Decision Process for Identification of Estuarine Benthic Impairments in Chesapeake Bay, USA

R. J. Llansó*, J. H. Vølstad
Versar, Inc., Columbia, Maryland
and
D. M. Dauer
Old Dominion University, Norfolk, Virginia

*llansorob@versar.com
Context

- States of Maryland and Virginia share the Chesapeake Bay and its tributaries
- Need to integrate monitoring and assessment efforts for reporting 303(d) impairment decisions under Clean Water Act
States of Maryland and Virginia share the Chesapeake Bay and its tributaries

Need to integrate monitoring and assessment efforts for reporting 303(d) impairment decisions under Clean Water Act

Integration underway for both
  - Freshwater streams
  - Chesapeake Bay estuarine waters
Context

- Integration issues include
  - Comparability of sampling methods
  - Comparability of indicators of condition (e.g., indices of biotic integrity)
  - Consistency in overall assessments and designation of impaired waters on 303(d) list
Context

Freshwater streams

- Maryland has biocriteria (based on Maryland Biological Stream Survey) supporting 303d listings
- Maryland and Virginia have different indicators, but comparability study is underway
Context

Freshwater streams

- Maryland has biocriteria (based on Maryland Biological Stream Survey) supporting 303d listings
- Maryland and Virginia have different indicators, but comparability study is underway

Chesapeake Bay

- Same sampling methods and indicator used by both states
- Need consistent method for impairment decisions

➤ Today’s presentation
Chesapeake Bay Benthic Monitoring Program

Restoration Goals

Benthic Index of Biotic Integrity

Survey of Condition (Status)

Probability Survey Design

Framework for application of B-IBI to the States’ water quality inventories

303(d) Lists
Benthic Index of Biotic Integrity

- Multi-metric, habitat-specific index of benthic community condition
- Selection of metrics and the values for scoring metrics developed separately for each of seven benthic habitat types in Chesapeake Bay

---

Objectives

- Develop a procedure for 303(d) impairment decisions based on the B-IBI
- Produce an assessment of Chesapeake Bay segments
Alternative approaches for 303(d) impairment decisions*

- Weighted mean approach
- Comparisons of cumulative frequency distributions and proportions

*using B-IBI scores
**Weighted mean approach**

<table>
<thead>
<tr>
<th>Reference</th>
<th>Segment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
</tr>
<tr>
<td>Hab1</td>
<td>4.1</td>
</tr>
<tr>
<td>Hab2</td>
<td>3.1</td>
</tr>
<tr>
<td>Hab3</td>
<td>3.5</td>
</tr>
<tr>
<td>Hab 1-3</td>
<td>3.56</td>
</tr>
</tbody>
</table>

*SE of the weighted mean

Example provided by Florence Faulk, US EPA ORD
Weighted mean approach

- One-sided t-test, the difference in weighted means divided by the pooled standard error

\[
t = \frac{\bar{X}_r - \bar{X}_s}{SE_p} = \frac{3.56 - 2.16}{0.461} = 3.04 > t_{0.05,18}
\]
Cumulative frequency distribution approach
Cumulative frequency distribution approach

$H_0: P_s = P_{ref}$

$H_A: P_s > P_{ref}$

$H_0: P_s - P_{ref} > 0.25$
Reference frequency distribution comparison among habitats

<table>
<thead>
<tr>
<th>Habitat Class</th>
<th>TF</th>
<th>OL</th>
<th>LM</th>
<th>HS</th>
<th>HM</th>
<th>PS</th>
<th>PM</th>
</tr>
</thead>
<tbody>
<tr>
<td>TF</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>OL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>LM</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HS</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HM</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PS</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PM</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Kolmogorov-Smirnov 2-sided test, X = p<0.05
Which method to use?

Cumulative frequency distributions

- Not appropriate to pool reference distributions across habitats if the distributions differ
Which method to use?

Cumulative frequency distributions

- Not appropriate to pool reference distributions across habitats if the distributions differ
- Tests based on exact binomial distributions such as Fisher’s exact test not valid for stratified data
Which method to use?

