

United States
Environmental Protection
Agency

Region III
Chesapeake Bay
Program Office

Region III
Water Protection
Division

EPA 903-R-10-002
CBP/TRS 301-10
May 2010

In coordination with the Office of Water/Office of Science and Technology, Washington, D.C., and the states of Delaware, Maryland, New York, Pennsylvania, Virginia and West Virginia and the District of Columbia



Ambient Water Quality Criteria for Dissolved Oxygen, Water Clarity and Chlorophyll *a* for the Chesapeake Bay and Its Tidal Tributaries: 2010 Technical Support for Criteria Assessment Protocols Addendum

May 2010

**Ambient Water Quality Criteria for Dissolved
Oxygen, Water Clarity and Chlorophyll *a* for the
Chesapeake Bay and Its Tidal Tributaries: 2010
Technical Support for Criteria Assessment
Protocols Addendum**

May 2010
U.S. Environmental Protection Agency
Region III
Chesapeake Bay Program Office
Annapolis, Maryland

and

Region III
Water Protection Division
Philadelphia, Pennsylvania

in coordination with

Office of Water
Office of Science and Technology
Washington, D.C.

and

The states of
Delaware, Maryland, New York
Pennsylvania, Virginia and
West Virginia and the District of Columbia

Contents

Acknowledgements	3
I. Introduction	5
Literature cited.....	6
II. Designated Use Boundaries: Episodic Pycnocline Application and Expanded Designated Uses	8
Background	8
Revising a Procedural Anomaly in the Designated Use Delineation.....	9
Identification of a Procedural Anomaly.....	9
Episodic Pycnocline Criteria Assessment Protocol Modification.....	10
Expanded Application of Deep-Water and Deep-Channel Designated Uses.....	11
Review of Designated Use Definitions.....	11
Mesohaline Segments Expanded Designated Uses.....	12
Literature Cited.....	14
III. Biologically-based Reference Curve: Revisions to the Methodology and Applications	15
Background.....	15
Issues with Dissolved Oxygen Criteria Assessment with the Available Previously Published Biologically-based Reference Curves.....	16
Updates to Dissolved Oxygen Biologically-based reference Curve Derivation Methodology.....	17
Restrict dataset to data collected beginning in 1996.....	18
Use sequential 3-year time periods rather than single years.....	18
Screening criteria sample size $n \geq 10$	19
Screening criteria standard deviation ≤ 1.0	19
Definition of healthy benthic macroinvertebrate reference community conditions.....	20
Use Grand Score in computations involving fixed station data	20
Summary of Recommendations	21
Application of a Reference Curve for Open-Water 30-day Mean Dissolved Oxygen Criteria: Summer Season.....	23
Application of a Reference Curve for Deep-Water Mean Dissolved Oxygen Criteria.....	23
Application of a Reference Curve for Deep-Channel Instantaneous Minimum Dissolved Oxygen Criterion.....	24
Comparisons of Degraded Reference Benthic Communities with the Published Deep Channel Reference Curve	25
Rationale for Acceptable Exceedances of the Deep Channel Instantaneous Minimum Dissolved Oxygen Criterion.....	26
Assessment of Summer Season Dissolved Oxygen Criteria.....	27

Literature Cited.....	28
IV. Revisions to the Chlorophyll <i>a</i> Criteria Assessment Methodology.....	31
Background.....	31
Review of the Current Chlorophyll <i>a</i> Criteria Attainment Assessment Procedure: Method and Assumptions.....	32
Chlorophyll <i>a</i> : Data Skewness, Log Transformation and the Seasonal Mean Calculation.....	34
Log normal character of chlorophyll <i>a</i> data.....	34
James River focused analyses of log transformed chlorophyll <i>a</i> data for normality.....	35
Chlorophyll <i>a</i> criteria Assessment Protocol Refinements using Log- Transformations.....	36
Implications of the Revised Assessment Protocol.....	37
Literature Cited.....	38
 Acronyms.....	 41
 Appendices	
A. IBI Sample Size and Standard Deviations on IBI Scoring When Screening Segments for Reference Community characterization.....	42
B. Shape of the Biologically-based Reference Curve.....	46
C. Derivation of the Deep-Water Biologically-Based Reference Curve.....	49
D. History of EPA Guidance Regarding the Deep-Channel Reference Curve.....	52
E. James River Chlorophyll <i>a</i> Data Normality Analysis Checking Normality of Log- Transformed Chlorophyll <i>a</i> Data.....	54
F. SAS Computer Code for James River VA Chlorophyll <i>a</i> Normality Tests, Spring and Summer Season	57

Acknowledgements

This fifth addendum to the EPA April 2003 publication of *Ambient Water Quality Criteria for Dissolved Oxygen, Water Clarity and Chlorophyll a for the Chesapeake Bay and Its Tidal Tributaries (Regional Criteria Guidance)* was developed and documented through the collaborative efforts of members of the Chesapeake Bay Program's (CBP) Criteria Assessment Protocols Workgroup and the Water Quality Goal Implementation Team.

PRINCIPAL AND CONTRIBUTING AUTHORS

The following are principal and contributing authors of this addendum: Peter Tango, U.S. Geological Survey/Chesapeake Bay Program Office; Jeni Keisman, University of Maryland Center for Environmental Science/Chesapeake Bay Program Office; Richard Batiuk, U.S. EPA Region 3 Chesapeake Bay Program Office; Elgin Perry, Statistical Consultant, and Jackie Johnson, Interstate Commission on the Potomac River Basin/Chesapeake Bay Program Office.

CRITERIA ASSESSMENT PROTOCOL WORKGROUP

Peter Tango, Chair, U.S. Geological Survey/Chesapeake Bay Program Office; Cheryl Atkinson, U.S. EPA Region 3, Water Protection Division; Harry Augustine, Virginia Department of Environmental Quality; Mark Barath, U.S. EPA Region 3, Water Protection Division; Tom Barron, Pennsylvania Department of Environmental Protection; Stephen Cioccia, Virginia Department of Environmental Quality; Richard Eskin, Maryland Department of the Environment; Sherm Garrison, Maryland Department of Natural Resources; Darryl Glover, Virginia Department of Environmental Quality; Rick Hoffman, Virginia Department of Environmental Quality; Jackie Johnson, Interstate Commission on the Potomac River Basin/Chesapeake Bay Program Office; Jeni Keisman, University of Maryland Center for Environmental Science/Chesapeake Bay Program Office; Susan McDowell, United States Environmental Protection Agency Region III; Larry Merrill, United States Environmental Protection Agency; Bruce Michael, Maryland Department of Natural Resources; Ken Moore, Virginia Institute of Marine Science; Shah Nawaz, District Department of the Environment; Jennifer Palmore, Virginia Department of Environmental Quality; Tom Parham, Maryland Department of Natural Resources; Elgin Perry, Statistics Consultant; Charlie Poukish, Maryland Department of the Environment; Tish Robertson, Virginia Department of Environmental Quality; Matt Rowe, Maryland Department of the Environment; John Schneider, Delaware Department of Natural Resources and Environmental Control; Gary Shenk, United States Environmental Protection Agency; Donald Smith, Virginia Department of Environmental Quality; Scott Stoner, New York State Department of Environmental Conservation; Matt Stover, Maryland Department of the Environment; Bryant Thomas, Virginia Department of Environmental Quality; Mark Trice, Maryland Department of Natural Resources; David Wolanski, Delaware Department of Natural Resources and Environmental Control.

WATER QUALITY GOAL IMPLEMENTATION TEAM

Dave Hansen, Co-Chair, University of Delaware; Robert Koroncai, Co-Chair, U.S. Environmental Protection Agency Region 3 Water Protection Division; Katherine Antos, Coordinator, Chesapeake Bay Program Office, U.S. Environmental Protection Agency; Victoria

Kilbert , Staff, Chesapeake Research Consortium; Rachel Streusand, Staff, Chesapeake Research Consortium; Rich Batiuk, U.S. Environmental Protection Agency Region 3 Chesapeake Bay Program Office; Steve Bieber, Metropolitan Washington Council of Governments; Patricia Buckley, Pennsylvania Department of Environmental Protection; Collin Burrell, District Department of the Environment; Monir Chowdhury, District Department of the Environment; Frank Coale, University of Maryland; Lee Currey, Maryland Department of the Environment; James Davis-Martin, Virginia Department of Conservation and Recreation; Chris Day, U.S. Environmental Protection Agency Region 3; Ron Entringer, New York Department of Environmental Conservation; Richard Eskin, Maryland Department of the Environment; Normand Goulet, Northern Virginia Regional Commission; Krista Grigg U.S. Navy; Mike Haire, U.S. Environmental Protection Agency Office of Water; Jeffrey Halka, Maryland Geological Survey; Carlton Haywood, Interstate Commission on the Potomac River Basin; Dave Heicher, Susquehanna River Basin Commission; Rick Hill, Virginia Department of Conservation and Recreation; Beth Horsey, Maryland Department of Agriculture; Ruth Izraeli, U.S. Environmental Protection Agency Region 2; Bill Keeling, Virginia Department of Conservation and Recreation; John Kennedy, Virginia Department of Environmental Quality; Teresa Koon, West Virginia Department of Environmental Protection; Felix Locicero, U.S. Environmental Protection Agency Region 2; Charles Martin, Virginia Department of Environmental Quality; Bruce Michael, Maryland Department of Natural Resources; Matt Monroe, West Virginia Department of Agriculture; Dave Montali, West Virginia Department of Environmental Protection; russel Morgan, U.S. Department of Agriculture Natural Resource Conservation Service; Matt Mullin, Chesapeake Bay Commission; Kenn Pattison, Pennsylvania Department of Environmental Protection; Russ Perkinson, Virginia Department of Conservation and Recreation; Alan Pollock, Virginia Department of Environmental Quality; Marel Raub, Chesapeake bay Commision; John Rhoderick, Maryland Department of Agriculture; John Schneider, Delaware Department of Natural Resources and Environmental Control; Mohsin Siddique, D.C. Water and Sewer Authority; Jennifer Sincock, U.S. Environmental Protection Agency Region 3; Randolph Sovic, West Virginia Department of Environmental Protection; Tanya Spano, Metropolitan Washington Council of Governments; Ann Swanson, Chesapeake Bay Commission; Jennifer Volk, Delaware Department of Natural Resources and Environmental Control; Robert Yowell, Pennsylvania Department of Environmental Protection.

CHAPTER 1

Introduction

In April 2003, the U.S. Environmental Protection Agency (EPA) published the *Ambient Water Quality Criteria for Dissolved Oxygen, Water Clarity and Chlorophyll a for the Chesapeake Bay and Its Tidal Tributaries* which was the foundation document defining Chesapeake Bay water quality criteria and recommended implementation procedures for monitoring and assessment (U.S. EPA 2003a). In October 2003, EPA published the *Technical Support Document for Identification of Chesapeake Bay Designated Uses and Attainability* which defined the five tidal water designated uses to be protected through the published Bay water quality criteria (U.S. EPA 2003b):

- Migratory fish spawning and nursery habitat;
- Open-water fish and shellfish habitat;
- Deep-water seasonal fish and shellfish habitat;
- Deep-channel seasonal refuge habitat; and
- Shallow-water bay grass habitat.

A total of six addendum documents have been published by EPA since April 2003. Three addenda were published documenting detailed refinements to the criteria attainment and assessment procedures (U.S. EPA 2004a, 2007a, 2008) previously published in the original April 2003 Chesapeake Bay water quality criteria document (U.S. EPA 2003a). One addendum published Chesapeake Bay numerical chlorophyll *a* criteria (U.S. EPA 2007b). Another addendum addressed detailed issues involving further delineation of tidal water designated uses (U.S. EPA 2004b) building from the original October 2003 tidal water designated uses document (U.S. EPA 2003b). Finally, one addendum addressed refinements to the Chesapeake Bay Program analytical segmentation schemes (U.S. EPA 2005) building from the original U.S. EPA 2004 document (U.S. EPA 2004c).

The detailed procedures for assessing attainment of the Chesapeake Bay water quality criteria continued to be advanced through the collective EPA, States and District of Columbia partnership efforts. These partners continue to develop and apply procedures that incorporate the most advanced state-of-the-science, magnitude, frequency, duration, space and time considerations with, as available, biologically-based reference conditions and cumulative frequency distributions. As a rule, the best test of any new method or procedure is putting it to application with partner involvement and stakeholder input. Through the work of its Criteria Assessment Protocols Workgroup, the Chesapeake Bay Program partnership has an established forum for resolving issues, factoring in new scientific findings, and ensuring implementation of consistent bay-wide criteria assessment procedure development and implementation. The Workgroup draws upon the talents and input from state, federal, river basin commission and academic partners as well as local government and municipal stakeholders. This EPA 2010 Chesapeake Bay Criteria addendum provides previously undocumented features of the present procedures as well as refinements and clarifications to the previously published Chesapeake Bay water quality criteria assessment procedures.

Chapter 2 documents refinements to the procedures for defining Chesapeake Bay designated uses and expands the application of the deep-water seasonal fish and shellfish designated use to two Chesapeake Bay segments in Maryland's tidal waters.

Chapter 3 documents refinements and additions to the previously published procedures for deriving biologically-based reference curves and recommendations for their application for Chesapeake Bay dissolved oxygen criteria assessments.

Chapter 4 documents refinements and provides recommendations for the procedures assessing the previously published numerical Chesapeake Bay chlorophyll *a* criteria.

Appendices to these three chapters provide more detailed documentation on derivation of the recommended refined criteria assessment procedures.

This document represents the fifth formal addendum to the original 2003 Chesapeake Bay water quality criteria document. As such readers should regard the sections in this document as new or replacement chapters and appendices to the original published Bay Criteria report (U.S. 2003a). The criteria assessment procedures published in this addendum also replace and otherwise supersede similar criteria assessment procedures published in the 2004, 2007 and 2008 addenda (U.S. EPA 2003a, 2004a, 2007a, 2007b, 2008). Publication of future addenda by EPA on behalf of the Chesapeake Bay Program watershed jurisdictional partners is likely as continued scientific research and management applications reveal new insights and knowledge that should be incorporated into revisions of state water quality standards regulations in upcoming triennial reviews.

LITERATURE CITED

U.S. Environmental Protection Agency. 2003a. *Ambient Water Quality Criteria for Dissolved Oxygen, Water Clarity and Chlorophyll a for the Chesapeake Bay and Its Tidal Tributaries (Regional Criteria Guidance)*. April 2003. EPA 903-R-03-002. Region III Chesapeake Bay Program Office, Annapolis, MD.

U.S. Environmental Protection Agency. 2003b. *Technical Support Document for Identification of Chesapeake Bay Designated Uses and Attainability*. October 2003. EPA 903-R-03-004. Region III Chesapeake Bay Program Office, Annapolis, MD.

U.S. Environmental Protection Agency. 2004a. *Ambient Water Quality Criteria for Dissolved Oxygen, Water Clarity and Chlorophyll a for the Chesapeake Bay and Its Tidal Tributaries – 2004 Addendum*. October 2004. EPA 903-R-04-005. Region III Chesapeake Bay Program Office, Annapolis, MD.

U.S. Environmental Protection Agency. 2004b. *Technical Support Document for Identification of Chesapeake Bay Designated Uses and Attainability – 2004 Addendum*. October 2004. EPA 903-R-04-006. Region III Chesapeake Bay Program Office, Annapolis, MD.

U.S. Environmental Protection Agency. 2004c. *Chesapeake Bay Program Analytical Segmentation Scheme: Revisions, Decisions and Rationales 1983-2003. October 2004.* EPA 903-R-04-008. Region III Chesapeake Bay Program Office, Annapolis, MD.

U.S. Environmental Protection Agency. 2005. *Chesapeake Bay Program Analytical Segmentation Scheme: Revisions, Decisions and Rationales 1983-2003. 2005 Addendum. December 2005.* EPA 903-R-05-004. Region III Chesapeake Bay Program Office, Annapolis, MD.

U.S. Environmental Protection Agency. 2007a. *Ambient Water Quality Criteria for Dissolved Oxygen, Water Clarity and Chlorophyll a for the Chesapeake Bay and Its Tidal Tributaries – 2007 Addendum. July 2007.* EPA 903-R-07-003. Region III Chesapeake Bay Program Office, Annapolis, MD.

U.S. Environmental Protection Agency. 2007b. *Ambient Water Quality Criteria for Dissolved Oxygen, Water Clarity and Chlorophyll a for the Chesapeake Bay and Its Tidal Tributaries – Chlorophyll a Addendum. October 2007.* EPA 903-R-07-005. Region III Chesapeake Bay Program Office, Annapolis, MD.

U.S. Environmental Protection Agency. 2008. *Ambient Water Quality Criteria for Dissolved Oxygen, Water Clarity and Chlorophyll a for the Chesapeake Bay and Its Tidal Tributaries–2008 Technical Support for Criteria Assessment Protocols Addendum. September 2008.* EPA 903-R-08-001. Region III Chesapeake Bay Program Office, Annapolis, MD.

