who need information on finer scales. The compiled Chessie BIBI database can provide much of the needed information because it contains the agencies' own data. It presently contains data for over 21,000 sampling events applicable after data preparation steps for family-level index calculations. The database and the Chessie BIBI index can meet CBP's reporting need for unbiased stream health estimates if certain technical issues are resolved. The following eleven recommendations provide a way forward.

1) The "2008" baseline is the six-year sampling period from 2006 to 2011.

After reviewing the sampling designs of many of the data providers present at the workshop, it was recommended that a six-year sampling period, 2006 – 2011, be the 2008 baseline because it encompasses at least one round of each program's sampling design, some of which are rotational. Future time periods being compared to the 2008 baseline period should also represent a six-year time span to provide consistent temporal variability. Post-workshop follow-up will determine if a) a minimum of 5 data points is sufficient to represent the stream miles of a HUC12 watershed or if more points (e.g., 10) are needed, and b) if data from "targeted" stations can be included in evaluations of a HUC12 without biasing the overall result.

2) Use the bioregion-specific, family-level Chessie BIBI indices to establish the baseline and to monitor trends.

Family-level region indices were recommended in Smith et al. (2017) for assessing the Chesapeake Bay watershed because they provided similar results to the family-level bioregion indices and represented a simpler assessment approach. However, the workshop participants preferred to use the bioregion indices for establishing a baseline and monitoring trends. Bioregion indices may be more sensitive to natural landscape differences, and therefore may detect change faster than the region indices. Workshop participants had several recommendations for future development of the Chessie BIBI and steps to enhance the usefulness and availability of stream datasets. Each of these recommended tasks requires additional effort and cannot be appended onto current efforts without additional funds or staff time.

3) Use the Proportional Watershed method of analyzing observed (monitoring) data to determine status.

Of the six methods explored, workshop participants recommended using the Proportional Watershed method for communication purposes because it is an intuitive way to assess data. The method first weights each rating in small watersheds (defined by HUC-12 and bioregion) by the percentage of stations in the watershed with that rating, then apportions the percentage of each rating to the watershed's total stream miles and sums the results for the 2,470 such watersheds across the entire Chesapeake watershed. (The Chesapeake has 1,967 HUC-12 watersheds; some are bisected by a bioregion boundary, which creates 2,470 HUC-12/bioregion watersheds.) Workgroup participants also recommended applying the similar Bootstrap method to the observed data because it provides an estimate of error, which is useful when measuring the statistical significance of trends. Aggregating and analyzing the observed data by the relatively small HUC-12/bioregion watersheds reduces but does not eliminate spatial bias due to uneven sampling densities. However, results in sparsely sampled regions are less likely to have undue influence on the overall result.

4) Use output from a predictive (Random Forest) model to fill gaps in the observed data.

Although the Proportional Watershed and Bootstrap analysis methods reduce spatial bias, problematic biases still can occur in the observed data. These issues are driven mostly by data gaps. In the 2006 – 2011 baseline period, approximately 1,145 (46.4%) of the 2,470 HUC-12/bioregion watersheds have fewer than 5 data points and would need to be reported as having insufficient data to determine status. To mitigate the gaps, results of a predictive (Random Forest) model developed by USGS staff in consultation with ICPRB (Maloney et al. in press) will be applied. Output from the predictive model will be used for HUC-12/bioregion watersheds that do not contain enough observed points to be accurately represented with the Proportional watershed or Bootstrap method. This mixed monitoring/modeling representation of the watershed is not ideal but it is currently the best recommendation for resolving the spatial biases. The proportions of stream miles with Excellent, Good, Fair, Poor and Very Poor ratings of the Chessie BIBI in the 2006 – 2011 baseline period will be estimated using this approach in a follow-up analysis by ICPRB and USGS.

5) Build consensus on what qualifies as 10% improvement

A consensus is needed in the Stream Health Workgroup on what qualifies as 10% improvement in stream miles. Workshop participants noted that an immediate, measurable improvement in biological communities does not always occur in severely degraded streams, even after substantial efforts to improve in-stream habitat or water quality. The merits of measuring progress by the overall change in the proportions of the five Chessie BIBI rating categories versus change in actual Chessie BIBI index scores (0 - 100) were discussed. The former requires BIBI ratings in 10% of all stream miles to improve by at least one rating unit (e.g., Poor to Fair, Fair to Good). The latter requires an overall 10% of stream miles to show an improvement in actual BIBI index scores. Index scores are expected to be more sensitive to stream improvements because they can register changes of 10% in the Very Poor and Poor ratings, each of which spans more than 10 points on the 0 – 100 point index scale. However, index scores would be inherently more variable over time, which would make it harder to confirm that lasting improvements occurred.

