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

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
- 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

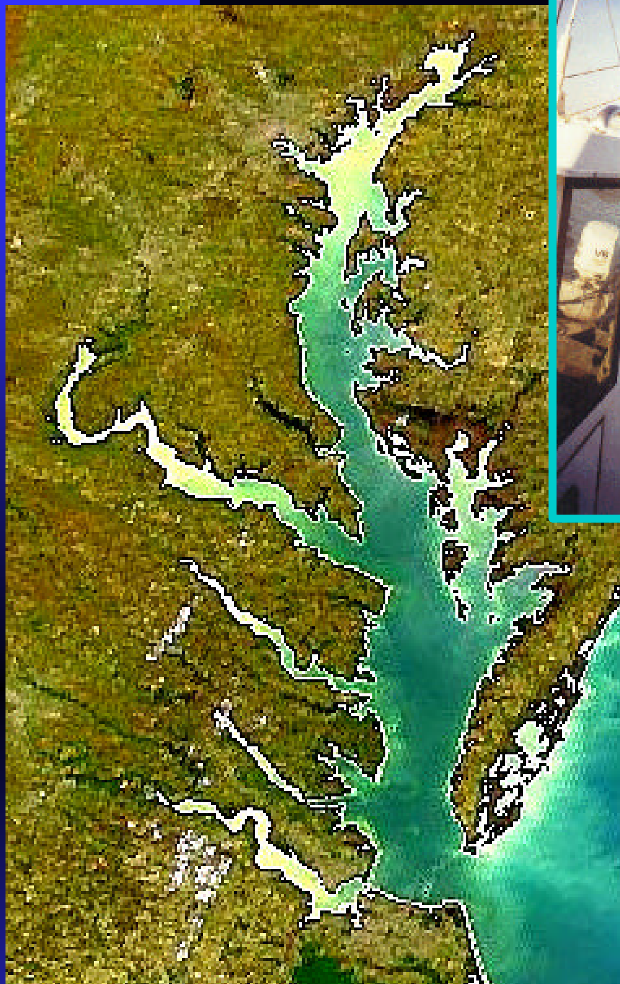
Context

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

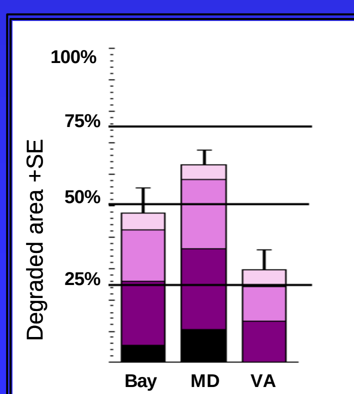
Probability Survey Design

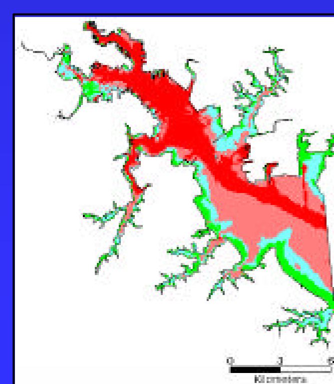
Benthic Index of Biotic Integrity

Survey of Condition (Status)

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

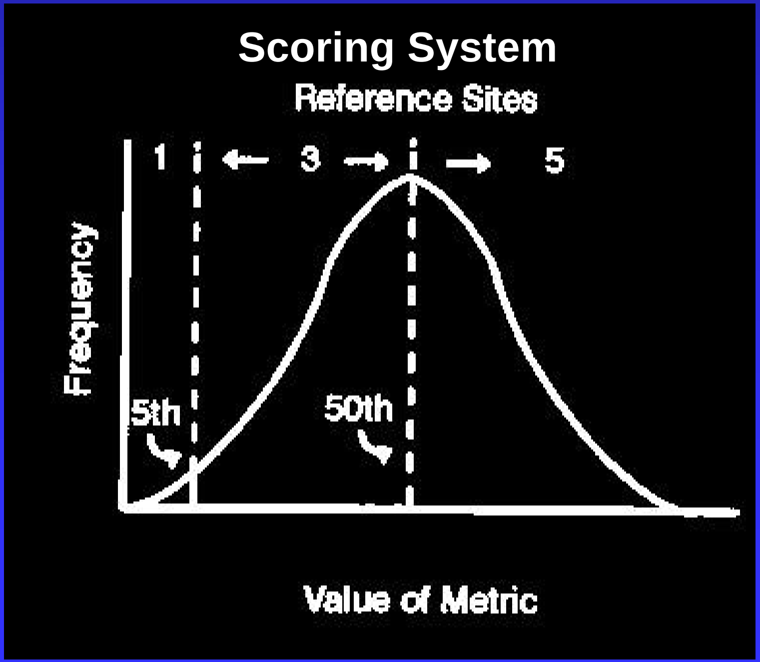
303(d) Lists





Benthic Index of Biotic Integrity[1]

- 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



[1]Weisberg et al. 1997, *Estuaries* 20:149-158

[1]Alden et al. 2002, *Environmetrics* 13:473-498

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

	Reference		Segment		
	Mean	SE	Mean	SE	Weight
Hab1	4.1	0.69	2.7	0.69	3/10
Hab2	3.1	0.58	2.1	0.58	3/10
Hab3	3.5	0.55	1.8	0.35	4/10
Hab 1-3	3.56	0.35*	2.16	0.30*	