CHAPTER 2

Designated Use Boundaries: Episodic Pycnocline Application and Expanded Designated Uses

BACKGROUND

In the *2003 Ambient Water Quality Criteria for Dissolved Oxygen, Water Clarity and Chlorophyll a for Chesapeake Bay and Its Tidal Tributaries*, EPA defined five tidal water habitats as designated uses providing the context for setting protective Chesapeake Bay water quality criteria (U.S. EPA 2003). Detailed dissolved oxygen criteria were established for Chesapeake Bay and its tidal tributaries and embayments tailored to each designated use accounting for its variations in space and time. EPA has published and Delaware, Maryland, Virginia and the District of Columbia have adopted into their state's water quality standards regulations dissolved oxygen criteria protective of the published migratory spawning, open-water, deep-water and deep-channel designated uses. These dissolved oxygen criteria include 30-day, 7-day and 1-day means along with instantaneous minima as needed to protect various species and life stages within the designated uses (U.S. EPA 2003).

Since the Chesapeake Bay dissolved oxygen criteria were published in 2003, refinements and updates to the criteria attainment assessment methodologies have been published. Most recently, the refined and expanded dissolved oxygen criteria assessment methodologies documented in Chapter 3 and associated appendices of the *Ambient Water Quality Criteria for Dissolved Oxygen, Water Clarity and Chlorophyll a for the Chesapeake Bay and its Tidal Tributaries - 2008 Technical Support for Criteria Protocols Addendum*, replaced the methodologies previously published by EPA (U.S. EPA 2008).

Critical to the dissolved oxygen criteria assessments are the pycnocline delineations defining the timing and vertical position of the open-water, deep-water and deep-channel designated use boundaries. The standardized method for calculating upper and lower boundaries of pycnoclines was originally published in *Ambient Water Quality Criteria for Dissolved Oxygen, Water Clarity and Chlorophyll a for the Chesapeake Bay and its Tidal Tributaries – 2004 Addendum* (U.S. EPA 2004a). U.S. EPA (2008), on pages 15-18 together with its Appendix A, provide a review of, and step by step details associated with, calculating upper and lower pycnoclines which, in turn, delineate the vertical boundaries for the open-water, deep-water and deep-channel designated uses.

The following outline lays out the assessment protocol steps for the 30-day mean criteria (open-water and deep-water dissolved oxygen criteria (see U.S. EPA 2008, Appendix A for details):

- 1) Compiling and formatting the data set
- 2) Interpolation of water quality monitoring data
 - 2.1 Vertical interpolation
 - 2.2 Horizontal interpolation
 - 2.3 30-day average interpolation by month

2.4 Apportioning results by designated use

2.5 Water quality criteria assessment, attainment and violations

Step 2.4 above, carried forward the Step 4-Pointwise Compliance considerations of a statistical decision-making framework originally published in U.S. EPA 2007a Chapter II: *Refinements to Chesapeake Bay Water Quality Criteria Assessment Methodology* (pp. 17-18) and revisited in U.S. EPA 2008 (Appendix A). This section on pointwise compliance states:

“While interpolation allows for standardization of many types of data, pointwise attainment allows for standardization of many criteria. Because attainment is determined at moments in time and points in space, it is possible to vary the criterion in time and space. If different levels of a water quality criterion are acceptable in different seasons, then the criterion can vary seasonally. It is possible to implement different criteria over space for a segment that bridges, for example, oligohaline and mesohaline, salinity zones. It might even be possible to let the criterion be a continuous function of some ancillary variable such as temperature or salinity, although this situation requires that such data exist for every interpolator cell. The only requirement is that the final attainment determination be “yes” or “no” for each interpolator cell.”

The implicit assumption of the Chesapeake Bay partners was that if no pycnoclines were found for a particular sampling event then the open-water designated use and its respective dissolved oxygen criteria were being applied, i.e. that water column dynamics including “episodic pycnoclines” were accounted for as part of the criteria assessment computations. The U.S. EPA Chesapeake Bay Program Office’s criteria assessment computer code, however, applied the long-term average pycnocline depth(s) to those water quality monitoring cruise sampling events when no pycnocline was found for those 13 segments, identified in U.S. EPA 2004 where deep-water and/or deep channel designated uses applied during the June-September time period. Therefore, under special cases, on the basis of pre-determined characterization, there were errors in designated use classification.

REVISING A PROCEDURAL ANOMALY IN THE DESIGNATED USE DELINEATION

Identification of a Procedural Anomaly

During 2009, a procedural anomaly was discovered between EPA published dissolved oxygen criteria assessment protocols through 2008 for pycnocline delineation that defined the boundaries for the open-water, deep-water and/or deep-channel designated uses and the assessment procedures as defined in the criteria assessment computer code developed by the EPA Chesapeake Bay Program Office and used by the states and the District. The published procedures set forth that attainment is determined at moments in time and space given that the designated uses, their boundaries and the applicable dissolved oxygen criteria will also vary in time and space (U.S. EPA 2003). U.S. EPA (2008) published details of the computations for identifying pycnoclines where they exist on a water quality monitoring cruise-by-cruise basis. EPA also identified 13 Chesapeake Bay segments where deep-water and (or) deep-channel designated uses applied during the June-September time period (U.S. EPA 2004, page 5, Figure

II-2, and Table II-1 below). The remaining tidal segments in Chesapeake Bay were characterized as having the open-water designated use year-round.

Table II-1 Chesapeake Bay segments with assigned designated uses.

Designated Use	Segment Code	Location
Deep Water and Deep Channel	CB3MH CB4MH CB5MH CHSMH EASMH PATMH POTMH RPPMH (portion S of UTM Y = 4185000)	Chesapeake Bay Mainstem Chesapeake Bay Mainstem Chesapeake Bay Mainstem Chester River Eastern Bay Patapsco River Lower Potomac River Lower Rappahannock River
Deep Water Only	CB6PH (portion north of UTM Y 4145) CB7PH (portion N/NW of UTM Y = UTM X + 3752745) PAXMH SBEMH YRKPH	Chesapeake Bay Mainstem Chesapeake Bay Mainstem Lower Patuxent River South Branch Elizabeth River Lower York River

Source: U.S. EPA 2004

The EPA Chesapeake Bay Program Office acknowledged this computation code improperly imposed pycnocline presence at times and places where none was found. Such applications of a long-term mean pycnocline instead of no pycnocline were, therefore, incorrectly applying dissolved oxygen criteria assessments in such situations. The EPA published procedures, as described below, allow for the presence of episodic pycnoclines.

Episodic Pycnoclines Criteria Assessment Protocols Modification

The dissolved oxygen criteria assessment methodology is now clarified to specifically allow the deep-water and deep-channel designated uses to occur “episodically” for those 13 segments that have been identified as having deep-water and (or) deep-channel designated uses (see Table II-1 in U.S. EPA 2004) When a pycnocline is observed during the tidal water quality monitoring cruise within one of the 13 segments during June 1 through September 30, the deep-water and (or) deep-channel designated uses exist and their respective numeric dissolved oxygen criteria are applied to those uses. When no pycnocline is observed, the open-water designated use applies to the entire water column. By definition, this approach eliminates the default use of long term pycnocline average when no pycnocline is observed.

Reassessment of previous dissolved oxygen assessments by EPA and its state and District partners showed only small changes in Chesapeake Bay dissolved oxygen criteria attainment results over time. Times and places where no pycnocline could be defined for summer season among the 13 Chesapeake Bay segments with previously defined deep-water and deep-channel designated uses were shown to be rare events.

EXPANDED APPLICATION OF DEEP-WATER AND DEEP-CHANNEL DESIGNATED USES

A total of 13 Chesapeake Bay segments characterized with deep-water and deep-channel designated uses were published in U.S. EPA 2004 (Table II-1). In a number of segments classified as having the open-water designated use only applied year-round, dissolved oxygen criteria assessments through time provided evidence of persistent criteria non-attainment. In a select set of these same Chesapeake Bay segments, results from numerous Chesapeake Bay water quality/sediment transport model scenarios, simulating dissolved oxygen concentrations across a wide range of nutrient load reductions, suggested lack of dissolved oxygen responses to nutrient load reductions due to physical constraints to re-oxygenation. Segments not previously classified with the deep-water and (or) deep-channel designated uses in mesohaline salinities but showing both stratification (presence of a pycnocline) and persistent dissolved oxygen criteria non-attainment were reviewed for possible expanded application of deep-water and deep-channel designated uses.

Review of Designated Use Definitions

The 2003 *Ambient Water Quality Criteria for Dissolved Oxygen, Water Clarity and Chlorophyll a for Chesapeake Bay and Its Tidal Tributaries* highlights two relevant guidelines-stratification (presence of pycnoclines) and evidence of a physical barrier restricting reoxygenation-for determining the need to apply the deep-water and (or) deep-channel designated uses (U.S. EPA 2003). Specifically, the following are published definitions for determining when and where the open-water, deep-water and (or) deep-channel designated uses apply within Chesapeake Bay tidal waters:

Open-Water Designated Use

“If the presence of a pycnocline prevents oxygen replenishment, the open-water fish and shellfish designated use extends only as far as the upper boundary of the pycnocline. If a pycnocline exists but other physical circulation patterns (such as the inflow of oxygen-rich oceanic bottom waters) provide oxygen replenishment to the deep waters, the open-water fish and shellfish designated use extends to the bottom water-sediment interface.” U.S. EPA 2003, Appendix A, page A-6.

(Also see U.S. EPA 2007, pages 37-38, *Dissolved oxygen assessments in shallow versus open waters*, for details regarding the open water designated use definition beyond vertical water column structure.)

Deep-Water Designated Use

“Tidally influenced waters located between the measured depths of the upper and lower boundaries of the pycnocline, where a measured pycnocline is present and presents a barrier to oxygen replenishment from June 1 to September 30...the deep-water designated use extends from the upper boundary of the pycnocline down to the sediment/water interface at the bottom, where a lower boundary of the pycnocline is not calculated.” U.S. EPA 2003, Appendix A, page A-6.

Deep-Channel Designated Use

“Tidally influenced waters at depths greater than the measured lower boundary of the pycnocline in isolated deep channels.” U.S. EPA 2003, Appendix A, page A-6.

Mesohaline Segments Expanded Designated Uses

Using the time period 1991-2000¹, depth profiles of change in density and dissolved oxygen concentrations from the Chesapeake Bay Water Quality Monitoring Program² were reviewed for both evidence of stratification and prevention of re-oxygenation. Chesapeake Bay segments in the mesohaline salinity zone, not previously classified with deep-water and (or) deep-channel designated uses, were evaluated for evidence of stratification and persistent dissolved oxygen criteria non-attainment under a range of different Chesapeake Bay water quality/sediment transport model loading scenarios. Ten segments meeting these characteristics were identified in Maryland and Virginia’s Chesapeake Bay tidal waters (Table II-2).

Table II-2. Ten Chesapeake Bay segments in the mesohaline salinity zone of Maryland and Virginia’s Chesapeake Bay tidal waters reviewed for possible expanded designated use classifications.

Chesapeake Bay Segment	Tidal Water Body
MAGMH	Magothy River
SOUMH	South River
EBEMH	East Branch Elizabeth River
WBEMH	West Branch Elizabeth River
CRRMH	Corrottoman River
FSBMH	Fishing Bay
WICMH	Wicomico River
SEVMH	Severn River
WSTMH	West River
YRKMH	York River

Only the South River (SOUMH) and Magothy River (MAGMH) segments met the deep-water designated use definition originally described in U.S. EPA 2003 where a measured pycnocline was present and presented a barrier to oxygen replenishment during the period June 1 to September 30.

In the South River segment, 39 of 43 depth profiles (91%) had an upper pycnocline and 19 of 43 depth profiles (44%) had a lower pycnocline. In the Magothy River, 16 of 40 depth profiles (40%) had an upper pycnocline and 0 of 40 depth profiles (0%) had a lower pycnocline. Evaluation of the Chesapeake Bay water quality/sediment transport model scenario results for both segments showed depression of dissolved oxygen concentrations with increasing depth suggesting a physical mixing constraint on re-oxygenation due to stratification.

¹ These years of Chesapeake Bay Water Quality Monitoring Program data were selected to be consistent with the hydrologic period for management application of the Chesapeake Bay Water Quality/Sediment Transport Model.

² www.chesapeakebay.net

In the presence of a pycnocline, the deep-water designated use will also apply to the Magothy River and South River mesohaline segments in the June 1 through September 30 time period. The application of the deep-water designated use to these two segments is fully consistent with previously published procedures which called for:

1. the physical exchange of higher oxygenated waters from the upper water-column is much reduced by density stratification, and
2. pycnocline waters are not reoxygenated by riverine or oceanic bottom waters

in order to apply the deep-water designated use (U.S. EPA 2003).

Previously, such segments including the deep-water designated use were only thought to be “located principally in the river channel at the lower reaches of the major rivers and along the spine of the middle mainstem of the Bay” (U.S. EPA 2003). These analyses conducted in support of the development of this addendum have demonstrated the deep-water designated use can occur in smaller tidal tributaries segments receiving limited freshwater flow from their surrounding watershed.

Given the South River segment has a lower pycnocline and 19 of 43 depth profiles (44%) over the 1991-2000 data record, consideration was given to whether a deep-channel designated use should apply to this segment as well as a deep-water designated use. The published procedures for delineating a deep-channel designated use included:

1. The very deep water-column and adjacent bottom surficial sediment habitats located principally in the river channel at the lower reaches of the major river and along the spine of the middle mainstem of the bay;
2. At depths below which seasonal anoxic to severe hypoxic conditions routinely set in and persist for extended periods of time under current conditions; and
3. At depths greater than the lower boundary of the pycnocline (U.S. EPA 2003).

The South River segment does not contain a “very deep water-column” given a total maximum depth of 5 meters and the segment does not have conditions where “seasonal anoxic to severe hypoxic conditions routinely set in and persist for extended periods of time under current conditions”. Therefore, even in the presence of a lower pycnocline, a deep-channel designated use will not be applied to this segment.

The initial review of stratification and dissolved oxygen data from the eight remaining segments identified in Table II-2 did not provide immediate evidence of where stratification appeared to be limiting oxygen replenishment. A more in-depth review of water column stratification conditions and identification of any needs for further adjustments to the applicable designated uses for the remaining segments is planned for completion prior to the 2012 303(d) listing cycle.

LITERATURE CITED

U.S. Environmental Protection Agency. 2003. *Ambient Water Quality Criteria for Dissolved Oxygen, Water Clarity and Chlorophyll a for Chesapeake Bay and Its Tidal Tributaries* April 2003 (Regional Criteria Guidance). EPA 903-R-03-002. Region III Chesapeake Bay Program Office, Annapolis, Maryland.

U.S. Environmental Protection Agency. 2004. *Technical Support Document for Identification of Chesapeake Bay Designated Uses and Attainability-2004 Addendum*. October 2004. EPA 903-R-04-006, Chesapeake Bay Program Office, Annapolis, MD.

U.S. Environmental Protection Agency. 2007. *Ambient Water Quality Criteria for Dissolved Oxygen, Water Clarity and Chlorophyll a for the Chesapeake Bay and Its Tidal Tributaries – 2007 Addendum*. July 2007. EPA 903-R-07-003. Region III Chesapeake Bay Program Office, Annapolis, MD.

U.S. Environmental Protection Agency. 2008. *Ambient Water Quality Criteria for Dissolved Oxygen, Water Clarity and Chlorophyll a for the Chesapeake Bay and its Tidal Tributaries – 2008 Technical Support for Criteria Protocols Addendum*. September 2008. EPA 903-R-08-001. Region III Chesapeake Bay Program Office, Annapolis, MD.

CHAPTER 3

Biologically-based Reference Curves: Revisions to the Methodology and Applications

BACKGROUND

The published dissolved oxygen criteria assessment methodology currently used for assessing Chesapeake Bay water quality criteria attainment involves the use of cumulative frequency distribution (CFD) curves in a two-dimensional space of percent time and percent space (U.S. EPA 2003). Minimum concentrations of dissolved oxygen must be present to support species and their various life stages requiring protection. Dissolved oxygen criteria provide threshold conditions established for the designated uses such that water quality conditions that exceed this threshold are considered impaired.

However, it is recognized that all water quality parameters are inherently variable in space and time. There will be small regions that persistently exceed the threshold due to poor flushing or other natural conditions. The Chesapeake Bay dissolved oxygen criteria have several durations reflecting the various tolerances of different life stages and effects (U.S. EPA 2003, 2008). Small regions or time periods of degraded condition should not lead to a degraded assessment for the segment (U.S. EPA 2003). Recognition that ephemeral exceedances of the threshold in both time and space do not represent persistent impairment of the segment leads to an assessment methodology that allows these conditions to be classed as acceptable while conditions of persistent and wide spread impaired condition will be flagged as unacceptable. (E. Perry, Pers. Comm. 2005).

During an independent scientific peer review of the EPA published CFD procedures, reviewers raised specific concerns about the method for deriving the biological reference curves (STAC 2006). At the time, there were no apparent solutions to resolve the concerns that were raised. However, during recent application of criteria assessment procedures to model simulated outputs, evaluation of the resultant model outputs put the spotlight back on the criteria assessment process and the underlying biological reference curve methodology.

Work by the EPA Chesapeake Bay Program Office and its partners suggested that application of the currently published application of the Benthic-Index of Biotic Integrity (B-IBI) (Weisberg et al. 1997) did not accurately distinguish between healthy and degraded communities with corresponding distinct sets of dissolved oxygen violations. EPA Chesapeake Bay Program Office analysts and partners worked with recognized Chesapeake Bay benthic community experts² to revise the published methods for identifying “healthy” and “degraded” benthic communities. During this process, it was determined that the B-IBI provides a robust delineation of healthy and degraded benthic communities with corresponding distinct dissolved oxygen violation rates.