Cumulative frequency distributions

- Not appropriate to pool reference distributions across habitats if the distributions differ
- Tests based on exact binomial distributions such as Fisher’s exact test not valid for stratified data

Weighted means

- Parametric test problematic for small sample size
Which method to use?

Cumulative frequency distributions

- Not appropriate to pool reference distributions across habitats if the distributions differ
- Tests based on exact binomial distributions such as Fisher’s exact test not valid for stratified data

Weighted means

- Parametric test problematic for small sample size
- Weights based on estimated proportion of each habitat
Which method to use?

Cumulative frequency distributions

- Not appropriate to pool reference distributions across habitats if the distributions differ
- Tests based on exact binomial distributions such as Fisher’s exact test not valid for stratified data

Weighted means

- Parametric test problematic for small sample size
- Weights based on estimated proportion of each habitat
- Does not measure areal extent of degradation
Frequency distribution approach using a stratified Wilcoxon rank sum test

- Test is robust even when small and unbalanced stratified data sets are used
- Can control for Type I and Type II errors
- Implemented with StatXact
Reference data set

- 243 Chesapeake Bay B-IBI development samples\(^1\)

\(^1\)Weisberg et al. 1997, *Estuaries* 20:149-158
Assessment data set

• Chesapeake Bay long-term benthic monitoring program 1998-2002 random samples:
  • Maryland, 750
  • Virginia, 500
  • Elizabeth River, 275
• 90 segments (including Virginia sub-segmentation)
Segmentation

• Assessments produced for each of 90 Chesapeake Bay Program segments and sub-segments containing benthic data
Segmentation

- Assessments produced for each of 90 Chesapeake Bay Program segments and sub-segments containing benthic data
- Segments are Chesapeake Bay regions having similar salinity and hydrographic characteristics
Segmentation

- Assessments produced for each of 90 Chesapeake Bay Program segments and sub-segments containing benthic data

- Segments are Chesapeake Bay regions having similar salinity and hydrographic characteristics

- In Virginia, segments were sub-divided into smaller units (sub-segments) to separate tributaries with no observed violations of water quality standards
Standardized classifications of B-IBI scores across habitats

- Maximum possible number of B-IBI scores differ by habitat
- B-IBI scores were classified into ordered response categories (‘condition categories’)

### Condition categories

<table>
<thead>
<tr>
<th>Condition Category</th>
<th>B-IBI Score</th>
<th>Benthic Community Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.0-2.0</td>
<td>Severely degraded</td>
</tr>
<tr>
<td>2</td>
<td>2.1-2.9</td>
<td>Degraded</td>
</tr>
<tr>
<td>3</td>
<td>3.0-5.0</td>
<td>Meets goal</td>
</tr>
</tbody>
</table>
Comparing B-IBI scores from segments and reference distributions

- Segment and reference scores represent two independent ordered multinomial distributions
- Test if the two populations have the same underlying multinomial distribution of B-IBI scores by condition category
Hypothesis test

- Stratified Wilcoxon rank sum test
- Question: Does segment have lower B-IBI scores than reference?
- One-sided Test:

\[ H_0: \text{Equal multinomial distributions} \]
\[ H_1: \text{Shift in location toward lower B-IBI responses in segment than in reference} \]
Type I and Type II errors

- Critical alpha level of 1% will be applied to test for impairment
- Only segments where power is $\geq 90\%$ and $p<0.01$ will be listed
- Minimum sample size for assessment of segment is $n \geq 10$ (same as for freshwater streams)
Results of assessment

- 26 of 90 Chesapeake Bay segments were considered degraded based on the B-IBI and identified as impaired under Section 303(d) of the Clean Water Act

