Evaluating the Validity of the Umbrella Criteria Concept for Chesapeake Bay Tidal Water Quality Assessment.

Findings of the Umbrella Criteria Action Team
Tidal Monitoring and Assessment Workgroup
2009-2011

December 2011

Contents

List of Tables

List of Figures

Appendices

Executive Summary

Dissolved oxygen (DO) water quality standards were established for the Chesapeake Bay and its tidal tributaries ("the Bay") under the authority of the Clean Water Act, in order to protect the reproduction, survival and growth of estuarine living resources. The criteria have spatially-specific habitat components (e.g. open water, deep water, deep channel, migratory and spawning) with different temporal and seasonal applications (i.e., 30-day, 7-day, 1-day and instantaneous minimum; spring, summer and "rest of year"). However, synoptic monitoring of the Bay has generally occurred on a temporal scale of only one or two measurements per month over the 25-year history of the Chesapeake Bay Program. As a result, insufficient monitoring data and/or appropriate analytical methods existed to support assessment of the shorter-term criteria.

In the course of developing Total Maximum Daily Loads (TMDLs) for pollutants entering the Chesapeake Bay, analysts at the USEPA CBP conducted an assessment of all water quality criteria using hourly output from the calibration run of the Phase 5.1 version of the estuarine Water Quality Sediment Transport Model (WQSTM). Based on this analysis, the CBP determined that evaluation of the 30-day mean criteria was sufficient to establish listing status for the open water and deep water designated uses of the Bay. We have called this determination the "**Umbrella Criteria Assumption**" because it surmises that one criterion (i.e. the 30-day mean) serves as an "umbrella" for the remaining criteria of a given designated use. The focus of this report is an evaluation of the relative protectiveness of the summer season (June – September) DO criteria.

The experimental nature of some of the analytical approaches used in assessing the validity of the Umbrella Criteria assumption produced a 'multiple lines of evidence' approach. Participants in the 2011 STAC-sponsored Umbrella Criteria Workshop explored the use of high frequency data to evaluate the assertion that the open water & deep water 30-day mean criteria are the most conservative criteria for their respective designated uses. Evaluation and validation of spectral analysis techniques for generating synthetic high-frequency datasets comprised an essential component of the ad-hoc Umbrella Criteria project. Shallow water DO dynamics and nearshore-offshore DO behavior comparisons were also investigated.

Key findings of the Umbrella Criteria assessment process include:

Methods

- Spectral analysis approaches recommended for short term DO criteria assessment in USEPA (2004) and USEPA (2007) were explored, updated and validated. The updated method is referred to as *Spectral Casting*.
- Uncertainty contributed by casting site DO characteristics affected criteria assessment outcomes less than uncertainty contributed due to low frequency sampling of the long term water quality monitoring program. These findings suggest Spectral Casting could be recommended to address certain previously unassessed short term water quality criteria (e.g. 7-day mean).
- Conditional probability analyses provided key insights with tests of protectiveness by the 30-day mean for 7-day means but more importantly, overcame challenges facing evaluation of Umbrella Criteria assessment for protectiveness of the 30-day mean against the instantaneous minimum.

Modeling-Monitoring comparisons

- Modeling-Monitoring comparisons showed seasonal patterns of trend in the monitoring data were reflected the CBP WQSTM output. For the month of August, with limited unmatched (2009 Vertical profiler vs. 1991-2000 Model output) comparisons, model variability tended to be lower than observations.
- A more restricted comparison using hydrologically similar years, model year 1993 had lower variability while 1996 had comparable variability to August 2009 vertical profiler DO observations.

Umbrella Criteria Protectiveness

- Analysis of the open water 30-day mean DO criterion was generally found to be consistent with and supporting of protecting the 7-day mean criterion. The value of this work was bolstered by the diversity of geography, tidal fresh to polyhaline habitats, shallow and offshore waters, a range of nutrient conditions and interannual variability taken into account in the evaluation.
- The weight of evidence tends to be inconsistent with the assumption of open water 30-day mean protecting the instantaneous minimum criterion.
- Limited analyses of deep water designated use from mesohaline habitats suggest general consistency with the 30-day mean protecting the 1-day mean criterion. There remains unlikely support for 30-day mean protecting the instantaneous minimum. Further assessments across geography, salinity zones and broader eutrophication gradient are encouraged.

Conditions for which the Umbrella Criteria assumptions appear to hold

- With the 7-day mean, consistent protectiveness by the 30-day mean was detected across geography, salinity zones, and Bay habitats (shallow water, open water).

Conditions for which the Umbrella Criteria assumptions may be violated

- Measurable violations of criteria can be found even at sites with good water quality. Understanding DO variability relative to a criterion value is key.
- Variation in dissolved oxygen measures around the mean criterion affects the risk of meeting or failing attainment. When dissolved oxygen variability increased the risk of failing a criterion increased.
    - Analyses have also shown that shallow waters may be more likely than open waters to exhibit large 24-hour (diel) fluctuations in DO concentrations
- Regions with strong spring-neap tide effects were raised as areas of concern relative to meeting umbrella criteria assumptions.
- Inter-annual variation in violations of DO criteria may be related to inter-annual variation in flow. Wind and temperature effects on shallow water DO behavior were further noted. Such findings pointed toward the concerns regarding further accounting for the effects of climate change on the relationship among Chesapeake Bay criteria and meeting the umbrella criteria assumptions.

Conditions currently available data that do not allow us to test the umbrella criteria assumption?

- The consensus of the workshop participants was that given sufficient time and analytical resources, we can provide more thorough answers to questions 1 and 2 with currently available data. To the extent that uncertainty remains, it is a function of both the amount of data available and the ongoing developmental nature of the analytical tools at hand. How *well* we answer these questions, and in particular the potential for decision error in our listing assessments, is affected by the amount of data collected.
- Nearshore-offshore variability was similar at the scale of a week but breaks down when looking at the scale of a day. Combining shallow water and offshore habitats appear supported at the monthly and weekly scale but in the assessments of the shortest time scale criteria (1-day, instantaneous) may be inappropriate for an umbrella criteria assessment and the habitats needing to be separated.
- Modeling-monitoring comparisons with high frequency data were limited. The Chesapeake Bay model outputs for the TMDL were focused on 1991-2000 while vertical profiler data and most shallow water CONMON is from more recent monitoring program efforts. There is some earlier work with CONMONs that should provide model-monitoring comparisons for the same time and place.

What are the data needed to test this assumption for all conditions?

- Shallow water CONMON has been essential to the evaluation of the umbrella criteria assumptions. Continued investment in shallow water high-frequency data across the tidal Bay habitats are supported.
- Our understanding of variability in mid-channel locations, and particularly in Deep Water regions, is hindered by the paucity of high-frequency vertical profiles available. Vertical profiler data were invaluable in providing verification of assessment for Spectral casting and reduces uncertainty of assessments in segments where it is available. Continuation of the ongoing effort to obtain high-frequency vertical profile data is supported.

Recommendations and Next Steps Emerging from Workshop

- Collect more high-frequency vertical profile data in Deep Water regions, and
- Further explore the concept of duration both as a component of the criteria and as a potential indicator of improving conditions.
- Generate a single common dataset so that every analyst who participates in the collaboration is using the same version of the same data. Expand the dataset to incorporate the most recent data collected using vertical profilers, buoy- and bottom-mounted sensors, and shallow water CONMON stations.
- Update the segment-by-segment quadrant analysis described in USEPA 2004.
- With regard to communicating the results of the quadrant and conditional probability analyses to managers and decision-makers, quantify and clearly communicate the risk of erroneously classifying segment-DUs as impaired or un-impaired ("false positives" or "false negatives") given currently available data and analytical methods in use. In particular, the uncertainty of current calculations of

the "30-day mean" (using only the long-term fixed station datasets) should be quantified and communicated.

- The instantaneous minimum criterion must be defined more precisely than is currently the case. This issue becomes paramount when one begins working with high-frequency datasets, for which there is sometimes a new measurement every 5-15 minutes.
- Convene an expert panel to review the adequacy of the spectral casting method for assessing short-duration criteria.
- Modify the CBP's criteria assessment programs as necessary to conduct a full regulatory assessment of high-frequency criteria for Open Water and Deep Water designated uses, using Spectral casting to produce the synthetic datasets for segments where sufficient data are available. Publish a new USEPA technical addendum to the ambient water quality criteria with the updated assessment approaches.

Related Questions and Topics of Importance Beyond Umbrella Criteria – Insights, Lessons Learned

Combining Shallow-Water and Mid-Channel Assessments?

- There was consensus among participants that when shallow, near-shore and deep, offshore waters are combined in a single volume-based DO assessment, the sheer volume of the offshore region may overwhelm signals of distress that occur in shallower waters. Workshop participants suggested that the partitioning of shallow, near-shore waters into their own assessment units (i.e. sub-segments) may more adequately represent the impact of DO criteria violation in these biologically active regions.

Duration of Hypoxic Events

- A central finding to emerge from the analyses of high frequency observations in shallow water datasets was that there is yet another temporal scale of hypoxia in the Bay in addition to the seasonal scale hypoxia chronic to the deeper portions of the Bay. CONMON data exhibit diel scale hypoxia. Since criteria levels are thresholds of significance to aquatic biota, analysis of *duration below criteria* is a potentially valuable measure of habitat suitability.
- Preliminary assessment suggests maximum duration of violation was later shown to be linearly and positively related to percent violation of a given DO criterion. A more complete development of this index could be conducted comparing segment level impairment results that are output during the CFD assessment on % failure and multiple CONMON sites within segments.

Eutrophication Gradients and DO Variability

- Suggestions for further research included better defining patterns and drivers of diel-cycling hypoxia, and the utility of shallow-water DO concentration variability as a signal of eutrophication. Such climate forcing affects DO dynamics and assessments of Bay health.

- The question of whether the characteristics of variability in DO concentrations are likely to change along a restoration trajectory was of particular interest to workshop participants. At least two conceptual models have been previously conceived for tracking ecosystem response to eutrophication based on high frequency data. The group emphasized that further exploration and development along these lines of inquiry would advance our understanding of the relationships between river flow, eutrophication, and the timing (e.g. day and/or night), duration, and variability of hypoxic events.

Acknowledgments

This workshop report was developed by the dedicated work of the Chesapeake Bay Program Tidal Monitoring and Analysis Ad Hoc Workgroup for Umbrella Criteria assessment and further supported by Chesapeake Bay Program Scientific and Technical Assessment Committee.

Principal and Contributing Authors

This report resulted from the collaborative expertise and talents of the Chesapeake Bay Program's state, federal agency, academic institutions, river commissions and contract partners. Principal authors on the summary report are: Peter Tango - USGS Chesapeake Bay Program Office, Walter Boynton - University of Maryland Chesapeake Biological Laboratory, Tish Robertson - Virginia Department of Environmental Quality, Elgin Perry Statistical Consultant, Claire Buchanan - Interstate Commission on the Potomac River Basin, Matt Hall Maryland Department of Natural Resources, Mike Lane – Old Dominion University and Jeni Keisman - University of Maryland College of Estuarine Science. Contributing authors and their appendices are Appendix 1: Claire Buchanan, Appendix 2: Elgin Perry, Appendix 3: Walter Boynton, Eva Bailey, Lisa Wainger, and Amy Drohan, Appendix 4: Walter Boynton, Eva Bailey, Matt Hall and Elgin Perry, Appendix 5: Mark Brush, Iris Anderson, Howard Kator, Appendix 6: Donna Bilkovic, Appendix 7: Marcia Olson, Appendix 8: Matt Hall, Appendix 9: Tish Robertson and Mike Lane, Appendix 10: Elgin Perry. Appendix 11: Elgin Perry. Appendix 12:Elgin Perry.

Umbrella Criteria Workshop Steering Committee

Walter Boynton, UMCES, Claire Buchanan, ICBPB, Aaron Gorka, CRC/CBPO, Marjy Friedrichs, VIMS/STAC, Matt Hall, MD DNR, Rick Hoffman, VADEQ, Jeni Keisman, UMCES/CBPO, Mike Lane, ODU, Elgin Perry, Statistical Consultant, Tish Robertson, VADEQ, Peter Tango, USGS/CBPO, Elizabeth Van Dolah CRC/STAC Coordinator

Umbrella Criteria Workshop Participants/Invitees
Invitees:
Rich Batiuk, United States Environmental Protection Agency; Donna Bilkovic, Virginia Institute of Marine Science; Walter Boynton, University of Maryland Center for Environmental Science; Denise Breitburg, Smithsonian Environmental Research Center; Mark Brush, Virginia Institute of Marine Science; Claire Buchanan, Interstate Commission on the Potomac River Basin; Marjy Friedrichs, Virginia Institute of Marine Science; Matt Hall, Maryland Department of Natural Resources; Robert Hirsch, United States Geological Survey; Rick Hoffman, Virginia Department of Environmental Quality; Fred Holland, Former Director of the NOAA Marine Lab in Charleston, SC and SC Department of Natural Resources Marine Resources Laboratory; Dave Jasinski, Chesapeake Research Consortium; Jeni Keisman, University of Maryland Center for Environmental Science; Mike Lane, Old Dominion University; Kenneth Moore, Virginia Institute of Marine Science; Andrew Muller, United States Naval Academy; Nagaraj

Neerchal, University of Maryland Baltimore County; Marcia Olsen, Statistical Consultant; Elgin Perry, Statistical Consultant; James Pierson, University of Maryland Center for Environmental Science; Tish Robertson, Virginia Department of Environmental Quality; David Rudders, Virginia Institute of Marine Science; David Secor, University of Maryland Center for Environmental Science; Gary Shenk, United States Environmental Protection Agency; Peter Tango, United States Geological Survey; Mark Trice, Maryland Department of Natural Resources; Liz Van Dolah, Chesapeake Research Consortium; Steve Weisberg, Southern California Coastal Research Project;

1. Introduction

1.1 The "Umbrella Criteria" Assumption

In order to ensure the protection of living resources, DO criteria were developed to apply to different time periods ranging from instantaneous (those measured at a single point in time) to 30-day average concentrations (Table 1). However, synoptic monitoring of the Bay has generally occurred on a temporal scale of only one or two measurements per month over the 25-year history of the Chesapeake Bay Program. As a result, insufficient monitoring data and/or appropriate analytical methods exist to support widespread assessment of the shorter-term criteria for the **92 individual management segments** of the Bay (USEPA 2004; see Figure 1).

Table 1. Chesapeake Bay Water Quality Criteria (from USEPA 2003).

Designated Use	**Criteria Concentration/Duration**	**Protection Provided**	**Temporal Application**
Migratory fish spawning and nursery use	7-day mean ≥ 6 mg/L (tidal habitats with 0-0.5 salinity)	Survival/growth of larval/juvenile tidal-fresh resident fish; protective of threatened/endangered species	February 1-May 31
	Instantaneous minimum ≥ 5 mg/L	Survival and growth of larval/juvenile migratory fish; protective of threatened/endangered species	
	Open-water fish and shellfish designated use criteria apply		June 1 – January 31
Shallow-water bay grass use	Open–water fish and shellfish designated criteria apply		Year-round
Open-water fish and shellfish use[1]	30-day mean ≥ 5.5 mg/L (tidal habitats with >0.5 salinity)	Growth of tidal-fresh juvenile and adult fish; protective of threatened/endangered species	Year-round
	30-day mean ≥ 5 mg/L (tidal habitats with >0.5 salinity)	Growth of larval, juvenile and adult fish and shellfish; protective of threatened/endangered species	
	7-day mean ≥ 4 mg/L	Survival of open-water fish larvae.	
	Instantaneous minimum ≥ 3.2 mg/L	Survival of threatened/endangered sturgeon species[1]	
Deep-water seasonal fish and shellfish use	30-day mean ≥ 3 mg/L	Survival and recruitment of bay anchovy eggs and larvae.	June 1 – September 30
	1-day mean ≥ 2.3 mg/L	Survival of open-water juvenile and adult fish	
	Instantaneous minimum ≥ 1.7 mg/L	Survival of bay anchovy eggs and larvae	
	Open-water fish and shellfish designated-use criteria apply		October 1 – May 31
Deep-channel seasonal refuge use	Instantaneous minimum ≥1 mg/L	Survival of bottom-dwelling worms and clams	June 1 – September 30
	Open-water fish and shellfish designated use criteria apply		October 1 – May 31

1. Note: At temperatures considered stressful to shortnose sturgeon (*Acipenser brevirostrum*) (>29°C) dissolved oxygen concentrations above an instantaneous minimum of 4.3 $mg*L^{-1}$ will protect survival of this list sturgeon species.