² Dr. Dan Dauer, Old Dominion University and Dr. Roberto Llanos, Versar, Inc.

Using the newly delineated “healthy” and “degraded” benthic communities, EPA Chesapeake Bay Program Office analysts worked to produce a set of revised biological reference curves that minimize the error in distinguishing between “healthy” and “degraded” segments. In this chapter and its associated appendices, updates to the methodology involving development of biologically-based reference curves with Chesapeake Bay benthic macroinvertebrate monitoring program data are provided. Further, directions on application of reference curves for open-water, deep-water and deep-channel designated uses are provided for completing the Chesapeake Bay dissolved oxygen criteria attainment computations.

ISSUES WITH DISSOLVED OXYGEN CRITERIA ASSESSMENT WITH THE PREVIOUSLY PUBLISHED BIOLOGICALLY-BASED REFERENCE CURVES

The current published method for assessing dissolved oxygen (DO) impairments in Chesapeake Bay incorporates the use of a cumulative frequency distribution as the final step of assessment (U.S. EPA 2003). In this step, a set of DO violation rates for a particular segment-designated use (e.g. “CB4MH Deep Water”) are plotted as a cumulative frequency distribution (CFD) and compared to a “biological reference curve” comprising a cumulative frequency distribution of “acceptable violation rates” of the DO criteria. If the assessment curve exceeds, *at any point*, the reference CFD, then the given segment is considered “impaired (Figure III-1).

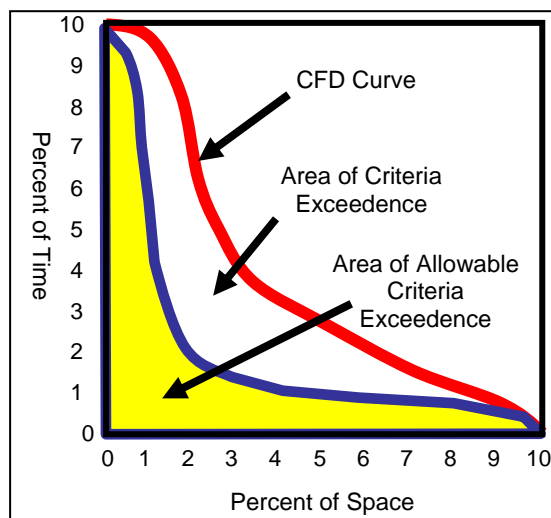


Figure III-1. Conceptual graph illustrating the CFD assessment procedure. The red line is an example of a hypothetical “healthy” assessment curve; the blue line is the hypothetical reference curve.

It has been recognized, however, that by combining violation rates from all healthy areas into one biologically-based reference curve, we create a curve that theoretically represents approximately the median of all curves included. Thus, a large percentage of the presumably “acceptable” violation rate CFDs that were pooled in order to generate the biologically-based reference curve may fail an assessment conducted against that same biologically-based reference curve. A more detailed evaluation confirmed this concern. In Figure III-2 below, the CFD for

CB3MH Deep Water 1987, a segment/designated use considered having a healthy B-IBI for that year, whose acceptable violation rates were included in the generation of the biological reference curve, fails assessment by that same biological reference curve.

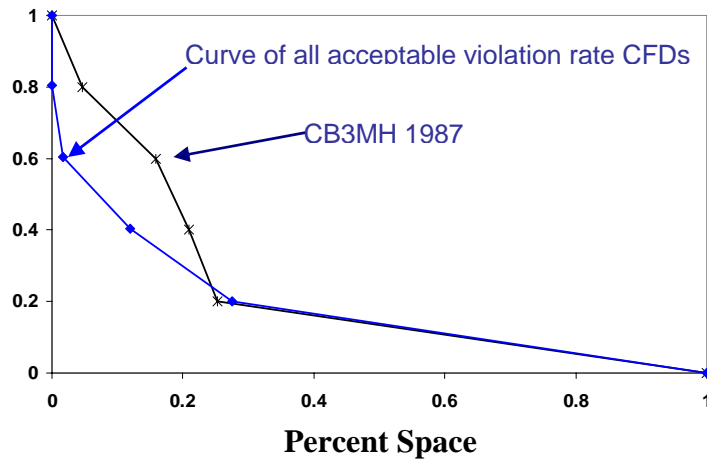


Figure III-2. An example of a 30-day mean deep-water dissolved oxygen criteria and the violation expressed by a healthy segment (CB3MH 1987) curve used in deriving the 30-day mean criterion biologically-based reference curve.

Further analyses revealed that the biological reference curves used for the deep-water and deep-channel dissolved oxygen criteria attainment assessments fail the majority of supposedly “healthy” segment-years used to construct those same curves.

As described in U.S. EPA 2003, the preferred methodology for defining the reference curve is to determine levels of allowable violation based on the demonstrated tolerance of the living resources for whose protection the water quality criteria were designed. Benthic habitat assessments were conducted with the updated methodology, which is described below, for assessing the appropriateness of biologically-based reference curves as indicators of water quality conditions.

UPDATES TO DISSOLVED OXYGEN BIOLOGICALLY-BASED REFERENCE CURVE DERIVATION METHODOLOGY

Based on the findings described above, the following revisions are recommended to the methodology for categorizing benthic communities as “healthy” for the purposes of providing a reference for allowable *frequency* of dissolved oxygen criteria exceedance. The intent of these revisions is to improve the accuracy with which benthic communities are categorized as healthy.

Revisions to the previously published methodology for developing dissolved oxygen biological reference curves include:

- 1) Restriction of the reference dataset to data collected beginning in 1996;
- 2) expansion of time period for classifying benthic community health from 1 year to sequential 3-year time periods;

- 3) restriction of reference segment-periods to those for which at least 10 observations are available;
- 4) refined definition of a “healthy” benthic community as one for which the mean B-IBI score is at least 3.0;
- 5) the standard deviation of the mean is less than 1.0; and
- 6) use Grand Score in computations involving fixed data.

The rationale underlying each of these six modifications is described in further detail below.

Restrict Dataset to Data Collected Beginning in 1996

Criteria violation results of dissolved oxygen criteria attainment assessments are compared with a reference CFD curve (e.g., standard 10% reference or biologically-based reference CFD curve), representing allowable amounts of criteria exceedance in a healthy habitat. When an appropriate biological reference community is identified and sufficient data are deemed available, a biological reference curve of acceptable percent exceedance is generated using a CFD of violation rates for “healthy” biological communities in that designated use. A review of the plotting methodology is provided in U.S. EPA 2008 (see Appendix A).

Historically, the benthic monitoring work of the Chesapeake Bay Benthic Monitoring Program consisted of fixed station monitoring with sampling usually taking place in August and September (Chesapeake Bay Program 1989). The sampling design was primarily intended to assess long-term trends in living resources over decadal, annual and seasonal time scales.

Derivation of the original dissolved oxygen biologically-based reference curves relied on the 1985-2005 Chesapeake Bay benthic monitoring program dataset in order to take advantage of the full two decades of monitoring results. However, data collection methods have undergone revision during the 21 years of monitoring. In 1996, a stratified random sampling component was added to the benthic monitoring program in order to provide confidence limits on estimates of impaired waters in Chesapeake Bay. In order to ensure adequate spatial resolution of benthic community health, STAC (2009), in accordance with recognized Chesapeake Bay benthic community experts, recommended truncating the reference data set to start in 1996 when the updated sampling procedures were established. The data period was extended one year to 2006 to include the most recently available data. The use of the 1996-2006 Chesapeake Bay benthic monitoring program data set is an update to previously published methods (U.S. EPA 2007, Chapter 4). The recommended data set represents a consistent period of improved assessments of Chesapeake Bay health condition.

Use Sequential 3-year Time Periods Rather Than Single Years

The biologically-based reference curve derivation methodology, as outlined in U.S. EPA 2007 (see Chapter 4), used single year assessments to determine the health of the benthic community for the purposes of identifying acceptable dissolved oxygen criteria exceedances. However, dissolved oxygen criteria assessments are conducted on sequential 3-year time frames for each segment (U.S. EPA 2003); two year time steps are used in reporting for 303d listing cycles (e.g. the 2008 303d listing cycle used 2004-2006 data, the 2010 303d listing cycle used 2006-2008

data) while benthic community assessments are conducted annually with annual time steps for a variety of purposes (e.g. indicator reporting for the Chesapeake Bay Barometer). Using sequential 3-year time periods to classify benthic community health, advancing the data in one year time steps (e.g. 1996-1998, 1997-1999, etc.), brings the reference community identification method into better alignment with the dissolved oxygen criteria assessment protocols for which reference communities are being identified. This modification addresses a concern raised by the Chesapeake Bay Program's Scientific and Technical Advisory Committee (STAC) review of the CFD approach (STAC 2006) which noted that sample sizes for reference and assessed conditions should be made similar to reduce the effect of sample size bias on the shape of the CFD. The combination of a segment and sequential 3-year assessment time periods is hereafter referred to as a "segment period".

Screening criteria: Sample Size ≥ 10

Keller and Cavallaro (2008) reported that listing decisions on the U.S. Clean Water Act 303a listing impairments of surface waters by states were often based on insufficient data, or that data were not sufficiently representative of temporal and spatial conditions for the water body being assessed. Llanso et al. (2009), however, require a minimum sample size of $n \geq 10$ for habitat health assessments using the Chesapeake Bay B-IBI. The EPA Chesapeake Bay Program Office and its partners examined the effects of relaxing the data screening criteria to accept segment-period combinations with sample size ≥ 8 to increase the number of "healthy" segment-periods available for reference community analysis.

The decision to eliminate segment-periods with fewer than 10 observations was based on analyses by EPA Chesapeake Bay Program Office and its partners, which showed that fewer than 10 observations weakened the ability of the reference CFD to appropriately classify segments. Llanso et al (2009) confirmed Keller and Cavallaro (2008)'s findings regarding sample size and temporal and spatial distribution. They found that analysis of Chesapeake Bay segments with less than 10 samples produced "inconclusive results relative to the (U.S. EPA) listing process." In their review of the proposed methodology, STAC (2009) determined that a minimum sample size of 10 is reasonable and has been applied elsewhere (Alden et al. 2002). Further details on the sample size analyses are available in Appendix A.

Screening Criteria: Standard Deviation < 1.0

The EPA Chesapeake Bay Program Office and its partners examined the isolated and combined effects of relaxing the data screening criteria to accept segment-periods with fewer samples ($n \geq 8$ instead of 10) and/or expanding the standard deviation criteria surrounding the B-IBI results from <1.0 to ≤ 1.2 in order to increase the number of "healthy" segment-periods for analysis. The relaxation of both the sample size and the standard deviation criteria (see Scenario D in Appendix A) increases the number of segment-periods classified as "healthy" from 10 to 16. However, four of these additional CFD curves extend into "degraded" CFD space to a degree that calls into question the accuracy of their classification as healthy (see Figure A-3 in Appendix A). Defining healthy benthic communities for deriving a benthic community based biological reference curve, therefore, relies on sample size $n \geq 10$ with a standard deviation $<$

1.0. Further details of the sample size and standard deviation analyses are available in Appendix A.

Definition of Healthy Benthic Macroinvertebrate Reference Community Conditions

The methodology described by U.S. EPA (2007, 2008) defined healthy segments (with respect to benthic communities) as those with a minimum B-IBI score ≥ 3.0 . However, no sample size restriction was introduced. As a result, a large segment could contain a single B-IBI score, and if that single score exceeded 3, then the segment was classified as healthy. The likelihood of a degraded segment containing 10 B-IBI scores (in any given 3 years) all of which are ≥ 3.0 is small. Furthermore, benthic community experts (Llanso et al. 2009) have more commonly defined a healthy community as one with a sample mean ≥ 3.0 , given an adequately large sample size and small variance. Thus, the EPA Chesapeake Bay Program Office and its partners now define “healthy” benthic reference communities as those with an average B-IBI score ≥ 3.0 and standard deviation (SD) < 1.0 . STAC (2009) supported the use of “healthy” benthic reference communities defined by those with an average B-IBI score ≥ 3.0 , rather than a minimum, and a standard deviation (SD) < 1.0 , ($n \geq 10$). A degraded benthic community is defined as having an average B-IBI score < 3.0 with a standard deviation < 1.0 , ($n \geq 10$).

The methodological refinements described above led to findings that provide ongoing support of the need for a hyperbolic curve that distributes allowable violations in CFD space, as do both the new deep-water biologically-based reference curve described below and the default 10% reference curve described in U.S. EPA (2007). A more in-depth discussion of the shape of the reference curve with respect to “healthy” and “degraded” CFD-space can be found in Appendix B.

STAC (2009) recommendations suggested that based on the assumptions of normality, the standard deviation criterion applied when classifying a healthy benthic community could alternatively be expressed as “no more than 16% of the sample observations should have a score less than 2.0”. This is a one-sided version of the screening criterion, and addresses concerns that clearly healthy segments with high variance could be excluded from the analyses. EPA Chesapeake Bay Program Office staff conducted an exploratory analysis to classify benthic communities using the following benthic community classification rules:

- 1) average B-IBI score ≥ 3.0 with no more than 16% of sample observations < 2.0 ($n \geq 10$) defines a healthy benthic community, and
- 2) average B-IBI score < 3.0 with no more 16% of sample observations > 4.0 B-IBI score, ($n \geq 10$) defines a degraded benthic community.

Results using these revised classification rules were consistent with the results of the biologically-based reference curve derivation methodology outlined in this chapter.

Use Grand Score in Computations Involving Fixed Station Data

The 1996-2006 Chesapeake Bay B-IBI sample results consist of both fixed station and random station data. These data are combined in the analyses but have different scoring categories within the CIMS database.

Specifically, for the “fixed station” samples both “total_score” and “grand_score” records are reported within the Chesapeake Bay Information Management System (CIMS)³. “Total_score” records are replicate measurements of the same community event; the average of these is reported as the “grand_score.” The Chesapeake Bay benthic experts (R. Llanso, Versar Inc., Pers. Comm. 2009) recommended using the “grand_score” in the Chesapeake Bay dissolved oxygen criteria assessment analyses to avoid errors not accounting for the replicate results of a sampling event. By comparison, random station records in the CIMS database report only a “total score” as the sampling event B-IBI measure; no “grand scores” will be found associated with random station data records.

SUMMARY OF RECOMMENDATIONS

Based on the findings of the analyses and in accordance with the STAC (2009) recommendations, Table III-1 summarizes the revisions to the methodology for identifying dissolved oxygen biologically-based reference curves for Chesapeake Bay water quality criteria attainment assessments (Table III-1).

Table III-1. Chesapeake Bay dissolved oxygen criteria biologically-based reference curve derivation recommendations.

U.S. EPA 2007, 2008 Addenda	2010 Addendum
Obtain dataset of all benthic index of biotic integrity (B-IBI) scores for time period 1985-2005	Restrict dataset to data starting in 1996; for random station data use 'total score' and for fixed station samples use "grand score" only. ¹
<p>For the relevant subset of Chesapeake Bay segments with open-water (OW) and deep-water (DW) and/or deep-channel (DC) designated uses:</p> <p>Match benthic stations and scores in a dataset with monthly open-water, deep-water, and deep-channel designated use boundaries.</p> <ul style="list-style-type: none"> Boundaries are derived using the standardized, automated method for identifying pycnocline boundaries documented in U.S. EPA 2008. <p>Pycnocline boundaries are then interpolated using the interpolator (Visual Basic program, Version 4.61, August 2006, Chesapeake Bay Program Office, as referenced in U.S. EPA 2008, Appendix A, p36).</p> <ul style="list-style-type: none"> Interpolator cells are matched with benthic station locations, and interpolated pycnocline boundaries are applied to each benthic station location. 	Pycnocline boundaries are interpolated using the episodic pycnoclines approach defined in Chapter 2, this addendum.
Benthic stations (and their associated B-IBI scores) are assigned to a designated use: OW, DW, or DC.	No modification recommended
To define the biological reference community for each designated use, all individual segment-years for which the minimum B-IBI was ≥ 3.0 are identified. (Minimum sample size within a segment-year is recognized as $n=1$.) These are denoted as 'healthy' segment-years.	<ol style="list-style-type: none"> Use 3-year rolling time periods rather than single years.² Require a B-IBI score sample size $n \geq 10$.³ "Healthy" reference communities are those with an average B-IBI score ≥ 3.0 rather than a minimum, and standard deviation (SD) < 1.0.⁴
For the 'healthy' segment-years, the monthly (in the case of OW and DW) or instantaneous (DC) dissolved oxygen criteria violation rates are obtained based on the water quality profiles of sampling data collected by the Chesapeake Bay long term water quality monitoring program.	No modification recommended.
These season- and designated use-specific Chesapeake Bay dissolved oxygen criteria violation rates (e.g. percentage of a segment-designated use volumes failing the DO criteria in a given month; thus 4 measures per summer for OW and DW – June thru September) are used to define "acceptable" exceedances of the dissolved oxygen criteria. This definition of acceptable exceedances in space and time is based on the logic that if a healthy benthic community existed in the segment-designated use in that summer, then the degree of DO criteria violation that occurred did not lead to an impaired benthic community.	No modification recommended.