Weighted Estimates

*SE of the weighted mean

Example provided by Florence Faulk, US EPA ORD

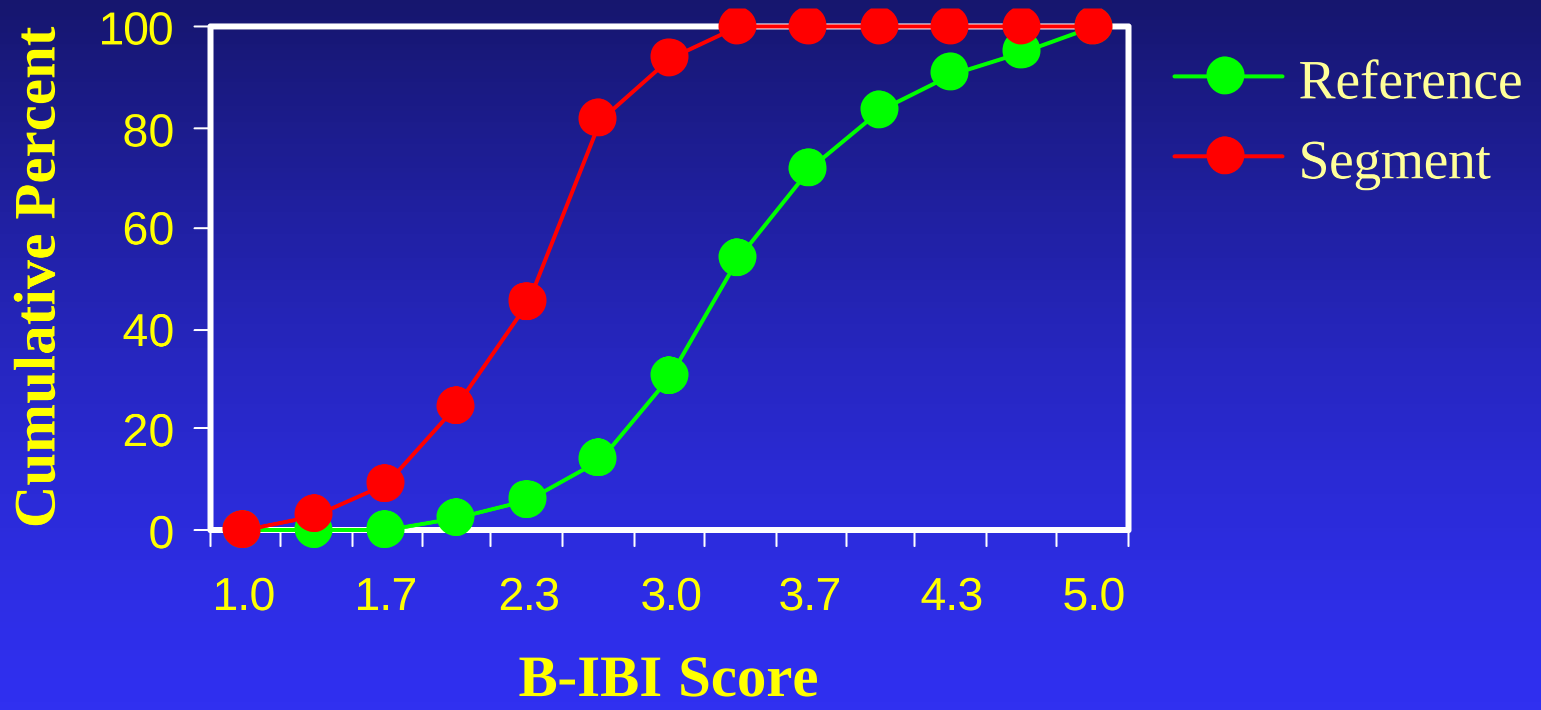
Weighted mean approach

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

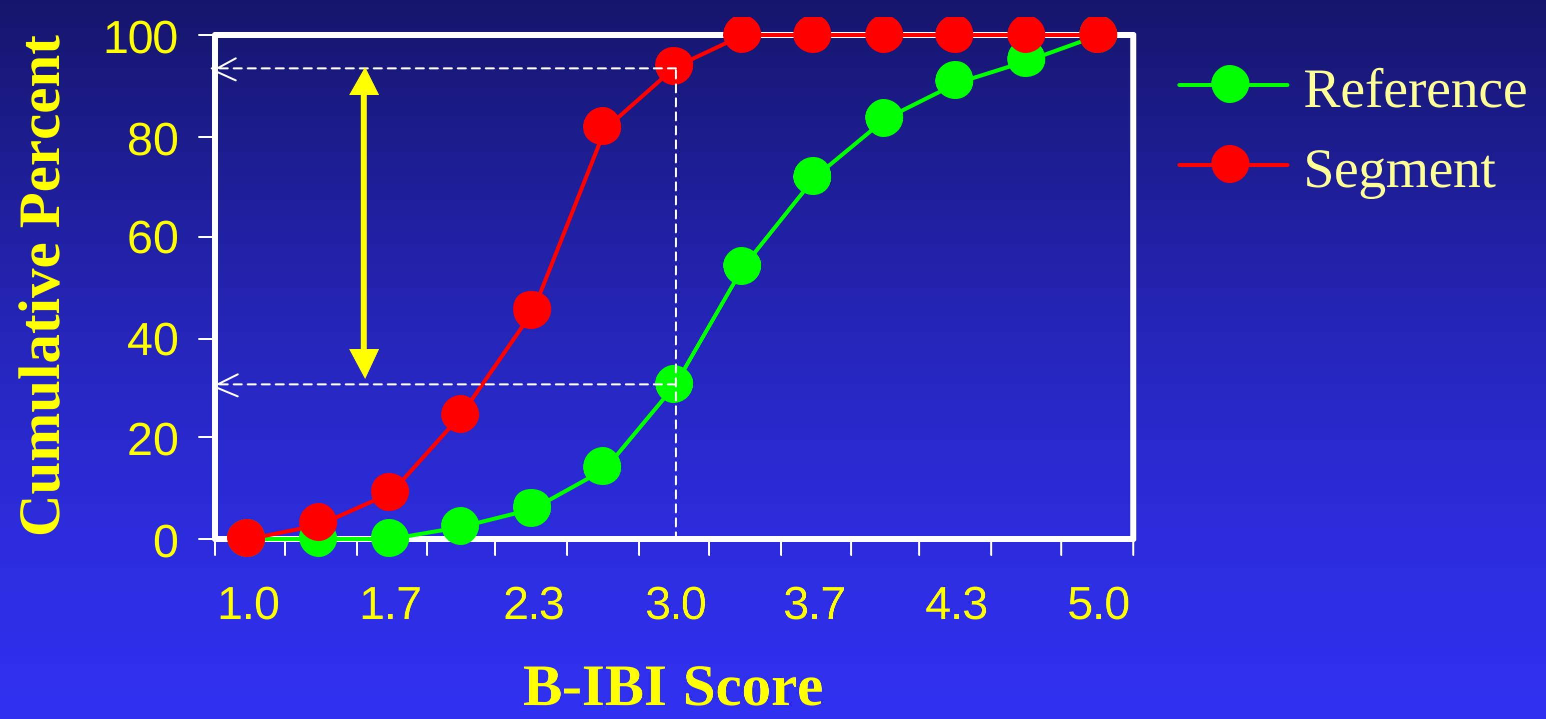
$$t = \frac{\overline{X}_r - \overline{X}_s}{SE_p} = \frac{3.56 - 2.16}{0.461} = 3.04 > t_{0.05,18}$$

Example provided by Florence Faulk, US EPA ORD

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

Habitat Class (columns); **Habitat Class** (rows)

	TF	OL	LM	HS	HM	PS	PM
TF			X	X	X	X	X
OL						X	
LM	X						
HS	X						
HM	X						
PS	X	X					
PM	X						

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

[1]Alden et al. 2002, *Environmetrics* 13:473-498

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

Condition Category	B-IBI Score	Benthic Community Condition
1	1.0-2.0	Severely degraded
2	2.1-2.9	Degraded
3	3.0-5.0	Meets goal

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: Equal multinomial distributions

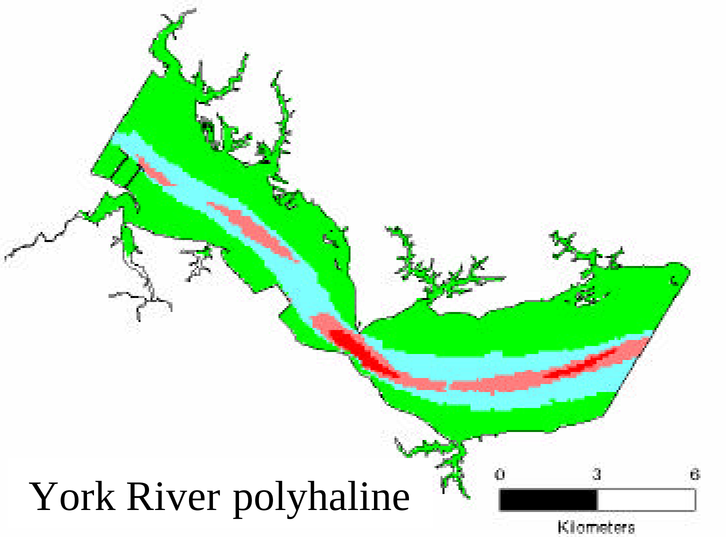
H_1: Shift in location toward lower B-IBI responses in segment than in reference