6) Verify sample type

Smith et al. (2017) questioned the accuracy of "sample type" identifications (probabilistic or targeted) in some of the previously submitted stream datasets. Stations intentionally co-located with a known pollution source (targeted) are excluded from calculations of stream health, or status; stations that were initially located randomly or systematically are included, regardless of whether they are revisited later. Errors in sample type identification affect the composition of the dataset used to calculate stream health. Therefore, accurate assignments of sample type are required to eliminate this potential bias. Several of the data providers in attendance at the workshop have already sent ICPRB tables containing the correct sample type assignments. An additional call will be made by ICPRB to verify or update the sample type information currently contained in the Chessie BIBI database. ICPRB recommends that the updated sample type information be collectively reviewed and verified by the data providers.

7) Refine poorly performing Chessie BIBI bioregion indices

Several of the family-level bioregion indices selected for CBP reporting purposes are known to perform poorly (Smith et al. 2017). Workshop participants agreed the indices for the Mid-Atlantic Coast (MAC), the Central Appalachians (CA), Northern Appalachian Plateau and Uplands (NAPU), and the Blue Ridge Mountains (BLUE) bioregions are comparatively weak and could be improved. The MAC, CA, and NAPU bioregion indices were developed using only data collected within the Chesapeake watershed. The Level III ecoregions that define these bioregions extend beyond the Chesapeake watershed. Incorporating data from outside the watershed would increase sample sizes available for index refinement and should lead to more sensitive, robust index performance. Workgroup participants concurred with ICPRB's recommendation that indices for the MAC, CA, and NAPU bioregions could be refined with larger datasets. Datasets used to develop the BLUE bioregion index had only seven sites that classified as degraded. This limited the ability to test the index's classification efficiency. Incorporating data from additional degraded sites in the BLUE bioregion, if they exist, should improve that index's sensitivity.

In sidebar conversations and at the workshop, a strong interest was expressed by representatives of several monitoring programs (New Jersey DEP, Delaware DNREC, Virginia DEQ, Fairfax Co. WAB) in collaboratively revisiting the Maxted et al. (2000) coastal plain indices. The objective is to refine the indices using a larger, updated dataset that spans the coastal plain region from New Jersey to South Carolina. A proposal to that effect has been drafted by ICPRB.

8) Perform power analyses

The Chessie BIBI database already contains a large number of benthic macroinvertebrate data (n > 21,000 sampling events), but workshop participants recommended performing a power analysis at various levels to verify there is enough data available to make statistically valid assessments. The HUC-12/bioregion watershed was recommended as the spatial unit for analyzing point data with the Proportional Watershed and Bootstrap methods (see above). A power analysis could be used to inform analysts on how many points are required to accurately represent a HUC-12/bioregion unit. After the points have been aggregated to the HUC-12/bioregion level, then another power analysis could be performed to identify the number of HUC-12/bioregion units need to accurately represent stream condition for the Chesapeake Bay basin. At this time, there are no resources to perform these analyses.

9) Establish more sentinel sites

As discussed above, overall stream conditions across the Chesapeake watershed in a specific period (e.g., the 2006 – 2011 baseline) are best estimated from randomly or systematically located sampling sites. However, trends are difficult to determine from these data. Sites that are periodically revisited (sentinel sites) are more suited for trend analysis. Most sampling sites currently in the Chessie BIBI database are not revisited. If a subset of these sites were to be periodically revisited, they would provide a more accurate method for determining stream health trends in the Chesapeake watershed. The small number of current sentinel sites, their uneven spatial distribution, and their poor representation of the whole stream disturbance gradient (most are reference sites) limits our ability to use them to measure trends for the Chesapeake Bay basin. Each monitoring agency has its own sampling design and the data collected at the sentinel sites are not standardized across the Chesapeake watershed (i.e., frequency and sampling period differ between data providers). A mixed model could be used to evaluate trends but to accurately represent the basin, and especially the stream disturbance gradient, more sentinel sites are required. The recommendation from the workshop participants is to identify the available sentinel site data, identify which sentinel sites are expected to be monitored in the future, and establish new sentinel site locations to fill in areas under-sampled in the basin. New sentinel sites should also encompass the range of the stream disturbance gradient. At this time, there are no resources to perform these steps.

