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

Findings of the Umbrella Criterion Action Team
Tidal Monitoring and Analysis Workgroup (TMAW)
2009-2011

August, 2012

STAC Publication 12-02

About the Scientific and Technical Advisory Committee

The Scientific and Technical Advisory Committee (STAC) provides scientific and technical guidance to the Chesapeake Bay Program on measures to restore and protect the Chesapeake Bay. As an advisory committee, STAC reports periodically to the Implementation Committee and annually to the Executive Council. Since its creation in December 1984, STAC has worked to enhance scientific communication and outreach throughout the Chesapeake Bay watershed and beyond. STAC provides scientific and technical advice in various ways, including (1) technical reports and papers, (2) discussion groups, (3) assistance in organizing merit reviews of CBP programs and projects, (4) technical conferences and workshops, and (5) service by STAC members on CBP subcommittees and workgroups. In addition, STAC has the mechanisms in place that will allow STAC to hold meetings, workshops, and reviews in rapid response to CBP subcommittee and workgroup requests for scientific and technical input. This will allow STAC to provide the CBP subcommittees and workgroups with information and support needed as specific issues arise while working towards meeting the goals outlined in the Chesapeake 2000 agreement. STAC also acts proactively to bring the most recent scientific information to the Bay Program and its partners. For additional information about STAC, please visit the STAC website at www.chesapeake.org/stac.

Publication Date:
August, 2012

Publication Number:
12-02

Cover photo courtesy of the Chesapeake Bay Program.

STAC Administrative Support Provided by:
Chesapeake Research Consortium, Inc.
645 Contees Wharf Road
Edgewater, MD 21037
Telephone: 410-798-1283; 301-261-4500
Fax: 410-798-0816
http://www.chesapeake.org

Contents

List of Figures

List of Tables

Appendices

Concept for the Instantaneous Minimum Criterion.

Executive Summary

Dissolved oxygen (DO) water quality criteria were established for the Chesapeake Bay and its tidal tributaries ("the Bay") under the authority of the Clean Water Act. The DO criteria, initially published by the U.S. Environmental Protection Agency (USEPA 2003) on behalf of the Chesapeake Bay watershed jurisdictions and supported by subsequent criteria addenda and technical documents, were adopted by Delaware, the District of Columbia, Maryland, and Virginia into their respective water quality standards regulations in 2004-2005. These Chesapeake Bay DO criteria were established to protect the reproduction, survival and growth of estuarine living resources. The DO criteria have spatially-specific habitat components or designated uses (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; spawning season, summer and "rest of year").

In the course of developing the Chesapeake Bay Total Maximum Daily Loads or TMDL, analysts at the USEPA's Chesapeake Bay Program Office (CBPO) conducted an assessment of how well DO criteria measured with the current Chesapeake Bay long term water quality monitoring program mutually protected the unmeasured criteria. Using hourly output from the calibration run of the Chesapeake Bay Water Quality Sediment Transport Model (WQSTM), the CBPO analysts determined that evaluation of the 30-day mean DO criteria was sufficient to determine attainment of the open-water and deep-water designated uses of the Bay. This "Umbrella Criterion Assumption" surmises that attainment assessment of one criterion serves as an "umbrella" assessment for the remaining criteria in a designated use. These findings have significant implications for whether and how data from the CBP partnership's Chesapeake Bay water quality monitoring program support assessment of the full suite of dissolved oxygen water quality standards applicable to the Bay's tidal waters.

Members of the Chesapeake Bay scientific and management communities raised questions concerning the model-based validity of the umbrella criterion assumption described above. An Umbrella Criterion Assessment Team was formed as a result of the CBP's Monitoring Realignment Action Team's efforts in 2008 and 2009. The Umbrella Criterion Assessment Team set out to characterize Bay conditions under which the umbrella criterion assumption were upheld and when and where it was violated. Team members explored both new methods and approaches for updating and validating previously proposed methods (USEPA 2004, 2007) to address short term (7-day, 1-day, instantaneous) Chesapeake Bay DO water quality criteria assessment.

The focus of this report supports 1) a demonstration of new methods as well as approaches for updating and validating previously proposed methods to address short duration (7-day, 1-day, instantaneous minimum) Chesapeake Bay DO water quality criteria assessment, and 2) an evaluation of the umbrella criterion assumption. The testing of the umbrella criteria assumption used applications of the new and updated analyses with Chesapeake Bay water quality monitoring data based on a CBP Scientific and Technical Advisory Committee (STAC)-sponsored workshop held in March 2011. The CBP STAC-sponsored workshop included a review of the advance analyses conducted by the Umbrella Criteria Assessment Team in preparation for the workshop.

Chapter 1 of this report reviews the basis of the umbrella criterion assumption. The chapter also provides background information on Chesapeake Bay DO water quality criteria, Chesapeake Bay criteria assessment methodology, and an historical perspective on assessments of relative protection provided between Chesapeake Bay DO criteria measured at different durations.

Members of the Umbrella Criterion Assessment Team conducted several independent analyses to evaluate the assumption. The experimental nature of some of the analytical approaches meant that significant time was invested in developing, testing and validating the methods. **Chapter 2** describes the data used in the analyses, provides insights into methods development and their evaluation, and provides a catalog of analyses and results from the umbrella criterion assessment. Appendices 2-12 document the details of methods and more complete results from individual analysts.

Participants in the STAC-sponsored Umbrella Criteria Workshop, held on March 15-16, 2011, addressed the original four questions submitted in the STAC proposal. STAC Workshop participants reviewed the analyses and provided discussion and recommendations for next steps in the umbrella criterion assessment. Outputs of the workshop are also summarized in Chapter 2.

Members of the Umbrella Criterion Assessment Team and attendees of the STAC Workshop developed additional questions and identified topics of importance beyond those specified in the original workshop proposal. In **Chapter 3**, insights, lessons learned and recommendations are provided regarding designated use definitions, the importance, significance and measurement of hypoxic event duration and a call for better defining patterns and drivers of diel cycling hypoxia.

Key findings and recommendation of the 2009-2011 umbrella criteria assessment process are summarized in terms of methods, modeling-monitoring comparisons, and umbrella criteria protectiveness.

Methods

Findings

- *Spectral casting - a statistical technique that creates a synthetic high density time series dataset at a low measurement frequency water quality monitoring site by integrating information from water quality measurements collected from separate high- and low-measurement frequency monitoring sites* - was applied to assess criteria attainment rates of the open-water 7-day mean DO criterion using the Chesapeake Bay DO water quality criteria attainment assessment framework.
    - The 7-day mean DO attainment results were compared to 30-day mean DO attainment results in evaluating the *Umbrella Criterion* assumption.

- It was deemed unfeasible to use *Spectral casting* to test protection of the instantaneous minimum DO criterion by the 30-day mean DO criterion under the current Chesapeake

Bay DO criteria attainment assessment framework. However *Spectral casting* is still a useful tool to assess other longer term criterion.

- A successful pilot test of an alternative approach (i.e., *conditional probability analysis*) to address the question of Umbrella Criterion protection was conducted by the Umbrella Criteria Assessment Team.
  - The application of this method is illustrated for evaluating the protection of the summer Open Water instantaneous minimum DO criterion by the summer Open Water 30-day mean DO criterion in Chesapeake Bay management segment CB4.
  - The probability approach also provided results applicable to testing the summer Open Water 30-day mean DO criterion protection for the summer OW 7-day mean DO criterion.

<u>Recommendations</u>

- *Spectral Casting* and *Conditional Probability Analysis* are recommended as useful tools for continuing umbrella criterion assumption evaluations.

Modeling-Monitoring comparisons

<u>Findings</u>

- Recognizing the modeling output-monitoring data comparisons were made at a single location but were not a 1:1 comparison in time (i.e., the model calibration data were from 1991-2000 while monitoring data were from 2009), modeling-monitoring comparisons showed:
  - Season-level patterns of trend in the DO monitoring measurements from May to November were accurately reflected in the CBP WQSTM output.
  - Comparisons using just the month of August showed model data variability tended to be lower than in the monitoring data.
  - A more restricted modeling-monitoring data comparison was conducted using two hydrologically similar years from the model calibration period to compare with the August 2009 monitoring results. Comparisons showed the model with either closely comparable or lower variability than DO monitoring observations.