Type I and Type II errors

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

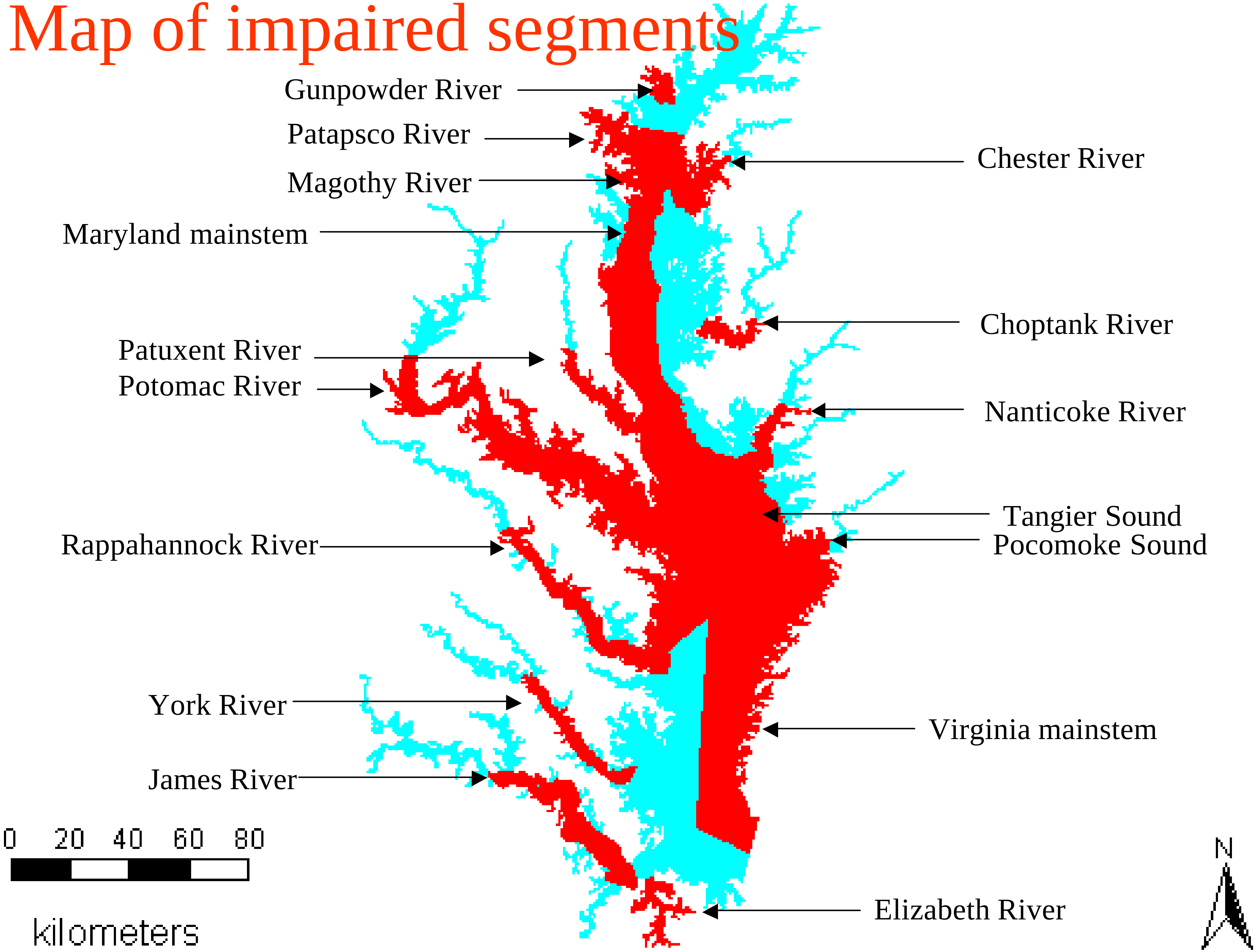
Results of assessment

- 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

Map of impaired segments

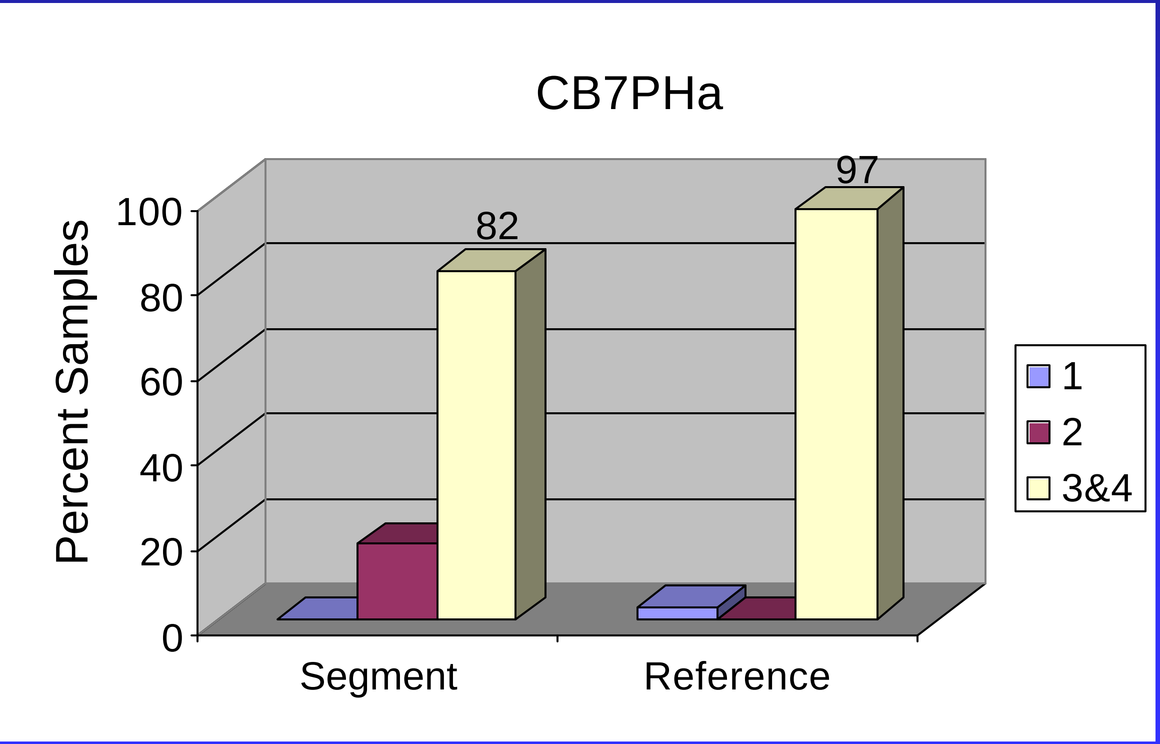


List of impaired segments

Segment	Name	Sample size	Weighted P less then 3.0			Seg-Ref
			Seg	Ref	Deg	
SBEMHa	Southern Branch Elizabeth River	116	0.93	0.04	0.99	0.89
EBEMHa	Eastern Branch Elizabeth River	32	0.88	0.08	0.98	0.79
WBEMHa	Western Branch Elizabeth River	39	0.82	0.04	0.99	0.78
POTMH	Potomac mesohaline	98	0.81	0.09	0.94	0.72
LAFMHa	Lafayette River	35	0.77	0.06	0.99	0.71
CB4MH	Maryland mainstem	30	0.73	0.09	0.98	0.65
PATMH	Patapsco River	45	0.69	0.07	0.89	0.62
YRKMHa	York River mesohaline	66	0.64	0.07	0.98	0.57
POCMH	Pocomoke River	11	0.64	0.07	0.99	0.56
RPPMHa	Rappahannock River mesohaline	96	0.60	0.08	0.95	0.53
ELIMHa	Elizabeth River mesohaline	36	0.56	0.03	0.99	0.52
CB5MH	Maryland mainstem	46	0.57	0.06	0.99	0.50
JMSMHa	James River mesohaline	40	0.55	0.05	0.93	0.50
YRKPHa	York River polyhaline	27	0.52	0.03	0.99	0.48
POTOH	Potomac River oligohaline	15	0.60	0.12	0.72	0.48
PAXMH	Patuxent River mesohaline	108	0.57	0.10	0.95	0.47
MAGMH	Magothy River	20	0.55	0.08	0.91	0.47
JMSOHa	James River oligohaline	29	0.55	0.13	0.75	0.42
GUNOH	Gunpowder River	10	0.50	0.09	0.75	0.41
TANMH	Tangier Sound	38	0.45	0.06	1.00	0.39
CB3MH	Maryland mainstem	55	0.48	0.10	0.89	0.38
CHOMH2	Choptank River	14	0.43	0.07	0.88	0.36
NANMH	Nanticoke River	11	0.45	0.09	0.87	0.36
CHSMH	Chester River	35	0.43	0.08	0.92	0.35
ELIPHa	Elizabeth River polyhaline	25	0.36	0.04	0.99	0.32
CB7PHa	Virginia mainstem	41	0.20	0.03	1.00	0.17

Segment CBP7PHa (Virginia mainstem)

- Listing of this segment as impaired is problematic, 80% of all B-IBI scores in the segment >= 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