Currently, Chesapeake Bay long term water quality monitoring data are deemed adequate to assess Clean Water Act (CWA) 303(d) listing status for only the 30-day mean criteria of the Open Water and Deep Water designated uses of the Bay (USEPA 2003). As the Deep Channel designated use contains only one criterion (an instantaneous minimum), it is assessed using data collected on the same temporal and spatial scale as those used to calculate monthly means for the OW and DW designated uses.

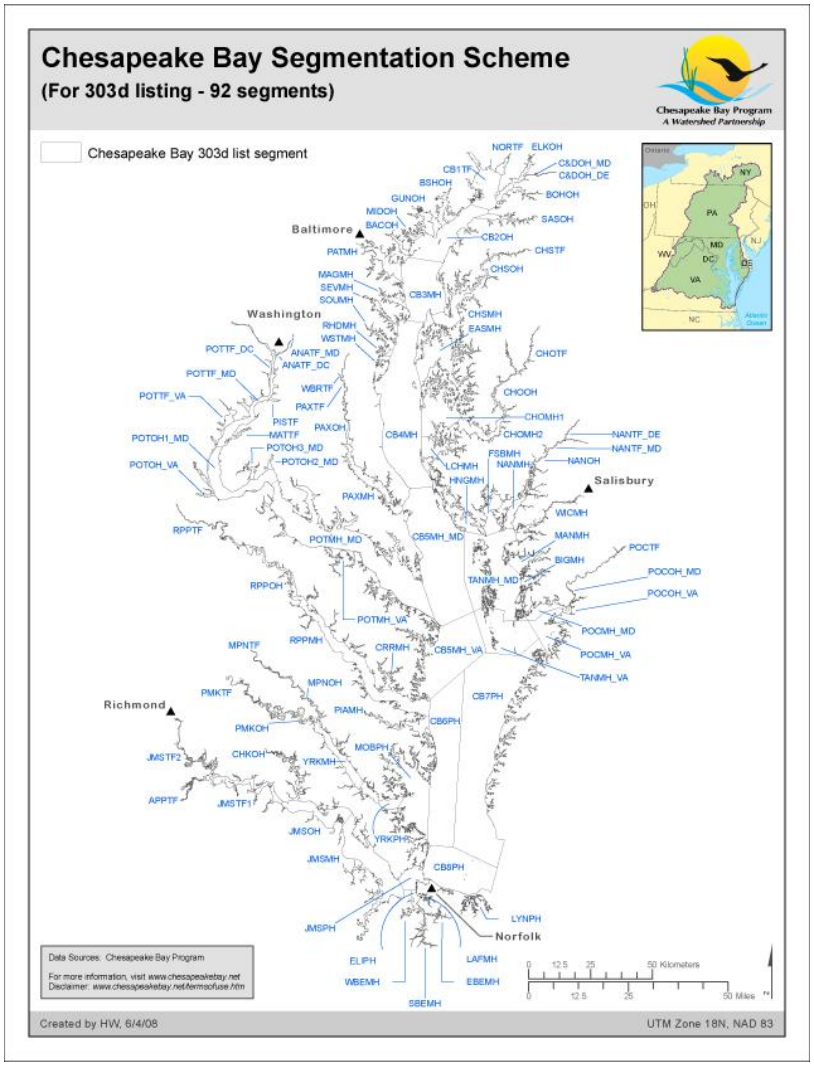


Figure 1. Chesapeake Bay 92 segment management grid of the tidal waters.

In the course of developing Total Maximum Daily Loads (TMDLs) for pollutants entering the Chesapeake Bay, it was necessary to ascertain if assessment of the 30-day mean DO criteria gave sufficient information to accurately determine the 303(d) listing status of the OW and DW designated uses. Because the estuarine water quality sediment transport model (WQSTM) being used to develop the TMDL simulates DO concentrations on an hourly time step, analysts at the USEPA CPB (Shenk, Keisman) conducted an assessment of all water quality criteria using hourly output from the calibration run of the Phase 5.1 version of the WQSTM. Note that for the purposes of developing the Chesapeake Bay TMDL, the summer season defined by the criteria (June – September) is assumed to be the limiting season in all designated uses being assessed for DO impairment. Thus efforts to evaluate the Umbrella Criteria assumption focused primarily on this timeframe.

The version of the WQSTM used to support management community decisions for Bay restoration produces simulated data at a high spatial (1 km^2) and temporal (1 hr time step) resolution, compared to actual monitoring data. This high-resolution output was used to evaluate whether the DO criteria being assessed by the CBP Water Quality Monitoring Program are more or less protective than those criteria that are currently not assessed. Model results indicated that when the summer 30-day mean criteria are attained in OW and DW designated uses, the associated higher frequency criteria (i.e. 7-day, 1-day or instantaneous as appropriate for a designated use) are also attained (Shenk and Batiuk, 2010). In other words, non-attainment of the 30-day mean water quality criteria in the Summer OW and DW designated uses occurred at a higher rate than for the instantaneous minimum, 1-day, or 7-day mean criteria for the same designated use. Furthermore, in segments containing a Summer DC designated use (8 of the 92 segments in Chesapeake Bay), non-attainment rates of the summer instantaneous minimum criterion for the DC were higher than for any other criterion in the OW and DW designated uses of the same segment. *Thus the criteria currently being assessed by the Chesapeake Bay long term water quality monitoring program appear to be "umbrella criteria" – the most restrictive of all available criteria*.

Based on this analysis, the CBP determined that evaluation of the 30-day mean criteria was sufficient to establish Clean Water Act 303d listing status for the OW and DW designated uses of the Bay. We have called this determination the **"Umbrella Criteria Assumption"** because it surmises that one criterion (i.e. the 30-day mean) serves as an "umbrella" for the remaining criteria of a given designated use.

Members of the Chesapeake Bay scientific and management community raised questions concerning the validity of this assumption, particularly in light of the understanding that the WQSTM is calibrated using only the long-term, biweekly to monthly, mid-channel monitoring dataset. The model thus may not adequately capture the true variability of dissolved oxygen concentrations relevant to the short-term criteria. In recent years, the development and application of technologies for near-continuous monitoring of water quality has allowed the collection of datasets of a high temporal density (as short as 15-minute intervals) in select regions of the Bay. While limited in spatial distribution and length of monitoring period, these datasets shed light on the short-term (minutes-to-months) variability of dissolved oxygen in both shallow and offshore locations within a subset of the Bay's 92 segments.

1.2 What are the Chesapeake Bay Water Quality Criteria?

Dissolved oxygen (DO) water quality standards were established for the Chesapeake Bay and its tidal tributaries ("the Bay") under the authority of the Clean Water Act, in order to protect the reproduction, survival and growth of estuarine living resources. Standards are defined by a *criterion*, which represents a threshold below which DO concentrations should not fall, and also by a representation of *duration* and *frequency* for the criterion. The concepts of duration and frequency serve to refine the criterion by defining the degree to which it may be exceeded occasionally, without inflicting unacceptable harm on the intended resource.

The primary focus of this report is an evaluation of the relative protectiveness of DO criteria that were established to safeguard aquatic life during the summer season

(June – September) in various habitats of the Bay. The duration and frequency concepts become relevant in the course of determining the degree to which observed violations of different criteria would result in non-attainment of the water quality standard overall. Further explanation of these details will be provided in the analytical results section of the report. For now, it is most important to understand that there are multiple DO criteria for the Bay, and that they vary with timeframe as well as location (Table 1).

Criteria for DO were set for five categories of "designated uses" (habitats used by various living resources in different seasons and life stages): migratory and nursery (MN), shallow water (SW) for Bay grasses, open water (OW), deep water (DW) and deep channel (DC) (Figure 2). The Chesapeake Bay is also divided up with a management grid of **92 individual management segments** (see Figure 1) within which one or more of the designated uses can apply. In the majority of tributary segments, the OW designated use applies to the entire volume of water in the segment – from surface to bottom and shore to shore. Shallow waters are not separated from mid-channel waters for dissolved oxygen assessments (USEPA 2003b). In those segments containing Summer DW and DC designated uses, the OW use comprises those well-mixed waters occurring above the upper boundary of the pycnocline, and again extends from shore to shore (USEPA 2003b). When stratification is observed, and only for those segments identified as having DW and DC designated uses, the DW designated use pertains to those waters located below the upper boundary of the pycnocline. The DC designated use (when it is present) pertains to the lower mixed zone occurring below the lower boundary of the pycnocline layer (USEPA 2003b).

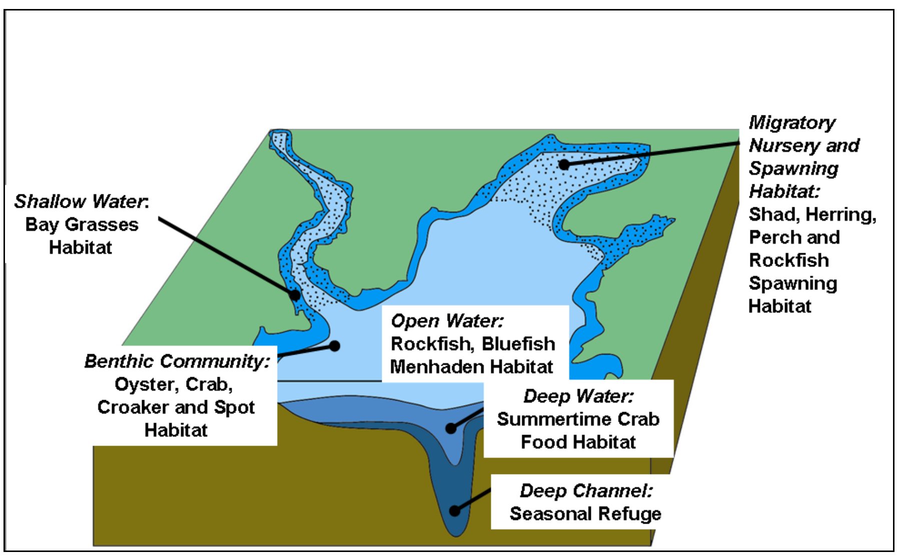


Figure 2: Chesapeake Bay water quality designated uses (USEPA 2003).

Migratory and nursery, OW, and SW DO criteria were set at concentrations designed to prevent impairment of growth, and to protect reproduction and survival of all aquatic organisms (USEPA 2003, Batiuk et al, 2009; Table 1). Deep water seasonal fish and shellfish designated use habitats were defined for the summer season only (June-September), when the water column is significantly stratified, and were set at levels to protect the survival of juvenile and adult fish and shellfish, as well as the recruitment success of the Bay anchovy. Criteria for the summer deep channel seasonal refuge designated use habitats were designed to protect the survival of bottom sediment dwelling worms and clams (USEPA 2003a).

When the DO criteria were developed, DO requirements for the species and communities inhabiting open water and shallow water habitats were considered similar enough that a single set of summer season criteria could protect both designated uses (Batiuk et al. 2009). Water quality of the shoreline habitats in the Bay was poorly characterized at the time. However, USEPA (2007) acknowledged the importance of acquiring better understanding of water quality dynamics for shallow Bay habitats. Preliminary evidence suggested conditions in these areas can differ greatly from the mid-channel environment. Many segments (47 of 91 with 1 segment still needing data) have substantial (>50%) shallow water habitat that can potentially exert significant influence on criteria assessment results (Table 2). There are multiple analyses contained in this report lending new insights into Chesapeake Bay shallow water behavior.

Table 2. Chesapeake Bay segments and segment areas ≤ 2 meters illustrating the amount of shallow water habitat in each segment.

2003 Bay Management Segment	SAV Tier III goal (acres <2m)	Total Segment acres	% Area ≤ 2m
CHOTF	0	2201	0.0
WBEMH	0	1484	0.0
SBEMH	0	2074	0.0
EBEMH	0	1427	0.0
LAFMH	0	1422	0.0
ELIPH**	0	5227	0.0
CB8PH	1053	101913	1.0
CB6PH	5130	183687	2.8
CB4MH	9301	224582	4.1
CB5MH	18691	364395	5.1
CB3MH	5510	89350	6.2
CB7PH	32575	375803	8.7
JMSPH	2266	18919	12.0
CB2OH	9212	68013	13.5
PATMH	3543	23130	15.3
C&DOH	170	881	19.3
POTMH	45807	219396	20.9
PMKOH	860	3483	24.7
TANMH	58024	221885	26.2
POTOH	15199	53119	28.6
SEVMH	2108	7262	29.0
CHOMH1	18424	59814	30.8
MPNOH	613	1965	31.2
MAGMH	2177	6559	33.2
PAXMH	8829	26584	33.2
CHOMH2	6222	18335	33.9
CHOOH	1285	3716	34.6
JMSOH	10954	31567	34.7
EASMH	20808	57961	35.9

MOBPH	30554	84687	36.1
POCMH	17969	48414	37.1
RPPMH	30035	80020	37.5
SOUMH	2288	5926	38.6
JMSMH	29138	75180	38.8
CHSMH	11510	29477	39.0
RHDMH	904	2251	40.2
YRKPH	7139	16906	42.2
POCOH	1515	3415	44.4
CRRMH	2612	5803	45.0
PIAMH	7789	17242	45.2
SASOH	3699	8176	45.2
POTTF	17838	37173	48.0
RPPTF	4515	9020	50.1
NANOH	2056	4066	50.6
RPPOH	2511	4828	52.0
LCHMH	11799	22135	53.3
YRKMH	12666	23375	54.2
CB1TF	20401	37466	54.5
JMSTF	12842	23550	54.5
ELKOH	5028	9210	54.6
WSTMH	1527	2793	54.7
PAXOH	2073	3520	58.9
MPNTF	1352	2293	59.0
BSHOH	4606	7547	61.0
MIDOH	2481	4007	61.9
MANMH	9338	15021	62.2
CHSOH	2310	3655	63.2
HNGMH	15481	24147	64.1
NANMH	7714	11949	64.6
BOHOH	1905	2947	64.6
PAXTF	707	1089	64.9
CHKOH	4505	6911	65.2
FSBMH	13633	20635	66.1
PMKTF	2654	4010	66.2
NORTF	2743	3909	70.2
BIGMH	5068	7183	70.6
BACOH	2861	3997	71.6
GUNOH	7460	10378	71.9
WICMH	6385	8677	73.6
POCTF	749	988	75.8
MATTF	1389	1799	77.2
NANTF	887	1139	77.9
APPTF	1604	1980	81.0
LYNPH	3961	4845	81.8
CHSTF	870	1009	86.2
WBRTF	32	32	98.5
PISTF	914	917	99.7
ANATF	-	842	-

1.3 How are Chesapeake Bay Dissolved Oxygen Criteria Assessed? CFD – The Chesapeake Bay Water Quality Criteria Assessment Methodology

The water quality criteria assessment methodology currently used by the USEPA Chesapeake Bay Program evaluates observed violations of the dissolved oxygen criteria against a cumulative frequency diagram (CFD) curve (USEPA 2003, 2007) that defines the degree of allowable violation, or "exceedance." This **reference curve,** a curve of compliance, is represented in a two dimensional plane of percent space and percent time (Figure 3). Historically, USEPA has provided for a 10% allowable exceedance in temporal or spatial assessments against thresholds. The application of the two-dimensional CFD is an innovative approach that is designed to better reflect the relative impact of varying degrees of violation. The curve may be a mathematically derived 10% curve or a "bioreference" curve. The bioreference curve is a reference CFD derived from observed criteria violations tolerated by healthy communities of a relevant biological resource. The bioreference curve could allow more or less than the 10% allowable exceedance defined by the mathematically derived curve.