<u>Recommendations</u>

- Conduct future assessments of model outputs-monitoring data comparisons using the real-time DO data from times coincident with some or all of the calibration period of the model. This will provide more of a 1:1 comparison in time and space for evaluating the model behavior against real time data. Example data sets for further investigating existing model output with monitoring data may include 1997-2000 MD DNR shallow water CONMON measurements associated with a Harmful Algal Bloom monitoring program.

- When the Water Quality Sediment Transport Model calibration period is extended beyond 2006, there will be greater opportunities for matching offshore, vertical water quality monitoring profiler data with model simulated output results. We view this as a particularly important testing procedure.

Testing Umbrella Criteria Assumption Protections

Findings

- The majority of work conducted by the Umbrella Criterion Assessment Team focused on one comparison: the mutual protection of the summer Open Water 7-day mean DO criterion by the summer Open Water 30-day mean DO criterion.

- The summer season, open water 30-day mean DO criterion protected the summer season 7-day mean DO criterion under the USEPA DO criteria assessment framework.
    - A diversity of habitats, tidal fresh to polyhaline waters, shallow and offshore waters, a range of nutrient conditions and interannual variability were taken into account in the evaluation.
    - This result has the following caveats: 1) analyses did not include all Bay segments, 2) occasional violations of the open water 7-day mean DO criterion occurred under conditions attaining the open water 30-day mean DO criterion, however, the violations were not excessive (< 10%) and could be deemed 'allowable exceedances' as described in USEPA regulatory assessments, and 3) DO variability is important to understanding the risk level of violating the assumption of protection provided by the 30-day mean DO criterion against other criteria.

- The summer season, open water 30-day mean DO criterion offered less than universal protection of the summer open water instantaneous minimum DO criterion. The finding is consistent with similar test results reported in USEPA (2004).

- The summer season deep water 30-day mean DO criterion mutually protected the summer deep water 1-day mean DO criterion in three, lower mesohaline Chesapeake Bay tributary site-specific assessments.
    - A caveat to the result is the new DO data assessment only reflected good water quality conditions; 30-day mean DO results did not closely approach or go below the 30-day mean DO criterion threshold.

- The summer season deep water 30-day mean DO criterion showed less than universal protection for the summer deep water 1-day mean DO criterion.

Recommendations

- Further analyses will be required to:
    - Complete a baywide assessment of summer season open-water 30-day mean protection for the summer open-water 7-day DO mean and instantaneous minimum.

- Complete a baywide assessment of summer deep water 30-day mean protection for the summer deep water 1-day mean DO and instantaneous minimum DO criteria.

STAC Workshop: Addressing the CBP STAC Workshop Proposal Questions

1. **Under what conditions do the Umbrella Criterion assumptions appear to hold?**
    - The summer season, open water 30-day mean DO criterion protected the 7-day mean DO criterion across salinity zones, and habitats (shallow water, open water).

2. **Under what conditions are the Umbrella Criterion assumptions likely violated?**
    - High frequency DO measurements in the Bay show violations of criteria thresholds occur even at sites with generally good water quality.

    - As variability in measurements of dissolved oxygen concentrations increased, the risk of failing the umbrella criterion protection assumption also increased.
        - Shallow water was more susceptible to failing a criterion than open water at short time scales (< 1 week) during the summer.
            - Umbrella Criteria Analysis Team results showed summer season shallow water frequently exhibits larger 24-hour (diel) fluctuations in DO concentrations than offshore open water.

    - Analyses conducted by team members raised interest in further investigating the effects of hydrodynamics (e.g. regions with strong spring-neap tide) and climate variability (e.g. interannual flows, wind, temperature) on the stability of the umbrella criterion assumption.

    - Recommendations:
        - Consideration can be given to separating shallow water (≤2m) and offshore water for DO criteria assessments. Implications of such a change on criteria assessment and attainment will require further analyses.
        - Further assess the effects of hydrodynamics and climate change impacts on the validity of the umbrella criteria protection assumption.

3. **Under what conditions do currently available data 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 regarding conditions where the umbrella criterion assumption is upheld or violated.
    - The further assessments could be based on currently available water quality monitoring data.
        - High frequency data sets (e.g. CONMON data) made the Umbrella Criteria assessment possible.
        - Vertical profiler data were invaluable in their utility but are a very rare commodity.

- Uncertainty surrounding the evaluations of the umbrella criterion assumption is a function of both the amount of data available and the ongoing developmental nature of the analytical tools at hand.
- The potential for decision error is affected by the amount of data available for the analyses.

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

- The Chesapeake Bay Program partner's shallow water CONMON data have proved essential to the evaluation of the umbrella criteria assumptions. Continued investment in shallow water, high-frequency data across the tidal Bay habitats is recommended.

- High-frequency vertical profiles of water quality are rare data sets, particularly in deep water regions of the Chesapeake Bay.
    - Our understanding of water quality variability in mid-channel locations is hindered by the paucity of high-frequency vertical water quality profiles available.
    - Vertical water quality profiler data were invaluable in providing verification of the spectral casting methodology, and reducing uncertainty in assessments for Chesapeake Bay management segments where it is available. More vertical water column continuous water quality monitoring measurements are needed and justified.

Recommendations for Next Steps Emerging from CBP STAC Workshop

- Collect high-frequency vertical profile data in deep -water regions of the Chesapeake Bay and its tidal tributaries.

- Further explore the concept of duration both as a component of the criteria and as a potential indicator of improving conditions.

- Generate a single dataset so that every analyst who participates in the collaborative analyses is using the exact same version of the water quality data.

- Expand the common 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 analysis addressing mutual protection between DO criteria as described in USEPA (2004).

- Quantify and clearly communicate the risk of erroneously classifying segment-designated uses as impaired or un-impaired ("false positives" or "false negatives").
    - In particular, quantify and communicate the uncertainty of current calculations of the "30-day mean" using only the long-term fixed station datasets.

- The instantaneous minimum criterion must be defined more precisely than is currently the case.
    - This issue becomes paramount when working with high-frequency datasets where measurements occur on scales of seconds to 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 for open water and deep water designated use assessments.
    - Use spectral casting where high-frequency shallow water data are available to conduct 7-day DO mean criterion assessments.
    - Publish a new USEPA technical addendum to the ambient water quality criteria with the updated DO criteria protection comparisons and advances in short-duration criteria assessment approaches.

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

Separating Shallow-Water and Mid-Channel Assessments

- There was consensus among workshop 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 shallow waters.

- Workshop participants suggested consideration for the partitioning of shallow, near-shore waters into their own assessment units to more adequately represent impacts of DO criteria violation in these biologically active regions.

Duration of Hypoxic Events

- In addition to the seasonal scale hypoxia chronic to the deeper portions of the Bay, the CONMON data exhibit diel scale hypoxia.
    - Since criteria levels are thresholds of significance to aquatic biota, analysis of *event scale duration below criteria* is a potentially valuable measure of habitat suitability.

- Seasonal maximum duration of DO violation of a criterion threshold, as measured with high frequency CONMON probes collecting data every 15 minutes, was shown to be linearly and positively related to percent violation of a given DO criterion based on a preliminary assessment.
    - A more complete development of this concept as an impairment or Bay recovery index is recommended.

Eutrophication Gradients and DO Variability

- Suggestions for further research included:
    - Better defining patterns and drivers of diel-cycling hypoxia.
    - Refining the utility of shallow-water DO concentration variability as a signal of eutrophication and for tracking Bay recovery.
    - Separating climate signals in the DO variability when defining the status of Bay health.
        - At least two conceptual models have been previously conceived for tracking ecosystem response to eutrophication based on high frequency DO data in Chesapeake Bay (see Chapter 3). The workshop participants emphasized that further exploration and development should be directed to advance our understanding of 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 Workgroup's (TMAW) Umbrella Criteria Assessment Team and further supported by Chesapeake Bay Program Scientific and Technical Advisory Committee (STAC).