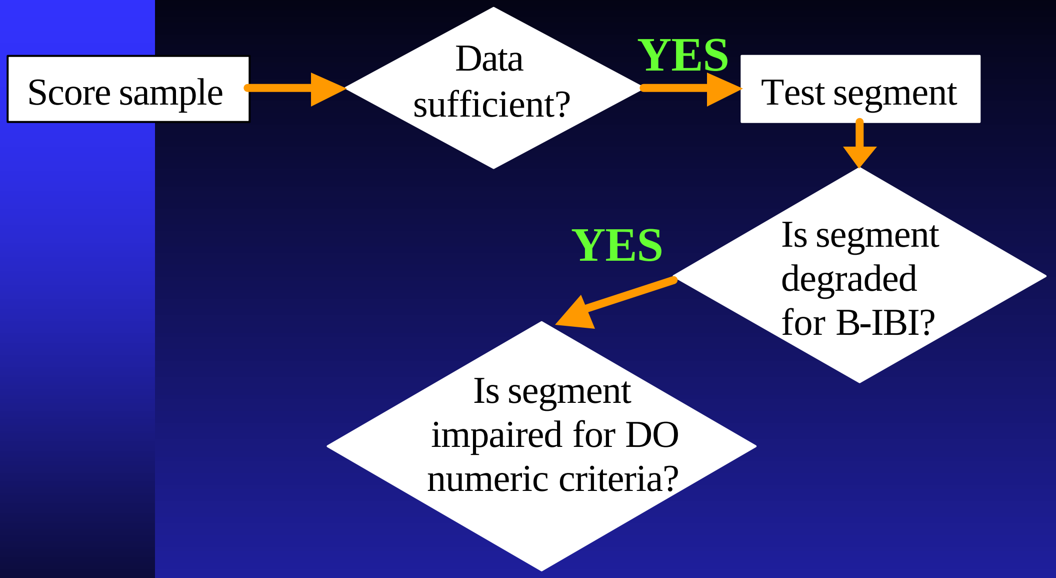
- Stratified Wilcoxon rank sum test may be too sensitive (detects significant differences for small shifts)
- It is not possible to estimate the magnitude of the shift in location (e.g., with a Hodges-Lehman confidence interval) for stratified data
- For stratified data, it is not possible to evaluate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

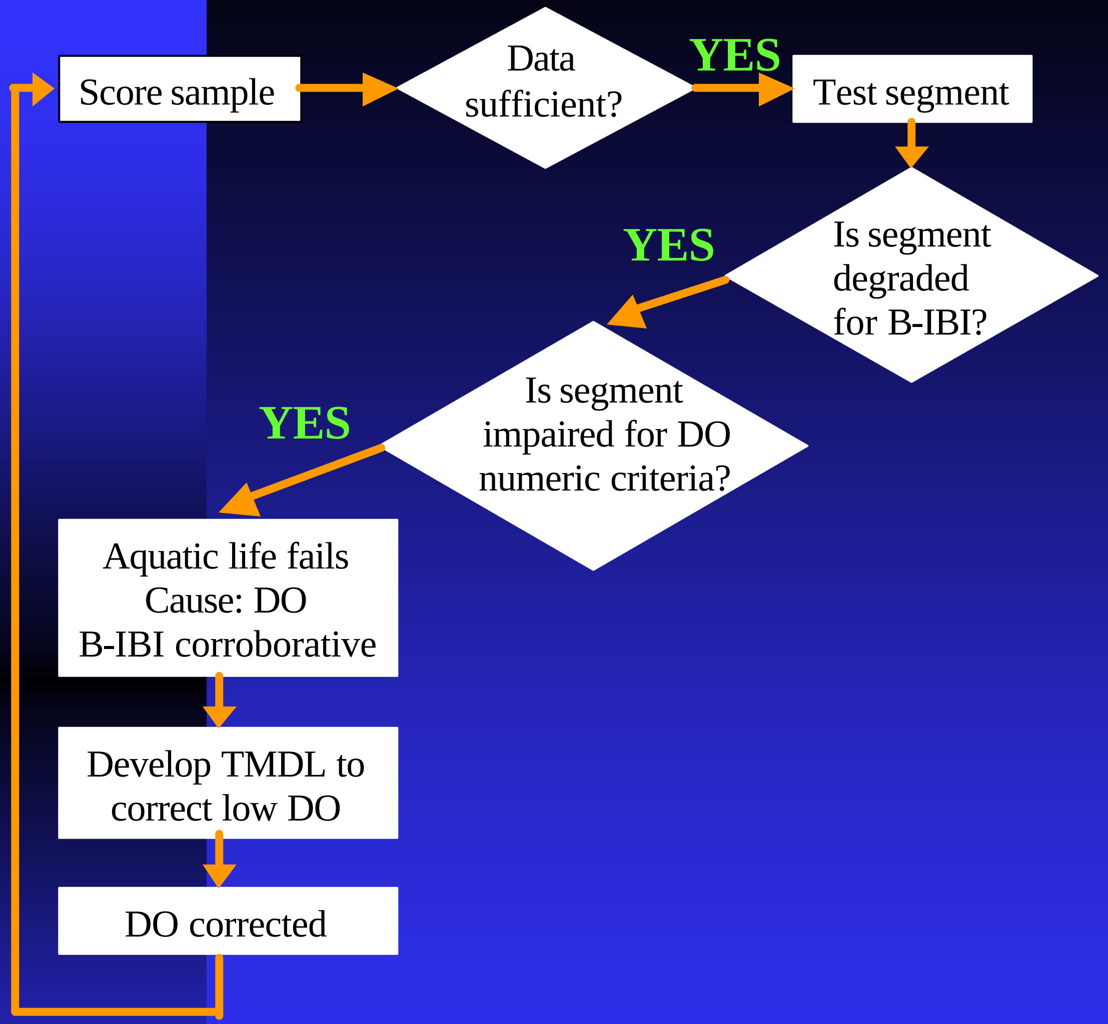
Limitations of current approach

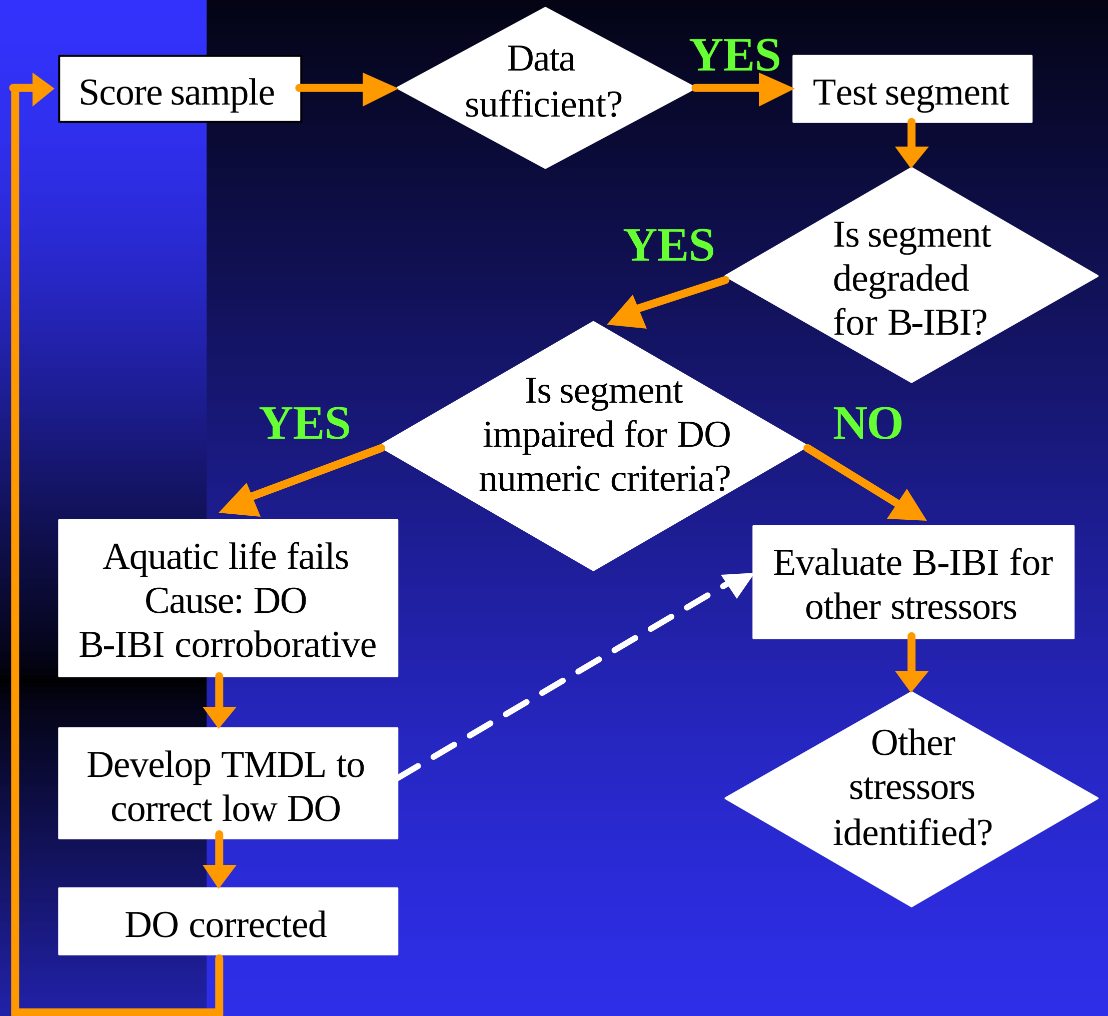
- Stratified Wilcoxon rank sum test may be too sensitive (detects significant differences for small shifts)
- It is not possible to estimate the magnitude of the shift in location (e.g., with a Hodges-Lehman confidence interval) for stratified data
- For stratified data, it is not possible to evaluate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