Source: U.S. EPA 2003, 2007, 2008.

1. Restrict dataset to 1996-2006 time period. For the fixed station samples included in the analyses, use the “grand score” only from the CIMS data base with these samples.
2. Use 3-year rolling time periods rather than single years. This brings the reference community ID method into better alignment with the DO criteria assessment method for which reference communities are being identified.
3. Require a B-IBI score sample size ≥ 10 . This improves the spatial representation of the B-IBI score.
4. “Healthy” reference communities are those with an average B-IBI score ≥ 3.0 , and a standard deviation (SD) < 1.0 , rather than a minimum. Using the average is consistent with published methods used by Chesapeake benthic experts to assess benthic communities (e.g. Llanso et al. 2009).

APPLICATION OF A REFERENCE CURVE FOR OPEN-WATER 30-DAY MEAN DISSOLVED OXYGEN CRITERIA: SUMMER SEASON

Reference curves for the 30-day mean open-water dissolved oxygen criterion (June 1-September 30 only) were based on criteria levels that would not impair biological communities (U.S. EPA 2003). Analyses conducted by the EPA Chesapeake Bay Program Office and its partners suggest that the B-IBI does not provide an appropriate reference community for assessment of open-water dissolved oxygen criteria violations. Even with the latest improvements in the assessment methodology to distinguish between healthy and degraded benthic communities, Figure III-3 illustrates that the health of the benthic community is not an appropriate indicator of open-water low dissolved oxygen conditions as defined by the summer season open-water 30-day mean dissolved oxygen criterion. This result is demonstrated by the cloudplot (Figure III-3) representing Chesapeake Bay Program segments deemed “healthy” and “degraded” according to the updated assessment methodology.

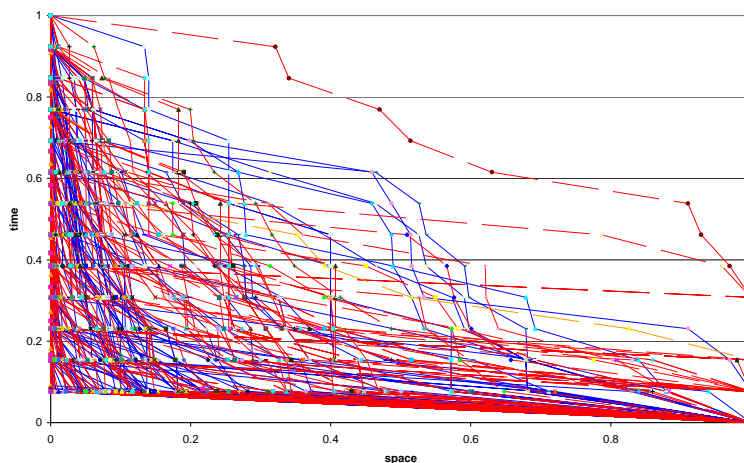


Figure III-3. Open water “healthy” and “degraded” benthic communities are not distinguished by violations of the summer open water DO 30-day mean criterion

APPLICATION OF A REFERENCE CURVE FOR DEEP-WATER MEAN DISSOLVED OXYGEN CRITERIA

Reference curves for the 30-day mean deep-water dissolved oxygen criteria (June 1-September 30 only) were based on criteria levels that would not impair biological communities (U.S. EPA 2003). Reference areas for derivation of the original 2003 published deep-water biologically-

based reference curves were identified using a measure of benthic community health – the Chesapeake Bay B-IBI (Weisburg et al. 1997). Using the revised methodology outlined in this addendum chapter, the EPA Chesapeake Bay Program Office and its partners identified two distinct sets of “healthy” and “degraded” (average B-IBI < 3.0, SD < 1.0) benthic communities, with correspondingly distinct violation rates (Figure III-4). The EPA Chesapeake Bay Program Office, in coordination with its partners, further determined that a reference curve constructed from the 100th percentile of healthy violation rates (x) for each point in time (y) accurately distinguished between healthy and degraded benthic communities with zero error in classification.

A step-by-step guide to the derivation of this curve (Figure III-4), including the x-y coordinate values for plotting the curve, is provided in Appendix C.

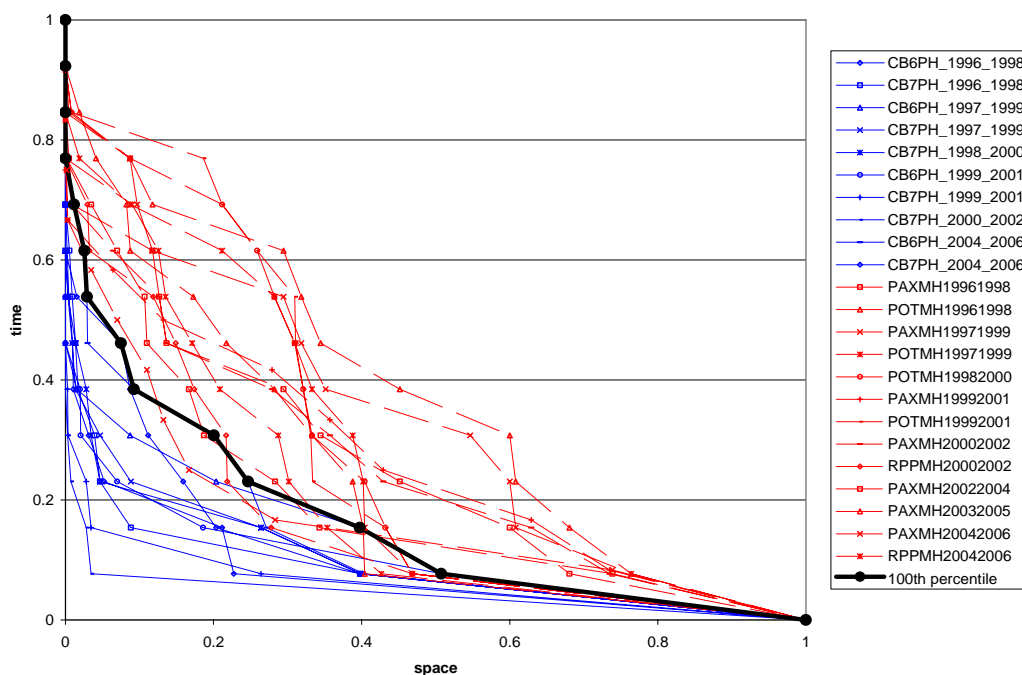


Figure III-4. Dissolved Oxygen deep-water criteria violation rates corresponding to healthy (blue) and degraded (red) benthic communities. Recommended new deep-water biologically-based reference curve represented by the 100th percentile of healthy violations is shown in black.

APPLICATION OF A REFERENCE CURVE FOR DEEP-CHANNEL INSTANTANEOUS MINIMUM DISSOLVED OXYGEN CRITERIA

In the case of the deep-channel instantaneous minimum dissolved oxygen criterion, the application of a biological reference curve was recommended in U.S. EPA 2007 (p. 43). Appendix D provides a more thorough review of history of EPA guidance regarding application of the deep-channel reference curve.

This 2007 recommendation for application of a biologically-based reference curve for assessment of the deep-channel dissolved oxygen criterion was based on the identification of a

small number of deep-channel segment-periods within which the benthic communities were categorized as “healthy” and, therefore, appropriate for use as a biological reference. These benthic communities were categorized using the methodology described on pp. 39-41 of U.S. EPA 2007.

The revised methodology published in this addendum was applied to derive a new deep-channel biologically-based reference curve. The revised method yielded no segment-periods meeting the revised criteria outlined in Table III-1 above. This suggests that the occurrence of healthy benthic communities in the deep-channel designated use are currently insufficient to identify a corresponding set of “acceptable” violations of the instantaneous minimum dissolved oxygen criteria to develop a biologically-based reference curve.

Comparisons of Degraded Reference Benthic Communities with the Published Deep-Channel Reference Curve

While no benthic communities could be categorized as “healthy” in the most recent review, 25 “degraded” reference benthic community segment-periods were identified. The EPA Chesapeake Bay Program Office and its partners conducted and reviewed analyses that showed that all 25 segment-periods (in the 1996-2005 time period) for which deep-channel benthic communities were categorized as “degraded” failed a dissolved oxygen criteria assessment conducted using the 10% default reference curve (Figure III-5).

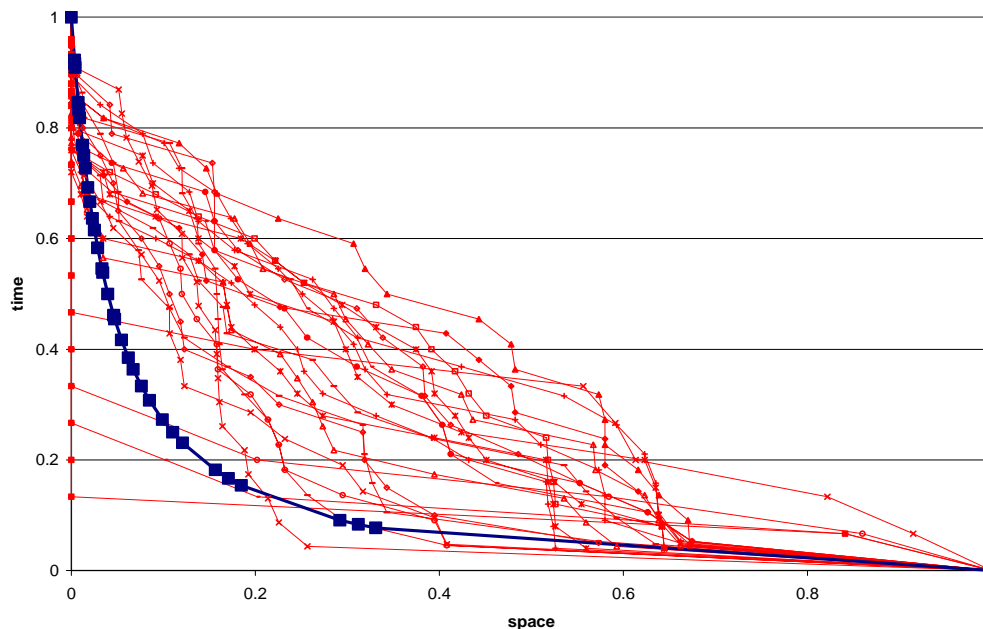


Figure III-5. CFD graph of deep-channel instantaneous minimum dissolved oxygen criterion violation rates corresponding to benthic communities categorized as “degraded” (red lines) in relation to the 10% default reference curve (blue line).

In the absence of a suitable reference community, a biological reference curve for the deep-channel instantaneous minimum dissolved oxygen criterion can not be constructed at this time. Under these circumstances, “a default reference curve such as the normal distribution curve

representing approximately 10 percent exceedance is appropriate in this case to account for anticipated natural criteria exceedances” (U.S. EPA 2003; p. 173).

Rationale for Acceptable Exceedances of the Deep Channel Instantaneous Minimum Dissolved Oxygen Criterion

EPA determined that there are allowable criteria exceedances that would not adversely effect protection of the designated use. As documented on p. 168 in U.S. EPA 2003:

“The recommended criteria attainment assessment approach is designed to protect the living resources as defined by the designated uses. The criteria levels themselves were largely based on scientific studies performed in laboratory settings or under controlled field conditions. The criteria establish the level of a given habitat condition that living resources need for survival. They do not account for many other environmental factors that could affect survival.

Reference curves were developed to provide a scientific-based, direct measure of the ‘allowable’ criteria exceedances. These exceedances are defined to be those that last a short enough time or cover a small enough area to have no adverse effects on the designated use. It is assumed that the designated uses can be attained even with some limited level of criteria exceedances and thus, the reference curves define those criteria exceedances deemed to be allowable—chronic in time but over small areas, or infrequent occurrences over large areas. Exceedances that occur over large areas of space and time would be expected to have significant detrimental effects on biological communities, which would imply nonattainment of designated uses.”

As reported in a recent paper on the Chesapeake Bay dissolved oxygen criteria by the key members of the original Chesapeake Bay dissolved oxygen (DO) criteria team (Batiuk et al. 2009):

“Unlike chemical contaminants or other more conventional pollutants, there were no clear, well established guidelines for deriving criteria for DO, particularly for estuarine waters inhabited by fresh-water and marine species. The goal in setting Chesapeake DO criteria was to use the best science possible to define conditions that would improve or sustain the suitability of Chesapeake Bay habitats for finfish and invertebrates, with the states ultimately factoring in consideration of attainability in adopting the criteria as water quality standards. Thus, we developed criteria that would greatly increase the spatial and temporal extent of Bay waters in which oxygen concentrations were not major limitations to growth and survival of organisms dependent on particular Bay habitats. We did not, however, derive criteria that would require oxygen concentrations high enough at all times and in all locations such that no organism would be negatively affected in any location in the Bay. The states and U.S. Environmental Protection Agency (EPA) determined that such conditions would not be achievable either economically nor technologically (U.S. EPA, 2003d) and may not, in fact, reflect pre-historical conditions of Chesapeake Bay, which

showed that low oxygen conditions, although not nearly as severe as today, may have been a historical feature in the deep channel of the bay (Cooper and Brush, 1991; Karlson et al., 2000; Adelson et al., 2001; Zimmerman and Canuel, 2002; Bratton et al., 2003; Colman and Bratton, 2003; Cronin and Vann, 2003; Zheng et al., 2003).”

In support of the deep-channel instantaneous minimum criterion of 1 mg/L, U.S. EPA 2003 summarized findings published in peer-reviewed literature sources indicating that several keystone benthic species “are resistant to dissolved oxygen concentrations as low as 0.6 mg/L,” and that “extensive mortality is likely only under persistent exposure to very low dissolved oxygen concentrations at high summer temperatures” (p. 61).

In light of both (1) the recognition that low dissolved oxygen conditions are a ‘pre-historical’ feature of these deep channel habitats, and (2) the observation that keystone benthic species of these deep channel habitats can tolerate small-scale occurrences of severe hypoxia (DO concentrations below 1 mg/L), EPA believes that an allowance for a small, limited set of exceedances in time and space is acceptable in assessment of the deep-channel designated use dissolved oxygen criterion.

ASSESSMENT OF SUMMER SEASON DISSOLVED OXYGEN CRITERIA

EPA recommends revising previously published dissolved oxygen criteria assessment guidance as described in Table III-2, including applying the default 10% reference curve for assessment of summer season (June 1- September 30) 30-day mean open-water and instantaneous minimum deep-channel dissolved oxygen criteria. The 30-day mean deep-water dissolved oxygen criterion biologically-based reference curve, as described in this addendum, is recommended for use assessing attainment of this criterion. Until EPA publishes methodologies for assessing the 7-day mean, 1-day mean and instantaneous minimum open-water and deep-water dissolved oxygen criteria, respectively, the Agency recommends that the states and the District of Columbia rely strictly on the assessment of the 30-day mean open-water and deep-water dissolved oxygen criteria for listing decisions (U.S. EPA 2007). The previously published non-summer open-water dissolved oxygen criteria reference curve remains unchanged as the 10% default reference curve (U.S. EPA 2007, p.42).

Table III-2. EPA recommended reference curves for conducting 303(d) list Chesapeake Bay dissolved oxygen criteria assessments.

Season and Designated Use	U.S. EPA 2007 July Addendum Reference Curve	U.S. EPA 2010 Addendum Reference Curve
Summer Open Water (30-day mean)	Published biologically-based reference curve. Refer to U.S. EPA 2007 p. 41, Figure IV-2.	Published default 10% reference curve Refer to U.S. EPA 2007, p13, Figure II-4 and Equation 1.
Non-summer Open Water (30-day mean)	Published default 10% reference curve. Refer to U.S. EPA 2007, p.13, Figure II-4 and Equation 1.	Published default 10% reference curve Refer to U.S. EPA July 2007, p.13, Figure II-4 and Equation 1.
Summer Deep Water (30-day mean)	Published biologically-based reference curve. Refer to U.S. EPA 2007, p. 41, Figure IV-3.	Revised Biologically-based reference curve, Figure III.4, and Appendix C. this document.
Summer Deep Channel (instantaneous minimum)	Published Biologically-based reference curve. Refer to U.S. EPA 2007 p. 42 Figure IV-4 and Appendix F and G.	Published default 10% reference curve Refer to U.S. EPA 2007, p13, Figure II-4 and Equation 1.

Sources: U.S. EPA 2003, 2007.

LITERATURE CITED

Adelson, J. M., G. R. Helz and C. V. Miller. 2000. Reconstructing the rise of recent coastal anoxia; molybdenum in Chesapeake Bay sediments. *Geochemica et Cosmochemica Acta* 65:237-252.

Alden III, R.A., D.M. Dauer, J.A. Ranasinghe, L.C. Scott, and R.J. Llanos. 2002. Statistical verification of the Chesapeake Bay benthic index of biotic integrity. *Environmetrics* 13:473-498.

Bratton, J. F., S. M. Colman, R. R. Seal and P. C. Baucom. 2003. In press. Eutrophication and carbon sources in Chesapeake Bay over the last 2,700 years: Human impacts in context. *Geochimica et Cosmochimica Acta*.

Batiuk, R.A., D.L. Breitburg, R.J. Diaz, T.M. Cronin, D.H. Secor and G. Thursby. 2009. Derivation of habitat-specific dissolved oxygen criteria for Chesapeake Bay and its tidal tributaries. *J. Exp. Mar. Biol. Ecol.* 381:S204-S215.