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 Center for Environmental Sciences 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 Environmental Sciences Chesapeake Biological Laboratory Chesapeake Bay Program Office. 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, David Rudders and David Jasinski, 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. Editorial assistance was further provided by Ashley Harvey, Chesapeake Research Consortium.

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 Olson, 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;

Chapter 1. Introduction

1.1 Chesapeake Bay Ambient Water Quality Criteria and The "Umbrella Criteria" Assumption

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 (USEPA 2003). The Chesapeake Bay is divided into a management grid of 92 individual management segments (Figure 1). DO criteria are applied to five categories of "designated uses" (habitats used by various living resources in different seasons and life stages, Figure 2) that further divide the Bay: migratory and nursery (MN), shallow water (SW) for Bay grasses, open water (OW), deep water (DW) and deep channel (DC).

Standards are defined by a *criterion*, which represents a threshold below which DO concentrations should not fall. Chesapeake Bay criteria (Table 1) are further defined in terms of *duration* and *frequency* of violations. Duration and frequency serve to refine the criterion by defining the degree a criterion may be exceeded without inflicting unacceptable harm on the intended resource. Criteria application periods range from instantaneous to a 30-day average with seasonal and temporal applications (Table 1).

Currently, Chesapeake Bay Program Office (CBPO) Long-term Water Quality Monitoring Program data are deemed adequate to assess Clean Water Act (CWA) 303(d) listing status for only the 30-day mean DO criteria of the OW and DW designated uses of the Bay (USEPA 2003). Synoptic monitoring of the Bay has generally occurred on a temporal scale of one or two measurements per month during the 27-year history of the partnerships' Chesapeake Bay Long-term Water Quality Monitoring Program. As a result, monitoring data and/or appropriate analytical methods have not existed to support Baywide assessment of the short-duration criteria (USEPA 2004). However, as the DC 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 (USEPA 2003).

In the course of developing the Chesapeake Bay Total Maximum Daily Loads or TMDL, analysts at the USEPA's CBPO conducted an assessment of how well DO criteria measured with the current Chesapeake Bay long term water quality monitoring program data protect as yet unmeasured criteria. Using hourly output from the calibration run of the Chesapeake Bay Water Quality Sediment Transport Model (WQSTM), the CBPO analysts determined that evaluation of the 30-day mean DO criteria was sufficient to determine attainment of the open-water and deep-water designated uses of the Bay. This **"Umbrella Criterion Assumption"** surmises that attainment assessment of one criterion serves as an "umbrella" assessment for the remaining criteria in a designated use. 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 criterion assumption focused on the summer season. These findings have significant implications for whether and how data from the CBP partnership's Chesapeake Bay water quality monitoring program support assessment of the full suite of dissolved oxygen water quality standards applicable to the Bay's tidal waters.

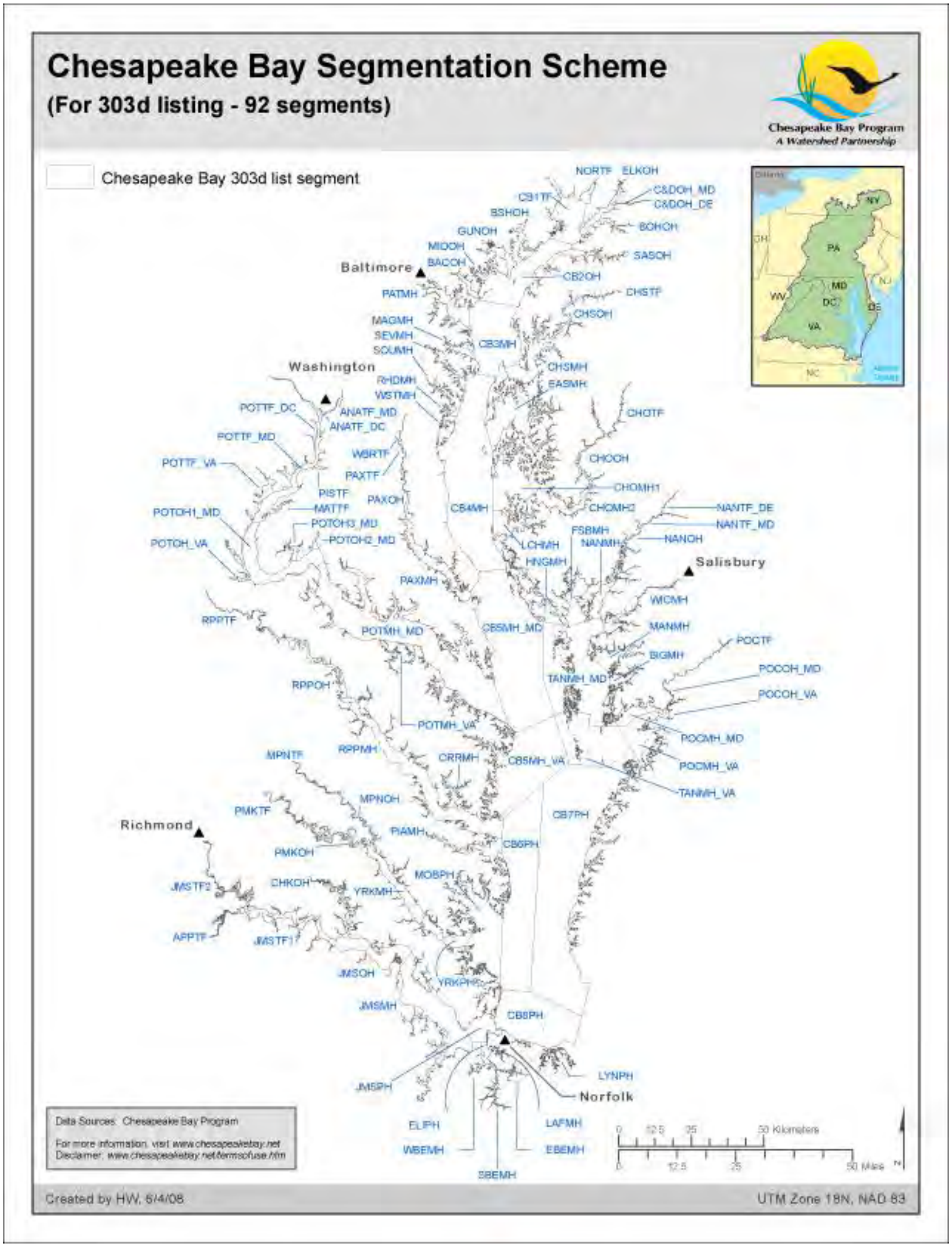


Figure 1. Chesapeake Bay 92 segment management grid of tidal waters

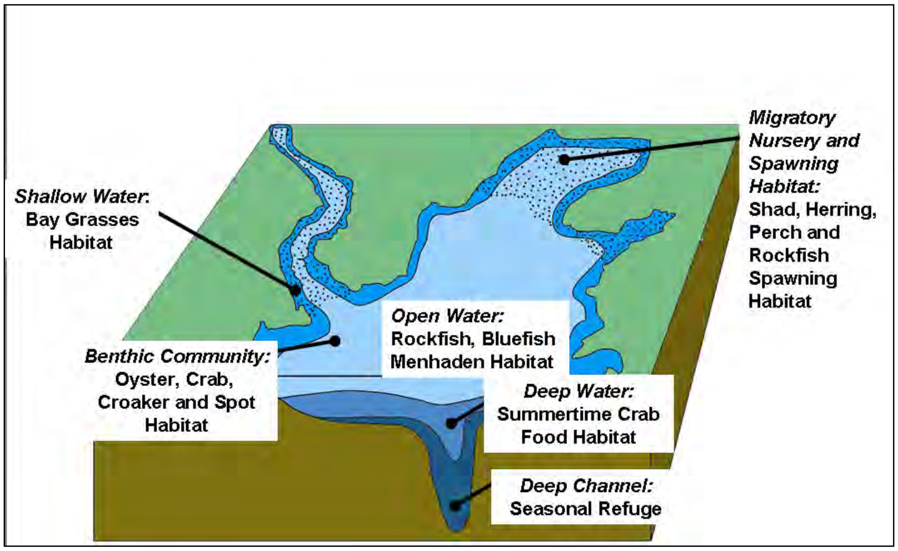


Figure 2. Chesapeake Bay water quality disignated uses (USEPA2003)

Table 1. Chesapeake Bay Water Quality Criteria (from USEPA2003)

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*) ($>29^{o}C$) dissolved oxygen concentrations above an instantaneous minimum of 4.3 $mg*L^{-1}$ will protect survival of this list sturgeon species.