10) Use CBP data protocols and the CEDR database

In the past, ICPRB performed the arduous process of acquiring (and sometimes computerizing) raw stream data from the data providers, quality assuring the data, and incorporating them into a standard (CBP) database structure. The CBP Data Center has nearly completed development of more streamline, semi-automated protocols for receiving and processing stream macroinvertebrate and habitat data into the Chesapeake Environmental Data Repository (CEDR). Data providers obtain several benefits using the CBP data protocols and CEDR database: (1) the process guarantees that all the data used to calculate the Chessie BIBI, and potentially other indices (e.g., state-specific macroinvertebrate IBIs), have gone through a standard QA and documentation process; (2) the process allows data providers to submit data on their own schedules; (3) if desired by the data provider, datasets from the CEDR database can be directly and automatically uploaded to the national Water Quality Exchange (WQX); (4) the large, diverse collection of Chesapeake stream datasets is readily available for future data analysis to guide multi-pronged restoration and resource management decisions. CBP and ICPRB recommend that monitoring agencies in the Chesapeake watershed use the CEDR protocols and data systems to help manage their stream data and address the 2014 Chesapeake Agreement goal of tracking and improving stream health.

11) Other indicators of stream condition

Participants explored other possible indicators of stream condition in a discussion led by the CBP Stream Health Workgroup co-chairs. Indicators of in-stream habitat, water quality, and riparian land use are valid ways to measure aspects of stream health. They can be combined into multi-metric indices to assess stream corridor conditions, e.g., the Upper Susquehanna Coalition's visual assessment tool to evaluate stream corridor functional health and flooding resiliency. Land use and other "far-field" metrics can also indicate stream conditions. Non-biological indicators are likely to respond to restoration activities (improve) faster than biological measures. For example, trees planted stream-side to increase the riparian buffer width are immediately recognizable as a stream habitat improvement. However, this activity does not necessarily result in immediate improvements in biology. Biology in degraded streams is expected to lag until more aspects of stream condition have improved. Workshop participants agreed that a suite of stream condition indicators would be useful while recognizing that biology provides the most robust indication of overall stream health. Discussions of how to develop and incorporate other indicators of stream condition are expected to continue in the Stream Health Workgroup.

Workshop Follow-up

Workshop participants agreed to future teleconferences/webinars to review follow-up products and make additional recommendations. ICPRB, USGS, the Stream Health Workgroup, and the CBP Data Center will work with data providers to implement the various recommendations where possible.

Workshop participants recognized that local and state managers are likely to use their existing stream macroinvertebrate indices for planning and reporting (e.g., CWA Integrated Reports). The Chessie BIBI is intended to provide a broad view of stream health across the Chesapeake Bay watershed, and assist in planning in large regions. With respect to CBP communication and reporting, workshop participants suggested a three-tiered approach relying on products with different information levels:

- Big picture (e.g., HUC12-based maps, pie-charts of the percentages of the five ratings for Chesapeake watershed)
- Diagnostic
- Case studies (e.g., watershed-based synthesis to highlight improvement and why it happened)

Workshop participants anticipate that data collection efforts by citizen scientists and nontraditional partners can assist in areas with insufficient data and help track changes over time at sentinel sites. These efforts will not contribute to establishment of the stream health 2008 baseline. However, filling in the data gaps removes the necessity for model estimates of IBI conditions in the future and coincidentally provides more data to help calibrate and improve model performance.

Link to Workshop Materials

The April 5-6, 2018, workshop background materials and presentations are available here.

Reference

Maloney, K., Z. M. Smith, C. Buchanan, A. Nagel, and J. A. Young. (in press) Predicting Stream Biological Conditions for Small Headwater Streams in the Chesapeake Bay Watershed. Freshwater Science.

Maxted, J. R., et al. 2000. Assessment framework for mid-Atlantic coastal plain streams using benthic macroinvertebrates. Journal of the North American Benthological Society 19:128–144.

Smith, Z. M., C. Buchanan, and A. Nagel. 2017. Refinement of the Basin-Wide Benthic Index of Biotic Integrity for Non-Tidal Streams and Wadeable Rivers in the Chesapeake Bay Watershed. ICPRB Report 17-2, Interstate Commission on the Potomac River Basin, Rockville, MD.