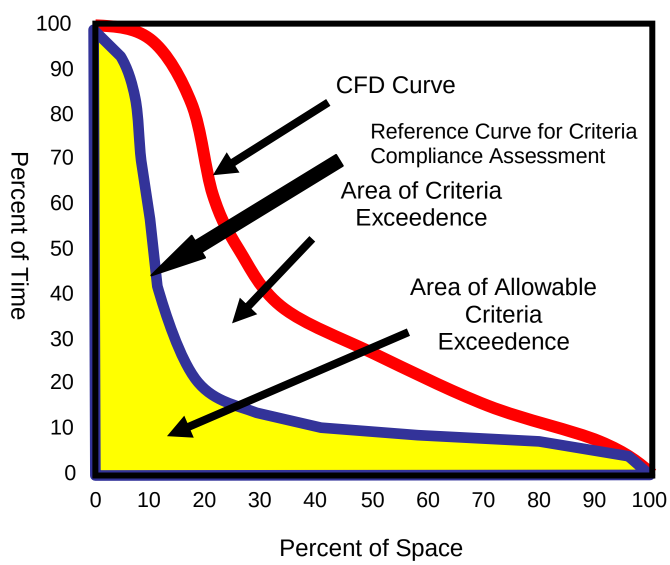


Figure 3. Reference curve, CFD compliance curve, and an illustration of the area of allowable exceedances.

The CFD approach is considered the best science currently available for assessment of the Bay's water quality criteria (USEPA 2007). Because the Umbrella Criteria concept is dependent upon the assessment methodology, the general procedural outline from data collection through compliance assessment is provided below (Box 1):

Box 1: DO criteria assessment procedure

Step 1. Collect data at known locations.

Step 2. Spatially interpolate the 30-day means across the entire segment.
- 2.1 Vertical interpolation first
- 2.2 Horizontal interpolation next
- 2.3 Interpolate the 30-day means by month
- 2.4 Apportion results by designated uses

Step 3. Determine the compliance status of each cell in the segment volume

Step 4. Produce a percent compliance matrix with sample period and percent space in compliance

Step 5. Rank the percent compliance in space from greatest to lowest values and assign percent of time associatied with the compliance values.

Step 6. Plot ranked percent space (x-axis) against percent time (y-axis).

Step 7. Evaluate habitat compliance against the reference curve.

Footnote: For background on reasoning leading to the development of this assessment tool, illustration of the technique, discussion of its properties and unresolved issues, and details to the procedural outline, see USEPA (2007 Appendix A).

This framework is currently applied to the assessment of the 30-day mean criteria in the OW and DW designated uses, and to the instantaneous minimum criterion in the DC designated use. It is currently considered desirable to extend the same method to assessment of the short term criteria. Historically, assessment of criteria for sub-30-day time periods has been limited by the availability of data at the appropriate temporal resolution.

1.4 History: Prior Insights on Comparative Criteria Protectiveness

The question of the relative protectiveness of water quality criteria designed for different temporal scales has been raised previously in the history of the Chesapeake Bay Program partnership. Early in the 1990s, when experts first identified the suite of DO concentrations necessary to protect the Bay's aquatic living resources, there was recognition that the temporal scale of the long-term fixed station monitoring program would, by itself, be insufficient to assess shorter-term criteria. By this time, several groups had already begun to experiment with near-continuous monitoring technologies, which allowed them to measure DO concentrations on timescales as short as 3-4 seconds (DATAFLOW) during spatial assessments and 5-15 minutes for days-to-months at a time at a fixed location (CONMON). In 1992, workers at the Maryland Department of Natural Resources and the EPA Chesapeake Bay Program used a combination of data from the nascent fixed-station monitoring program and a limited number of near-continuous datasets from local researchers to analyze the relationships between average seasonal DO concentrations, monthly mean concentrations, and "instantaneous" measures (i.e. individual observations) of DO (see Jordan et al 1992). Their objective was twofold: to identify the range of seasonal means within which DO was considered problematic, and to describe the relationship between these seasonal means and violations of the targeted monthly DO thresholds. They developed regression equations to derive the seasonal mean concentrations that could be presumed protective of target shorter-duration concentrations in a given segment. They concluded that knowing the seasonal mean DO concentration for a given region in the Bay permitted "a good estimate of what proportion of actual DO observations are likely to meet, or fail to meet, each of the target concentrations."

A decade later, with the publication of the official Chesapeake Bay DO criteria (USEPA 2003), the USEPA acknowledged that the fixed station monitoring program – designed to capture long-term trends, as well as seasonal and inter-annual variation – was "poorly suited for assessing" the 7-day mean, 1-day mean, and instantaneous minimum DO criteria (p. 177, USEPA 2003). At the same time, the high cost of direct monitoring on these timescales – particularly in near-shore waters where spatial variability was assumed to be high – precluded direct assessment of the shorter-duration criteria. It was suggested that eventually, assessment of these criteria could be accomplished using "statistical methods that estimate probable attainment (ibid, p. 179)." Two statistical approaches for assessing the sub-30-day criteria were discussed. A "spectral analysis" approach, first introduced by Neerchal et al (1994), was recommended for assessment of the 7-day and 1-day mean criteria, with the caveat that insufficient availability of high-frequency data and insufficient validation of the approach prevented its immediate application. For the instantaneous minimum criterion, the logistic regression approach (an updated version of the method applied by Jordan et al in their 1992 analysis) was recommended, again with the caveat that further development and validation would be required before any formal assessment using the approach could be accepted.

In 2004, the CBP revisited this question of comparative protectiveness among criteria once again with its publication of an addendum to the original Ambient Water Quality Criteria publication of 2003. Olson et al. (Chapter 5, USEPA 2004) compiled a database of 147 buoy data sets collected between 1987-1995 (where dates were noted), primarily by the EPA's EMAP program, to explore this question. They explored the relationship between the first percentile of DO concentrations (approximating the idea of instantaneous minimum) and the 30-day mean DO concentration within the same time period. They further developed and demonstrated the potential utility of using regression models to predict attainment of short-term criteria. More importantly, they showed that the relative protectiveness of the 30-day mean criterion varies across segments. They provided a catalogue of segments where they postulated that the 30-day mean criterion was protective of the instantaneous minimum criterion, as well as a list of key segments where this assumption should be made with caution, or not at all. In contrast to their findings for the instantaneous minimum criterion, Olson et al. found that the 30-day mean was generally protective of the 7-day mean criterion in those segments where both criteria applied.

With its publication of the 2007 Addendum to the Ambient Water Quality Criteria publication of 2003, the EPA CBP again expressed its support for continued development and eventual application of adequate statistical approaches for assessing the instantaneous minimum, 1-day mean, and 7-day mean DO criteria. Appendix E of USEPA (2007) detailed further advances in the development of the logistic regression approach, including the addition of models for each station in the long-term fixed station record. It also recommended eventual application of the spectral analysis approach, if further development showed it to be at least as robust as the logistic regression approach.

Additional exploratory analyses were conducted by Chesapeake Bay community researchers and analysts and described in the Chesapeake Bay Program Monitoring Realignment (MRAT) report (MRAT 2009). The results questioned but did not refute the protectiveness of the umbrella criterion concept. Further development of the MRAT analyses were requested to provide a better assessment of the umbrella criterion, ascertain conditions where the umbrella criterion assumptions may be violated, identify gaps in our ability to test the assumptions and recommend future data analysis and monitoring needs.

1.5 2010 Umbrella Criteria Workshop & Report Proposal

In early 2010, the Chesapeake Bay Program's Tidal Monitoring and Analysis Workgroup (TMAW) convened an ad hoc team of scientists, analysts and managers to analyze the available high frequency data and investigate, where spatiotemporal comparisons were possible, the relative protectiveness of the 30-day mean criterion to the instantaneous, 1-day, and 7-day mean criteria in OW and DW habitats. Their objective was to characterize the conditions under which the "umbrella criteria assumption" was upheld, and when and where the assumption was likely to be violated. Members of ad hoc team conducted several independent analyses to evaluate the validity of the assumption in regions with varying physical and ecological characteristics, ranging from shallow, tidal fresh regions to the mainstem of the Chesapeake Bay.

Another implicit objective of the Umbrella Criteria investigation was to evaluate analytical approaches and identify those that could increase the utility of the new, high frequency data and facilitate assessments of the short-term (7-day, 1-day) DO criteria. Section 2 of this report summarizes the data and methods used to perform these analyses, their results, and conclusions and recommendations regarding the Umbrella Criteria assumption.

Section 3 provides discussion of related questions and concerns that arose in the process of conducting this evaluation. Section 3 also describes a new shallow water habitat characterization for water quality dynamics previously un-described for Chesapeake Bay. The topics described in this section may lead to more recommendations for future work to support the accurate listing and de-listing of the 92 Chesapeake Bay management segments.

2. Umbrella Criteria Analyses

2.1 The Data

High frequency data collection on season-length scales has grown increasingly common in the Chesapeake Bay since 2000, and received a boost in 2003 with the start-up of the Chesapeake Bay Program Shallow Water Monitoring Network. These datasets, coupled with short term deployments of high frequency monitoring in Chesapeake Bay (e.g. USEPA EMAP data) provide the strongest available basis for testing new methods of data analysis and integration to address assessment of short term DO criteria. In addition to the Chesapeake Bay Program long-term fixed station monitoring and short-term buoy datasets already described above, recent data collected as part of the states' continuous monitoring (CONMON) and underway surface water mapping (DATAFLOW) programs, as well as data from recent VIMS and NOAA vertical profiler and VIMS ACROBAT deployments, were included in analyses evaluating the umbrella criterion assumptions (MRAT 2009, Table 3 this report,).

Table 3. Data sources serving the Umbrella Criteria analyses.

Program Description	Data Collection and Availability	Sampling Locations and Habitats
CBP long-term water quality monitoring program: Low temporal frequency and spatial spatial resolution, good vertical profile resolution of the data collection.	1985-present. Biweekly to monthly sampling. Water column profiles taken with grab samples and sensors. Web accessible data: *CBP CIMS* accessible.	Fixed site, mid-channel, approximately 150 stations. Covers tidal fresh to polyhaline habitat conditions.
USEPA EMAP primarily: Historical short-term buoy deployments with high temporal frequency at a station. Single depth sensor evaluations.	Mix of short term (days to weeks) time series with high temporal frequencies by sensor. See USEPA 2004.	Fixed site, off shore locations, varied depths represented across many dozens of studies. Tidal fresh to polyhaline habitat conditions.
CBP Shallow Water Monitoring Program, continuous monitoring (CONMON): High temporal frequency at moored locations.	Approximately 2000-present. Mostly seasonally, near continuous (15 min interval) time series April-October. Fixed depth sensor, usually 1m off bottom. Web accessible data: *Eyes on the Bay* in MD, *VECOS* in Virginia.	Fixed site, shallow water, nearshore locations, approximately 70 sites Baywide with 1-9 yrs of data. Tidal fresh to mesohaline conditions.
VIMS, MD DNR Vertical Profilers: High temporal frequency in 2 dimensions. **VIMS**: Bottom sonde .	Approximately varies, 2006-present. Limited seasons. Sensors provide water column profiles at sub-daily scales. Bottom sonde. Web accessible data: MD DNR and VADEQ.	Fixed sites (n<5), offshore locations in MD (Potomac River) and VA (York and Rappahannock Rivers). Dominantly mesohaline data in these analyses.
CBP Shallow Water Monitoring Program, surface water quality mapping with DATAFLOW: High Spatial resolution along temporally dense collection track.	Approximately 2000-present. Biweekly to monthly assessments within April-October season. Multi-year assessments (3 yr sets). Sensor 0.5m below surface Web accessible data: *Eyes on the Bay* in MD, *VECOS* in Virginia.	Chesapeake Bay Program management segments. Approximately 40 of 92 segments assessed to date. Tidal fresh to polyhaline habitats.
VIMS Volumetric Assessment with ACROBAT (towed sensor underwater at variable depths). High spatial resolution -	Approximately 2003-present Limited seasons. 3-dimensional sensor assessment of water column water quality. *VIMS data*, Brush et al.	York and Rappahannock Rivers (VA) study sites, deep water reaches. Dominantly mesohaline habitat.

To date, the development and assessment of DO criteria has relied primarily on the CBP long-term, low frequency data set and a few high frequency data sets from buoys deployed for short periods in the Bay (Jordan et al. 1992, USEPA 2004). The CBP long term water quality monitoring program has more than 25 years of data in vertical profile of the water column, collected biweekly to monthly ("low frequency"), at approximately 150 fixed stations in the mainstem Bay and its tidal tributaries. One or more stations are located in each of the CBP management segments (Figure 1). Data are available through the Chesapeake Information Management System (CIMS) database located at the CBP Office in Annapolis, MD.

Technological advancements and longer buoy deployments now make near-continuous or high frequency measurements of several water quality parameters – including dissolved oxygen – easy, reliable and accessible (McCaffrey 2004). Since 2003, fixed station nearshore CONMON and DATAFLOW sampling programs have been adopted as elements of the overall CBP long term water quality monitoring program strategy. The CONMON program collects data of high temporal frequency at static, nearshore, shallow-water locations, while the DATAFLOW program conducts surface water sampling cruises in a spatially intensive manner. Both approaches have fixed depth assessment limitations.

Due to cost, time and personnel constraints, CONMON and DATAFLOW are performed within a subset of CBP management segments each year. Each segment is monitored for a three-year period, in order to support criteria assessment procedures developed by the CBP partnership (USEPA 2003, Batiuk et al. 2009). Once three full years of data have been collected in a given segment, CONMON equipment is moved to a new segment. This rotation of the states' CONMON resources is intended to eventually provide high-frequency data on shallow-water regions for all 92 segments, in support of a comprehensive Baywide assessment of water quality conditions. In some years, over 50 CONMON stations are being maintained and operated throughout the Chesapeake Bay tidal system.

Research efforts to evolve water quality assessment applications in Chesapeake Bay using such advances as vertical profilers, ACROBAT and Autonomous Underwater Vehicle (AUV) technology are currently underway. Further development of these monitoring technologies will increase the temporal resolution of vertical profiles at fixed stations, and volumetric resolution (using a 'random walk' methodology), respectively.

2.2 Approaches Used to Evaluate the Umbrella Criteria Assumption.

Due to the experimental nature of some of the analytical methodologies available for this work, the Chesapeake Bay Program's Tidal Monitoring and Analysis Workgroup ad hoc Umbrella Criteria Analysis Team pursued a "multiple lines of evidence" approach to evaluating the validity of the Umbrella Criteria assumption. Evaluation approaches included reference to bi-plot evaluations with short term high frequency data sets previously published (USEPA 2004), "point-based" comparisons of violation rates (i.e. percent of measurements violating criterion) on instantaneous, 7-day, and 30-day timescales using shallow-water CONMON datasets, to a statistical evaluation of the probability of violating short-term criteria given estimates of the variability of short-term (i.e., instantaneous, 7-day mean) concentrations around the 30-day mean, to full 3-year

CFD assessments of synthetic datasets produced using the spectral casting method. Each approach provided additional levels of insight.