The version of the WQSTM used to support Chesapeake Bay's management community decisions for Bay restoration produces simulated data at a high spatial (1 km^2) and temporal (1 hr time step) resolution. This high-resolution output was used to evaluate how well water quality measures attaining one criterion are also protecting water quality for a second short-duration criterion (e.g. DO measurements supporting the attainment of the open water 30-day mean DO criterion also result in attainment of the OW 7-day mean criterion). WQSTM results indicated that when the summer 30-day mean DO 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). 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 DO 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*.

Members of the Chesapeake Bay scientific and management community raised questions concerning the validity of this assumption. The WQSTM is calibrated using the long-term, biweekly to monthly, water quality data collected in the mid-channel of tributaries and mainstem of the Bay. There were concerns expressed that the WQSTM thus may not adequately capture the true variability of dissolved oxygen concentrations relevant to the short-duration criteria or in habitats away from the mid-channel waters. 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). However, water quality of the shoreline habitats in the Bay has historically been poorly characterized. Jordan et al. (1992) focused their criteria protection work on management segments with seasonal hypoxia and not those with large areas of shallow water. Under the present Chesapeake Bay water quality criteria assessment framework, shallow waters are not separated from mid-channel waters when assessing dissolved oxygen criteria attainment (USEPA 2003b). However, USEPA (2007) acknowledged the importance of acquiring better understanding of water quality dynamics for shallow Bay habitats.

To improve our understanding of shoreline habitat conditions, shallow water monitoring was added to the CBP partnership's tidal water quality monitoring program in 2004. The development, application, availability and affordability of emerging technologies for near-continuous monitoring of water quality has allowed the collection of datasets of a high temporal density (seconds to minutes intervals) and segment level spatial distributions in the Bay. These datasets shed new light on previously uncharacterized short-duration (seconds-to-weeks) variability for dissolved oxygen in Bay shallow water and offshore habitats.

1.2 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 CBP evaluates observed violations of the dissolved oxygen criteria against a cumulative frequency distribution (CFD) curve (USEPA 2003, 2007). Historically, USEPA has provided for a 10% allowable exceedance in temporal or spatial assessments against thresholds. The application of a two-dimensional CFD simultaneously accounting for violations in time and space is an innovative approach designed to better reflect the relative impact of varying degrees of violation. A **reference curve** is the curve of compliance, represented in a two dimensional plane of percent space and percent time (Figure 3). The curve may be a mathematically derived 10% curve or a "**bioreference curve"**. The bioreference curve is a reference CFD derived from observed criterion 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.

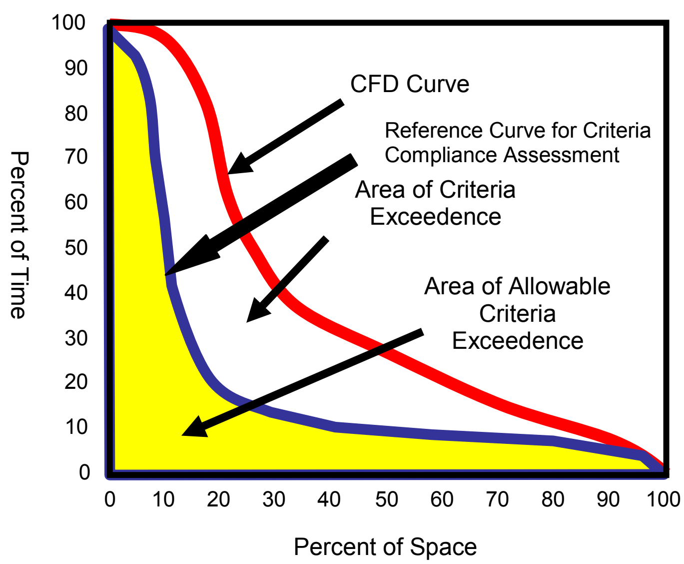


Figure 3. Reference curve, CFD compliance curve, and an illustration of the area of allowable exceedances

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 water quality standards compliance assessment is provided below (Box 1). This framework is currently applied to the assessment of the 30-day mean DO criteria in the OW and DW designated uses, and to the instantaneous minimum DO criterion in the DC designated use. It is currently considered desirable to extend the same method to assessment of the short duration criteria. Historically, assessment of criteria for sub-30-day time periods has been limited by the availability of data at the appropriate temporal resolution.

Box 1: Chesapeake Bay 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

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).

1.3 History: Prior Insights on Comparative Criteria Protectiveness

The question of the relative protection offered between water quality criteria measured at different temporal scales has been raised previously in the history of the CBP partnership. Early in the 1990s, when experts first identified a 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-duration criteria. By this time, several groups had already begun to experiment with near-continuous monitoring technologies. Chesapeake Bay scientists began measuring *in-situ* DO concentrations on timescales as short as 3-4 seconds (DATAFLOW) during spatial habitat assessments and 5-15 minutes for days-to-months at a time at a fixed site monitoring locations (e.g. CONMON or continuous monitoring). In 1992, workers at the Maryland Department of Natural Resources and the USEPA CBP used a combination of low and high frequency DO measurement data 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. Jordan et al. (1992) developed regression equations to derive the seasonal mean concentrations that could be presumed protective of target, shorter-duration DO thresholds in a given Bay management 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 DO 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" 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 (p. 179, USEPA 2003).

Two statistical approaches for assessing the sub-30-day duration DO 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. The caveats identified were that insufficient high-frequency DO data sets were available from the Bay and there had been insufficient validation of the approach preventing its immediate application. For the instantaneous minimum DO 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 use of this approach could be accepted.

In 2004, the CBP revisited this question of mutual protection among criteria with different duration applications. With the publication of the USEPA (2004) addendum to USEPA (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. Each dataset contained short duration (days to weeks), high frequency DO measurements. 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. The potential utility of using regression models to predict attainment of short-duration criteria was demonstrated. More importantly, they showed that the relative protection offered by the 30-day mean DO criterion for the instantaneous minimum DO criterion varied across segments. USEPA (2004) provides a catalogue of segments where they postulated that a 30-day mean DO criterion was also protective of the 7-day mean, 1-day mean or instantaneous minimum DO criterion. In contrast to their findings for the instantaneous minimum DO criterion, Olson et al. found that the Open Water (OW) 30-day DO mean was generally protective of the OW 7-day mean DO criterion in those segments where both criteria applied.

In the USEPA (2007) addendum to USEPA (2003), the USEPA CBP again expressed its support for continued development and eventual application of adequate statistical approaches for assessing the short-duration criteria (i.e., 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 prediction models for each station in the CBP long-term fixed station water quality monitoring program record. It also recommended eventual application of the spectral analysis approach for criteria assessment if further development showed it to be at least as robust as the logistic regression approach.

During 2008 and 2009, as part of the background work supporting the CBP Monitoring Realignment (MRAT) process, Chesapeake Bay researchers and analysts advanced analyses of comparative criteria protection (MRAT 2009). The Umbrella Criterion protection assumption was raised and preliminary results questioned but did not refute the protection offered by the umbrella criterion concept.