- Chesapeake Bay Program. 1989. *Chesapeake Bay Basin Monitoring Program Atlas. Volume II. Biological and Living Resource Monitoring Programs*. CBP/TRS 35/89. Annapolis, Maryland.
- Cooper, S. R. and G. S. Brush. 1991. Long-term history of Chesapeake Bay anoxia. *Science* 254:992- 996.
- Colman, S. M. and J. F. Bratton. 2003. Anthropogenically induced changes in sediment and biogenic silica fluxes in Chesapeake Bay. *Geology* 31(1):71-74.
- Cronin, T. M. and C. Vann. 2003. The sedimentary record of anthropogenic and climatic influence on the Patuxent estuary and Chesapeake Bay ecosystems. *Estuaries* 26 (2A).
- Karlsen, A. W., T. M. Cronin, S. E. Ishman, D. A. Willard, R. Kerhin, C. W. Holmes and M. Marot. 2000. Historical trends in Chesapeake Bay dissolved oxygen based on benthic foraminifera from sediment cores. *Estuaries* 23:488-508.
- Keller, A.A and L. Cavallaro. 2008. Assessing the US Clean Water Act 303d listing process for determining impairment of a waterbody. *J. Env. Manage.* 86:699-711.
- Llanso, R.J., D.M. Dauer and J.H. Volstad. 2009. Assessing ecological integrity for impaired waters decisions in Chesapeake Bay, USA. *Marine Pollution Bulletin* 59:48-53
- Scientific Technical Advisory Committee. 2006. *The Cumulative Frequency Diagram Method for Determining Water Quality Attainment: Report of the Chesapeake Bay Program STAC Panel to Review of Chesapeake Bay Program Analytical Tools*. STAC Publication 06-003. 74 pp.
- Scientific Technical Advisory Committee. 2009. Application or reference curves in dissolved oxygen criteria assessment. STAC Review and recommendations for the Chesapeake Bay Program. STAC Publ. 09-005. <http://www.chesapeake.org/stac/Pubs/biorefcuresreview.pdf>
- U.S. Environmental Protection Agency. 2003. *Ambient Water Quality Criteria for Dissolved Oxygen, Water Clarity and Chlorophyll a for the Chesapeake Bay and Its Tidal Tributaries (Regional Criteria Guidance)*. April 2003. EPA 903-R-03-002. Region III Chesapeake Bay Program Office, Annapolis, MD.
- U.S. Environmental Protection Agency. 2007. *Ambient Water Quality Criteria for Dissolved Oxygen, Water Clarity and Chlorophyll a for the Chesapeake Bay and Its Tidal Tributaries – 2007 Addendum*. July 2007. EPA 903-R-07-003. Region III Chesapeake Bay Program Office, Annapolis, MD.
- U.S. Environmental Protection Agency. 2008. *Ambient Water Quality Criteria for Dissolved Oxygen, Water Clarity and Chlorophyll a for the Chesapeake Bay and Its Tidal Tributaries-2008 Technical Support for Criteria Assessment Protocols Addendum*. September 2008. EPA 903-R-08-001. Region III Chesapeake Bay Program Office, Annapolis, MD.

Weisburg, S.B., J.A. Ranasinghe, D.M. Dauer, L.C. Schaffner, R.J. Diaz and J.B. Frithsen. 1997. An estuarine benthic index of biotic integrity (B-IBI) for Chesapeake Bay. 1997. *Estuaries* 20:149-158.

Zheng, Y., Weinman, B., Cronin, T., Fleisher, M.Q., Anderson, R.F., 2003. A rapid procedure for thorium, uranium, cadmium, and molybdenum in small sediment samples by inductively coupled plasma-mass spectrometry: application in Chesapeake Bay. *Appl. Geochem.* 18, 539–549.

Zimmerman, A.R., Canuel, E.A., 2002. Sediment geochemical records of eutrophication in the mesohaline Chesapeake Bay. *Limnol. Oceanogr.* 47, 1084–1093.

CHAPTER 4

Revisions to the Chlorophyll *a* Criteria Assessment Methodology

BACKGROUND

In the 2003 *Ambient Water Quality Criteria for Dissolved Oxygen, Water Clarity and Chlorophyll *a* for Chesapeake Bay and Its Tidal Tributaries*, EPA published narrative chlorophyll *a* criteria that states chlorophyll *a*

“...shall not exceed levels that result in ecologically undesirable consequences—such as reduced water clarity, low dissolved oxygen, food supply imbalances, proliferation of species deemed potentially harmful to aquatic life or humans or aesthetically objectionable conditions – or otherwise render tidal waters as unsuitable for designated uses balanced aquatic plant life populations and against the overgrowth of nuisance, potentially harmful species” (U.S. EPA 2003).

From 2004 to 2006, Virginia and the District of Columbia adopted numerical chlorophyll *a* criteria for application in the tidal James River (Virginia) and across the District’s jurisdictional tidal waters. In the 2007 publication *Ambient Water Quality Criteria for Dissolved Oxygen, Water Clarity and Chlorophyll *a* for Chesapeake Bay and its Tidal Tributaries-2007 Addendum*, EPA published chlorophyll *a* criteria assessment procedures (U.S. EPA 2007a, p. 62). With the establishment of numerical chlorophyll *a* concentration-based criteria promulgated by the states into their water quality standards regulations within Chesapeake Bay tidal waters, it was necessary to establish a reference curve for use in the published criteria attainment process (U.S. EPA 2003).

A biologically-based reference curve with which to assess chlorophyll *a* criteria attainment in Chesapeake Bay is not yet available. A dataset has not been identified from which there is confidence that a biological reference curve can be derived (U.S. EPA 2007b). The EPA Chesapeake Bay Program Office and its partners, in consultation with regional experts in phytoplankton and chlorophyll *a* monitoring and research, have explored the published work of Buchanan et al. 2005 and Lacouture et al. 2006 conducted during development of the phytoplankton index of biotic integrity (P-IBI).

In its current form, the published P-IBI work does not provide for a suitable representation of the integrated seasonal biological community conditions necessary to inform appropriate seasonal reference conditions for Chesapeake Bay chlorophyll *a* criteria attainment assessments. Benthic macroinvertebrates, for example, have life spans that integrate temporally variable environmental conditions over space, and the effects of multiple types of environmental stress and habitat alteration as used with the B-IBI for Chesapeake Bay (Llanos et al. 2009). However, standing crops of phytoplankton communities will respond to nutrient perturbations in 10-14 days (Heiskary and Walker 1995). Tracking the P-IBI results indicates any given segment can and

does move in and out of degradation within a single spring or summer season. Thus, the P-IBI does identify instances of high quality conditions, but currently does not provide the characteristics of a season-long “healthy” condition in terms of allowable exceedances that could be used to support derivation of a biologically-based reference for chlorophyll *a* criteria assessments.

Further work is needed to specify a metric that can provide *a priori* identification of an unimpaired system on the relevant timescale, from which allowable exceedance of the chlorophyll *a* criteria can then be inferred. EPA, therefore, recommends a default 10% reference curve for assessing the chlorophyll *a* criteria (U.S. EPA 2007a, Figure II-4, and Equation 1).

REVIEW OF THE CURRENT CHLOROPHYLL A CRITERIA ATTAINMENT ASSESSMENT PROCEDURE: METHOD AND ASSUMPTIONS

In Table IV-1, the current Chesapeake Bay chlorophyll *a* attainment assessment procedure is outlined for developing a seasonal mean for a Chesapeake Bay management segment to compare with published numerical criteria (e.g., Virginia tidal James River and the District of Columbia’s tidal waters, U.S. EPA 2007a, Appendix C).

Table IV-1. Outline of the previously published Chesapeake Bay chlorophyll *a* criteria attainment assessment methodology.

Outline of chlorophyll <i>a</i> attainment assesment steps	Comments
Chlorophyll <i>a</i> data used for scenario assessments comprise all chlorophyll <i>a</i> values in the CIMS water quality database with layer flagged “S” for surface.	For Virginia chlorophyll <i>a</i> assessments, use all publically available and appropriate surface data, i.e., CIMS data plus VIMS/HRSD DATAFLOW data. (U.S. EPA 2008 p.30).
Data are organized into individual “cruise” files for interpolation.	
Individual cruise files are interpolated using the Chesapeake Bay Interpolator (version 4.61), with the “ln-transform” and the “2-D Inverse-Distance Squared” options selected.	The Interpolator automatically back-transforms chlorophyll <i>a</i> values in its output files. (U.S. EPA 2008 p.30).
Interpolated chlorophyll <i>a</i> surfaces are averaged for an entire season (on a cell-by-cell basis).	The current methodology calculates an arithmetic mean on the back-transformed chlorophyll <i>a</i> values.
Seasonal means are assessed (cell-by-cell) against the criterion for the relevant river segment-season. Assessment curves were compared against a default reference curve. Non-attainment is calculated by subtracting the area of the reference curve from the area under the chlorophyll <i>a</i> criteria assessment curve*.	* If the assessment curve exceeds, at any point, the reference CFD, then the given segment is considered “impaired”.

Source: U.S. EPA 2008

To review the method details, U.S. EPA (2007a, p. 62) first states in the assessment procedure:

“Assessments of seasonal mean chlorophyll *a* criteria should be based on seasonal averages of interpolated data sets. To calculate seasonal averages, each interpolated cruise within a season should be averaged on a point-by-point basis in matching interpolator grid cells. Spatial violation rates should be calculated for each seasonally aggregated interpolation in an assessment period. For example summer open water seasonal chlorophyll *a* criteria

assessment of a three-year assessment period, three seasonal average interpolations representing each season (Year 1 summer, Year 2 summer, Year 3 summer) should be used.”

In the publication *Ambient Water Quality Criteria for Dissolved Oxygen, Water Clarity and Chlorophyll a for Chesapeake Bay and its Tidal Tributaries–2008 Technical Support for Criteria Assessment Protocols Addendum*, EPA provided further details documenting the chlorophyll *a* criteria assessment procedures (U.S. EPA 2008). Chapter 5 (U.S. EPA 2008, pp. 30-32) reviews the chlorophyll *a* criteria procedural steps to assess attainment while Appendix G (U.S. EPA 2008) provides a highly detailed step-by-step process for completing the chlorophyll *a* criteria assessments. The application of data transformations to the chlorophyll *a* assessment data sets occurs during analyses in the process of calculating the seasonal mean (U.S. EPA 2008). Chapter 5, p. 30, Step 4 (U.S. EPA 2008) highlights the use of such a transformation on chlorophyll *a* data and states:

“Data sets are imported into the Chesapeake Bay interpolator and transformed (natural log) prior to interpolation, as chlorophyll *a* measurements tend to follow a log-normal distribution. The program defaults for search area (25 km²) and maximum sample size (4) are used, and the ‘2D Inverse Distance Squared’ algorithm is chosen. The Interpolator automatically back-transforms interpolated estimates before creating the output files.”

Table IV-1 above shows the next step of computing a seasonal mean requires computation of an arithmetic mean over time at each point in the spatial interpolations represented by the 30-day means for the appropriate chlorophyll *a* criteria assessment season.

First, while the mean is often used to report central tendency, for skewed data the arithmetic mean may not be in accord with the notion of ‘middle’. Skewed data make it unsuitable to estimate quantiles, proportions or means by normal distribution expectations (Gilbert 1987), i.e. an arithmetic mean. Tett and Wallis (1995) cite Barnes (1952) as indicating it is common for the variance of measurements on phytoplankton to be dependent on the mean. Sokal and Rohlf (1969) recommend logarithmic transformation of data exhibiting such characteristics.

The previously published protocols for assessing Chesapeake Bay chlorophyll *a* criteria attainment were inconsistent in carrying out the seasonal mean computations since spatial interpolations are conducted on log transformed data while temporal averaging is conducted on untransformed data (U.S. EPA 2003, 2007a, 2008). Bland and Altman (1996) recommend that once data are transformed, carrying out all calculations on the transform scale and transform back once one has calculated the confidence intervals of the sample mean.

Transformations on data provide the ability to approximate a statistical distribution based on the analyses to be performed using established inferential statistical procedures. When there is substantial skew in the data it is common to transform the data to a symmetric distribution. Analyses conducted with data approximating a normal distribution throughout the calculations then support the use of a wide array of well known statistical inference procedures based on well established statistics of the normal distribution.

Second, there is an underlying assumption to the calculations conducted as defined in the U.S. EPA 2007a and 2008 chlorophyll *a* criteria assessments that Chesapeake Bay chlorophyll data show log-normal tendencies. Based on this assumption, analyses depend on log-transforming of chlorophyll *a* data to provide a reasonable approximation of the normal distribution and support the use of normal distributional inference procedures. There is use of log-transformation chlorophyll *a* data in the Chesapeake Bay criteria literature cited in U.S. EPA 2007b, and there is a suggestion for positive skewness for chlorophyll *a* data shown with a hypothetical chlorophyll *a* data distribution (U.S. EPA 2007b). However, there is little background documenting the statistical distributional characteristics of Chesapeake Bay chlorophyll *a* data within the *Ambient Water Quality Criteria for Dissolved Oxygen, Water Clarity and Chlorophyll a for Chesapeake Bay and its Tidal Tributaries* publication series (U.S. EPA 2003, 2007a, 2007b, 2008).

The following sections address: 1) peer-reviewed supporting literature regarding skewness and non-normality issues of chlorophyll *a* data; 2) log-normal transformation applications during analyses of Chesapeake Bay and other chlorophyll *a* data; and 3) recommended refinements to the published criteria assessment procedures. All these sections are directed towards providing consistency in computing the season mean of the 3-year assessments in logarithmic-space, thereby providing a sound estimate of central tendency for the final chlorophyll *a* assessment measures with the seasonal mean criteria.

CHLOROPHYLL A: DATA SKEWNESS, LOG TRANSFORMATION AND THE SEASONAL MEAN CALCULATION

Log-normal Character of Chlorophyll *a* Data

Support for the log-normal characteristics of chlorophyll *a* data have been published in the peer-reviewed scientific literature across a diversity of ecosystems. Harris (1986, Figure 9.7) illustrates seasonally dependent log-normal chlorophyll results for Hamilton Harbor (Lake Ontario). Vollenweider and Krekes (1980), as cited in Harris (1986), noted that algal biomass data from lakes was log-normally distributed. Recent work on Colorado lakes ($n = 20$) showed 19 of 20 lakes chlorophyll measurements were well fit with log-normal transformations to approximate the normal distribution⁴.

Within Chesapeake Bay, Jordan et al. (1991) describe correlations between watershed discharges and chlorophyll concentrations as complicated by non-normal distributions. Jordan et al. (1991) used the Box-Cox method (Sokal and Rohlf 1981) to identify the best power transformation for normalizing the data which was a log transformation. Harding (1994) showed that frequency distributions of chlorophyll and nutrient concentrations in Chesapeake Bay data were skewed; logarithmic transformations of the data produced normal distributions.

⁴<http://www.chatfieldwatershedauthority.org/pdf/Characterizing%20Chlorophyll%20Distributions%20in%20Colorado.pdf>

Log-transforming Chesapeake Bay water quality indicator data (including chlorophyll) was integral to improvements of the Relative Status Indicator during its evolution (Olson 2009). Initially, Olson (2009) reports that positive skewed data led to unequal data distributions

affecting the outputs resulting in too many areas characterized as “good” when they were clearly unsatisfactory. Modifications applied to indicator calculations from 1998-2000, benchmark and status data sets (3-year windows) were log-transformed prior to analysis to address data skewness issues negatively impacting equality of data distribution characterizations. It was thus noted “that for water quality parameters the log and square root transformations are about equal in effecting a normal distribution of the data, and more effective than inverse transformations or using untransformed data” (Olson 2009).

U.S. EPA (2007b) extended the published analyses of Harding (1994) and Harding and Perry (1997) modeling historical chlorophyll *a* data using a Generalized Linear Model (GLM) for log₁₀ (chlorophyll *a*). In deriving reference chlorophyll *a* criteria thresholds for Chesapeake Bay, thresholds were recommended as being derived by a model for the desired mean level of chlorophyll *a* in log space (U.S. EPA 2007b, page 17). Tables III-2 and III-3 in U.S. EPA 2007b, page 18) illustrate reference condition recommendations in log transform space mean chlorophyll and back transformed means. Recommendations for harmful algal bloom based chlorophyll *a* criteria in tidal fresh and oligohaline waters of Chesapeake Bay were further dependent upon log-transformed chlorophyll *a* analyses in their development (U.S. EPA 2007b).

James River Focused Analyses of Log-transformed Chlorophyll *a* Data for Normality

Tidal James River chlorophyll *a* data (1991-2000, n = 828) were log-transformed; natural logarithms were used. A Generalized Linear Modeling (GLM) approach was used to test chlorophyll *a* data for normality. Statistical Analysis Software (SAS) was used in the analysis. Seven Chesapeake Bay segments were included in the analysis: Mouth of Chesapeake Bay (CB8PH), Mouth to mid-Elizabeth River (ELIPH), Southern Branch Elizabeth River (SBEMH), Mouth of the James River (JMSPH), Lower James River (JMSMH), Middle James River (JMSOH) and Upper James River (JMSTF). Segments were grouped into one of four groups depending on similarity of their variances:

"JMSPH"	then SegGrp = 1;
"JMSMH" "SBEMH"	then SegGrp = 2;
"JMSOH" "CB8PH" "ELIPH"	then SegGrp = 3; and
"JMSTF"	then SegGrp = 4.