2.2.1 Focus Method Update: Validation of the Spectral Casting Method for Application to Assessing Short-Term Chesapeake Bay Dissolved Oxygen Water Quality Criteria Attainment.

USEPA (2003) and USEPA (2007) criteria publications recommended the further development and eventual application of a statistical technique called spectral analysis for assessing short-term dissolved oxygen criteria. However, as both publications stated, further validation of this tool was required in order to demonstrate its adequacy for DO criteria assessment. Thus evaluation and validation of spectral analysis techniques comprised an essential component of the ad-hoc Umbrella Criteria project.

The updated and validated method described in this section is referred to as **spectral casting**. Spectral Casting uses spectral analysis as an interpolation device to create a synthetic high frequency data set for monitored locations where high frequency data are not available (Perry Appendix 10, 11). The **sending site** is a monitoring location with high frequency data (e.g., 5-, 15-, 30-min intervals) which might be measured by an automated DO sensor (e.g. on a buoy). The **receiving site** is a monitoring location with low frequency data (e.g., 1 or 2 measures/month) which might be one of the fixed station monitoring sites of the CBP Long-term Water Quality Monitoring fixed station network. High frequency synthetic data for the receiving site are formulated by combining the low frequency data variability signal from the receiving site with the high frequency data variability signal for the sending site (Box 2).

Box 2: Spectral Casting Definitions

Definitions related to Spectral Casting
(Elgin Perry 2010).

Umbrella Criterion: the most protective criterion. When compliance with one criterion insures compliance at others, the one (i.e. most protective) is termed "the umbrella" criterion.

Spectral Casting: In the early 1990's it was proposed (Neerchall, 1994) that spectral analysis could be used to create a synthetic high frequency data set at a monitoring location with only low frequency data. Because the technique involves transporting the high frequency signals in observations from one monitoring location to a nearby location, Perry proposes we call this technique 'Spectral Casting' - an analogy to casting with a spinning rod.

Sending Site: Spectral casting involves combining the high frequency data variability signal from one site with the low frequency data variability signal of a second monitoring site. The result is a synthetic high frequency data record for the low frequency-monitoring-only site. Because the high frequency signal is being transported from one site to another, Perry proposes the site that generates the high frequency signal the 'Sending Site'.

Receiving Site: Following up on the preceding definition, the low frequency site that receives the high frequency signal is called the 'Receiving Site'.

High/Low Frequency Criterion: High (Low) frequency criteria are criteria which require high (low) frequency data for assessment.

Three elements of the spectral casting method were evaluated and validated. First, the use of the Fast Fourier transform for spectral decomposition of sending and/or receiving sites was compared to the cubic spline and linear interpolation techniques. Second, the concern that high-frequency data from any given sending site may not adequately represent the variability at the low-frequency receiving site was addressed. Finally, some preliminary analyses were conducted to evaluate potential in the variability of DO concentrations with depth.

Spectral Casting Validation: Comparison of spectral decomposition and interpolation methods

The first computation step in spectral casting uses a Fast Fourier Transform (FFT) to obtain a spectral decomposition at both monitoring locations (sites). The benefits of using the FFT interpolation approach with spectral casting is that it is computationally fast, allows cycle trimming, deals with cyclical prediction and preserves autocorrelation structure in the data. However, limitations on the technique include assumptions of cyclical behavior, equally spaced inputs in time and equally spaced outputs (Perry Appendix 10). The cubic spline approach has fewer implementation constraints than the FFT method.

To explore the most effective interpolation technique, Robertson and Lane evaluated results of full Chesapeake Bay water quality CFD criteria assessments for DO in a management segment with FFT versus a cubic spline (Appendix 9, Part A). Overall, the cubic spline interpolation produced lower violation rates than when using the FFT analysis. The degree of difference between the techniques appeared to relate to the underlying nature of the long-term data. However, cubic spline interpolation did not change the support for criteria assessment results (OW 30-day mean as an umbrella to the 7-day mean) compared to the FFT.

Perry compared the results FFT, spline and linear interpolations of time series (Appendix 11). His comparisons of cubic spline interpolation with FFT interpolation supported the findings of Robertson and Lane, with little difference in results. However, **comparisons of FFT and cubic spline interpolation results to linear interpolation results for the low-frequency component of the casting process found that in this step, linear interpolation provided a better fit than either the FFT or spline techniques in the examples tested to date.**

Spectral Casting Validation: Evaluating Sources of Uncertainty

In the first step of the spectral casting method a low frequency sample is interpolated in the time domain to estimate central tendency over time. In the second part of the process, short term high frequency variability is borrowed from a sending site in an effort to fill in the extremes of variability around the estimated central tendency. The resulting high frequency synthetic data then allows for estimates of short term temporal assessments of the frequency of excursions beyond criteria thresholds.

It is clear that there are two sources of error in this estimation procedure. On the one hand there is error because low frequency point samples may not capture the long term trends at the receiving location. On the other hand, high-frequency variation borrowed from a sending site may not represent high-frequency variation at the receiving site. It has generally been accepted that the Chesapeake Bay Program long-term, low-frequency fixed station water quality monitoring program adequately captures monthly-to-seasonal conditions and long-term trends. Thus concerns regarding the adequacy of the spectral casting method have focused primarily on the second type of error described above. In particular, questions regarding the maximum distance that should be allowed between a sending site and a receiving site, and regarding whether variability from a shallow-water sending site can be cast to a mid-channel receiving site, have reduced workers' initial comfort with application of the spectral casting method.

In an effort to understand the relative contributions of these two sources of error, Dr. Perry constructed a validation exercise (Appendix 11). He sub-sampled a high frequency time series to create a low frequency subsample that could serve as a receiving site of the spectral cast. The low frequency subsample was interpolated as if it were an actual low frequency time series from the fixed station data. Using FFT analysis, the high frequency variation from a sending site was cast in and convoluted with the low frequency interpolation to form synthetic data. The validation step compared the percent violation in the synthetic data to the true percent violation in the original high frequency time series. Two variations on this validation exercise allowed Dr. Perry to differentiate uncertainty due to low frequency sampling from uncertainty due to casting variability between monitoring sites. He performed these validation exercises first on data from several buoys moored at depths ranging from approximately 2 to 20 meters in the Chesapeake Bay mainstem and lower tributaries of the southern Bay, and then on datasets from two shallow water CONMON deployments in the lower Potomac River.

Results were consistent across both validation tests: **the variability due to casting site was less than the variability due to the selection of the low-frequency sampling at the receiving site** (see Figure 4). In contrast with prior conventional wisdom, these findings suggest that **the uncertainty contributed by casting site characteristics affects criteria assessment outcomes for a high-frequency synthetic dataset less than does the uncertainty contributed by the low sampling frequency of the long-term fixed station dataset. These findings then suggest Spectral Casting can be recommended to address certain previously unassessed short term water quality criteria (e.g. 7-day DO) where high and low frequency data sets exist to conduct segment level assessments.**

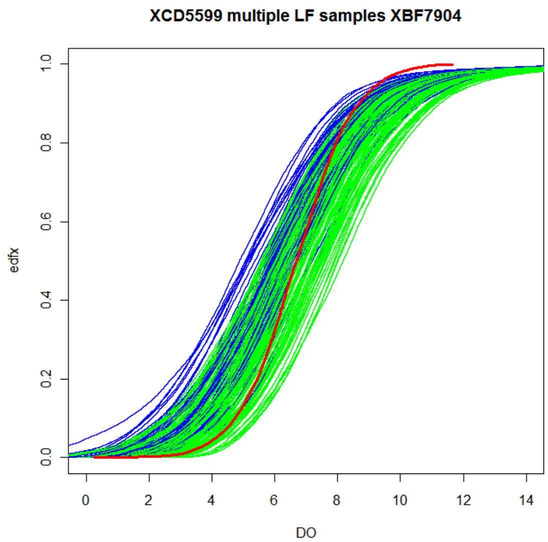
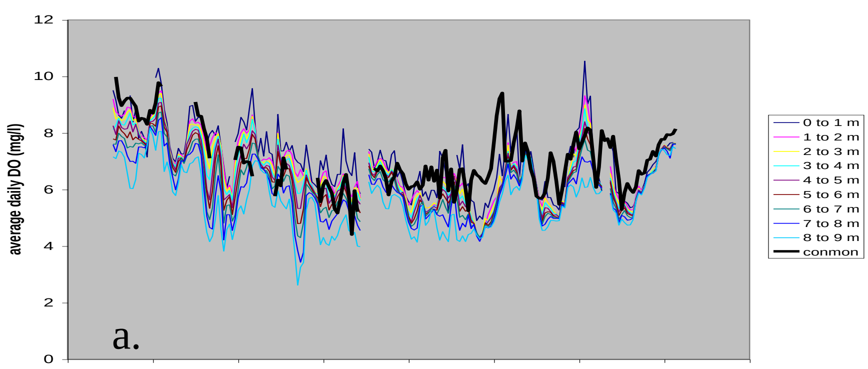


Figure 4. Variation due to multiple low-frequency samples from the receiving site with Fourier Series interpolation. The sending data set is held constant at one two-week interval. Blue curves synthetic data based on a series of night samples. Green curves are from a series of day samples. The red curve is the receiving site high frequency data

Spectral Casting Validation: Evaluating water column patterns of DO variability

Vertical water column patterns in short-term DO variability must be explored before vertical homogeneity of the water column can be assumed in the application of spectral casting to create synthetic data for a water column layer. CONMON stations have typically provided information about surface conditions only. If surface (i.e. above pycnocline) conditions are significantly less or more variable than bottom (i.e. below pycnocline) conditions, then using the short-term high frequency variability signal generated by nearshore CONMON stations may result in an inaccurate assessment of DO.

In comparisons of variability in deep and shallow waters, preliminary analyses of data from a vertical profiler deployed in the York River showed that on a weekly basis, DO varies similarly throughout the water column (T. Robertson, VADEQ. See Figure 5). However, Perry found that 24-hour periodicity explains a much greater proportion of the diel variability in surface water than in deep water layers. This suggests that one must be careful when extrapolating nearshore variability to offshore, deep water locations.



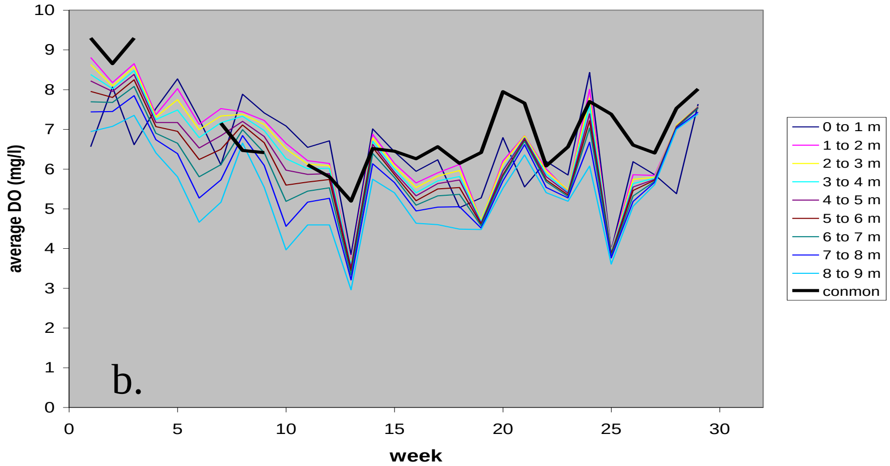


Figure 5. Comparison of DO from a Vertical Profiler and a continuous monitoring station (conmon) in YRKPH. A) DO represented by daily averages. The variance and trend are significantly different between continuous monitor and Profiler time-series. B) DO represented by weekly averages. The variance and trend are statistically similar between continuous monitor and Profiler time-series. The Fligner-Killeen test was used to test equal variance, while a test of parallelism was done using *sm* ancova.

In another test of the utility of shallow water continuous monitoring data for spectral analysis with mid-channel deep water data, variability from continuous monitors at St. George's Creek and in the Yeocomico River were cast upon data from a vertical profiler deployed in the lower Potomac River in 2009. For the analysis, the profiler dataset was sampled monthly at each depth, yielding a single dissolved oxygen concentration at each depth for each month. This approximated an open water fixed station dataset. Next, spectral analysis was used to create synthetic datasets using the monthly sampled vertical profiler data as a "receiving" site with data from nearby continuous monitors at St. George's Creek and in the Yeocomico River as "sending" sites. The synthetic datasets were tested for the frequency of each criterion (30-day mean, 7-day mean, instantaneous minimum) failing at each depth. The resulting frequencies were then compared to the failure frequencies for the continuous vertical profiler data at each depth. The synthetic data for the continuous monitor/modeled vertical profiler station did not violate the 30-day mean open water or deep water criteria. However, the real data from the vertical profiler did not violate the 30-day mean criteria either. Comparison of the 7-day (open water) and 1 day (deep water) mean criteria produced similarly inconclusive results. These results were similar regardless of depth. Under the instantaneous minimum criteria, many discrepancies between the modeled and real data occurred. However, outside of depths close to the pycnocline (7 and 9 meters), the differences in failure percentage were not large. This might be expected in modeled data where extreme highs and lows might not influence individual data points (i.e., instantaneous minima) as much as they do means. A greater number of nearshore-offshore comparisons are needed to more thoroughly examine applications of spectral casting from nearshore to layers of the water column (above pycnocline, below pycnocline) offshore.

2.2.2 A Parametric Simulation Approach to Assessing the Umbrella Criterion Concept for the Instantaneous Minimum Criterion.

High frequency samples of DO at fixed locations show that there is considerable serial dependence or autocorrelation in these DO time series. This lack of independence makes it difficult to analytically compute the probability that an instantaneous criteria will be violated when an umbrella criterion (e.g. weekly or monthly mean) is satisfied. Perry (Appendix X) develops and shows results from a simulation approach to addressing this question.

The basic approach of the simulation is to generate time series that have properties similar to observed DO time series. The data used for this exercise are the open water buoy data compiled by Olson. In these data, time series that are more than 1 week in length were parsed into 1 week time series. The time series that run less than 1 week are typically the 3-day data sets collected under the EMAP program. A simple AR(2) model that included structural terms for the mean, linear trend, and diel cycle was fitted to each of these time series using Proc AutoReg in SAS.

Output of the AR(2) model effort was analyzed with a MANOVA model including terms for Month, Total Water Depth, Sensor Depth, Latitude and Longitude. Some results from this overall model are presented.

For the simulation it seemed appropriate to focus on just one assessment unit. Thus only data from CB4 in the surface layer (sensor < 10 m depth) were used because CB4 is one of the best represented segments in these data.

2.2.3 Catalog of Analyses Approaches used for the Umbrella Criteria Assessment Assumptions.

Due to the complexities involved in spectral casting, significant testing was conducted to establish support for its suitability for short term dissolved oxygen criteria assessment. Additional analyses were typically less complex, more straightforward but no less insightful. The full complement of analyses used in the umbrella criteria validation assessment is summarized in Table 4a, 4b, 4c and 4d. The details of analyses and expanded findings are referenced in the Appendices at the end of this report.

Table 4a. Catalog of analyses conducted evaluating the Umbrella Criteria assumption. Nearshore, Shallow-water only.