1.4 2010 Umbrella Criteria Workshop & Report Proposal

The CBP partnership requested the continued development of the MRAT analyses to improve the assessment of the umbrella criterion assumption. In early 2010, the Chesapeake Bay Program's Tidal Monitoring and Analysis Workgroup (TMAW) convened the Umbrella Criteria Assessment Team of scientists, analysts and managers. The objective of the Umbrella Criteria Assessment Team was to use available Chesapeake Bay monitoring data to characterize the conditions under which the umbrella criteria assumption was upheld, and when and where the assumption was likely to be violated. The application of emerging monitoring technologies to Chesapeake Bay water quality monitoring (e.g. CONMON, DATAFLOW, vertical water quality profiling instruments) since the analyses conducted supporting DO criteria development (USEPA 2003, 2004) produced many new, often season-long, high-frequency DO measurement datasets. The new datasets were suitable to recommended applications needed for testing a variety of statistical techniques for measuring short-duration water quality criteria (USEPA 2003, 2004, 2007). Members of the Team conducted analyses evaluating the validity of the assumption across physical and ecological gradients ranging from shallow, tidal fresh regions to the polyhaline mainstem of the Chesapeake Bay.

Chapter 1 of this report reviews the basis of the umbrella criterion assumption. The chapter also provides background information on Chesapeake Bay DO water quality criteria, Chesapeake Bay criteria assessment methodology, and an historical perspective on analyses of relative protection provided by Chesapeake Bay DO criteria measured at different durations.

Members of the Umbrella Criterion Assessment Team conducted several independent analyses to evaluate the assumption. The experimental nature of some of the analytical approaches meant that significant time was invested in developing, testing and validating the methods. **Chapter 2** describes the data used in the analyses, provides insights into methods development and their evaluation, and provides a catalog of analyses and results from the Umbrella Criterion assessment.

Members of the Umbrella Criterion Assessment Team and attendees of the STAC Workshop developed additional questions and identified topics of importance beyond those specified in the original workshop proposal. In **Chapter 3**, insights, lessons learned and recommendations are provided regarding designated use definitions, the importance, significance and measurement of hypoxic event duration and a call for better defining patterns and drivers of diel cycling hypoxia.

Chapter 2. Umbrella Criterion Analyses

2.1 The Data

To date, the development and assessment of Chesapeake Bay DO criteria has relied primarily on the CBP long-term, low frequency (i.e. biweekly to monthly) water quality monitoring program dataset and short duration, high frequency datasets from assorted buoy deployments (e.g. USEPA EMAP) in the Bay (Jordan et al. 1992, USEPA 2004). The CBP long term water quality monitoring program has 27 years of data in the form of vertical water column profiles, collected at approximately 178 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 CBPO 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(15 minute intervals) at static, nearshore, shallow-water locations, while the DATAFLOW program conducts surface water sampling cruises mapping water quality in a spatially intensive manner (samples every 3-4s). 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 Clean Water Act water quality criteria assessment procedures developed by the CBP partnership (USEPA 2003, Batiuk et al. 2009). Once a three year assessment has been completed in a management segment, CONMON equipment is moved to a new segment. This rotation of CONMON resources is intended to eventually provide high-frequency data on shallow-water habitats for all 92 segments supporting 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. Table 3 catalogs the datasets used in the Umbrella Criterion Assumption assessment.

Research efforts to evolve water quality assessment applications in Chesapeake Bay using 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, improve volumetric resolution in monitoring programs (using a 'random walk' methodology), and reduce uncertainty in our understanding of Bay water quality dynamics.

Table 2. Data sources serving the Umbrella Criterion Assumption 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 178 stations. Covers tidal fresh to polyhaline habitat conditions.
USEPA EMAP: Historical short-	Mix of short term (days to weeks)	Fixed site, off shore locations,

term buoy deployments with high temporal frequency at a station. Single depth sensor evaluations.	time series with high temporal frequencies by sensor. See USEPA (2004).	varied depths. Tidal fresh to polyhaline habitat conditions.
CBP Shallow Water Monitoring Program, **Con**tinuous **Mon**itoring (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 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 lower tidal tributary data.
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 mapping 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.

2.2 Approaches Used to Evaluate the Umbrella Criteria Assumption

2.2.1 Catalog of Analyses Approaches used for the Umbrella Criteria Assessment Assumptions

The analyses used in the umbrella criteria assumption assessment are summarized in Table 4a-d. The suite of approaches documented here reference the work previously published in USEPA (2004, Chapter 5) as well as many analyses with new, shallow water high frequency CONMON and offshore, vertical water quality profiler data sets. The new analyses include previously recommended approaches (i.e. spectral analysis), but also direct time series assessments comparing violation rates between criteria across the summer season. The details of analyses and expanded findings are referred to in the tables and referenced in the Appendices at the end of this report.

The umbrella criteria assumption was tested for the summer season as defined by USEPA 2003 for regulatory water quality standards assessments in OW and DW habitats (designated uses). "Open water" was viewed three ways: 1) shallow water (≤2m deep), 2) offshore water (>2m deep) and 3) shoreline-to-shoreline combining shallow water and offshore water reflecting the present criteria assessment methodology definition of open water (USEPA 2003). Deep water was also evaluated according to the regulatory definition as water between the upper pycnocline and the bottom or lower pycnocline depending on measures of stratification.

Table 3. Catalog of analyses conducted evaluating the umbrella criteria assumption, nearshore, shallow-water (<2m) only

Nearshore, Shallow-water only			
Author	**Analysis**	**Data**	**Appendix**
Buchanan	**Violation rate plots** Attainment/nonattainment rates compared between 30-day mean and short duration criteria. Also regression analyses 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 4. 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 duration 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** Attainment/nonattainment rates compared between 30-day mean and short duration criteria. Also 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** Attainment/nonattainment rates compared between 30-day mean DO and short duration DO criteria with bi-plots for Bay segments.	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. Attainment/nonattainment compared between criteria. 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 were 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: Conditional Probability Analysis** The probability analysis evaluated the protection of the instantaneous minimum DO criterion from 30-day mean DO.	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 5. 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 6. 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** Attainment/nonattainment rates compared between 30-day mean DO and short duration DO criteria with bi-plots for Bay segments.	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** Attainment/nonattainment rates compared between 30-day mean and short duration criteria. Also regression 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 Focus Method Update: Validation of the Spectral Casting Method for Application to Assessing Short-Duration 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-duration 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 Umbrella Criteria Assessment Team effort.

The updated and validated spectral analysis method described here 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) such as one of the CBP long-term water quality monitoring fixed station network sites. Vertical profiles of 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 water quality criteria 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 effects from DO concentration variability with water column 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.

Robertson and Lane compared 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, using cubic spline interpolation did not change support illustrated for the Umbrella Criterion Assumption for OW 30-day mean DO criterion protection of the 7-day DO mean criterion identified using FFT.

Perry compared criteria violation results using FFT, spline and linear interpolations of time series (Appendix 11). His comparisons of cubic spline interpolation with FFT interpolation supported Robertson and Lane. 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 CBP long-term, low-frequency fixed station water quality monitoring program adequately captures -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 whether variability from a shallow-water sending site can be cast to a mid-channel receiving site, reduced workers' initial comfort with the application of the spectral casting method.

In an effort to understand the relative contributions of these two sources of error, Perry constructed a validation exercise (Appendix 11). Perry 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 integrated with the low frequency interpolation to form synthetic water quality 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 Perry to differentiate uncertainty due to low frequency sampling from uncertainty due to casting variability between monitoring sites. Validation exercises were performed 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. A second set of validation exercises were performed on datasets from two shallow water CONMON deployments in the lower Potomac River.

Results were consistent across both validation tests: **the variability due to the high frequency data 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 Chesapeake Bay management segment-level criteria assessments.**

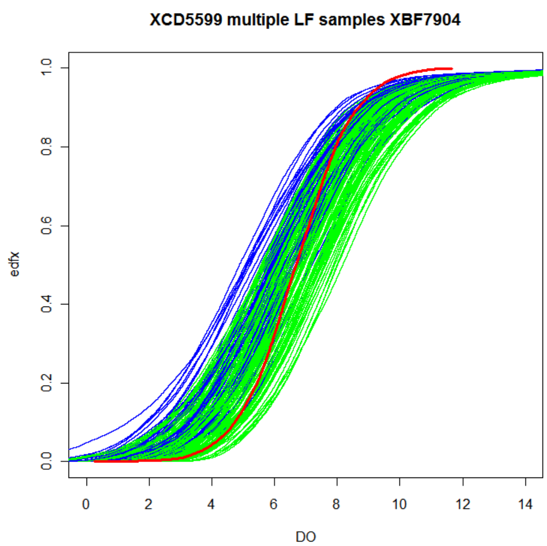


Figure 4. Variation due to multiple low-frequency samples from the receiving site with Fourier Series interpolation

Empirical distribution functions illustrate variation in dissolved oxygen measurements due to multiple low-frequency samples from the receiving site using a Fourier Series interpolation. The sending site data set was held constant as a single two-week interval. Blue curves are synthetic DO data frequency distributions based on a repeated sampling of night-time data. Green curves represent DO frequency data distributions from a repeated sampling of daytime DO measurements. The red curve is the actual DO frequency distribution based on the two weeklong receiving site high frequency data set. X-axis is DO values, Y axis is cumulative frequency of the DO measurements in the 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.