The GLM model was $\ln(\text{chlorophyll}) = \text{year, segment}$. (Equation 1)

Data was analyzed by season. Spring was defined as March, April and May with summer defined as July, August and September. Normality diagnostics were reviewed for the raw residuals.

For Spring and Summer seasons within the tidal James River, even without standardizing for heterogeneous variance, the $\ln(\text{chl}a)$ residuals from the GLM model results show a fairly close approximation to a normal distribution. The normal probability plot shows very high concordance between the expected residuals and the observed residuals except for two outlier points in the extreme tails of the sample. These outliers probably reflect a failure of the simple model to capture some extreme event rather than a failure of log normality. The Shapiro-Wilk statistic of 0.994 (spring) and 0.988 (summer) shows that the residuals are very highly correlated with the expected residuals for approximating a normal distribution (see Appendix C). The

Shapiro-Wilks statistic ranges from 0 to 1 where 0 is farthest from normality and 1 is high fidelity with a normal distribution. The normality test p-value suggests a statistically significant departure from normality but this is not surprising with a sample size $n = 828$. The Shapiro-Wilks test is sensitive to small departures from normality with large sample sizes. The large sample size gives one the power to detect very small statistical differences from normality that, for analysis of the transformed data, are of low practical significance. Further details of the test output are provided in Appendix D. The SAS programs are included in Appendix E.

CHLOROPHYLL A CRITERIA ASSESSMENT PROTOCOL REFINEMENTS USING LOG-TRANSFORMATIONS

Statistical treatment of chlorophyll *a* data from a review of non-Chesapeake Bay and Chesapeake Bay specific peer reviewed scientific literature and U.S. EPA Chesapeake Bay criteria documentation: 1) supports a common recognition of skewness with chlorophyll *a* data sets; and 2) shows a long history with the application of log-transformations for analyses. Bland and Altman (1996) recommend carrying out all calculations on the transform scale and transform back once one has calculated the confidence intervals of the sample mean. Log transformation of data during analyses to better reflect a normal distribution then better support the inference procedures based on normal distributions. The chlorophyll *a* criteria assessment protocol modifications described here (Table IV-2) constitute a more consistent and technically sound calculation than the currently published EPA methods (U.S. EPA 2003, 2007a, 2008). Analyses conducted with data approximating a normal distribution throughout the calculations supports the use of a wide array of statistical inference procedures based on normal distributions. Tidal James River chlorophyll *a* data was evaluated and showed fidelity to the normal distribution.

Table IV-2. Previously published Chesapeake Bay chlorophyll *a* criteria assessment methods and recommended modifications.

U.S. EPA 2008 Addendum	U.S. EPA 2010 Addendum
1. Chlorophyll <i>a</i> data used for scenario assessments comprise all chlorophyll <i>a</i> values in the CIMS water quality database with layer flagged “S” for surface.	No modification recommended.
2. Data are organized into individual “cruise” files for interpolation.	No modification recommended.
3. Individual cruise files are interpolated using the Chesapeake Bay Interpolator (version 4.61), with the “ln-transform” and the “2-D Inverse-Distance Squared” options selected. <i>The Interpolator automatically back-transforms chlorophyll a values in its output files.</i>	No modification recommended.
4. Interpolated chlorophyll <i>a</i> surfaces are averaged for an entire season (on a cell-by-cell basis). <i>The current methodology calculates an arithmetic mean on the back-transformed chlorophyll a values</i>	4a. Interpolated chlorophyll <i>a</i> surfaces are ln-transformed 4b. Seasonal means are calculated on ln-transformed chlorophyll <i>a</i> values.
5. Seasonal arithmetic means are assessed (cell-by-cell) against the criterion for the relevant river segment-season.	5. Ln-transformed seasonal means are assessed (cell-by-cell) against the ln-transformed criterion for the relevant river segment-season.

Source: U.S. EPA 2008.

IMPLICATIONS OF THE REVISED ASSESSMENT PROTOCOL

Conducting the spatial and temporal analyses in log-space produces geometric means. Geometric means will be less than the arithmetic means of the raw data, i.e. bias low for the estimator of the arithmetic mean, for all data sets with at least one pair of nonequal values (Bland and Altman 1996). When all values in the data set are the same value and only then will the arithmetic mean equal the geometric mean. However, while geometric means may be less than arithmetic means, the values will always be above the minimum observed value and below the observed maximum value in both approaches. For log-normally distributed data such as the chlorophyll *a* data, the geometric mean is further a more efficient measure of central tendency, efficiently estimating the median which might be considered more typical of observations from the sampled population (E. Perry, 2010, Pers. Comm.).

Given the very small number of data points in the tidal James River data analyses that influence the statistical measure of departure from normality, then this departure occurs in a small percentile of the distribution. Overall, the data align very well with the expected up through the 10th percentile (see Appendix D). Because the CFD assessment method is defining the upper bound chlorophyll *a* criteria somewhere around the 10th percentile, it is fair to conclude that the log-normal is adequate for that purpose. While there may be another distribution that matches the data better than the normal distribution, one would, however, have to weigh the benefits of improved estimation against the costs of developing a suite of estimation procedures for this other distribution. One clear advantage of working with the log-normal is that the log transformation provides for a normal metric where one has many choices of well developed and well tested statistical methods (E. Perry 2010, Pers. Comm.).

The present Virginia water quality standards for tidal James River and the District of Columbia's water quality standards for its tidal waters, stated as seasonal chlorophyll *a* means, reflect the importance of the assessment in measuring central tendency compared to an acceptable upper bound for acceptable water quality conditions. Chlorophyll *a* is a parameter whose measures repeatedly show skewed distributions appropriate to log transformation to approximate a normal distribution for making inference with well developed, well tested statistical methods. It is therefore appropriate to use a statistic that addresses the central tendency respecting the appropriate statistical properties of such data, i.e., the geometric mean.

The EPA Chesapeake Bay Program Office and its partners tested the recommended revised assessment methodology for Chesapeake Bay data (e.g., tidal James River) and compared the results with the application of the promulgated Virginia water quality standards' chlorophyll *a* criteria. Results showed almost universally greater levels of chlorophyll *a* attainment using the recommended revised methodology compared with the previously EPA published criteria assessment method (and adopted into Virginia's water quality standards). Acknowledging these findings, the revisions to the published criteria assessment method are recommended for ensuring consistency within the assessment procedures with acknowledged the statistical properties of the chlorophyll *a* data.

LITERATURE CITED

Barnes, H. 1952. The use of transformations in marine biological statistics. *J. du Conseil* 18:61-71.

Bland, J.M and D.G. Altman. 1996. Transformations, means and confidence intervals. *British Medical Journal* 312:1079.

Buchanan, C., R. V. Lacouture, H. G. Marshall, M. Olson, J. Johnson. 2005. Phytoplankton reference communities for Chesapeake Bay and its tidal tributaries. *Estuaries* 28(1):138-159.

Colorado lakes study.
<http://www.chatfieldwatershedauthority.org/pdf/Characterizing%20Chlorophyll%20Distribution%20in%20Colorado.pdf>

Gilbert, R.O. 1987. *Statistical methods for environmental pollution monitoring*. Van Nostrand Reinhold, New York, NY.320 pp.

Harding, L. Jr. 1994. Long term trends in the distribution of phytoplankton in Chesapeake Bay: roles of light, nutrients and streamflow. *Mar. Ecol. Prog. Ser.* 104:267-291.

Harding, L.W., and E.S. Perry. 1997. Long-term increase of phytoplankton biomass in Chesapeake Bay, 1950–1994. *Mar. Ecol. Prog. Series.* 157:39-52.

Harris, G.P. 1986. *Phytoplankton ecology: Structure, function and fluctuation*. Chapman and Hall, New York, NY. 384 pp.

Heiskary, Steven A and William W. Walker, Jr. 1995. Establishing a chlorophyll a goal for a run-of-the-river reservoir. *Lake and Reserv. Manage.* 11(1):67-76.

Jordan, T.E., D.L. Correll, J. Miklas, and D.E. Weller. 1991. Long-term trends in estuarine nutrients and chlorophyll, and short term effects of variation in watershed discharge. *Mar. Ecol. Progr. Ser.* 75:121-132.

Lacouture, R.V., Johnson, J.M., Buchanan, C., and Marshall, H.G. 2006. Phytoplankton index of biotic integrity for Chesapeake Bay and its tidal tributaries. *Estuaries Coasts* 29: 598–616.

Llanso, R.J., D.M. Dauer and J.H. Volstad. 2009. Assessing ecological integrity for impaired waters decisions in Chesapeake Bay, USA. *Marine Pollution Bulletin* 59:48-53

Olson, M. 2009. *Relative Status Indicator: Development and evolution of a relative measure of condition for assessing the status of water quality and biological parameters tracked in the U.S. EPA Chesapeake Bay Program long term monitoring programs*. Final Report. September 2009. Chesapeake Bay Program Office, Annapolis, MD.

Sokal, R.R. and F.J. Rohlf. 1981. *Biometry*, 2nd ed. W.H. Freeman, San Francisco, CA.

Tett, P. and A. Wallis. 1978. The general annual cycle of chlorophyll standing crop in Lock Creran. *J. Ecol.* 66:227-239.

U.S. Environmental Protection Agency. 2003. *Ambient Water Quality Criteria for Dissolved Oxygen, Water Clarity and Chlorophyll a for the Chesapeake Bay and Its Tidal Tributaries (Regional Criteria Guidance) April 2003*. EPA 903-R-03-002. Region III Chesapeake Bay Program Office, Annapolis, MD.

U.S. EPA 2007a. *Ambient Water Quality Criteria for Dissolved Oxygen, Water Clarity and Chlorophyll a for the Chesapeake Bay and its Tidal Tributaries – 2007 Addendum*. July 2007. EPA 903-R-07-003. Region III Chesapeake Bay Program Office, Annapolis, MD.

U.S. EPA 2007b. *Ambient Water Quality Criteria for Dissolved Oxygen, Water Clarity and Chlorophyll a for the Chesapeake Bay and its Tidal Tributaries – Chlorophyll Criteria Addendum*. November 2007. EPA 903-R-07-005. Region III Chesapeake Bay Program Office, Annapolis, MD.

U.S. EPA 2008. *Ambient Water Quality Criteria for Dissolved Oxygen, Water Clarity and Chlorophyll a for the Chesapeake Bay and its Tidal Tributaries – 2008 Technical Support for Criteria Protocols Addendum*. September 2008. EPA 903-R-08-001. Region III Chesapeake Bay Program Office, Annapolis, MD.

Vollenweider, R.A. and J. Krekes. 1980. The loading concept as a basis for controlling eutrophication: philosophy and preliminary results of the OECD programme on eutrophication. *Prog. Wat. Technol.* 12:5-38.

ACRONYMS

2-D	two-dimensional
B-IBI	benthic index of biotic integrity
CBP	Chesapeake Bay Program
CIMS	Chesapeake Information Management System
CFD	cumulative frequency distribution
CHLA	chlorophyll <i>a</i>
DC	deep channel
DO	dissolved oxygen
DW	deep water
EPA	U.S. Environmental Protection Agency
GLM	Generalized Linear Model
HRSD	Hampton Roads Sanitation District
m	meters
mg/L	milligrams per liter
OW	open water
P-IBI	phytoplankton index of biotic integrity
S	surface
SAS	Statistical Analysis Software
SD	Standard Deviation
STAC	Scientific and Technical Advisory Committee
VIMS	Virginia Institute of Marine Science

Appendix A.

B-IBI Sample Size and Standard Deviations on B-IBI Scoring when Screening Segments for Reference Community Characterization

The EPA Chesapeake Bay Program Office in cooperation with its partners, examined the effects of relaxing the data screening criteria to accept segment-period combinations as “healthy” when defining reference communities with sample size ≥ 8 (instead of the recommended $n > 10$) and/or standard deviation ≤ 1.2 (instead of the recommended < 1.0). Data were 1996-2006 from the CIMS database.

For the “fixed station” samples both “total_score” and “grand_score” records were included. “Total_score” records are replicate measurements of the same sampling event; the average of these is reported as the “grand_score.” Benthic experts (Llanso, Versar, Inc.) recommend using the “grand_score” in these analyses. Four scenarios were explored (Table A-1). The EPA accepted screening criteria is the default under Scenario A. Scenarios B, C and D relaxed the standard deviation, sample size and both sample size and standard deviation, respectively.

Table A1. Healthy deep-water segments as characterized with four scenarios of screening criteria. The accepted screening criteria is Scenario A.

	Scenario A (Default) B-IBI ≥ 3.0 $n \geq 10$ S.D. < 1.0	Scenario B B-IBI ≥ 3.0 $n \geq 10$ S.D. < 1.2	Scenario C B-IBI ≥ 3.0 $n \geq 8$ S.D. < 1.0	Scenario D B-IBI ≥ 3.0 $n \geq 8$ S.D. < 1.2
Total number of “healthy” deep water segment-periods	10	11	13	16

Relaxation of the criteria results in moderate increases (ranging from 1 to 6) in the number of segment-periods classified as “healthy.” Due to the increased risk of inaccurate classification, it is important to examine not just the number of additional segment-periods, but also the shape of these curves. If a curve is classified as “healthy” but its location in CFD space is consistent with DO violation CFDs of segment-periods classified as “degraded,” then it is reasonable to question whether an inaccurate classification has occurred.

In the case of Scenario B (relaxing the standard deviation criterion from a maximum of 1.0 to a maximum of 1.2), a single curve (CB5MH 1999-2001) is added to the group of “healthy” segment-periods. In Figure A-1 below, this curve is visible as a light blue line, while the population of 10 curves identified in Scenario A are presented by dark blue lines. Degraded

segment-periods are visible as red lines. The biologically-based reference curve generated from the 100th percentile of “Scenario A” violations at each time step is visible as a yellow line.

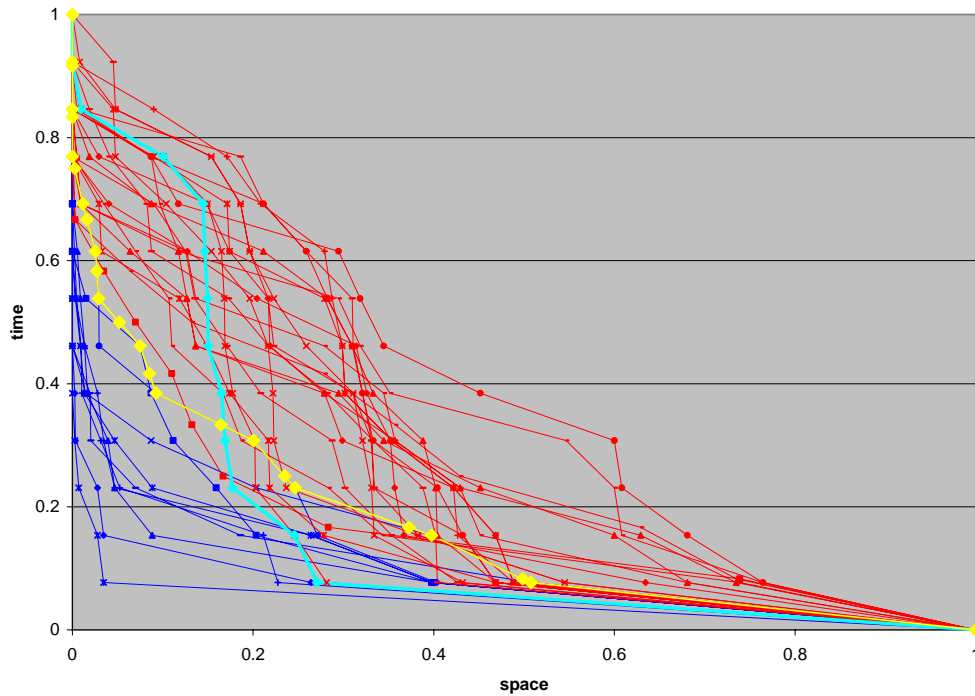


Figure A-1. Scenario B – illustrates the impact of maintaining the sample size criterion of $n \geq 10$ while relaxing the standard deviation criterion from a maximum of 1.0 to a maximum of 1.2

The shape of the CB5MH 1999-2001 curve (light blue line in Figure A-1) raises the question of whether increasing the uncertainty of the screening criteria resulted in erroneous classification of this segment-period as healthy. In particular, the location of the top half of this curve in CFD space that is dominated by degraded curves decreases confidence in the accuracy of its classification. The addition of this curve, particularly in combination with the methodology of taking the 100th percentile of each curve at each point in time, would increase the potential for the resulting biologically-based reference curve to allow rates of hypoxia that result in degradation of the benthic community.

In the case of Scenario C, the standard deviation is kept consistent with the recommended screening criteria but the sample size criterion is relaxed from 10 to 8. This relaxation of the recommended criteria results in the classification of 3 additional segments as “healthy.” The CFD curves for these additional segments are shown as light blue curves in Figure A-2.

While two of the additional curves (CB6PH 1998-2000 and CB6PH 2000-2002) fall within the cloud of violation rates deemed “acceptable,” one curve (CB3MH 1996-1998) once again extends into the cloud of data dominated by CFDs associated with degraded segment-periods (Figure A-2). As described earlier, this raises the concern that relaxation of the criteria has resulted in the inaccurate classification of a degraded segment-period as healthy.