Nearshore, Shallow-water only			
Author	**Analysis**	**Data**	**Appendix**
Buchanan	**Violation rate plots** Failure rates of the 7-day mean and instantaneous minimum DO criteria for Open Water computed from shallow water CONMON data; 7-day mean failure rates compared to the 30-day mean, and instantaneous minimum failure rates compared to the 7-day and 30-day means; probability of failing the instantaneous minimum and the 7-day mean criteria determined as a function of daily & weekly mean DO and the diel (daily) magnitude of change in DO.	Potomac River 2004-08 CONMON sites combined across tidal fresh to mesohaline salinity zones. Potomac-specific results.	Appendix 1
Perry	**Risk analysis** Combined point data from continuous monitoring time series. The probability analysis provided a risk assessment given the statistical properties of residuals for short term criteria means against the monthly mean criterion.	CONMON – generalized results across Chesapeake Bay.	Appendix 2
Boynton et al.	**Violation tables and plots** Frequencies of failure for criteria evaluated across the range of open water designated use summer criteria. Violation tables and plots were produced.	CONMON, non-Potomac data across salinity zones, nutrient enrichment gradient and duration (years) of monitoring (1-9 yrs per site)	Appendix 3, 4

Table 4b. Catalog of analyses conducted evaluating the Umbrella Criteria assumption. Open Water, Offshore data only assessments.

Open Water (Offshore only)			
Author	**Analysis**	**Data**	**Appendix**
Brush et al.	**Time series of site specific vertical profiler assessments** Vertical profiler time series data in Virginia tributaries was expressed as rolling averages plotted against short term DO criteria.	One vertical profiler site, mesohaline location on the York River. 2007-08.	Appendix 5
Bilkovic et al.	**Time series for vertical profiler assessments** 1. Criteria violation means versus criteria failure frequency analyses. 2. Regression analyses.	Two vertical profiler sites, one mesohaline location on the York River. 2007-09, one mesohaline location in the Rappahannock River 2009.	Appendix 6
Olson in USEPA 2004	**Generalized Baywide analysis** Bi-plots assessing violations	147 short term bouy deployments data sets, e.g. EMAP, etc.	Appendix 7
Hall	**Multiple analyses approaches, Site specific evaluations** 1. Frequencies of failure for open water summer criteria using synthetic data modeled from two mid-channel stations with nearby continuous monitors in the lower Potomac River. Violation rate plots were created for assessment. 2. **Spectral casting in time only, partial summer season. Verification of casting results.** Frequencies of failure for open water summer criteria comparing synthetic data modeled from mid-channel vertical profiler sampled monthly (to simulate monthly sampling) and nearby continuous monitors with real vertical profiler data.	1. Synthetic data time series from two locations in mesohaline Potomac River 2006-2008. Piney Point and Popes Creek. **2.** Synthetic data created for a vertical profiler site using two continuous monitors in mesohaline Potomac River, Yeocomico River and St. Georges Creek CONMON, Potomac River, June 22-Aug 16th, 2009.	Appendix 8
Perry	**Generalized Risk Assessment** The probability analysis provided a risk assessment for protectiveness of the instantaneous minimum DO criterion using the 30 day mean	Segment CB4 analysis only. Combined point data from EMAP short term continuous monitoring buoy deployment time series based on Olson in USEPA 2004.	Appendix 12

Table 4c. Catalog of analyses conducted evaluating the Umbrella Criteria assumption.
Open water – USEPA regulatory definition.

Open Water (USEPA regulatory definition: Offshore + Shallow water, i.e. shoreline to shoreline)			
Author	**Analysis**	**Data**	**Appendix**
Robertson and Lane	**Spectral casting with full 3-year CFD regulatory assessment of criteria** Recreated the USEPA published CFD assessment approach for Chesapeake Bay Open Water criteria assessment on synthetic data interpolated in time and space. The "Phase 1" spectral analysis approach (Neerchal 1994) was used to combine long term CBP water quality monitoring summer dissolved oxygen data collected from mid-channel, fixed station network with the temporally-intensive data gathered at nearshore continuous monitoring stations. Mid-channel daily or weekly averages are then interpolated spatially then assessed using the CFD approach described above.	12 CBP Bay management segments, VA-only waters. Summer, three year regulatory evaluations of either 2006-08 or 2007-09 for the: James River Tidal Fresh (TF1, TF2) Oligohaline (OH), Mesohaline (MH), Polyhaline (PH) Rappahannock River TF, OH York River MH Mobjack Bay PH Appomattox TF Chickahominy OH Corrotoman PH	Appendix 9

Table 4d. Catalog of analyses conducted evaluating the Umbrella Criteria assumption. Deep Water Criteria comparisons assessments.

Deep Water			
Author	**Analysis**	**Data**	**Appendix**
Brush et al.	**Time series from a single location against criteria**. Deep water assessment of time series of data against deep water criteria.	Two months from one bottom-mounted sonde at the vertical profiler site, mesohaline location on the York River.	Appendix 5
Olson in USEPA 2004	**Generalized Baywide analysis** Bi-plots assessing criteria violations	Evaluated from the 147 short term bouy deployments data sets, e.g. EMAP, etc.	Appendix 7
Bilkovic et al.	**Criteria violation means versus criteria failure frequency analyses**	Vertical profiler, 1 location on the York River. 2007-09 (3 yrs), 1 location in the Rappahannock River, 2009.	Appendix 6
Hall	**Spectral casting, partial summer time series, verification against vertical profiler data.** Frequencies of failure for deep water summer criteria comparing synthetic data modeled from mid-channel vertical profiler sampled monthly (to simulate monthly sampling) and nearby continuous monitors with real real vertical profiler data. **Not a direct test of umbrella criteria**; inference only.	Synthetic data from vertical profiler and two continuous monitors in mesohaline Potomac River 2009.	Appendix 8

2.3 Results

2.3.1 Single site, Limited Comparison of Chesapeake Bay Program Water Quality Sediment Transport Model Output to Select Depth Layers of High Frequency Vertical Profile Dissolved Oxygen Concentration Data.

The Umbrella Criteria assumption was derived from an analysis of hourly output generated from the calibration run of the EPA CBPO's Water Quality Sediment Transport Model (WQSTM). An assessment of the direct model output indicated that in the Open Water and Deep Water designated uses, violations of the 30-day mean criterion exceeded violations of all shorter-duration criteria. Thus the 30-day mean criteria in OW and DW could be presumed to be the most conservative, or protective of these designated uses.

This premise rests on a key assumption: that "the temporal variability of dissolved oxygen in the Bay is reasonably well-characterized by the Bay model" (USEPA 2010).

In order to evaluate the ability of the WQSTM to capture the range of short-term variability that has been observed in recent high-frequency data collections, we compared output from the WQSTM to data from a deployment of the MD DNR vertical profiler in a mid-channel location of the lower (mesohaline) Potomac River.

Comparisons are based on the following:

- We matched the hourly model output for the years 1991-2000 to the same location and days-of-year as the 2009 vertical profiler deployment.
- We binned profiler observations by depth into groups matching vertical cells from the WQSTM
- Compared empirical frequency distributions (EDFs) of observations with WQSTM simulations for the same range of days, for each year of the WQSTM calibration run.

Hydrologically, 2009 annual flows to Chesapeake Bay were below average (http://md.water.usgs.gov/waterdata/chesinflow/). Dissolved oxygen dynamics have some relationship to flows that deliver loads to the Bay where typically wetter years produce more hypoxia than drier years. While it is good to look across all years in the calibration for comparisons, because we are comparing water quality in a year that is not part of the calibration period, we might at least pay particular attention to water quality behavior in the model and monitoring data in years with similar flows. In 1993 and 1996 annual flows were below average and most similar to 2009.

In general, we found that across a range of years and hydrological conditions, the seasonal level patterns of trend present in the monitoring data are reflected in the model output (see Figure 4). Higher average levels of dissolved oxygen tend to occur earlier in the year and declined into the summer.

For the month of August, comparisons within individual years from 1991 to 2000, depending on depth, the variability in model's output tended to be lower than the observed variability of measured concentrations (see Figure 5). If we consider hydrologically similar years, model output for August 1993 showed much lower variability than August 2009 profiler data in both the surface and deep water layers while August 1996 data were more comparable – model and observed data show high variability in the surface compared with other years and were the most similar for measures of central tendency and range in the deep water.

In upcoming years the WQSTM model calibration period will be extended and overlap in time with many more years and stations that collected high frequency continuous monitoring data across the Bay. Further analyses should compare WQSTM output with all other offshore vertical profiler data available. There are a few nearshore high-frequency shallow-water monitoring data sets with continuous monitoring data that coincide with the present calibration period (1999 and 2000) on the lower Eastern Shore of MD (Tangier Sound tributaries such as the lower Pocomoke River). Nearshore, shallow water regions are likely to be less well-characterized in the model than mid-channel data and lower Eastern Shore of Maryland waters tend to be characterized by blackwater conditions that will challenge the model even further. This does, however, provide further opportunity to make direct modeling-monitoring comparisons and gain insight into model and ecosystem behavior.

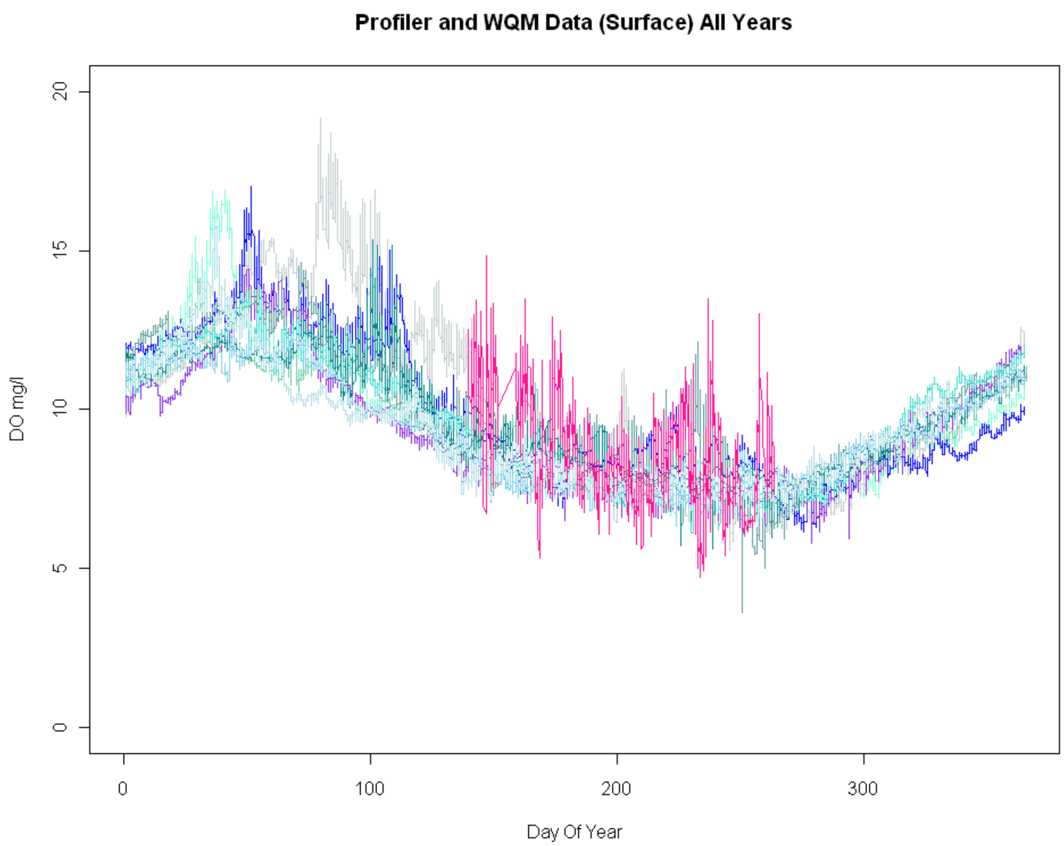
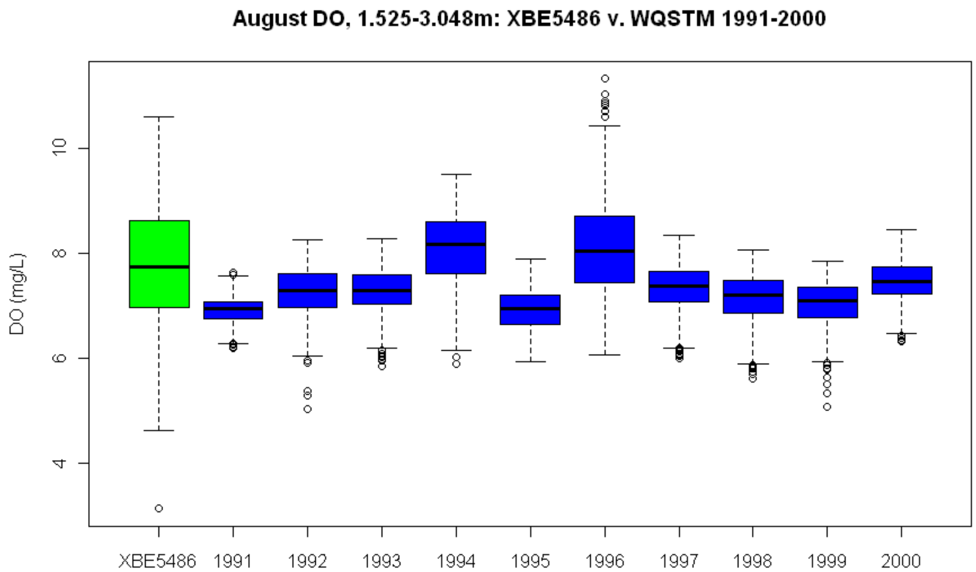


Figure 6: Time series of 2009 May-November deployment of the POTMH profiler (pink) compared to all 10 yrs of WQSTM output combined (various other colors). Surface cell only

(a) 1.525-3.048 meters depth



(b) 7.63 – 9.144 meters depth

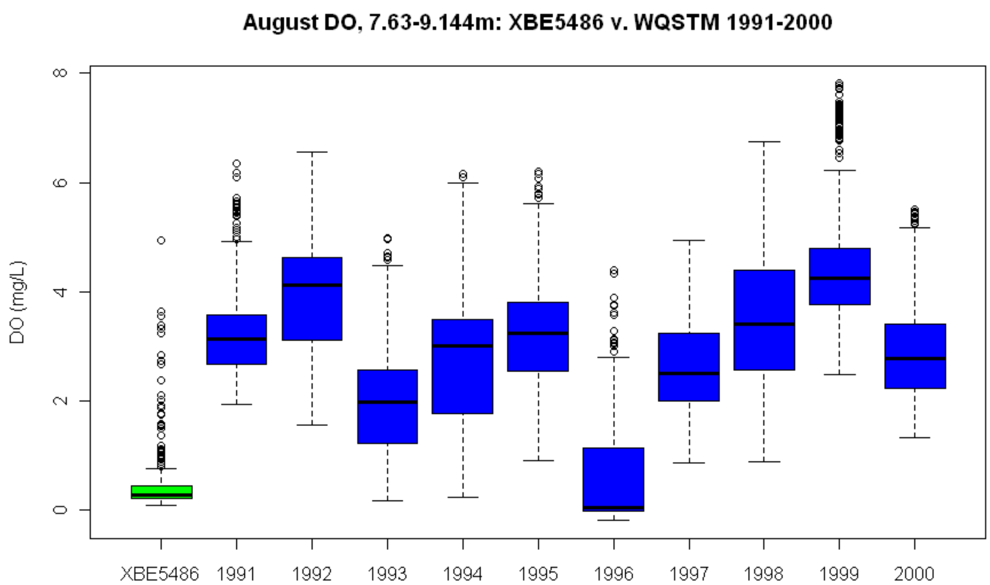


Figure 7: WQSTM simulated DO concentrations for August of individual years, 1991-2000 (blue boxes) compared with observed DO concentrations for the same approximate depth and location from August, 2009 (green box).

2.3.2 Open Water: Protectiveness of the 7-day Dissolved Oxygen Mean Criterion by the 30-day Mean Dissolved Oxygen Criterion.