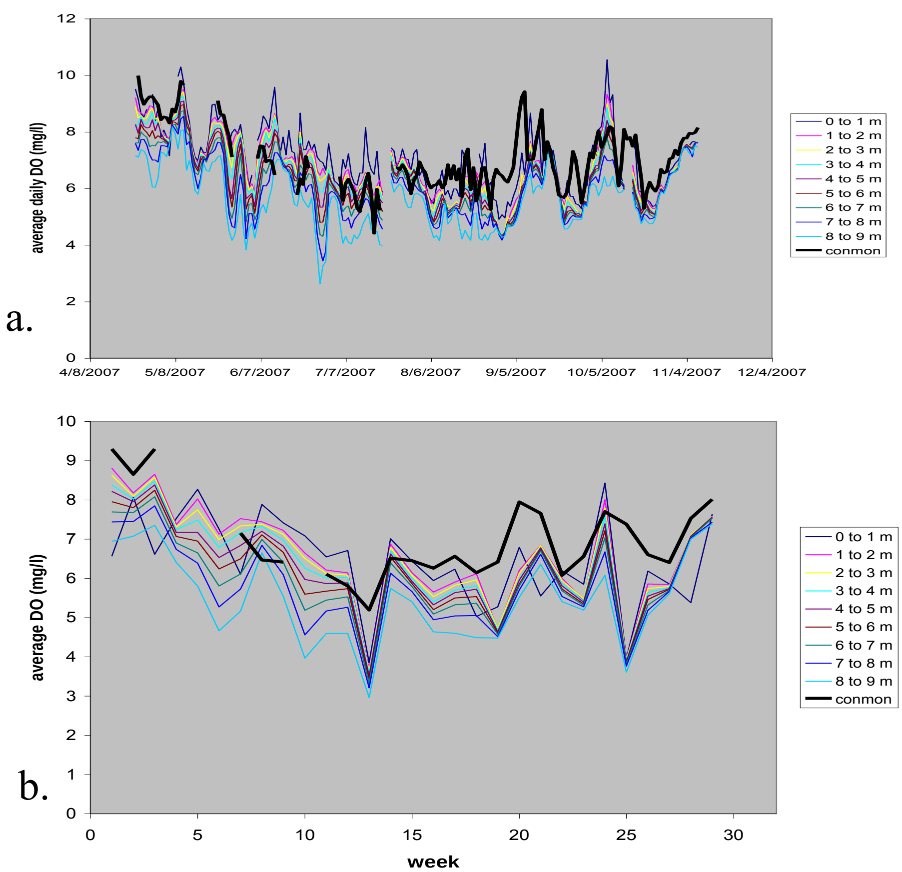


Figure 5. Comparison of dissolved oxygen from York River vertical profiler and continuous monitoring station d trend are significantly different between continuous monitor and profiler time-series. B) DO represented by weekly averages. The variance and trend are statistically similar between continuous 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, a nonparametric form of ANCOVA analysis.

In another test of spectral casting variability from St. George's Creek and the Yeocomico River CONMONs were cast upon a 2009 lower Potomac River vertical profiler site. For the test, the high frequency profiler dataset was sampled monthly at each depth, yielding a single dissolved oxygen concentration at each depth for each month. The subsampling effort approximated the sampling protocol and resulting vertical DO concentration water column profile from a CBP partnership long term water quality fixed station. Next, spectral analysis was used to create synthetic datasets using the monthly sampled vertical profiler data as a “receiving” site with data from the nearby CONMONs as “sending” sites. The resulting midchannel synthetic time series were tested for failure frequency of each DO criterion (30-day mean, 7-day mean, instantaneous minimum) according to habitat. The synthetic data failure frequencies were compared in a validation exercise to the failure frequencies for the continuous vertical profiler data. The tests supported the application of the approach for certain short duration criteria attainment assessments. Neither the synthetic nor raw data produced results violating the umbrella criteria assumption; results were similar regardless of depth for 30-day versus 7-day mean DO criterion. Under the instantaneous minimum criterion, however, there was less comparability. Many discrepancies were observed between the synthetic and real data but they appeared to be in the pycnocline transition depths; otherwise, the differences in failure percentages were not large.

2.3.2 Focus Method: Pilot Study of a Parametric Simulation Approach to Assessing the Umbrella Criterion Concept for the Instantaneous Minimum Criterion

High frequency DO concentration time series data collected from fixed stations showed considerable serial dependence or autocorrelation. This lack of independence makes it difficult to analytically compute the probability of instantaneous criterion violation when an umbrella criterion (e.g. weekly or monthly mean) is satisfied. The Umbrella Criterion Assessment Team (Perry Appendix 12) used a simulation approach in a pilot study attempting to address this question.

The basic approach of the simulation is to generate time series that has characteristic properties similar to observed DO concentration time series. The data used for this exercise were 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 autoregressive (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. The simulation test focused on one assessment unit, the surface layer of Chesapeake Bay management segment CB4 (sensor < 10 m depth) because CB4 is one of the best represented segments in these data.

Results showed the summer OW instantaneous minimum DO criterion was not protected by the 30-day mean DO criterion. A much higher 30-day mean criterion threshold would be required to mutually protect the OW instantaneous minimum DO criterion. A secondary output of the simulation, however, provided additional support for the summer OW 30-day mean DO criterion protecting the summer OW 7-day mean DO criterion.

2.3.3 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 USEPA CBPO's WQSTM. An assessment of the direct model output indicated that in the OW and DW 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.

Model-monitoring comparisons are based on the following:

- Hourly model output for the years 1991-2000 was matched with the Potomac River mesohaline vertical profiler monitoring location and the days-of-the-year from the MD DNR 2009 vertical profiler deployment;
- For seasonal comparisons of water quality behavior, surface water quality measurements from May through November 2009 were compared with the output from the surface water cell containing that site in the model;
- For the month of August, box plots show the statistical characteristics of DO vertical profiler measurements from 2009 compared with binned 1991-2000 August model output for two water column depth layers represented in the WQSTM.

Hydrologically, 2009 annual flows to Chesapeake Bay were below average (http://md.water.usgs.gov/waterdata/chesinflow/). Dissolved oxygen dynamics have been correlated with flows that deliver loads to the Bay where typically wetter years produce more hypoxia than drier years. While it is important to make general model output and monitoring data comparisons here given the caveat of non-overlapping data in this analysis, attention can be focused further on years between 1991 and 2000 with similar annual river flows. In 1993 and 1996 annual flows to the Bay were below average and most similar to 2009.

The seasonal level patterns of trend present in the monitoring data are reflected in the model output across a range of years and hydrological conditions (see Figure 4). Higher average levels of dissolved oxygen tend to be measured earlier in the year and decline into the summer and autumn seasons.

Comparisons for the month of August, across years from 1991 to 2000 (and depending on depth), show the variability in model DO output tended to be lower than the observed variability from measured DO concentrations (see Figure 5). Focusing on the most hydrologically similar years to 2009, however, the model output for August 1993 showed much lower variability than August profiler data in both the surface and deep water layers while August 1996 data were most comparable with the monitoring measurements. Model 1993 and 1996 output and observed 2009 data show high variability in the surface compared with other years; these years were however the most similar for measures of central tendency and range in the deep water. In the deep water, August 2009 would have ranked as the lowest average DO compared to all years between 1991 and 2000. By comparison, Years 1993 and 1996 DO also show the lowest DO levels and had the two lowest measures of central tendency in the 1991-2000 time series making them most similar in DO behavior to 2009.