The relaxation of both the sample size and the standard deviation criteria (Scenario D) increases the number of segment-periods classified as “healthy” from 10 to 16. However, 4 of these additional CFD curves extend into “degraded” CFD space to a degree that calls into question the accuracy of their classification as healthy (Figure A-3).

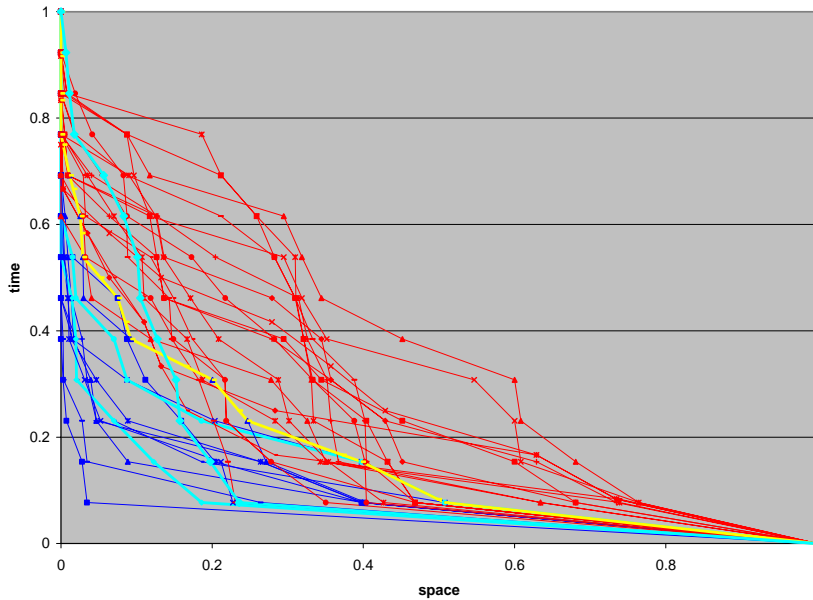


Figure A-2. Scenario C - illustrates the impact from relaxing the sample size criterion from $n \geq 10$ to $n \geq 8$ while maintaining the standard deviation criterion of $S.D. < 1.0$.

Relaxing the screening criteria for defining healthy segments based on the B-IBI with respect to minimum sample size and maximum standard deviation increases the number of healthy segments that can be used to generate the biologically-based reference curve. However, the increased uncertainty of accurate classification resulting from relaxation of the criteria far outweighs the potential benefit of increased sample size. For the reference CIMS dataset, the EPA recommended methodology results in a total sample size of 24 segment-periods, of which 10 are classified as healthy and 14 are classified as degraded. Accounting for the trade offs with segment classification risks, this present method is supported as sufficient in generating a low risk sample size for elucidating the boundary between acceptable (i.e. those which allow a healthy benthic community to persist) and unacceptable violations of the deep-water DO criteria.

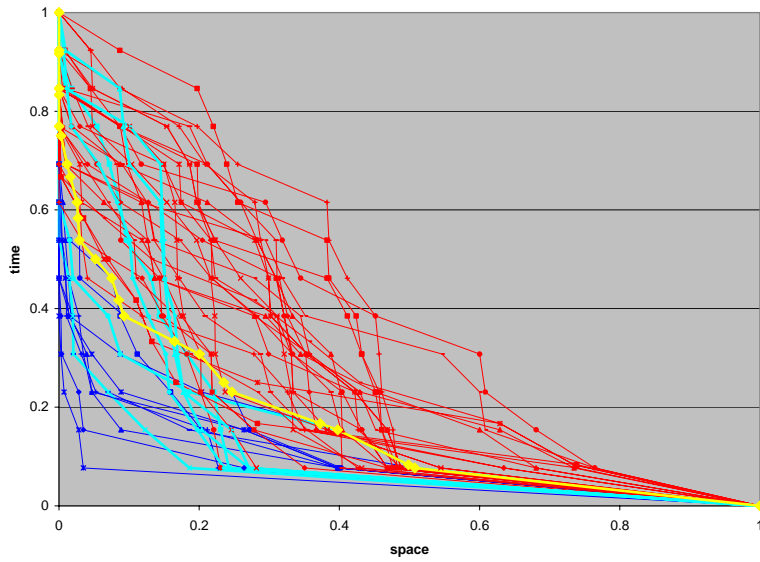


Figure A-3. Scenario D - illustrates the impact from relaxing the sample size criterion from $n \geq 10$ to $n \geq 8$ and the standard deviation criterion of $SD < 1.0$ to $SD < 1.2$.

Appendix B.

Shape of the Biologically-based Reference Curve

The shape of the biologically-based reference curve is an important factor in identifying acceptable violations of the Chesapeake Bay water quality criteria for dissolved oxygen. The shape of biologically derived reference curve has thus far reinforced the suitability of the hyperbolic 10% default reference curve when a biologically-based reference curve is unavailable. An alternative hypothesis, however, is that comparing the total area under a CFD assessment curve to the total area under the biologically-based reference curve is a better measure of the degree to which healthy biological communities can tolerate violations of the DO criteria than the existing “point” method. Arguments put forth to support this proposal include: (1) a segment-period may exceed the biologically-based reference curve in one area of CFD space while the overall area of its exceedance is within that represented by the biologically-based reference curve; (2) there is high variability in the shape of CFD curves and the data do not allow identification of combinations of time and volume that lead to poor B-IBI scores in a segment; and (3) the proposed “area” method has lower error rates than the published “point” method, even with the modifications proposed by EPA to the latter method.

With regard to arguments 1 and 2, application of the method modifications outlined in this addendum, Chesapeake Bay benthic communities are now being accurately classified as “healthy” or “degraded” when there is sufficient data to do so. As a result, the Chesapeake Bay data support a rather specific combination of time and volume that forms the boundary between healthy and degraded benthic communities in the deep-water designated use (Figure B-1).

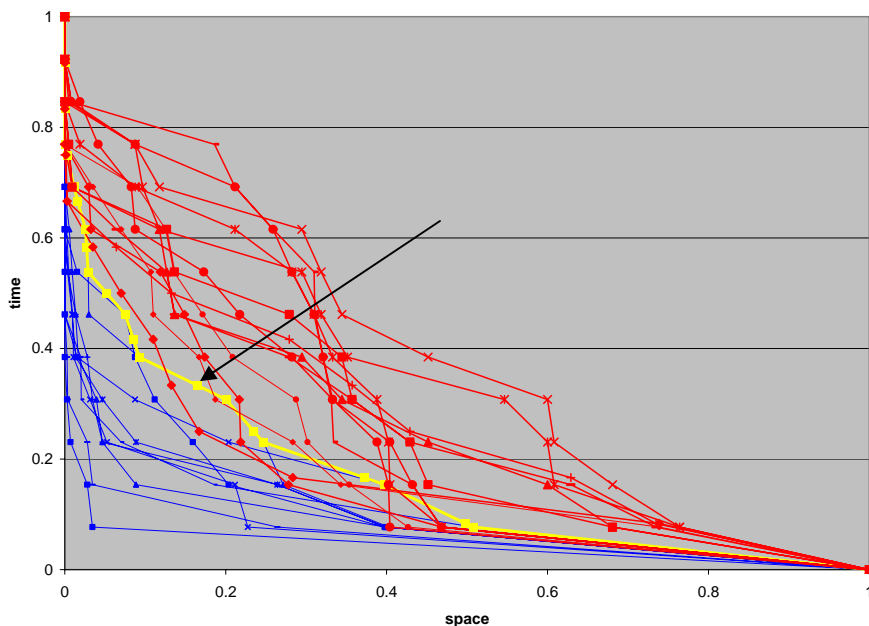


Figure B-1. Dissolved oxygen violation curves associated with healthy (blue) and degraded (red) benthic communities in deep-water designated use habitats. The deep-water biologically-based reference curve (yellow) is also shown.

Using a dataset with all duplicate records appropriately removed and all appropriate screening criteria applied, the error rate for the “Point Method” is zero. In this case, all segment-periods classified as “healthy” using the recommended screening criteria ($n \geq 10$, $SD < 1.0$) pass the EPA recommended deep-water biologically-based reference curve, and all segment-periods classified as “degraded” fail the recommended biologically-based reference curve (Tables B-1 and B-2).

Table B-1. Segment classifications using the recommended screening criteria: deep-water designated use.

Method	Correct		Incorrect	
	Healthy Segments Passing	Degraded Segments Failing	Healthy Segments Failing	Degraded Segments Passing
Published “Point” Method	100%	100%	0%	0%
Proposed “Area” Method	100%	100%	0%	0%

Table B-2. Segment-period classifications under the recommended method: deep-water designated use.

Method	Correct		Incorrect	
	Healthy Segments Passing	Degraded Segments Failing	Healthy Segments Failing	Degraded Segments Passing
Published “Point” Method	CB6PH_1996_1998 CB7PH_1996_1998 CB6PH_1997_1999 CB7PH_1997_1999 CB7PH_1998_2000 CB6PH_1999_2001 CB7PH_1999_2001 CB7PH_2000_2002 CB6PH_2004_2006 CB7PH_2004_2006	PAXMH_1996_1998 POTMH_1996_1998 PAXMH_1997_1999 POTMH_1997_1999 POTMH_1998_2000 PAXMH_1999_2001 POTMH_1999_2001 PAXMH_2000_2002 RPPMH_2000_2002 PAXMH_2001_2003 PAXMH_2002_2004 PAXMH_2003_2005 PAXMH_2004_2006 RPPMH_2004_2006		
Proposed “Area” Method	CB6PH 1996-1998 CB6PH 1997-1999 CB6PH 1999-2001 CB6PH 2004-2006 CB7PH 1996-1998 CB7PH 1997-1999 CB7PH 1998-2000 CB7PH 1999-2001 CB7PH 2000-2002 CB7PH 2004-2006	POTMH19992001 POTMH19982000 RPPMH20022004 PAXMH19992001 PAXMH20012003 PAXMH20042006 POTMH19971999 PAXMH20032005 PAXMH20002002 POTMH19961998 RPPMH20002002 PAXMH20022004 PAXMH19961998 PAXMH19971999		

Both methods result in the same error rates when duplicate records are removed and EPA's criteria are applied to the classification of benthic communities. However, in contrast to Argument 2 as described above, it is EPA's position that this dataset does provide convincing biological information with regard to the degree and distribution of deep-water DO criteria violations that can be tolerated by the benthic community. Furthermore, by using the worst violation rate allowed by any healthy community at each point in time, EPA has allowed for greater violation rates in regions of CFD-space where CFD curves from healthy and degraded communities overlap. It is reasonable to postulate, based on the distribution in CFD-space of curves associated with healthy and degraded benthic communities, that violations occurring in the CFD-space circled in black in Figure B-2 lead to degradation of the benthic community.

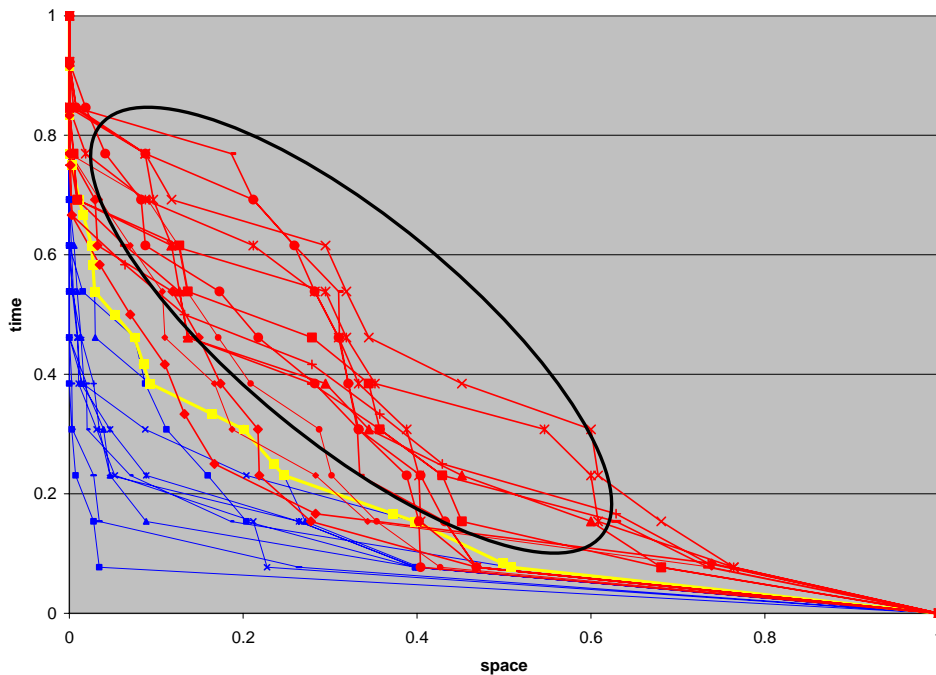


Figure B-2. Violations occurring in the CFD-space circled in black are postulated to lead to degradation of the benthic community.

It is suggested from the multiple lines of evidence that the shape of the biologically-based reference curve is an important factor in identifying acceptable violations of the DO criteria. The shape of this biological reference curve also provides further support for the suitability of the hyperbolic 10% reference curve, in that it illustrates the sensitivity of biological communities to chronic violations of DO criteria.

Appendix C.

Derivation of the Deep-Water Biologically-Based Reference Curve

Step 1. We obtained a dataset of benthic scores for the Chesapeake Bay and tidal tributaries. Data used comprised benthic communities sampled between 1996 and 2006, from both the “fixed station” and “random strata” sampling programs. Only “grand score” values – which are an average of replicate samples – were included from the fixed station program. For the purposes of deriving a deep-water biological reference curve, we restricted the dataset to only those samples taken in segments that contain a deep-water designated use.

Step 2. We removed from the dataset any samples obtained from the following segments: PATMH, SBEMH, and CB5MH. Benthic communities in PATMH and SBEMH are widely understood to be impacted by chemical contaminants (pers. comm, Roberto Llanso, Versar Inc.); a complication that confounds the relationship between hypoxia and benthic community health in these areas.

In the case of CB5MH, areas greater than 12 meters in depth – which account for 35 percent of the bottom surface area of CB5MH – are excluded from the benthic sampling program because they are assumed to be azoic or nearly azoic. For their analyses of benthic health, Llanso et al. 2009 assume that all areas greater than 12 m in depth are degraded, and perform a post-hoc correction to factor this assumption into their benthic assessment. For purposes of developing a biological reference curve, the exclusion from sampling of such a large portion of CB5MH calls into question our ability to accurately characterize the health of its deep-water benthic communities.

Step 3. We obtained water quality data from the Chesapeake Bay Program Water Quality database for the time period 1996-2006. Using the standardized method for locating pycnocline boundaries (see U.S. EPA 2008), we determined the depth of the upper and lower pycnocline boundaries for all sampling events in this time period.

Step 4. From this dataset, we selected the sampling event that was closest in space (at a minimum, within the same segment) and time (at a minimum, within the same month) to each benthic sampling event.

Step 5. We then classified each benthic sample as an “open-water,” “deep-water,” or “deep-channel” benthic sample based on its depth relative to the upper and lower pycnocline boundaries of the paired water quality sampling event. Benthic samples that were taken at depths between the upper and lower boundaries of the pycnocline were classified as “deep-water” samples. When no lower boundary was identified, benthic samples from depths below the upper boundary of the pycnocline were classified as “deep-water.” Benthic samples that could not be paired with a pycnocline boundary were discarded.

Step 6. For each 3-year time window from 1996-2006, we applied the following criteria to classify deep-water benthic communities as “healthy” for the purposes of generating a biological reference curve for the dissolved oxygen criteria assessment: at least 10 benthic IBI scores; mean score ≥ 3.0 ; and standard deviation of the mean < 1.0 . Segment-periods (e.g. “CB6PH 1996-1998”) that met the above criteria were classified as “healthy.”

Step 7. We obtained the “dissolved oxygen violation rates” for each healthy segment-period. These rates are an intermediate product of the dissolved oxygen criteria assessment procedure (see U.S. EPA 2007 and 2008). They represent the fraction of deep-water in a given segment that violates water quality criteria for dissolved oxygen in a given time period. Using these violation rates, we generated a CFD curve for each healthy segment-period. Most segment-periods contained 12 violation rates, but some contained only 11 rates. To account for segment-periods with different numbers of violation rates, all violation rates were interpolated to a common set of plotting positions (y values). To generate a biological reference curve that represented the “100th percentile” of healthy violation rates, we then used the largest violation rate (across healthy segment-periods) for each “y” value of the violation CFD. The resulting set of violation rates represents the largest of all healthy violation rates for each plotting position. See chapter 3 for more details on the selection of the 100th percentile curve.