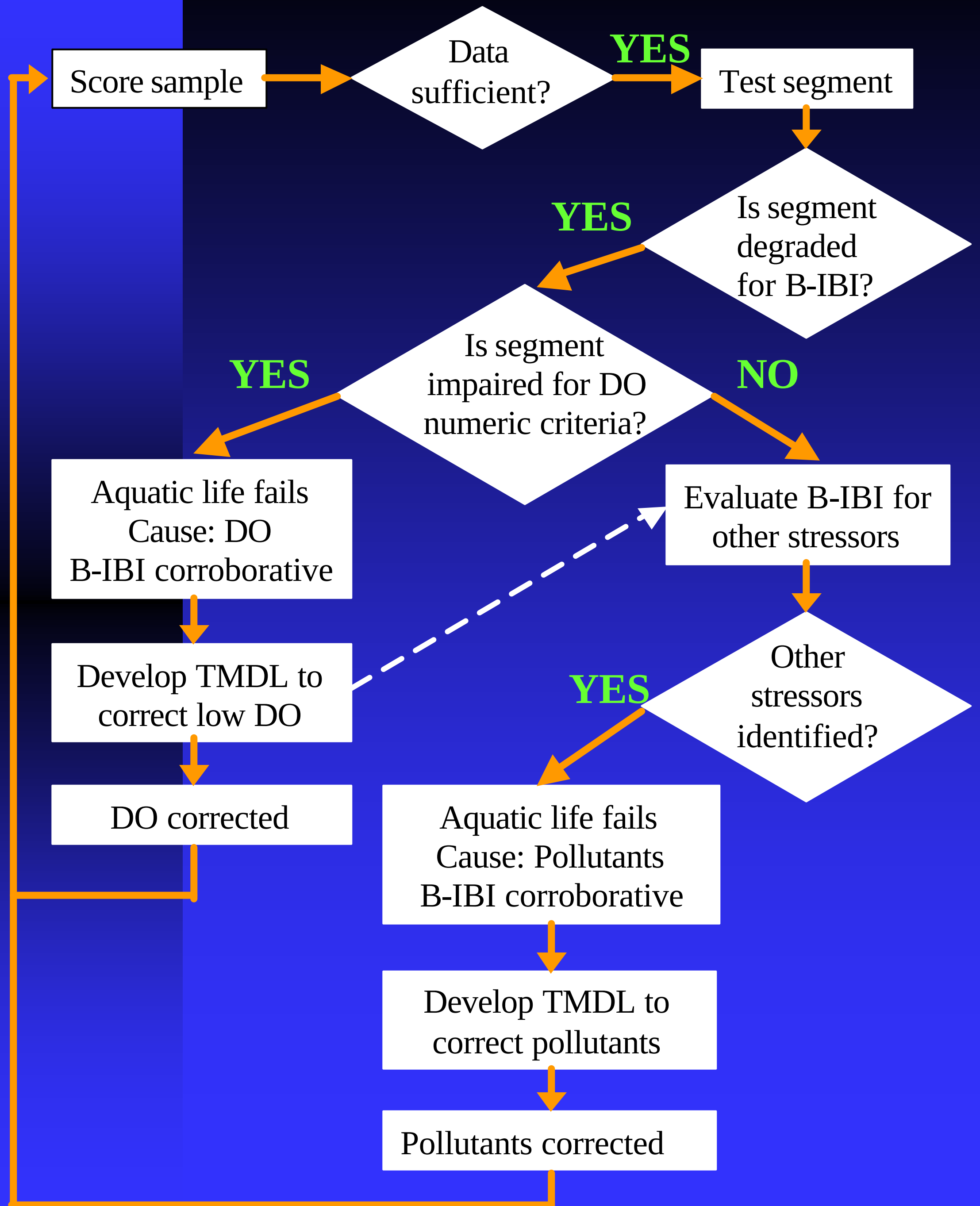
How is this approach used by the States to evaluate aquatic life use support?


Score sample
Data sufficient?
YES
Test segment

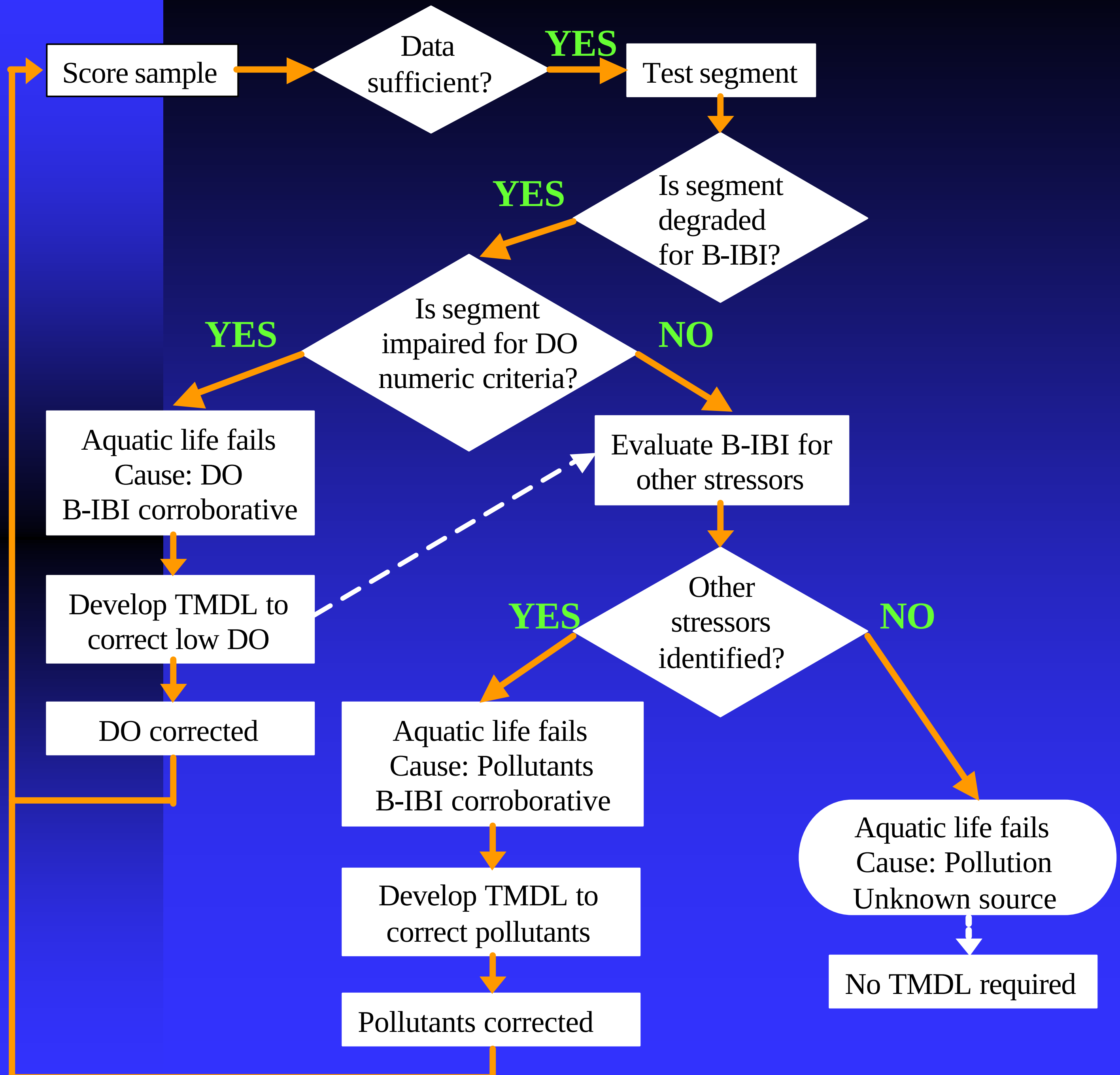

Score sample
Data sufficient?
YES
Test segment
Is segment degraded for B-IBI?
YES
Is segment impaired for DO numeric criteria?