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. The Umbrella Criteria Assessement Team recommends further analyses should expand comparisons of WQSTM output with 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 (e.g., 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 since calibrations rely on mid-channel monitoring data. Also, lower Eastern Shore of Maryland waters tend to be characterized by blackwater conditions that will further challenge the model in modeling-monitoring data comparisons. Such data sets, however, provide further opportunity to make direct modeling-monitoring comparisons and gain insight into model and ecosystem behavior.

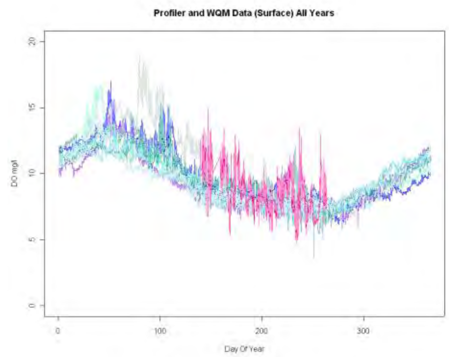
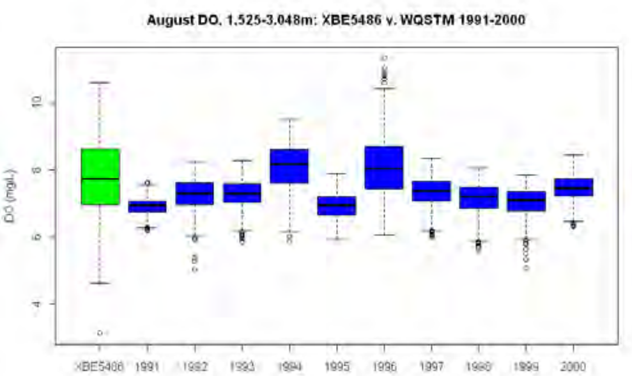


Figure 6. Time series of 2009 May-November deployment of the POTMH profiler (pink) compared to all 10 yrs of WQSTM output combined (various other colors). Surface cell only.

(a) August dissolved oxygen: 1.525-3.048 meters water depth at Chesapeake Bay long term water quality monitoring station XBE5486.



(b) August dissolved oxygen 7.63 – 9.144 meters water depth at Chesapeake Bay long term water quality monitoring station XBE5486.

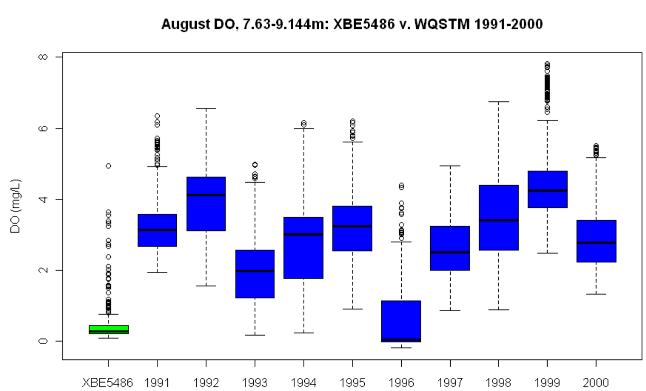


Figure 7. WQSTM simulated DO concentrations for August of individual years, 1991-2000 compared with observed DO concentrations for the same approximate depth and location from August, 2009.

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

The OW 30-day mean DO criterion of 5.0 mg/L (5.5 mg/L when salinity ≤ 0.5 ppt) was protective of the OW 7-day mean criterion of 4.0 mg/L. The Umbrella Criteria Assessment Team analyses derived support for the Umbrella Criteria assumption in OW habitat using water quality data from mainstem Chesapeake Bay and tributary locations, all salinity regimes (i.e. tidal fresh-to-polyhaline habitats), shallow and offshore waters, diverse nutrient conditions, and included inter-annual variability via multi-year evaluations for the summer season. The results have the following caveats: 1) analyses were not exhaustive of all Bay segments, 2) occasional violations of the OW 7-day mean criterion occurred under conditions attaining the OW 30-day mean criterion, however, the violations were not excessive (< 10%) and could be deemed 'allowable exceedances' as described in USEPA regulatory assessments, and 3) DO variability is important. Changes in DO variability relative to the measured DO means will affect the risk of failing water quality standards attainment; measurements showing greater DO variability will enhance failure risk.

Analyses by Perry (Appendix 2, 12) suggest that the risk of violating protection for the OW 7-day mean criterion may increase beyond 10% when DO measurements are consistently near the OW 30-day mean criterion level. Because measurements for the OW 30-day mean rarely exhibit this behavior of hovering around the criterion value for long periods of time, it appears safe to conclude that in most cases the OW 30-day criterion acts as an umbrella for the OW 7-day mean criterion. However, the umbrella does not seem to be a broad one. Slight increases in the variation of the 7-day mean DO about the monthly mean DO without corresponding increases in the monthly mean could increase the violation rate unacceptably to above 10%.

Table 7. Habitat-specific declarations on the protectiveness of the 30-day mean as the umbrella criterion: shallow-water

Shallow water, nearshore habitat only		
Author/Appendix	**Is the 30-day mean protective of…**	
	7-day mean?	**Instantaneous minimum**
Buchanan Appendix 1	**Support**	**Low Levels of Support**
Perry Appendix 2,	**Support**	**NA**
Boynton et al. Appendix 3, 4	**Support**	**Not supported**

Table 8. Habitat-specific declarations on the protectiveness of the 30-day mean as the umbrella criterion: above pycnocline open water not including shallow-water habitat

Open Water (Offshore habitat only)		
Author/Appendix	**Is the 30-day mean protective of…**	
	7-day mean?	**Instantaneous minimum?**
Brush et al. Appendix 5	**Support**	**Support**

Bilkovic et al. Appendix 6	**Support**	**Support**
USEPA 2004 Appendix 7	**Support**	**Intermediate level of Support**
Hall Appendix 8	**Support**	**Not Supported**
Perry Appendix 12	**Support**	**Not supported**

Table 9. Designated use specific declarations on the protectiveness of the 30-day mean as the umbrella criterion: 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	**Supported**	NA

Table 10. Habitat-specific declarations on the protectiveness of the 30-day mean as the umbrella criterion: 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	**Not supported**	**Not supported**
Bilkovic et al. Appendix 6	**Supported**	**Supported (York River site)** **Not supported (Rappahannock River site)**
USEPA 2004 Appendix 7	NA	**Not supported**
Hall Appendix 8	**Supported**	NA

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

Analysis of the Chesapeake Bay monitoring data suggest the OW 30-day mean DO criterion offers less than universal protection of the OW instantaneous minimum DO criterion (Tables 5a, 5b, 5c) as has been suggested by the WQSTM. Results are consistent with analyses conducted by Olson (USEPA 2004) focusing on offshore buoy data in the mainstem Bay and tributaries which showed the 30-day mean protected the instantaneous minimum (estimated by the 1st percentile of the data) approximately 85% of the time (n=94 cases). New vertical water quality profiler data sets from offshore locations in the mesohaline lower York River, lower Rappahannock River and lower Potomac River produced results consistent with the umbrella criteria assumption for protecting the OW instantaneous minimum. However, the Virginia and MD tributary vertical profiler analysis results cannot be deemed conclusive support for the Umbrella criteria assumption. Bilkovic et al. (Appendix 6) found low violation rates of the OW instantaneous minimum DO criterion (0.04-0.4%) in the lower York River during 3 summer seasons when observed 30-day means were >5.5 mg O_2/L; in 2009, there were 0 to 3.3%

violations of the OW instantaneous minimum DO criterion per month were observed in the lower Rappahannock River when OW 30-day DO means were >5.8 mg O_2/L. The DO measurements from these two sites produced no data for assessing the critical data region where the OW 30-day DO means are less than 5.5 mg O_2/L and equal to or above the OW 30-day DO mean criterion value of 5.0 mg O_2/L. Hall's results (Appendix 8), using a spectrally-cast synthetic DO dataset developed for the lower Potomac River, were similarly consistent with umbrella criteria assumptions.