The following segment-periods comprised the set of “healthy” segment-periods:

- CB6PH 1996-1998
- CB7PH 1996-1998
- CB6PH 1997-1999
- CB7PH 1997-1999
- PAXMH 1997-1999
- CB7PH 1998-2000
- CB6PH 1999-2001
- CB7PH 1999-2001
- CB7PH 2000-2002
- CB7PH 2003-2005
- CB6PH 2004-2006
- CB7PH 2004-2006

The resulting deep-water biologically-based reference curve for dissolved oxygen assessment is illustrated in Chapter 3, Figure III-4 of this addendum and defined as:

X (violation rate)	Y (plotting position)
0	1
0	0.923077
0	0.846154
0	0.769231
0.025641	0.692308
0.029132	0.615385
0.051185	0.538462
0.200524	0.461538
0.246642	0.384615
0.271513	0.307692
0.356639	0.230769
0.402786	0.153846
0.555376	0.076923
1	0

LITERATURE CITED

Llanso, R.J., D.M. Dauer and J.H. Volstad. 2009. Assessing ecological integrity for impaired waters decisions in Chesapeake Bay, USA. *Marine Pollution Bulletin* 59:48-53

U.S. Environmental Protection Agency. 2007. *Ambient Water Quality Criteria for Dissolved Oxygen, Water Clarity and Chlorophyll a for the Chesapeake Bay and Its Tidal Tributaries – 2007 Addendum*. July 2007. EPA 903-R-07-003. Region III Chesapeake Bay Program Office, Annapolis, MD.

U.S. Environmental Protection Agency. 2008. *Ambient Water Quality Criteria for Dissolved Oxygen, Water Clarity and Chlorophyll a for the Chesapeake Bay and Its Tidal Tributaries 2008 Technical Support for Criteria Assessment Protocols Addendum*. September 2008. EPA 903-R-08-001. Region III Chesapeake Bay Program Office, Annapolis, MD.

Appendix D.

History of EPA Guidance Regarding the Deep-Channel Reference Curve

In April 2003, the EPA published the guidance document, *Ambient Water Quality Criteria for Dissolved Oxygen, Water Clarity and Chlorophyll a for the Chesapeake Bay and Its Tidal Tributaries* (U.S. EPA 2003). In this publication, EPA documented the derivation of the dissolved oxygen criterion protective of the seasonal deep channel designated use. For seasonal deep-channel designated use, an instantaneous minimum criterion of 1 mg/L was determined to protect benthic organisms residing in the:

“deep water-column and adjacent bottom surficial sediment habitats located principally in the river channel at the lower reaches of the major rivers and along the spine of the middle mainstem Chesapeake Bay at depths below which seasonal anoxic (< 0.2 mg/L dissolved oxygen) to severe hypoxic conditions (< 1 mg/L dissolved oxygen) routinely set in and persist for extended periods of time under current conditions” (p. 60 in U.S. EPA 2003).

In support of the instantaneous minimum criterion of 1 mg/L, U.S. EPA (2003) summarized findings published in peer-reviewed literature sources indicating that several keystone benthic species “are resistant to dissolved oxygen concentrations as low as 0.6 mg/L,” and that “extensive mortality is likely only under persistent exposure to very low dissolved oxygen concentrations at high summer temperatures” (p. 61).

U.S. EPA (2003) also reported that in the mesohaline Chesapeake Bay (the primary location of the seasonal deep-channel designated use), “dissolved oxygen concentrations of less than 1 mg/L lead to mortality for even tolerant species (p. 61) and that “when dissolved oxygen drops significantly below 1 mg/L for even short periods of time (on the order of hours) mortality increases, even for tolerant species” (p. 65). Furthermore, it was stated that “States and other users must recognize that the deep-channel dissolved oxygen criterion is stated as an instantaneous minimum, thus *any* exceedance is assumed to have direct consequences to the survival of the bottom-dwelling community” (p. 151).

Regarding the definition of a water quality standard, it is explained in U.S. EPA 2003 and in Chapter 3 of the U.S. EPA’s *Water Quality Standards Handbook, 2nd Edition* (U.S. EPA 1994) that water quality criteria definition and assessment comprises not just the *magnitude* of a water quality criterion (i.e. “the quantifiable condition,” in this case the concentration of dissolved oxygen), but also the *duration* and *frequency* of that condition.

In this context, *duration* is addressed by restricting the applicability of the criterion to the summer period (June – September) when stratification and severe hypoxia occur in deep-channel regions of the Chesapeake Bay, and by defining the assessment period as “the most recent three consecutive years for which relevant monitoring data are available” (U.S. EPA 2003; p. 150-1).

The *frequency* component of the criterion “is directly addressed through comparison of the generated cumulative frequency distribution (CFD) with the applicable criterion reference curve” (U.S. EPA 2003; p. 151).

In summary, statements were made in U.S. EPA 2003 suggesting that the benthic community can tolerate small violations of the deep-channel instantaneous minimum criterion, but statements were also made suggesting that any violation of this criterion has negative effects on the survival of deep-channel benthic species. However, as also described in U.S. EPA 2003, national guidelines define a water quality standard as comprising not only the *magnitude* of a given condition, but also the *duration* over which that condition is assessed and the *frequency* of violation allowed within the given assessment *duration*. For the case of the Chesapeake Bay, the *frequency* of allowable violation is defined by the location of a reference CFD, more commonly called a “reference curve” (both the rationale for use of a biological reference curve and the development of the 10 percent reference curve are also well documented in U.S. EPA 2003).

LITERATURE CITED

U.S. Environmental Protection Agency. 1994. *Water Quality Standards Handbook, 2nd Edition*. August 1994. EPA 823-8-94-005a. Washington D.C.

U.S. Environmental Protection Agency. 2003. *Ambient Water Quality Criteria for Dissolved Oxygen, Water Clarity and Chlorophyll a for the Chesapeake Bay and Its Tidal Tributaries (Regional Criteria Guidance)*. April 2003. EPA 903-R-03-002. Region III Chesapeake Bay Program Office, Annapolis, MD.

Appendix E.

James River Chlorophyll *a* Data normality Analysis

Checking Normality of Log-transformed Chlorophyll *a* Data

Summary Notes regarding results of test log-normal assumption for James River chlorophyll. E. Perry 2/24/2010.

SUMMER:

Even without standardizing for heterogeneous variance, the ln(chl) residuals from the Year X Segment model seem to be fairly close to a normal distribution. The normality test show significant departure from normality

Tests for Normality for un-standardized residuals

Test	--Statistic--	-----p Value-----
Shapiro-Wilk	W 0.988201	Pr < W <0.0001
Kolmogorov-Smirnov	D 0.059704	Pr > D <0.0100

but this is not surprising with a sample size of 828. The large sample size gives you the power to detect very small differences from normality. The Shapiro-Wilk statistic of 0.988 shows that the residuals are very highly correlated with the expected residuals from a normal distribution. The normal probability plot shows very high concordance between the expected residuals and the observed residuals except for two outlier points in the extreme tails of the sample. These outliers probably reflect a failure of our simple model to capture some extreme event rather than a failure of log normality.

Levene's test shows that the data do exhibit heterogeneous variances even in the log-metric. This heterogeneity seems to be associated with changing variance over segments.

Levene's test for un-standardized residuals

12:18 Wednesday, February 24, 2010

The GLM Procedure

Dependent Variable: aRChl

Source	DF	Sum of Squares	Mean Square	F Value	p-value
Model	15	96.2326827	6.4155122	23.79	<.0001
Error	812	218.9548618	0.2696488		
Corrected Total	827	315.1875445			

Source	DF	Type III SS	Mean Square	F Value	p-value
year	9	3.87412641	0.43045849	1.60	0.1120
CBSEG_2003	6	91.23339617	15.20556603	56.39	<.0001

Standardizing the residuals by estimates of standard deviation by segment-group and year leads to improvement in both normality and homogeneous variance. However, both non-normality and heterogeneous variance remain statistically significant.

Tests for Normality for standardized residuals

Test	--Statistic--	-----p Value-----
Shapiro-Wilk	W 0.99529	Pr < W 0.0121
Kolmogorov-Smirnov	D 0.032238	Pr > D 0.0361

Levene's test for standardized residuals

Dependent Variable: aStdResChl

Source	DF	Sum of Squares	Mean Square	F Value	p-value
Model	15	12.9100323	0.8606688	2.26	0.0041
Error	812	309.7902578	0.3815151		
Corrected Total	827	322.7002901			

Source	DF	Type III SS	Mean Square	F Value	p-value
year	9	3.48204541	0.38689393	1.01	0.4267
CBSEG_2003	6	9.43256796	1.57209466	4.12	0.0004

Again the heterogeneity seems to be associated with segments which suggests that the grouping algorithm could be improved.

SPRING:

Similar to summer results, without standardizing for heterogeneous variance, the ln(chl) residuals from the Year X Segment model seem to be fairly close to a normal distribution. The normality test show significant departure from normality but the p-value is larger than for Summer.

Tests for Normality for un-standardized residuals

Test	--Statistic--	-----p Value-----
Shapiro-Wilk	W 0.994437	Pr < W 0.0070
Kolmogorov-Smirnov	D 0.034024	Pr > D 0.0323

The Shapiro-Wilk statistic of 0.994 shows that the residuals are very highly correlated with the expected residuals from a normal distribution. The normal probability plot shows very high concordance between the expected residuals and the observed residuals and like the result for summer, the departure from normality appears as outlier points in the extreme tails of the sample.

Levene's test shows that the data do exhibit heterogeneous variances even in the log-metric. This heterogeneity seems to be associated with changing variance over both segments and years.

Levene's test for un-standardized residuals

12:18 Wednesday, February 24, 2010

Dependent Variable: aRChI

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	15	28.9931483	1.9328766	5.88	<.0001
Error	742	243.8524755	0.3286422		
Corrected Total	757	272.8456238			

Source	DF	Type III SS	Mean Square	F Value	Pr > F
year	9	12.18939188	1.35437688	4.12	<.0001
CBSEG_2003	6	16.81566976	2.80261163	8.53	<.0001

Standardizing the residuals by estimates of standard deviation by segment-group and year appears to resolve the heterogeneous variance issue but yields little improvement on normality.

Tests for Normality for standardized residuals

Test	--Statistic--	-----p Value-----
Shapiro-Wilk	W 0.994055	Pr < W 0.0044
Kolmogorov-Smirnov	D 0.038676	Pr > D <0.0100

Levene's test for standardized residuals

Dependent Variable: aStdResChI

Source	DF	Sum of Squares	Mean Square	F Value	p-value
Model	15	3.7421584	0.2494772	0.73	0.7499
Error	742	251.8744219	0.3394534		
Corrected Total	757	255.6165803			

Appendix F.

SAS Computer Code for James River Chlorophyll *a* Normality Tests, Spring and Summer Season

```
*****
* PROGRAM: JAMES_RIVER.sas
* This program will TEST CHLOROPHYLL DATA FOR NORMALITY
* Base code from Elgin Perry 02/16/2010
* additional code written by Jackie Johnson 02/17/2010
*****;
libname ALGAE "G:\LR\OTHER_LR_DATA\Criteria_work\chlorophyll\2010";
*libname ALGAE "C:\Projects\CBP\CHLCRIT\LogNormal\";
options ls=72;
*OPTIONS LS=120 PS=55 REPLACE NOCENTER;
OPTIONS formchar = '|----|+|---+=|-\<>*' ;
*PROC IMPORT OUT= ALGAE.JAMES_SPRING_CHL
          DATATABLE= "JAMES_SPRING_CHL"
          DBMS=ACCESS2000 REPLACE;

*
DATABASE="G:\LR\OTHER_LR_DATA\Criteria_work\chlorophyll\2010\james_river.mdb"
;
*RUN;
*PROC CONTENTS DATA=ALGAE.JAMES_SPRING_CHL;RUN;
data one;
set ALGAE.JAMES_SPRING_CHL;
logE_Ch1=log(reported_value);
label logE_Ch1="LOG_E UG/LITER";
  if cbseg_2003 = "APPTF"
  or cbseg_2003 = "CHKOH"
  or cbseg_2003 = "EBEMH"
  or cbseg_2003 = "LAFMH"
  or cbseg_2003 = "WBEMH" then delete;

  if cbseg_2003 = "JMSPH" then SegGrp = 1;
  if cbseg_2003 = "JMSMH"
  or cbseg_2003 = "SBEMH" then SegGrp = 2;
  if cbseg_2003 = "JMSOH"
  or cbseg_2003 = "CB8PH"
  or cbseg_2003 = "ELIPH" then SegGrp = 3;
  if cbseg_2003 = "JMSTF" then SegGrp = 4;
  sdate = DatePart(SAMPLE_DATE);
  year = year(SDATE);
RUN;
*Proc Contents;
run;
title "Spring James River Data 1991-2000";
proc glm data=one;
  class year cbseg_2003;
  model logE_Ch1=year cbseg_2003;
  output out= reschl r=rchl;
run;
proc Univariate normal plot data=work.reschl;
  title2 "Normality test on raw residuals";
```

```

var rchl;
run;
*( step to get rid of heterogeneous variances;
Proc Sort data=reschl;
  by SegGrp year;
  run;
Proc Means data=reschl noprint;
  by SegGrp year;
  var rchl;
  output out=sdchl StdDev = sdchl n=n;
run;
*(proc print data=sdchl;
*( title "standard deviation results";
*( var SegGrp year sdchl n;
data ResCh1;
  merge ResCh1 sdchl;
  by SegGrp year;
  StdResCh1 = rchl/sdchl;
run;

proc Univariate normal plot data=work.reschl;
  title2 "Normality test on standardized residuals";
  var StdResCh1;
run;
*(proc contents data=work.reschl;
*(run;
data four;
  set work.reschl;
  aRchl= abs(Rchl);
  aStdResCh1= abs(StdResCh1);
run;
proc glm data=work.four;
  title2 "Levene's test for both types of residuals";
  class year cbseg_2003;
  model aRchl aStdResCh1=year cbseg_2003;
  *(means cbseg_2003/snk;
  *(lsmeans cbseg_2003;
run;

*****
* PROGRAM: JAMES_RIVER.sas
* This program will TEST CHLOROPHYLL DATA FOR NORMALITY
* Base code from Elgin Perry 02/16/2010
* additional code written by Jackie Johnson 02/17/2010
*****;
libname ALGAE "G:\LR\OTHER_LR_DATA\Criteria_work\chlorophyll\2010";
*libname ALGAE "C:\Projects\CBP\CHLCRIT\LogNormal\";
options ls=72;
*OPTIONS LS=120 PS=55 REPLACE NOCENTER;
OPTIONS formchar = '|----|+|----+=|-\<>*';
*PROC IMPORT OUT= ALGAE.JAMES_SPRING_CHL
  DATATABLE= "JAMES_SPRING_CHL"
  DBMS=ACCESS2000 REPLACE;

```

```

*
DATABASE="G:\LR\OTHER_LR_DATA\Criteria_work\chlorophyll\2010\james_river.mdb"
;
*RUN;
*PROC CONTENTS DATA=ALGAE.JAMES_SUMMER_CHL;RUN;
data one;
set ALGAE.JAMES_SUMMER_CHL;
logE_Ch1=log(reported_value);
label logE_Ch1="LOG_E UG/LITER";
  if cbseg_2003 = "APPTF"
  or cbseg_2003 = "CHKOH"
  or cbseg_2003 = "EBEMH"
  or cbseg_2003 = "LAFMH"
  or cbseg_2003 = "WBEMH" then delete;

  if cbseg_2003 = "JMSPH" then SegGrp = 1;
  if cbseg_2003 = "JMSMH"
  or cbseg_2003 = "SBEMH" then SegGrp = 2;
  if cbseg_2003 = "JMSOH"
  or cbseg_2003 = "CB8PH"
  or cbseg_2003 = "ELIPH" then SegGrp = 3;
  if cbseg_2003 = "JMSTF" then SegGrp = 4;
  sdate = DatePart(SAMPLE_DATE);
  year = year(SDATE);
RUN;
*Proc Contents;
run;
title "Summer James River Data 1991-2000";
proc glm data=one;
  class year cbseg_2003;
  model logE_Ch1=year cbseg_2003;
  output out= reschl r=rchl;
run;
proc Univariate normal plot data=work.reschl;
  title2 "Normality test on raw residuals";
  var rchl;
run;
*( step to get rid of heterogeneous variances;
Proc Sort data=reschl;
  by SegGrp year;
  run;
Proc Means data=reschl noprint;
  by SegGrp year;
  var rchl;
  output out=sdchl StdDev = sdchl n=n;
run;
*(proc print data=sdchl;
*( title "standard deviation results";
*( var SegGrp year sdchl n;
data ResCh1;
  merge ResCh1 sdchl;
  by SegGrp year;
  StdResCh1 = rchl/sdchl;
run;

proc Univariate normal plot data=work.reschl;
  title2 "Normality test on standardized residuals";

```

```
var StdResCh1;  
run;  
*(proc contents data=work.reschl;  
*(run;  
data four;  
  set work.reschl;  
  aRCh1= abs(RCh1);  
  aStdResCh1= abs(StdResCh1);  
run;  
proc glm data=work.four;  
  title2 "Levene's test for both types of residuals";  
  class year cbseg_2003;  
  model aRCh1 aStdResCh1=year cbseg_2003;  
  *(means cbseg_2003/snk;  
  *(lsmeans cbseg_2003;  
run;
```



May 2010
U.S. Environmental Protection Agency
Region III
Chesapeake Bay Program Office
Annapolis, Maryland

and

Region III
Water Protection Division
Philadelphia, Pennsylvania

in coordination with

Office of Water
Office of Science and Technology
Washington, D.C.

and

The states of
Delaware, Maryland, New York
Pennsylvania, Virginia and
West Virginia and the District of Columbia