York River polyhaline

Kilometers
## List of impaired segments

<table>
<thead>
<tr>
<th>Segment</th>
<th>Name</th>
<th>Sample size</th>
<th>Weighted P less then 3.0</th>
<th>Deg</th>
<th>Seg-Ref</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Seg</td>
<td>Ref</td>
<td>Deg</td>
</tr>
<tr>
<td>SBEMHa</td>
<td>Southern Branch Elizabeth River</td>
<td>116</td>
<td>0.93</td>
<td>0.04</td>
<td>0.99</td>
</tr>
<tr>
<td>EBEMHa</td>
<td>Eastern Branch Elizabeth River</td>
<td>32</td>
<td>0.88</td>
<td>0.08</td>
<td>0.98</td>
</tr>
<tr>
<td>WBEMHa</td>
<td>Western Branch Elizabeth River</td>
<td>39</td>
<td>0.82</td>
<td>0.04</td>
<td>0.99</td>
</tr>
<tr>
<td>POTMH</td>
<td>Potomac mesohaline</td>
<td>98</td>
<td>0.81</td>
<td>0.09</td>
<td>0.94</td>
</tr>
<tr>
<td>LAFMHa</td>
<td>Lafayette River</td>
<td>35</td>
<td>0.77</td>
<td>0.06</td>
<td>0.99</td>
</tr>
<tr>
<td>CB4MH</td>
<td>Maryland mainstem</td>
<td>30</td>
<td>0.73</td>
<td>0.09</td>
<td>0.98</td>
</tr>
<tr>
<td>PATMH</td>
<td>Patapsco River</td>
<td>45</td>
<td>0.69</td>
<td>0.07</td>
<td>0.89</td>
</tr>
<tr>
<td>YRKMHa</td>
<td>York River mesohaline</td>
<td>66</td>
<td>0.64</td>
<td>0.07</td>
<td>0.98</td>
</tr>
<tr>
<td>POCMH</td>
<td>Pocomoke River</td>
<td>11</td>
<td>0.64</td>
<td>0.07</td>
<td>0.99</td>
</tr>
<tr>
<td>RPPMH</td>
<td>Rappahannock River mesohaline</td>
<td>96</td>
<td>0.60</td>
<td>0.08</td>
<td>0.95</td>
</tr>
<tr>
<td>ELIMHa</td>
<td>Elizabeth River mesohaline</td>
<td>36</td>
<td>0.56</td>
<td>0.03</td>
<td>0.99</td>
</tr>
<tr>
<td>CB5MH</td>
<td>Maryland mainstem</td>
<td>46</td>
<td>0.57</td>
<td>0.06</td>
<td>0.99</td>
</tr>
<tr>
<td>JMSMHa</td>
<td>James River mesohaline</td>
<td>40</td>
<td>0.55</td>
<td>0.05</td>
<td>0.93</td>
</tr>
<tr>
<td>YRKKPHa</td>
<td>York River polyhaline</td>
<td>27</td>
<td>0.52</td>
<td>0.03</td>
<td>0.99</td>
</tr>
<tr>
<td>POTOH</td>
<td>Potomac River oligohaline</td>
<td>15</td>
<td>0.60</td>
<td>0.12</td>
<td>0.72</td>
</tr>
<tr>
<td>PAXMH</td>
<td>Patuxent River mesohaline</td>
<td>108</td>
<td>0.57</td>
<td>0.10</td>
<td>0.95</td>
</tr>
<tr>
<td>MAGMH</td>
<td>Magothy River</td>
<td>20</td>
<td>0.55</td>
<td>0.08</td>
<td>0.91</td>
</tr>
<tr>
<td>JMSOHa</td>
<td>James River oligohaline</td>
<td>29</td>
<td>0.55</td>
<td>0.13</td>
<td>0.75</td>
</tr>
<tr>
<td>GUNOH</td>
<td>Gunpowder River</td>
<td>10</td>
<td>0.50</td>
<td>0.09</td>
<td>0.75</td>
</tr>
<tr>
<td>TANMH</td>
<td>Tangier Sound</td>
<td>38</td>
<td>0.45</td>
<td>0.06</td>
<td>1.00</td>
</tr>
<tr>
<td>CB3MH</td>
<td>Maryland mainstem</td>
<td>55</td>
<td>0.48</td>
<td>0.10</td>
<td>0.89</td>
</tr>
<tr>
<td>CHOMH2</td>
<td>Choptank River</td>
<td>14</td>
<td>0.43</td>
<td>0.07</td>
<td>0.88</td>
</tr>
<tr>
<td>NANMH</td>
<td>Nanticoke River</td>
<td>11</td>
<td>0.45</td>
<td>0.09</td>
<td>0.87</td>
</tr>
<tr>
<td>CHSMH</td>
<td>Chester River</td>
<td>35</td>
<td>0.43</td>
<td>0.08</td>
<td>0.92</td>
</tr>
<tr>
<td>ELIPHa</td>
<td>Elizabeth River polyhaline</td>
<td>25</td>
<td>0.36</td>
<td>0.04</td>
<td>0.99</td>
</tr>
<tr>
<td>CB7PHa</td>
<td>Virginia mainstem</td>
<td>41</td>
<td>0.20</td>
<td>0.03</td>
<td>1.00</td>
</tr>
</tbody>
</table>
Segment CBP7PHa (Virginia mainstem)