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

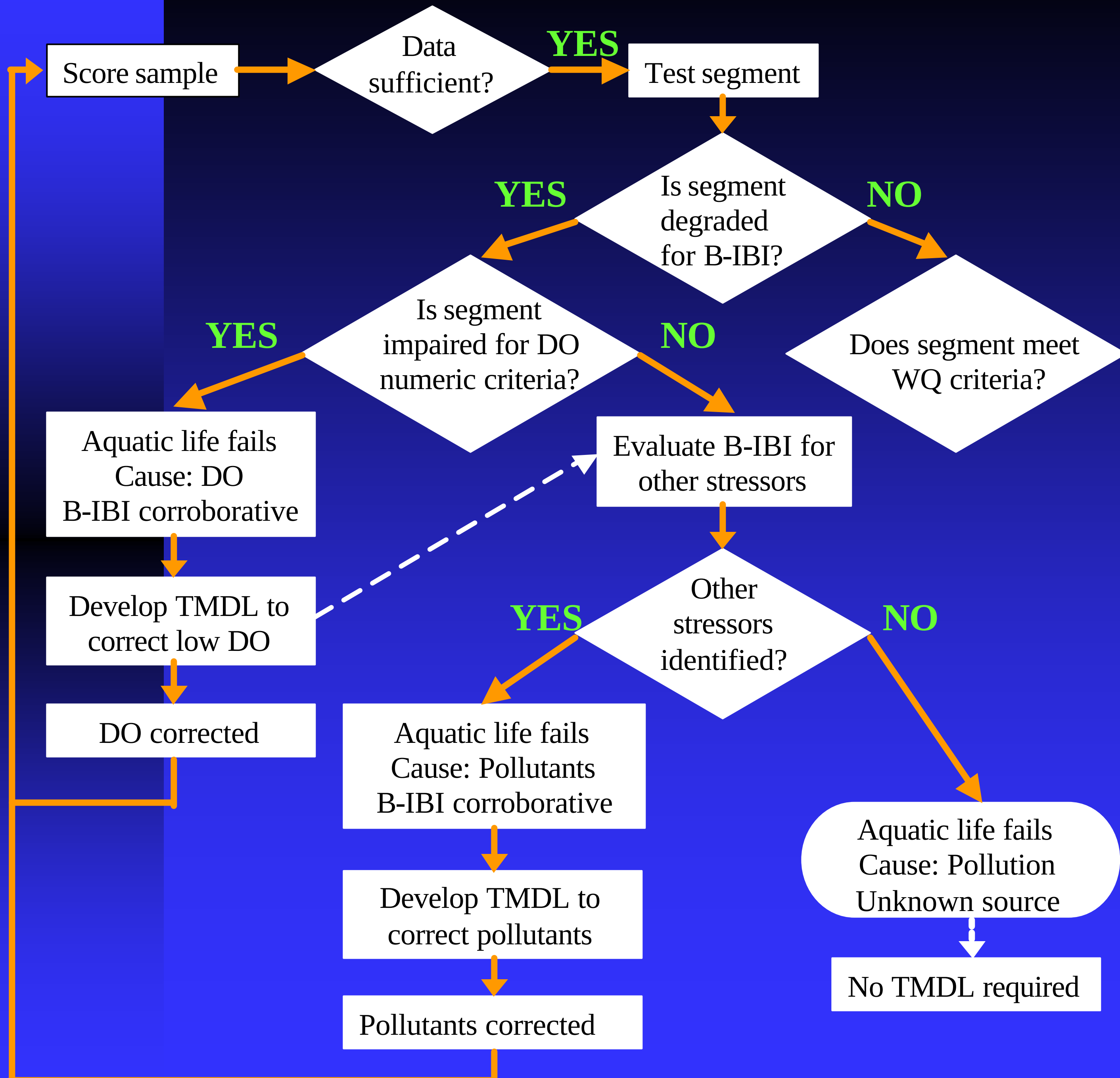

Score sample
Data sufficient?
YES
Test segment
Is segment degraded for B-IBI?
YES
Is segment impaired for DO numeric criteria?
YES
NO
Aquatic life fails
Cause: DO
B-IBI corroborative
Evaluate B-IBI for other stressors
Develop TMDL to correct low DO
Other stressors identified?
DO corrected