The Open Water 30-day mean DO criterion of 5.0 mg/L (5.5 mg/L when salinity ≤ 0.5 ppt) was generally found to be consistent and supportive (see Box 3) **protection for the 7-day mean criterion of 4.0 mg/L** in analyses conducted by Buchanan (Appendix 1), Perry (Appendix 2 & 12), Boynton et al. (Appendix 3, 4), Brush et al. (Appendix 5), Bilkovic et al. (Appendix 6), USEPA 2004 (Appendix 7), and Hall (Appendix 8), (Table 5a, 5b, 5c). For point sample locations, Boynton et al. and Buchanan reported occasional violations of the 7-day mean under conditions attaining the 30-day mean, however, the violations were not excessive (< 10%) and could be deemed 'allowable exceedances' as described in EPA regulatory assessments. Because Robertson and Lane used the full regulatory CFD assessment procedure, their results took the regulatory "allowable exceedance" into account (per USEPA 2003). **The value of this body of support is bolstered by the fact that assessments were made across mainstem and tributary locations that captured tidal fresh-to-polyhaline habitats, shallow and offshore waters, spanned diverse nutrient conditions, and included inter-annual variability via multi-year evaluations for the summer season.**

The conditional probability analysis (Perry, Appendix 2, 12) suggested that the risk of violating protectiveness for the 7-day mean may increase beyond 10% when data are consistently near the criteria level. Because the 30-day mean rarely exhibits this behavior in the field data, it seems safe to conclude that in most cases the 30-day criterion acts as an umbrella for the 7-day criterion. However, under the conditional probability analysis, the umbrella does not seem to be a broad one. Slight increases in the variation of the weekly mean about the monthly mean without corresponding increases in the monthly mean could start to increase the violation rate for the 7-day criterion to above 10 percent.

Box 3. Declarations of Umbrella Criterion Protectiveness.

Definitions related to descriptions of analyses finding concordance or discordance with the umbrella criterion protectiveness assumptions.

Umbrella Criterion: the most protective criterion. When compliance with one criterion insures compliance at others, the one (i.e. most protective) is termed "the umbrella" criterion.

Consistent with Assumptions – Providing weight of evidence with the Umbrella Criteria assumptions. Site or watershed specific analyses with any or all of the following: habitat, season or length of time series limitations relative to the regulatory 3-year assessments for listing and delisting decisions. An example of 'consistency' includes comparisons of 30-day mean time series for a month passing the open water criteria coincident with the 7 day mean always passing the criteria. The Umbrella Criteria assumption is not violated but there were no violations of either criteria and therefore limited information to base full support or lack of support for the assumptions from the data set.

Support for Assumptions – The strongest analyses. Generalized analyses with Baywide applications or results implementing methods such as full 3-year CFD assessments, Conditional Probability analyses or USEPA 2004 bi-plot evaluations.

Table 5a. Habitat-specific declarations on the protectiveness of the 30-day mean as the Umbrella Criteria: Shallow-water.

Shallow water, nearshore habitat only		
Author/Appendix	**Is the 30-day mean protective of…**	
	7-day mean?	**Instantaneous minimum**
Buchanan Appendix 1	**Consistent** with Umbrella Criteria assumptions.	**Inconsistent** with Umbrella Criteria assumptions. Little protection in time series. Violations of instantaneous minimum criteria per month range from 0% - 60.4% and 65.4% of the months sampled had instantaneous minimum criterion violations of some degree.
Perry Appendix 2,	**Support** Umbrella Criteria assumptions. Small risk of violation if dissolved oxygen levels persist near the criteria threshold	**Not assessed**
Boynton et al. Appendix 3, 4	**Tentatively Consistent** with Umbrella Criteria assumptions**.** Occasional criterion failures observed but otherwise consistent with Umbrella Criteria assumptions	**Tentatively Inconsistent** with Umbrella Criteria assumptions. Common failures observed. Generally fewer failures at better quality sites. High interannual variability.

Table 5b. Habitat-specific declarations on the protectiveness of the 30-day mean as the Umbrella Criteria. Above pycnocline open water not including shallow-water.

Open Water (Offshore habitat only)		
Author/Appendix	**Is the 30-day mean protective of…**	
	7-day mean?	**Instantaneous minimum?**
Brush et al. Appendix 5	**Consistent** with Umbrella Criteria assumptions	**Consistent** with Umbrella Criteria assumptions
Bilkovic et al. Appendix 6	**Consistent**. No violations observed	**Consistent** York River Profiler site assessment. **Consistent** Rappahannock River profiler site assessment.
USEPA 2004 Appendix 7	**Support**	**Does not support** Umbrella Criteria assumptions. The protection rate was 85% of cases tested (n=94)
Hall Appendix 8	**1. Consistent** with Umbrella Criteria assumptions. 2. **Consistent** with Umbrella Criteria assumptions.	1. **Inconsistent** with Umbrella Criteria assumptions. 2. **Inconsistent** with Umbrella Criteria assumptions.
Perry Appendix 12	**Support** Umbrella Criteria assumptions. Small risk of violation if dissolved oxygen levels persist near the criteria threshold	**Does not support** Umbrella Criteria assumptions for Bay management segment CB4.

Table 5c. Designated use specific declarations on the protectiveness of the 30-day mean as the Umbrella Criteria: Regulatory assessments of Open Water.

Open Water Designated Use (Regulatory definition, i.e. shoreline to shoreline)		
Author/Appendix	**Is the 30-day mean protective of…**	
	7-day mean?	**Instantaneous minimum?**
Robertson and Lane Appendix 9	**Support** Umbrella Criteria assumption. Violations were detected, not excessive.	Not assessed

Table 5d. Habitat-specific declarations on the protectiveness of the 30-day mean as the Umbrella Criteria: Deep water.

Deep Water habitat only		
Author/Appendix	**Is the 30-day mean protective of…**	
	1-day mean?	**Instantaneous minimum?**
Brush et al. Appendix 5	**Inconsistent** with Umbrella Criteria assumptions suggested.	**Inconsistent** with Umbrella Criteria assumptions suggested.
Bilkovic et al. Appendix 6	**Consistent** York and Rappahannock River profiler data sets	**Consistent** York River Profiler site assessment. **Inconsistent** Rappahannock River profiler site assessment.
USEPA 2004 Appendix 7	Not assessed	**Inconsistent** 57% of cases tested showed 30-day mean protective on the 1st-%ile as estimate for instantaneous minimum (n=26)
Hall Appendix 8	**Consistent** with Umbrella Criteria assumptions	Not assessed

2.3.3 Open Water: Protectiveness of the Instantaneous Minimum Criterion by the 30-day Mean Dissolved Oxygen Criterion.

The weight of evidence across analyses tends to be inconsistent with the Umbrella Criterion assumption for the 30-day mean protection against the instantaneous minimum criterion in open water (Table 5a, 5b, 5c). Conditional probability analyses (Perry, Appendix 12) conducted for offshore waters of mesohaline segment CB4 further suggest that even providing a significant increase in 30-day mean DO levels above the 30-day mean criterion may not provide sufficient protectiveness of the instantaneous minimum criterion. Evaluations involving individual monitoring sites did however vary in water quality behavior from no violations to considerable violation of the Umbrella Criteria concept within months.

Buchanan (Appendix 1) focused on shallow water CONMON sites in the Potomac River and found little protection of the instantaneous minimum criteria by the 30-day mean DO criterion. Over the course of 211 summer monthly means evaluated, Buchanan found that the 30-day mean was completely protective of the instantaneous minimum criterion only about 35% of the time. Boynton et al. (Appendices 3,4) evaluated shallow-water CONMON data and reported common exceedances of instantaneous minimum criterion in shallow water habitat that showed acceptable 30-day

DO means; there were, however, generally fewer failures at better quality sites. Bilkovic et al. (Appendix 6) looked at vertical profiler data for two locations and found low violation rates of the instantaneous minimum (0.04-0.4%) in the York River over 3 summer seasons when observed 30-day means were >5.5 mg O_2/L; in 2009, 0 to 3.3% violations of the instantaneous minimum per month were observed in the Rappahannock River when means were >5.8 mg O_2/L. However, without data to assess violation rates when means are close to or below the 30-day DO mean criterion we cannot yet conclude the profiler data is fully consistent with the assumption of protectiveness. Hall's results, using a spectrally-cast synthetic dataset for the lower Potomac River, were similarly consistent with Umbrella Criteria assumptions; again, few violations of the criteria occurred for the available time series assessed (Appendix 8).

Historical analyses are further consistent with these recent findings. As described in Section 1 of this report, USEPA (2004) analysis results suggested protectiveness of the instantaneous minimum (measured then as the 1%-ile of the data) approximately 85% of the time in offshore Open Water habitats. Thus even considering the results in the regulatory context of 'allowable exceedances,' the Perry, USEPA, Buchanan and Boyton et al. findings provide a weight of evidence that continues to tend towards insufficient protectiveness of the instantaneous minimum criterion by the 30-day mean criterion in OW.

2.3.4 Deep Water: Protectiveness of the 1-day mean criterion by the 30-day mean criterion.

Bilkovic et al. (Appendix 6) and Hall (Appendix 8) suggest general consistency with umbrella criteria assumptions for protectiveness of the 1-day mean criterion by the 30-day mean DO criterion (Table 5d). Results were based on one site each for vertical profiler stations in mesohaline habitats on the York and Rappahannock Rivers (Bilkovic et al.) and on the lower Potomac River (Hall). Brush et al (Appendix 5) had earlier suggested the 30 day mean criterion may be inconsistent with protecting the 1-day mean at the York River vertical profiler site based on 2 months of deep water time series from a bottom mounted sonde for the summer of 2009. Further work is recommended to better establish the level of support for the umbrella criteria assumption in deep water regarding the 1-day mean across salinity zones, geography and eutrophication gradient.

2.3.5 Deep Water: Protectiveness of the Instantaneous Minimum Criterion by the 30-day Mean Dissolved Oxygen Criterion.

Tests of the protectiveness of the summer season Deep Water 30-day mean criterion (3.0 mg/L) continue to suggest unlikely support for being protective of the Deep Water instantaneous minimum criterion (1.7 mg/L) (Table 5d). USEPA (2004) first suggested that the 30-day mean criterion was only protective of the instantaneous minimum in the DW designated use 57% of the time. The most recent analyses from three vertical profiler sites, the mesohaline York River (Brush et al. Appendix 5, Bilkovic et al. Appendix 6), the Rappahannock River (Bilkovic et al. Appendix 6) of and the mesohaline lower Potomac River (Hall Appendix 8), were 2:1 favoring results inconsistent with the Umbrella Criteria protection assumption for the 30-day mean protecting the instantaneous minimum. Further work is again recommended here to better establish the level of support for the umbrella criteria assumption across salinity zones, geography and eutrophication gradient.

2.4 Discussion and Recommendations

The analyses described in the attached appendices, and results summarized above and in Tables 5a, 5b, 5c and 5d, were presented to a broader group of researchers and analysts at a workshop held on March 15-16, 2011 in Annapolis, MD.

With regard to the umbrella criteria assumption and the 4 questions posed in the proposal:

1. Under what conditions does the "Umbrella Criteria" assumption appear to be accurate?
2. Under what conditions does this assumption appear to be violated?
3. For what conditions do currently available data not allow us to test this assumption?
4. What are the data needed to test this assumption for all conditions?

The group came to the following conclusions:

2.4.1 Conditions for which the Umbrella Criteria Assumption appears to hold

The Open Water 30-day mean criterion is generally protective of the Open Water 7-day mean criterion in both mid-channel and nearshore shallow water regions, across salinity zones and across years with varying flows and nutrient conditions. According to Dr. Perry's conditional probability analysis, the risk of violating the OW 7-day mean criterion while attaining the 30-day mean criterion is just over 10%. Thus "the umbrella is not a broad one." Increasing the 30-day mean to 5.3 reduces the risk of violating the 7-day mean to 6-9%, depending on the standard deviation of the dataset.

There is some preliminary evidence that the Deep Water 30-day mean may be protective of the Deep Water 1-day mean criterion; further analyses should be conducted to further confirm these results, and to define the conditions under which this result this holds true.

2.4.2 Conditions for which the Umbrella Criteria Assumption may be violated

1. Measurable violations of criteria can be found even at sites with good water quality. Understanding DO variability relative to a criterion value is key.
2. Variation in dissolved oxygen measures around the mean criterion affects the risk of meeting or failing attainment. When dissolved oxygen variability increased the risk of failing a criterion increased. Perry (Appendix 2, 12) pointed out this issue in the context risk of criterion failure with data hovering around the criterion and concerns that if DO variability increased slightly there would be considerable risk in further violating the assumptions of the 30-day mean being protective of the 7-

day mean in Open water. Buchanan (Appendix 1) and Boynton et al. (Appendix 3,4) suggest similar issues concerning DO variability.
    a. A corollary to this observation is that criterion values should be matched with declarations of acceptable variance in the measure.

3. Analyses have also shown that shallow waters may be more likely than open waters to exhibit large 24-hour (diel) fluctuations in DO concentrations (e.g., Buchanan, Appendix 1, Robertson and Lane Appendix 9. See Section 3.1). The availability of shallow water CONMON data sets has allowed for this previously unavailable assessment. Because the percentage of a Chesapeake Bay management segment in shallow water habitat varies from 0 to nearly 100% with roughly 50% of segments having 50% or more of their habitat in shallow water, segments with large shallow water acreages may have additional challenges to meeting their criteria than those segments dominated by offshore, open water volumes.
4. Regions with strong spring-neap tide effects were raised as areas of concern relative to meeting umbrella criteria assumptions (Brush et al. Appendix 5 results).
5. Boynton et al (Appendix 4) showed that inter-annual variation in violations of DO criteria may be related to inter-annual variation in flow. Wind and temperature effects on shallow water DO behavior were further noted. Such findings pointed toward the concerns regarding further accounting for the effects of climate change on the relationship among Chesapeake Bay criteria and meeting the umbrella criteria assumptions.

2.4.3 For what conditions do currently available data not allow us to test this assumption?

1. The consensus of the workshop participants was that given sufficient time and analytical resources, we can provide more thorough answers to questions 1 and 2 with currently available data. To the extent that uncertainty remains, it is a function of both the amount of data available and the ongoing developmental nature of the analytical tools at hand. How *well* we answer these questions, and in particular the potential for decision error in our listing assessments, is affected by the amount of data collected.
2. Robertson and Lane (appendix 9) indicated that nearshore-offshore variability was similar at the scale of a week but breaks down when looking at the scale of a day. Combining shallow water and offshore habitats appear supported at the monthly and weekly scale but in the assessments of the shortest time scale criteria (1-day, instantaneous) may be inappropriate for an umbrella criteria assessment and the habitats needing to be separated.
3. Modeling-monitoring comparisons with high frequency data were limited. The Chesapeake Bay model outputs for the TMDL were focused on 1991-2000 while vertical profiler data and most shallow water CONMON is from more recent monitoring program efforts. There is some earlier work with

CONMONs (e.g. lower Pocomoke River 1997-2000 MD DNR) that can provide model-monitoring comparisons for the same time and place.

2.4.4 What are the data needed to test this assumption for all conditions?

1. It was agreed that the investment in shallow water CONMON has been tremendously valuable, and has greatly enhanced our understanding of shallow water habitats and the ability to conduct these analyses. Continued investment in shallow water high-frequency data across the tidal Bay habitats will improve the strength and accuracy of our analyses with regard to questions about inter-annual variability within local, nearshore locations.
2. Equally important is the information to be gained on relative differences in patterns of short-term variability across nearshore, shallow-water systems with different levels of impairment.
3. Our understanding of variability in mid-channel locations, and particularly in Deep Water regions, is hindered by the paucity of high-frequency vertical profiles available. Vertical profiler data were invaluable in providing verification of assessment for Spectral casting by Robertson and Lane (appendix 9) and Hall (appendix 8). Vertical profiler data in place of long term monitoring site water quality profiles reduces uncertainty of assessments in segments where it is available. Continuation of the ongoing effort to obtain high-frequency vertical profile data in these regions will reduce the uncertainty associated with analytical results. In 2011, deployment of the MD DNR profiler was planned for a region of the mainstem where hydrodynamic models predict high variability in physical parameters that contribute to variability in dissolved oxygen concentrations in the summer season (*personal communication*, Aaron Bever, VIMS). It is hoped that data from such deployments can simultaneously serve as a validation test for the estuarine models while providing key insights into the bounds of variability in the Deep Water regions of the Bay.