- Listing of this segment as impaired is problematic, 80% of all B-IBI scores in the segment \( \geq 3.0 \)
- Shift in distribution for pooled (un-stratified) data was 0.33 B-IBI units
Limitations of current approach

- Stratified Wilcoxon rank sum test may be too sensitive (detects significant differences for small shifts)
Limitations of current approach

- Stratified Wilcoxon rank sum test may be too sensitive (detects significant differences for small shifts)
- It is not possible to estimate the magnitude of the shift in location (e.g., with a Hodges-Lehman confidence interval) for stratified data
Limitations of current approach

- Stratified Wilcoxon rank sum test may be too sensitive (detects significant differences for small shifts)
- It is not possible to estimate the magnitude of the shift in location (e.g., with a Hodges-Lehman confidence interval) for stratified data
- For stratified data, it is not possible to evaluate power for a range of sample sizes
Limitations of current approach

- Stratified Wilcoxon rank sum test may be too sensitive (detects significant differences for small shifts)
- It is not possible to estimate the magnitude of the shift in location (e.g., with a Hodges-Lehman confidence interval) for stratified data
- For stratified data, it is not possible to evaluate power for a range of sample sizes
- Reference sites are “best of the best”, and may not be representative of typical distribution of scores for good condition
How is this approach used by the States to evaluate aquatic life use support?
Score sample → Data sufficient? Yay → Test segment
Score sample → Data sufficient? → Test segment → Is segment impaired for DO numeric criteria?

Is segment degraded for B-IBI? → YES → YES

Data sufficient? → YES
Score sample

Data sufficient? 

Test segment

Is segment degraded for B-IBI?

Is segment impaired for DO numeric criteria?

Aquatic life fails
Cause: DO
B-IBI corroborative

Develop TMDL to correct low DO

DO corrected

YES

YES

YES
Score sample

Data sufficient?

Test segment

Is segment degraded for B-IBI?

YES

Is segment impaired for DO numeric criteria?

YES

Aquatic life fails
Cause: DO
B-IBI corroborative

Develop TMDL to correct low DO

DO corrected

NO

Evaluate B-IBI for other stressors

Other stressors identified?

YES NO
Score sample

Data sufficient? YES Test segment

Is segment degraded for B-IBI? YES

Is segment impaired for DO numeric criteria? YES

Aquatic life fails Cause: DO B-IBI corroborative

Develop TMDL to correct low DO

DO corrected

Aquatic life fails Cause: Pollutants B-IBI corroborative

Develop TMDL to correct pollutants

Pollutants corrected

Evaluate B-IBI for other stressors

Other stressors identified? YES

YES

NO
Score sample

Test segment

Is segment impaired for DO numeric criteria?

Is segment degraded for B-IBI?

Evaluate B-IBI for other stressors

Other stressors identified?

Aquatic life fails
Cause: DO
B-IBI corroborative

Develop TMDL to correct low DO

DO corrected

Aquatic life fails
Cause: Pollutants
B-IBI corroborative

Develop TMDL to correct pollutants

Pollutants corrected

Aquatic life fails
Cause: Pollution
Unknown source

No TMDL required

Data sufficient?
Score sample

Data sufficient?

Test segment

Is segment impaired for DO numeric criteria?

Aquatic life fails
Cause: DO
B-IBI corroborative

Develop TMDL to correct low DO

DO corrected

Is segment degraded for B-IBI?

Evaluate B-IBI for other stressors

Does segment meet WQ criteria?

Other stressors identified?