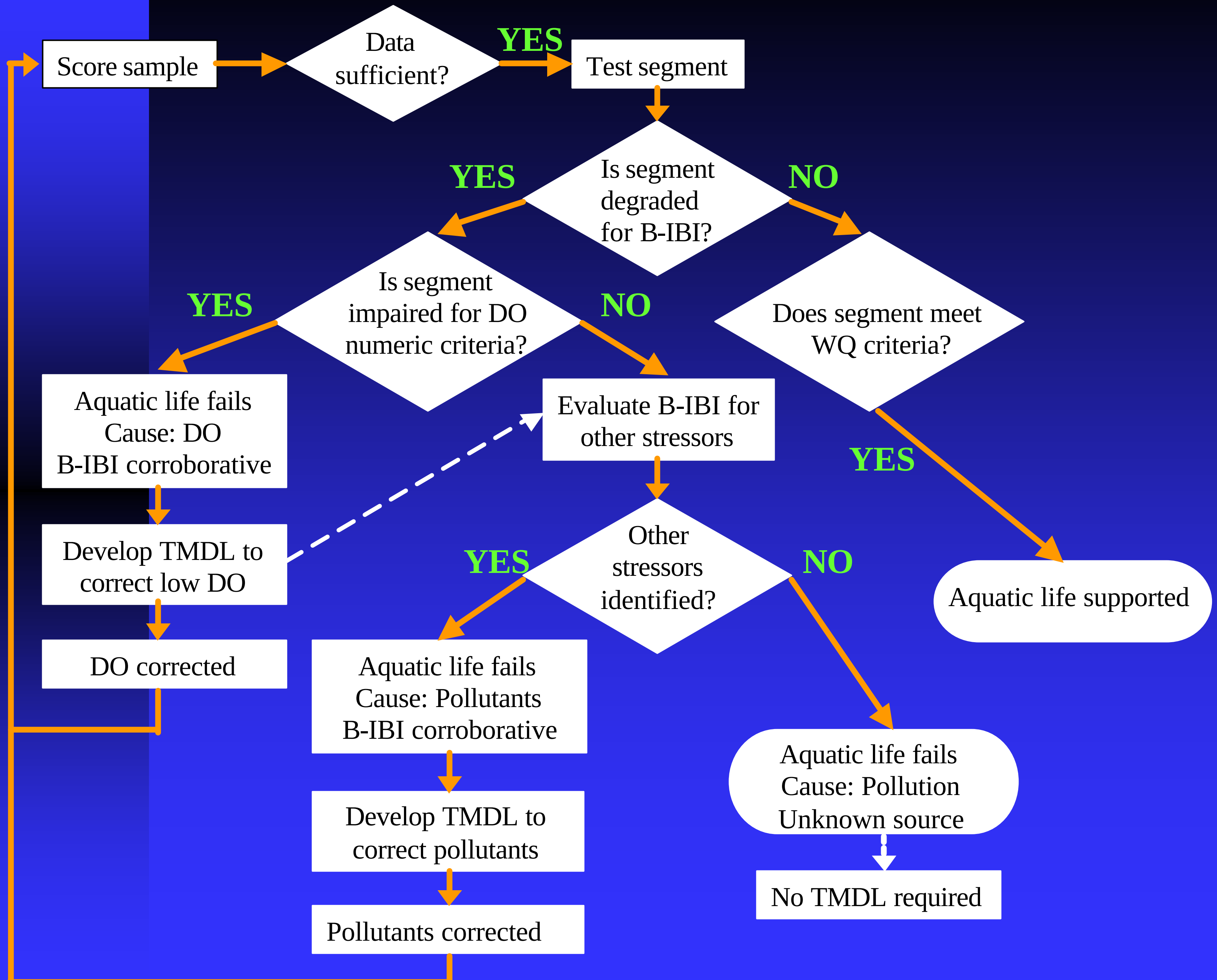


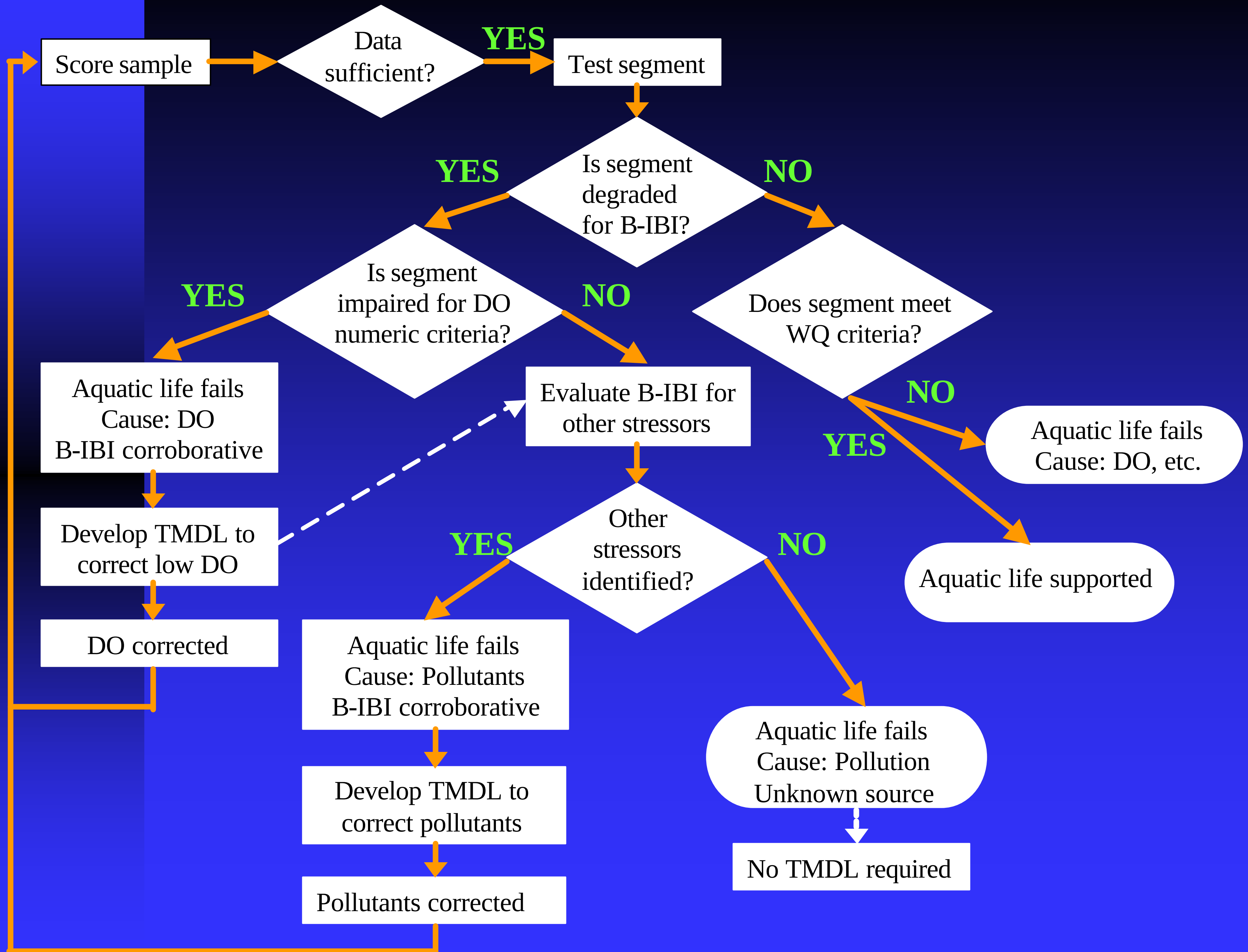
Score sample
Data sufficient?
YES
Test segment
Is segment degraded for B-IBI?
YES
Is segment impaired for DO numeric criteria?
YES
NO
Aquatic life fails
Cause: DO
B-IBI corroborative
Evaluate B-IBI for other stressors
Develop TMDL to correct low DO
Other stressors identified?
YES
DO corrected
Aquatic life fails
Cause: Pollutants
B-IBI corroborative
Develop TMDL to correct pollutants
Pollutants corrected

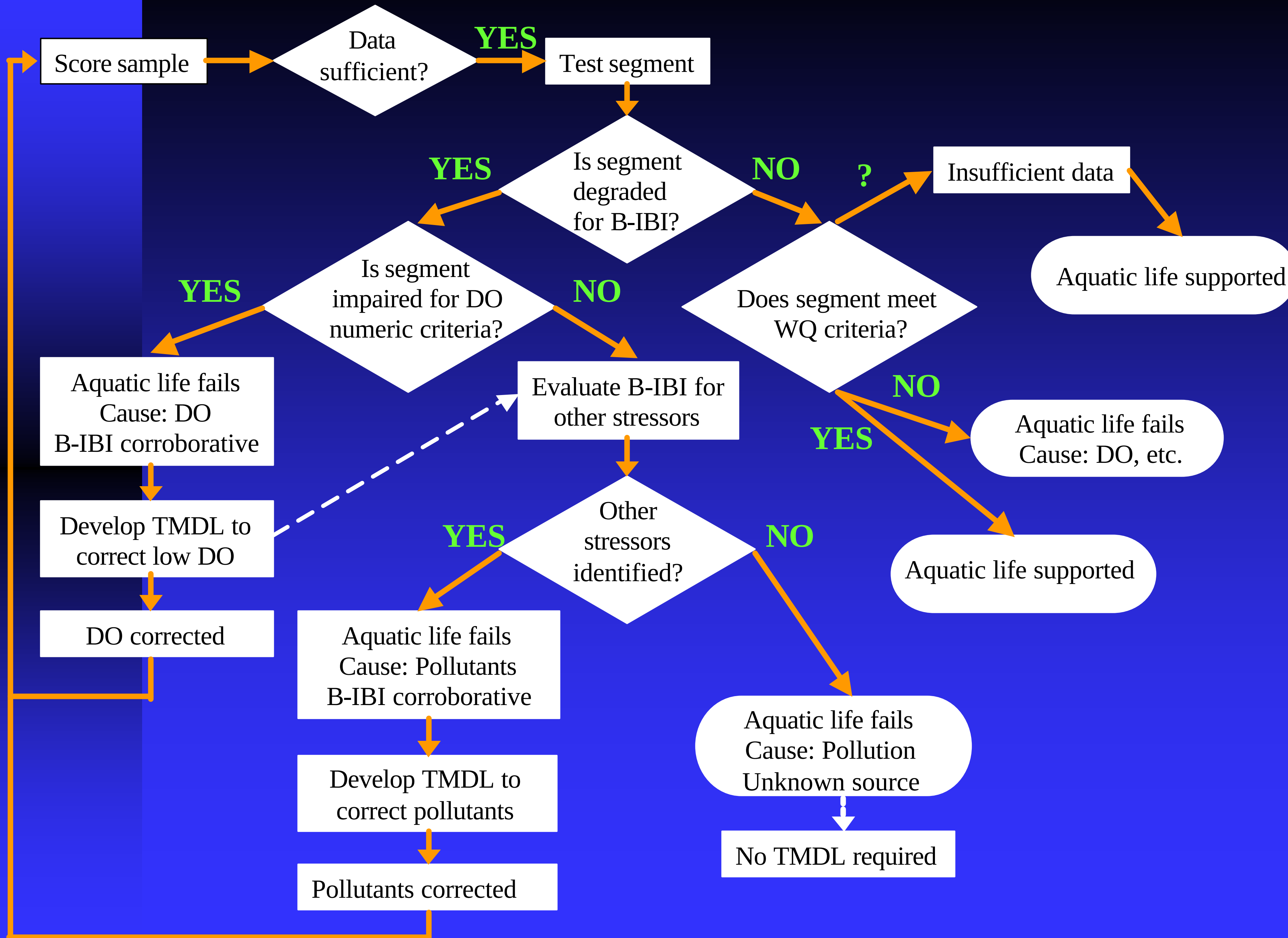


Score sample
Data sufficient?
YES
Test segment
Is segment degraded for B-IBI?
YES
Is segment impaired for DO numeric criteria?
YES
NO
Aquatic life fails
Cause: DO
B-IBI corroborative
Evaluate B-IBI for other stressors
Develop TMDL to correct low DO
Other stressors identified?
YES
NO
DO corrected
Aquatic life fails
Cause: Pollutants
B-IBI corroborative
Aquatic life fails
Cause: Pollution
Unknown source
Develop TMDL to correct pollutants
No TMDL required
Pollutants corrected