2.4.5 Recommendations and Next Steps Emerging from Workshop

Recommendations emerged both for concrete “next steps” aimed at achieving further development and application of a short-duration criteria assessment method, and for further research and conceptual development of the dissolved oxygen criteria themselves.

In order document those times and regions for which it is most important that short-duration criteria be assessed, and then to assess short-duration criteria, participants recommended the following activities:

1. Generate a single common dataset so that every analyst who participates in the collaboration is using the same version of the same data. Expand the dataset to incorporate the most recent data collected using vertical profilers, buoy- and bottom-mounted sensors, and shallow water CONMON stations.

2. The analyses in this report were conducted on a subset of the 92 Chesapeake Bay Program management segments. Using an updated version of the logistic regression method described in USEPA 2004 and/or the conditional probability analysis recently applied by Dr. Perry (Appendix 12), update the segment-by-segment quadrant analysis described in USEPA 2004. Include quantification of a 30-day mean umbrella *threshold*, as well as the probability of violating the 7-day mean, 1-day mean, and instantaneous minimum criteria in Open Water and Deep Water designated uses given the actual 30-day mean *criterion* (as opposed to the threshold) and the observed data distribution. The quadrant analysis can then be used to target regions of greatest concern for high-frequency data collection and short-duration criteria assessment.

3. With regard to communicating the results of the quadrant and conditional probability analyses to managers and decision-makers, quantify and clearly communicate the risk of erroneously classifying segment-DUs as impaired or un-impaired ("false positives" or "false negatives") given currently available data and analytical methods in use. In particular, the uncertainty of current calculations of the "30-day mean" (using only the long-term fixed station datasets) should be quantified and communicated. Workshop participants recommended that a group focus at first on those uncertainties that could be quantified and communicated most easily. Participants also suggested that a Venn diagram would be a useful visual aid for communicating the degree to which longer-duration assessments are protective of short-term durations (see Figure 8).

4. In order to properly assess the instantaneous minimum and to compare its violation rates to those of other criteria, the instantaneous minimum criterion must be defined more precisely than is currently the case. This issue becomes paramount when one begins working with high-frequency datasets, for which there is sometimes a new measurement every 5-15 minutes. While a consensus was not reached regarding the most appropriate application of the instantaneous minimum, two recommendations were discussed. First, there was extensive discussion regarding the importance of duration with regard to the instantaneous minimum. Participants agreed that further research should address the question of "allowable duration below the instantaneous minimum" for a single violation. Some participants proposed that the instantaneous minimum criterion should be assessed as the *daily minimum*, with the violation rate defined as the percentage of days when the daily minimum violated the instantaneous minimum. These two suggestions are not mutually exclusive. Indeed, adoption of the "daily minimum" proposal would still leave open the question of whether every measurement taken in an hours-long hypoxic event should be considered a violation, and how variation in the timing of high-frequency sampling affects the result.

5. Convene an expert panel to review the adequacy of the spectral casting method for assessing short-duration criteria. Through the Umbrella Criteria assessment effort we have already gained insight in uncertainties due to spectral casting

relative to information available from monitoring sites in the Chesapeake Bay long term water quality monitoring program used in water quality criteria attainment assessments (Section 2.2.1 this report). Emphasize that the question with respect to adequacy is a pragmatic one that recognizes the incremental nature of progress. For example, one could ask whether implementation of spectral casting for DO criteria assessment results in an analysis that is at least as accurate as the method currently in use.

6. Modify the CBP's criteria assessment programs as necessary to conduct a full regulatory assessment of high-frequency criteria for Open Water and Deep Water designated uses, using Spectral casting to produce the synthetic datasets for segments where sufficient data are available. If possible, target "marginal" segments where some degree of violation is probable, but where some attainment of the 30-day mean is likely. In order to conduct such an assessment, decision rules will have to be codified that translate the essential intent of the general prescriptions contained in the ambient water quality criteria addenda into a practically applicable method. These decision rules should be carefully documented and reviewed by the Criteria Assessment Protocols Workgroup and published in a new USEPA technical addendum to the ambient water quality criteria.

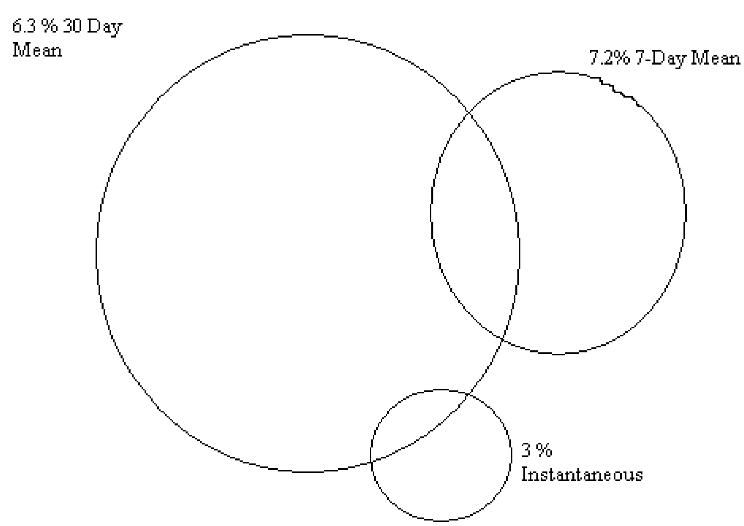


Figure 8: Hypothetical example of usage of a Venn diagram for communicating relative protectiveness of each criterion to managers.

Participants emphasized in general that the dissolved oxygen criteria would benefit from further research and conceptual development. Section 3 below will describe some of these more long-term, development recommendations in further detail. However, two of these recommendations can also be considered "next steps" that may relate directly to the action items described above. These are: (1) to collect more high-frequency vertical profile data in Deep Water regions, and (2) to further explore the concept of duration both as a component of the criteria and as a potential indicator of improving conditions. Specifically, researchers and analysts should explore whether *duration* of hypoxic events is a more sensitive indicator of restoration progress than are DO criteria violation rates.

SECTION 3: Related Questions and Topics of Importance Beyond Umbrella Criteria – Insights, Lessons Learned.

Participants in the umbrella criteria project and workshop repeatedly raised concerns regarding whether the existing interpretation and assessment of the ambient DO criteria adequately reflect the intent of the criteria with respect to protection of aquatic living resources. Concerns revolved primarily around three issues:

- Designated Use Boundaries: The potential effects of combining near-shore, shallow water volumes with offshore volumes for assessment of the Open Water designated use, rather than assessing shallow-water volumes separately;

- Criteria definitions and implications: The importance of duration with hypoxic events (now measurable with high frequency data), with regard to both the assessment of criteria and the question of how duration of hypoxic events varies with pollutant loads.

- Questions regarding the potential for patterns of ***variability*** in DO concentrations to change along eutrophication gradients

3.1 Combining Shallow-Water and Mid-Channel Assessments?

There was consensus among participants that when shallow, near-shore and deep, offshore waters are combined in a single volume-based DO assessment, the sheer volume of the offshore region may overwhelm signals of distress that occur in shallower waters. This is of particular concern because the biological importance of hypoxia in near-shore waters may be disproportionate to its relative volume. For example, there is a body of evidence showing that fish kills were associated not directly with seasonal deep water DO conditions, but rather frequently with episodic shallow water events (Figures 9, 10).

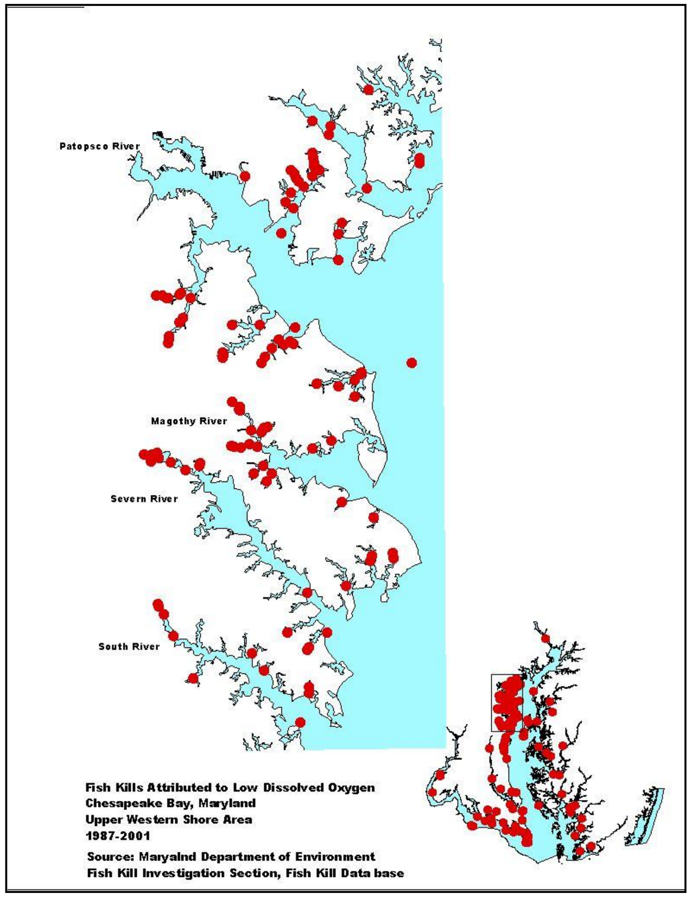


Figure 9. Distribution of fish kills attributed to low dissolved oxygen, Chesapeake Bay, MD. 1987-2001. Source: Maryland Department of Environment.

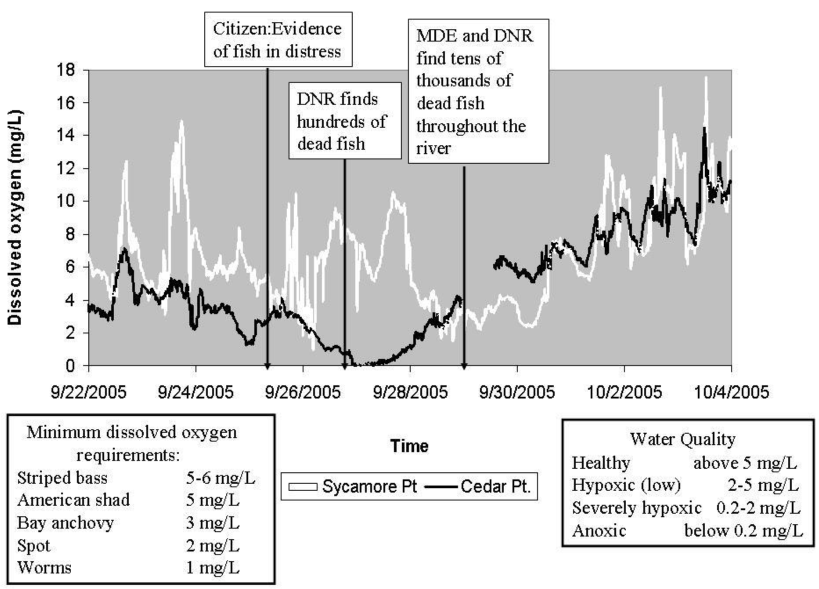


Figure 10. Corsica River 2005, Chronology of a fish kill from a highly eutrophic location in Maryland, Chesapeake Bay. Graphic from Mark Trice, MDDNR originally in P. Tango, 2005 Waterman's Gazette.

Analyses have also shown that shallow waters may be more likely than open waters to exhibit large 24-hour (diel) fluctuations in DO concentrations (Buchanan, Appendix 1). The presence of abundant and diverse primary producer communities (i.e. phytoplankton, SAV, benthic algae, macroalgae, epiphytic algal communities) can rapidly generate large amounts of oxygen during daylight hours. Conversely, the close proximity of oxygen-consuming benthic organisms and sediment processes can rapidly drive down oxygen levels at night.

Robertson and Lane (Appendix 9) have further documented differences in the characteristics of DO variability between shallow and mid-channel sampling locations. In a spectral analysis of data from the lower York River, they found that **daily averages of DO values from a mid-channel YSI Vertical Profiler were statistically different, in terms of both variability and overall trend, from daily averages at a shallow-water continuous monitoring station approximately 3 km away**. However, when weekly averages were compared, the variability of the two datasets was statistically similar.

For the reasons described above, and in more detail in the appendices, **workshop participants suggested that the partitioning of shallow, near-shore waters into their own assessment units (i.e. sub-segments) may more adequately represent the impact of DO criteria violation in these biologically active regions.**

3.2 Duration of Hypoxic Events

A central finding to emerge from the analyses of high frequency observations in shallow water datasets was that **there is yet another temporal scale of hypoxia in the Bay in addition to the seasonal scale hypoxia chronic to the deeper portions of the Bay. CONMON data exhibit diel scale hypoxia** (Figure 11). Since criteria levels are thresholds of significance to aquatic biota, analysis of *duration below criteria* is a potentially valuable measure of habitat suitability.

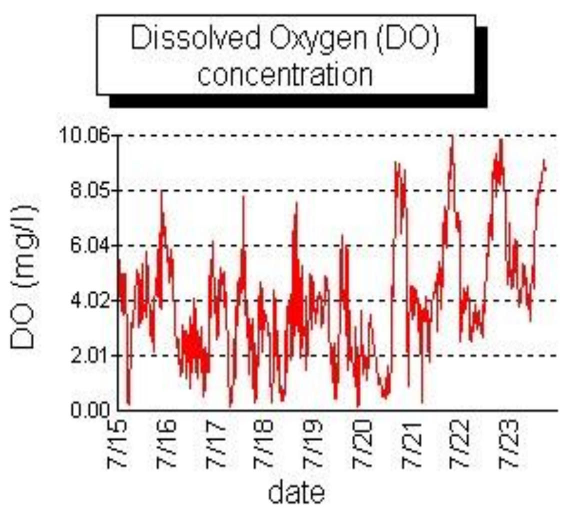


Figure 11. Ben Oaks, Severn River, MD example of diel hypoxia in shallow water. Data collected every 15 minutes. Graphic attributed to Maryland Department of Natural Resources.

At a 2011 Tidal Monitoring and Analysis Workgroup meeting, Matt Hall (MD DNR) presented analyses demonstrating that for a summer season using CONMON data, the maximum duration of DO criteria violation varied across shallow water in the Potomac River. Maximum duration below most criteria generally followed a gradient of putative eutrophic condition. In other words, the most eutrophic stations had the longest maximum durations below the DO criteria and vice versa. Boyton et al. (Appendix 3) further support this tendency across non-Potomac stations. Maximum duration of violation was later shown to be linearly and positively related to percent violation of a given DO criterion (P. Tango, Pers. Comm.). Figure 12 below illustrates this relationship using the Open Water 30-day mean; a similar relationship was found for the OW instantaneous DO criterion. The EPA-supported CFD water quality criteria assessment methodology does not explicitly address low DO event duration. However, if total violations are correlated to maximum duration of violation, then the CFD approach implicitly captures measures of event duration. We can therefore postulte that habitat stress is not only decreased in the seasonal condition, but in sub-season, short duration measures of habitat condition as well.