Aquatic life fails
Cause: Pollutants
B-IBI corroborative

Develop TMDL to correct pollutants

Pollutants corrected

Aquatic life fails
Cause: Pollution
Unknown source

No TMDL required
Score sample

Data sufficient?

Test segment

Is segment degraded for B-IBI?

YES

Is segment impaired for DO numeric criteria?

YES

Aquatic life fails
Cause: DO
B-IBI corroborative

Develop TMDL to correct low DO

DO corrected

NO

Aquatic life fails
Cause: Pollutants
B-IBI corroborative

Evaluate B-IBI for other stressors

Other stressors identified?

YES

Aquatic life fails
Cause: Pollution
Unknown source

Develop TMDL to correct pollutants

Pollutants corrected

NO

Aquatic life supported

Does segment meet WQ criteria?

YES

NO

No TMDL required
Score sample

Test segment

Data sufficient?

Is segment degraded for B-IBI?

Yes

Is segment impaired for DO numeric criteria?

Yes

Aquatic life fails
Cause: DO
B-IBI corroborative

Develop TMDL to correct low DO

DO corrected

Aquatic life fails
Cause: Pollutants
B-IBI corroborative

Evaluate B-IBI for other stressors

Other stressors identified?

Yes

Aquatic life fails
Cause: Pollutants
B-IBI corroborative

Develop TMDL to correct pollutants

Pollutants corrected

No

Aquatic life fails
Cause: DO, etc.

Aquatic life supported

NO

Does segment meet WQ criteria?

NO

Aquatic life fails
Cause: Pollution
Unknown source

No TMDL required

YES

YES

YES

NO

NO

YES

NO
Score sample

Data sufficient?

Test segment

Is segment impaired for DO numeric criteria?

 YES

Is segment degraded for B-IBI?

 YES

Evaluate B-IBI for other stressors

 NO

Does segment meet WQ criteria?

 NO

Insufficient data

 Aquatic life supported

 Aquatic life fails Cause: DO, etc.

 YES

Other stressors identified?

 NO

 Aquatic life supported

 Aquatic life fails Cause: DO, etc.

 YES

Aquatic life fails Cause: Pollutants B-IBI corroborative

Develop TMDL to correct pollutants

Pollutants corrected

 Aquatic life fails Cause: Pollutants B-IBI corroborative

Develop TMDL to correct low DO

DO corrected

 NO

 Aquatic life fails Cause: Pollutants B-IBI corroborative

Develop TMDL to correct pollutants

Pollutants corrected

 NO

Aquatic life fails Cause: Pollution Unknown source

 No TMDL required
Score sample

Data sufficient?

Test segment

Is segment impaired for DO numeric criteria?

YES

NO

Is segment degraded for B-IBI?

YES

NO

Evaluate B-IBI for other stressors

YES

NO

Other stressors identified?

YES

NO

Aquatic life fails Cause: Pollutants B-IBI corroborative

Develop TMDL to correct pollutants

Pollutants corrected

Aquatic life fails Cause: Pollution Unknown source

No TMDL required

Aquatic life supported

Does segment meet WQ criteria?

YES

NO

Insufficient data

Additional monitoring information needed

Aquatic life unknown

NO

YES

Aquatic life fails Cause: DO

B-IBI corroborative

Develop TMDL to correct low DO

DO corrected

NO
What’s next?

- Research into alternative methods
  - Ray Alden et al. confidence limit approach\(^1\)

What’s next?

- Research into alternative methods
  - Ray Alden et al. confidence limit approach\(^1\)
- Develop methods that take into account magnitude of difference between segment and reference distribution

What’s next?

• Research into alternative methods
  • Ray Alden et al. confidence limit approach¹

• Develop methods that take into account magnitude of difference between segment and reference distribution

• Diagnose causes of benthic community degradation (See Dauer’s presentation, Thursday 4:30-5:00)

What’s next?

• Research into alternative methods
  - Ray Alden et al. confidence limit approach\(^1\)

• Develop methods that take into account magnitude of difference between segment and reference distribution

• Diagnose causes of benthic community degradation (See Dauer’s presentation, Thursday 4:30-5:00)

• Determine what an ecological meaningful difference should be

Acknowledgments