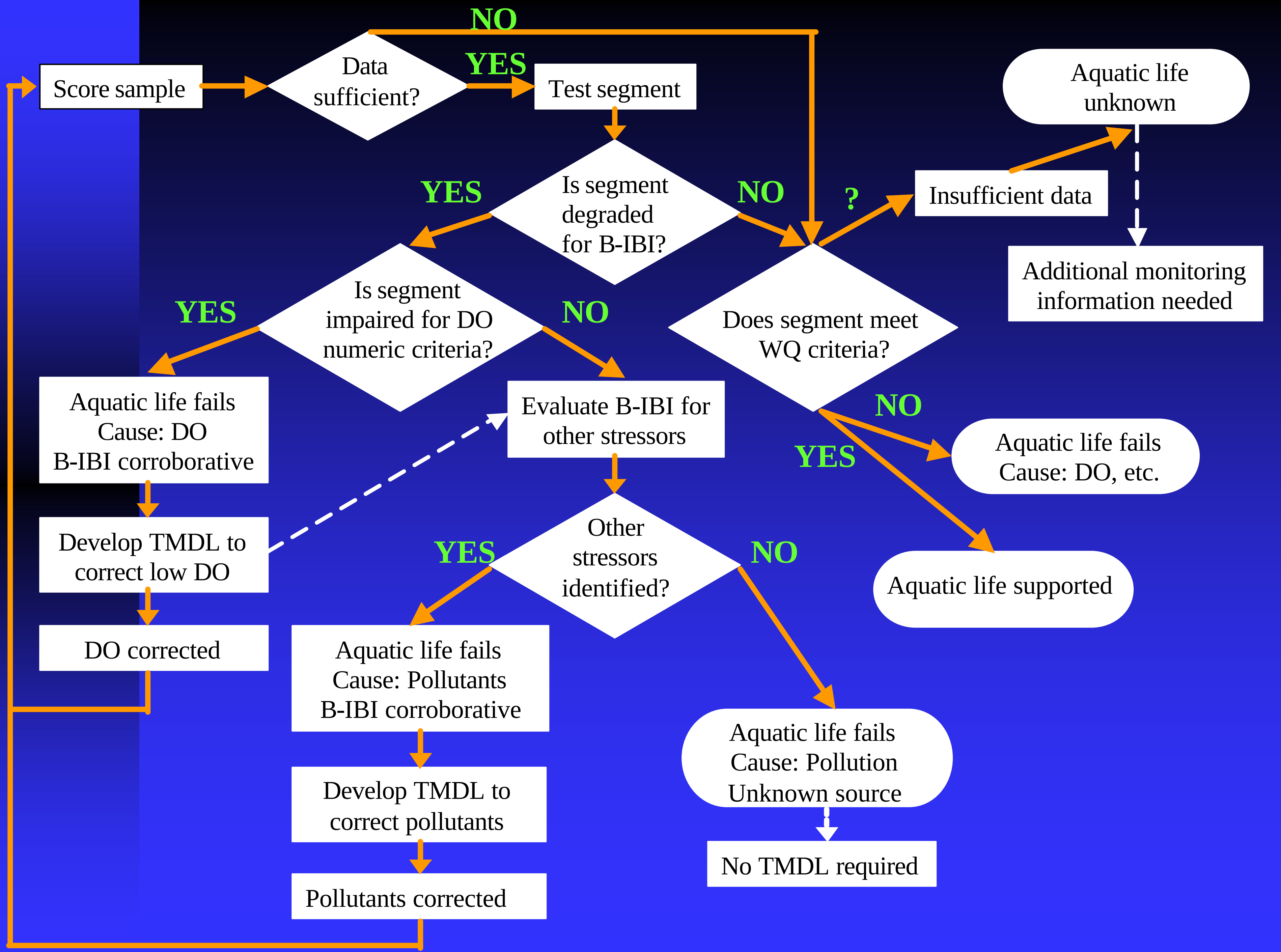

Score sample
Data sufficient?
YES
Test segment
Is segment degraded for B-IBI?
YES
NO
Is segment impaired for DO numeric criteria?
YES
NO
Does segment meet WQ criteria?
Aquatic life fails
Cause: DO
B-IBI corroborative
Evaluate B-IBI for other stressors
Develop TMDL to correct low DO
DO corrected
Other 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


Score sample
Data sufficient?
YES
Test segment
Is segment degraded for B-IBI?
YES
NO
Is segment impaired for DO numeric criteria?
YES
NO
Does segment meet WQ criteria?
Aquatic life fails
Cause: DO
B-IBI corroborative
Evaluate B-IBI for other stressors
YES
Develop TMDL to correct low DO
Other stressors identified?
YES
NO
Aquatic life supported
DO corrected
Aquatic life fails
Cause: Pollutants
B-IBI corroborative
Aquatic life fails
Cause: Pollution
Unknown source
Develop TMDL to correct pollutants
No TMDL required
Pollutants corrected


Score sample
Data sufficient?
YES
Test segment
Is segment degraded for B-IBI?
YES
NO
Is segment impaired for DO numeric criteria?
YES
NO
Does segment meet WQ criteria?
Aquatic life fails
Cause: DO
B-IBI corroborative
Evaluate B-IBI for other stressors
NO
YES
Aquatic life fails
Cause: DO, etc.
Develop TMDL to correct low DO
Other stressors identified?
YES
NO
Aquatic life supported
DO corrected
Aquatic life fails
Cause: Pollutants
B-IBI corroborative
Aquatic life fails
Cause: Pollution
Unknown source
Develop TMDL to correct pollutants
No TMDL required
Pollutants corrected



Score sample
Data sufficient?
YES
Test segment
Is segment degraded for B-IBI?
YES
NO
?
Insufficient data
Aquatic life supported
Is segment impaired for DO numeric criteria?
YES
NO
Does segment meet WQ criteria?
Aquatic life fails
Cause: DO
B-IBI corroborative
Evaluate B-IBI for other stressors
NO
YES
Aquatic life fails
Cause: DO, etc.
Develop TMDL to correct low DO
Other stressors identified?
YES
NO
Aquatic life supported
DO corrected
Aquatic life fails
Cause: Pollutants
B-IBI corroborative
Aquatic life fails
Cause: Pollution
Unknown source
Develop TMDL to correct pollutants
No TMDL required
Pollutants corrected


Score sample
Data sufficient?
NO
YES
Test segment
Is segment degraded for B-IBI?
YES
NO
Is segment impaired for DO numeric criteria?
YES
NO
Does segment meet WQ criteria?
?
Insufficient data
Aquatic life unknown
Additional monitoring information needed
Aquatic life fails Cause: DO B-IBI corroborative
Develop TMDL to correct low DO
DO corrected
Evaluate B-IBI for other stressors
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
NO
YES
Aquatic life fails Cause: DO, etc.
Aquatic life supported

What's next?

- Research into alternative methods
  - Ray Alden et al. confidence limit approach[1]

[1]Alden et al. 2002, *Environmetrics* 13:473-498

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

[1] Alden et al. 2002, *Environmetrics* 13:473-498

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)

[1]Alden et al. 2002, *Environmetrics* 13:473-498

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

[1] Alden et al. 2002, *Environmetrics* 13:473-498

Acknowledgments

UNITED STATES ENVIRONMENTAL PROTECTION AGENCY

MARYLAND DEPARTMENT OF NATURAL RESOURCES

mde

DEQ VIRGINIA DEPARTMENT OF ENVIRONMENTAL QUALITY