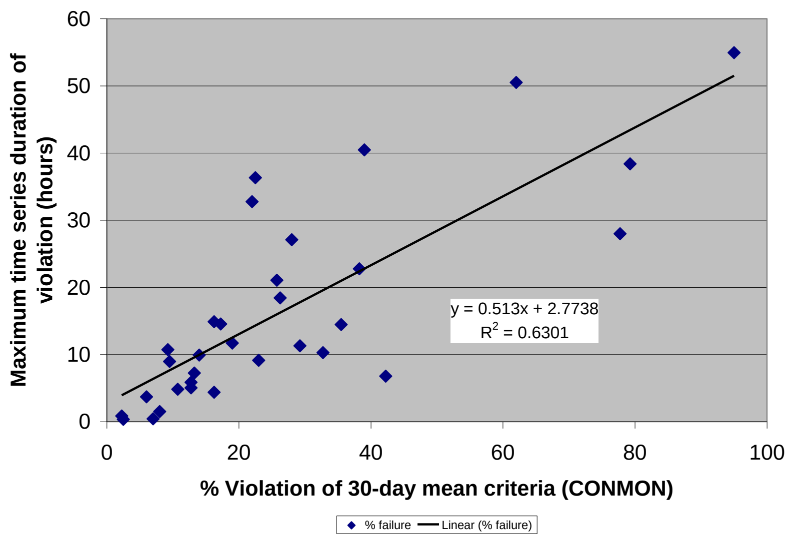


Figure 12. Maryland ConMon assessment of site specific % ConMon criteria violations against the maximum period of duration of criteria violation (hours) in a 30-day period. Data: Matt Hall MD DNR, Graphic: Peter Tango USGS@CBPO.

Results of the EPA CFD analyses are reported as pass-fail for CWA 303d listing purposes. However, the maximum time interval measured below the criterion value (in hours) shows promise as an index to the 30 day mean violation rate measured at a point. A more complete development of this index could be conducted comparing segment level impairment results that are output during the CFD assessment on % failure and multiple CONMON sites within segments.

3.3 Eutrophication Gradients and DO Variability

Suggestions for further research included better defining patterns and drivers of diel-cycling hypoxia, and the utility of shallow-water DO concentration variability as a signal of eutrophication. Wind driven seiching introducing anoxic waters into shallow water is known from historical reference (de Jonge et al. 1994) and shown further in recent data examples (Figures 13). Such climate forcing affects DO dynamics and assessments of Bay health.

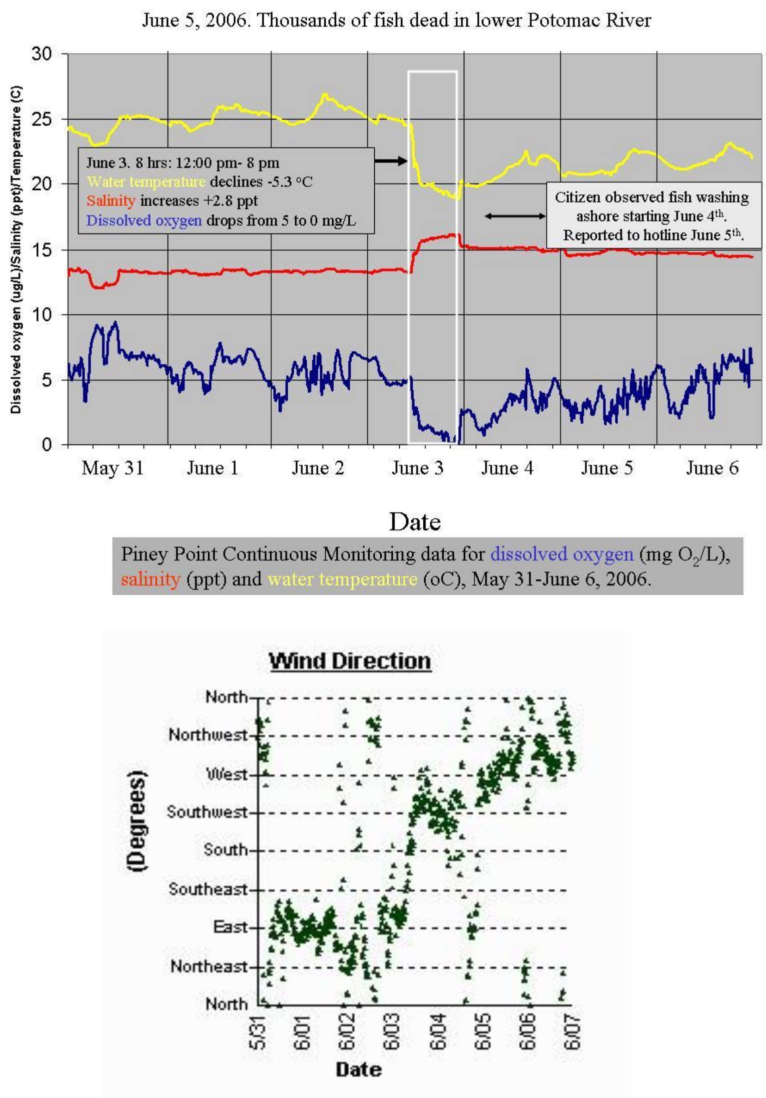


Figure 13. Top panel: Lower Potomac River Piney Point ConMon data (MD DNR) from May 31-June 6, 2006 shows intrusion of anoxic waters from the Bay. Such an intrusion affecting nearshore dissolved oxygen resources was linked with climate forcing effects of wind direction changes on 6/3/06 (Lower panel: Sandy Point data, MD DNR) and a resulting seiche of bottom waters of the mainstem Bay. Graphics from P. Tango.

As described in the appendices of this report, preliminary investigations by workshop participants and their collaborators; Boynton et al, Appendix 4) have shown that climate forcing (e.g. river flow – Boyton et al. Appendix 4, Figure 14 below), and temperature and solar angle (Buchanan, Appendix 1) plays a role in shallow water DO concentrations. There was also qualitative evidence that the most severe diel-scale hypoxia is observed at sites experiencing severe eutrophication (Boynton et al, Appendix 3). In contrast to this finding, Buchanan observed increased diel-cycling hypoxia in improving locations with thriving seagrass and/or benthic algal communities. These potentially conflicting observations highlight the need to further investigate whether diel-cycling hypoxia is a natural characteristic of seagrass beds in restored systems, or whether it represents a transitional stage in the restoration trajectory.

The question of whether the characteristics of variability in DO concentrations are likely to change along a restoration trajectory was of particular interest to workshop participants. Buchanan (Appendix 1) showed that daily mean dissolved oxygen concentrations differ significantly from year-to-year. Boynton et al (Appendix 4) demonstrated that inter-annual variation in violations of DO criteria may be related to inter-annual variation in flow, as evidenced by 2006-2008 results for the St. George Island ConMon site in Chesapeake Bay (Figure 14).

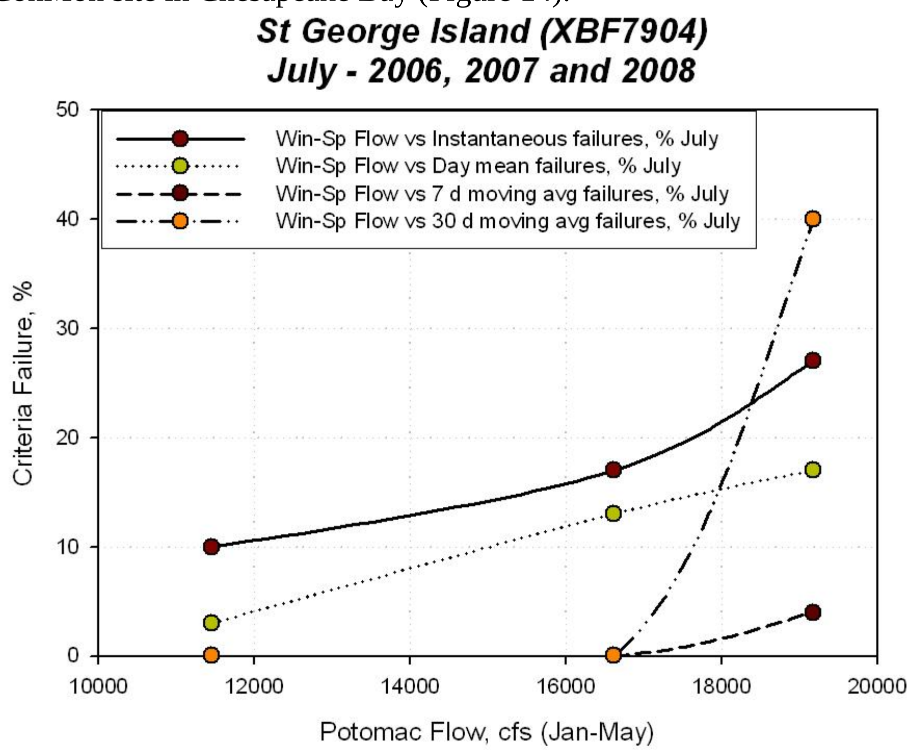


Figure 14. A multiple scatter plot of July DO % criteria non-attainment as a function of Potomac River flow (Jan-May flow period). Different DO % non-attainment calculation methods are indicated on the diagram.

At least two conceptual models have been previously conceived for tracking ecosystem response to eutrophication based on high frequency data. Jordan et al. (1992) used a plot of low DO duration and numbers of low DO events to illustrate a means of tracking changes in water quality conditions (Figure 15). In a similar vein, Boynton et al.

(2009) postulated a response to nitrogen load reductions in the Corsica River in the form of reduced duration of hypoxic events, as well as benefits to available bottom plant habitat (Figure 16).

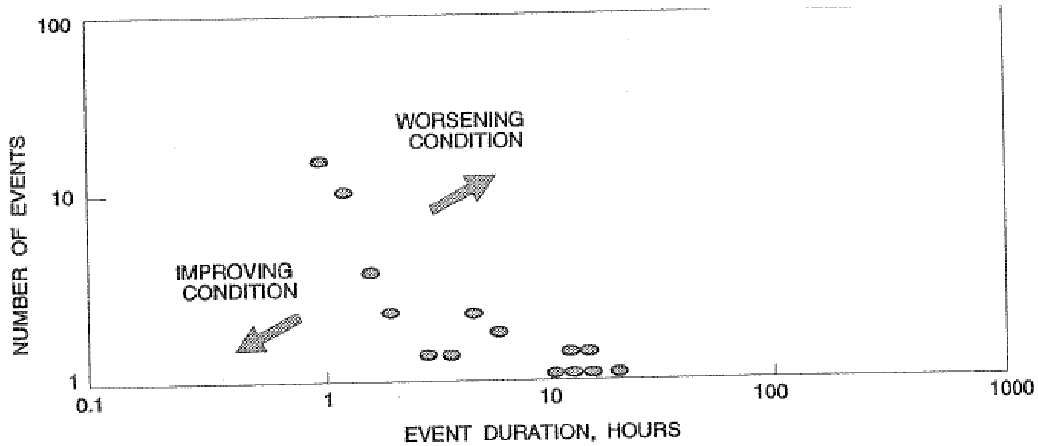


Figure 15. Number of events where DO fell below 3 mg/L in St. Leonard Creek, MD, 1988. Jordan et al. 1992

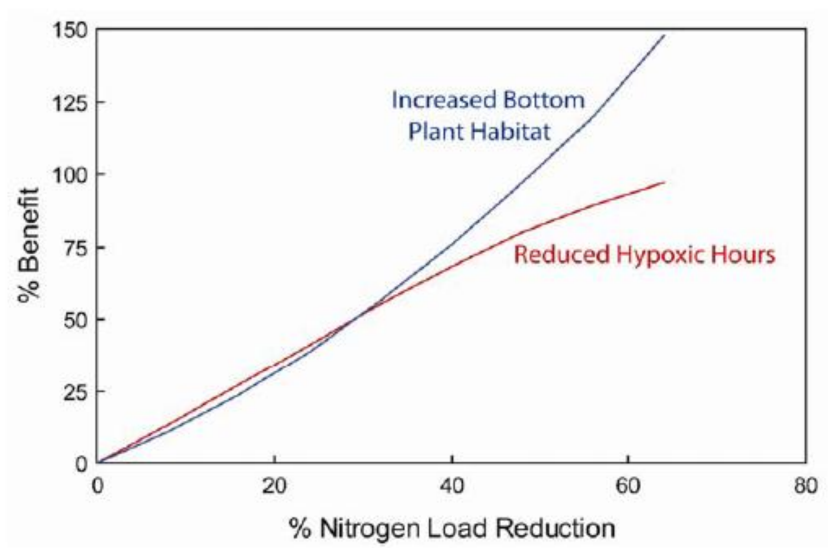


Figure 16. Potential Nitrogen load reductions and simulated % change in summer hypoxic hours and habitat for benthic algae in the Corsica River, Chester River system, MD. Boynton et al. 2009

The group emphasized that further exploration and development along these lines of inquiry would advance our understanding of the relationships between river flow, eutrophication, and the timing (e.g. day and/or night), duration, and variability of hypoxic events.

References

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.

Boynton, W.R., J. Testa and M. Kemp. 2009. An ecological assessment of the Corsica River estuary and watershed: Scientific advice for future water quality management. Final Report to the Maryland Department of Natural Resources. Technical Report Series No. TS-587-09 of the University of Maryland Center for Environmental Science. 54pp.

Brush, M.J., I.C. Anderson, and H. I. Kator. 2009. Comments on Addressing Optimization Needs of the STAC Review: Optimization of the Chesapeake Bay Water Quality Monitoring Program toward meeting Management Effectiveness needs in the Watershed, 9/17/2009. Appendix 9 *in* Wardrop, D. and C. Haywood. Monitoring Re-alignment Action Team Final Report to the Chesapeake Bay Program Management Board on behalf of the Monitoring Re-alignment Synthesis Team. October 27, 2009.

de Jonge, V.N., W.Boynton, C.F. D'Elia, R. Elmgren and B. Welsh. 1994. Responses to developments in four different North Atlantic estuarine systems. Pp 179-196 *in* Dyer, K.R and R.J. Orth (eds). Changes in fluxes in estuaries: implications from science to management. Olson and Olson Publ. Fredenberg, Denmark.

Jordan, S., C. Stenger, M. Olson, R. Batiuk, and K. Mountford. 1992. Chesapeake Bay dissolved oxygen goal for restoration of living resource habitats: A synthesis of living resource habitat requirements with guidelines for their use in evaluating model results and monitoring information. CBP Re-evaluation report #7c. CBP/TRS 88/93. Annapolis MD. Published by the MD Department of Natural Resources Tidewater Administration, Chesapeake Bay Research and Monitoring Division. 81pp.

McCaffrey, J. 2004. Factors controlling net ecosystem metabolism in U.S. estuaries. Estuaries 27(1):90-101.

MRAT 2009. Wardrop, D. and C. Haywood. Monitoring Re-alignment Action Team Final Report to the Chesapeake Bay Program Management Board on behalf of the Monitoring Re-alignment Synthesis Team. October 27, 2009

Neerchal, N.K., G. Papush, and R.W. Shafer. 1994. Statistical method for measuring DO restoration goals by combining monitoring station and buoy data. Transactions of Ecology and the Environment 5: 1743-3541.

Shenk, G. and R. Batiuk. 2010. *Evaluation of the Most Protective Bay Dissolved Oxygen Criteria.* Appendix D in USEPA (U.S. Environmental Protection Agency)

Chesapeake Bay Total Maximum Daily Load. EPA-R03-OW-2010-0736. U.S. Environmental Protection Agency, Region 3, Philadelphia, PA.

Tango, P. 2005. Diagnosing the Corsica River Fish Kill. Maryland's Waterman's Gazette, October 2005.

USEPA. 2003. *Ambient water quality criteria for dissolved oxygen, water clarity and chlorophyll a for the Chesapeake Bay and its tidal tributaries*. EPA 903-R-03-002, Washington, D.C.

USEPA 2003b designated use delineations.

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

USEPA. 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.