Common failures of the OW instantaneous minimum DO criterion were also observed in shallow water habitat when the OW 30-day mean DO criterion was met. By evaluating 211 summer monthly means from Potomac River CONMON results, Buchanan found that the OW 30-day DO mean was completely protective of the OW instantaneous minimum DO criterion about 35% of the time. Boynton et al. (Appendix 3,4) evaluated shallow water CONMON data and reported common exceedances of instantaneous minimum criterion in shallow water habitat coincident with 30-day DO means above the OW DO mean criterion; there were, however, generally fewer failures at better quality sites.

The Umbrella Criteria Assessment Team was successful in applying the full USEPA criteria assessment methodology with the addition of spectral casting to conduct the OW 30-day mean criteria comparison with the water quality data analyzed for 7-day DO means. The Team, however, deemed it unfeasible with available staff and computing resources to apply the full USEPA criteria assessment methodology for OW DO criteria at the CBP segment-level using high frequency measurements of the instantaneous minimum DO. Perry (Appendix 12) demonstrated the use of conditional probability analysis to overcome the computational challenges of the regulatory criteria assessment method evaluating OW 30-day mean DO criterion protection for the instantaneous minimum DO criterion. Applying conditional probability analysis using high frequency buoy data from Chesapeake Bay management segment CB4, Perry showed the OW instantaneous minimum was not protected when data met the OW 30-day mean DO criterion. Further, Perry's analyses showed that even a significant increase in OW 30-day DO mean measurements above the OW 30-day mean DO criterion did not provide sufficient protection for the OW instantaneous minimum DO criterion in segment CB4.

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

The DW 30-day mean DO criterion protected the DW 1-day mean DO criterion in site specific assessments. Testing, however, was limited to three new DO data time series from vertical water quality profiler stations located in the lower York, Rappahannock and Potomac River tributaries to Chesapeake Bay. Expanded testing is recommended as data becomes available, for example, from mainstem Chesapeake Bay management segments that have Deep Water designated uses.

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

The Deep Water 30-day mean DO criterion showed less than universal protection for the Deep Water 1-day mean DO criterion in site-specific assessments. Analyses in USEPA (2004) indicated that the DW 30-day mean DO criterion protected the DW instantaneous minimum DO criterion in 57% of the cases reviewed. Among the new analyses from three vertical profiler sites - the mesohaline York River (Brush et al. Appendix 5, Bilkovic et al. Appendix 6), the mesohaline Rappahannock River (Bilkovic et al. Appendix 6) and the mesohaline lower Potomac River (Hall Appendix 8) – results were mixed with 2 showing violation of the Umbrella Criteria protection assumption. Further work, however, is again recommended 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 four key questions posed in the original CBP STAC Workshop proposal regarding the Umbrella Criteria assumption assessment were:

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 Workshop participants came to the following conclusions:

2.4.1 Conditions for which the Umbrella Criterion Assumption appears to hold

The OW 30-day mean criterion is generally protective of the OW 7-day mean criterion in Bay habitats. This finding was consistent with a previous analysis summarized in USEPA (2004). Workshop participants also recognized that violations of the OW 30-day mean criterion threshold existed in DO time series even where water quality was deemed good. Analyses showed that with the level of variability we observe in DO measurements in the Bay, the umbrella protecting the 7-day mean is not a broad one. The risk of violating the assumption is approximately 10%, a key value for exceedance thresholds in meeting or not meeting DO criteria attainment standards. Analyses showed that as the 30-day mean DO criterion threshold value was increased even a small amount (e.g. 5.0 to 5.3) the risk of violating the OW 7-day mean declined to 6-9%, depending on the standard deviation of the dataset.

With limited analyses, the preliminary evidence suggested the DW 30-day DO mean protected the DW 1-day DO mean criterion; further analyses were recommended to support these results, and to further define the conditions under which this result this holds true.

2.4.2 Conditions for which the Umbrella Criterion Assumption may be violated

Understanding DO variability relative to a criterion value was considered key to understanding the level of confidence behind the Umbrella Criterion Assumption. Measurable violations of criteria can be found even at sites with good water quality. Variation in dissolved oxygen measures relative to 30-day mean DO was observed to affect the risk of failing the water quality standards attainment. For a given 30-day mean DO value, when dissolved oxygen variability increases the resulting risk of failing attainment for shorter term criteria also increased. A corollary to this observation was discussed at the workshop suggesting that an improvement to the criteria framework would have criterion values matched with declarations of acceptable measures of DO variance.

Analyses showed that shallow waters may be more likely than open waters to exhibit large 24-hour (diel) fluctuations in DO concentrations. For periods less than one week, shallow water DO variation was statistically different than offshore DO variation. Concern was expressed that shallow water may violate criteria at a different rate than offshore, open water. Workshop recommendations suggested that strong consideration be given for separating shallow water and offshore water in the OW DO criteria attainment assessments.

A variety of stressors that may affect in situ DO variability were presented and discussed in terms of their potential effect on DO criteria attainment assessments including 1) regions with strong spring-neap tide effects, 2) inter-annual variation in river input (i.e., flow and nutrients), 3) wind, 4) temperature and 5) algal biomass levels. Further analyses were recommended for separating and accounting for the impacts of the variety of stressors on meeting DO attainment standards.

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

The consensus of the workshop participants was that given sufficient time and analytical resources, more thorough answers could be provided on conditions where the umbrella criteria assumption holds or is violated using currently available data. However, uncertainty surrounding the boundaries of these condition statements is a function of both the amount of data available and the ongoing developmental nature of the analytical tools at hand. Decision error in our Clean Water Act listing assessments is affected by the amount of data collected.

Significant differences were identified in DO dynamics between shallow and offshore waters. The workshop participants discussed the present USEPA DO criteria assessment framework which combines shallow and offshore habitats in the OW criteria evaluations. Shallow water and offshore waters were statistically similar for weekly and monthly means but significantly different at sub-weekly time scales. Separating these habitats for criteria assessment was recommended. The separation of the habitats would be most critical for the short-duration (<7 day) OW criteria assessments.

The basis of the umbrella criteria assumption rested on CBP WQSTM model output. The CBP community was interested in direct modeling-monitoring comparisons with high frequency data. Comparisons in this assessment were limited due to available data from emerging technologies coming online for Bay monitoring primarily in the decade after the 1991-2000 focus years of model output. The Chesapeake Bay model outputs for the TMDL were focused on 1991-2000 while vertical profiler data and most shallow water CONMON measurements are from more recent (2009) monitoring program efforts. There is some earlier work with CONMONs (e.g. lower Pocomoke River 1997-2000 MD DNR) that might be provided for additional model-monitoring comparisons until model outputs are available for years after 2000.

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

It was agreed that the investment in the shallow water CONMON program has been tremendously valuable and has greatly enhanced our understanding of shallow water tidal Bay habitats. CONMON datasets were critical in the ability of the Umbrella Criteria Assessment Team to conduct analyses evaluating the Umbrella Criterion assumption. Continued investment in collecting shallow water high-frequency water quality measurements across the tidal Bay habitats will improve the strength and accuracy of our analyses and understanding of water quality in local, nearshore environments. Equally important was the information gained on relative differences in short-term variability for dissolved oxygen measures across nearshore, shallow-water systems with different levels of impairment.

The workshop participants further discussed that our understanding of water quality variability in mid-channel locations, and particularly in deep water regions, remains very limited. Vertical profiler data were invaluable and essential in providing verification of the Spectral casting method assessment and garnering support for recommending adoption of the technique into the USEPA criteria assessment framework. The paucity of high-frequency vertical water quality profile data sets limited DW designated use testing of the Umbrella Criterion assumption as well as additional verification of approaches suitable for the analyses. Vertical profiler water quality data for criteria attainment assessments can reduce uncertainty in attainment measurement results for CBP management segments. Continuation of the ongoing effort to obtain high-frequency vertical profile water quality data in these regions is recommended. 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 validation data sets for new statistical approaches to water quality criteria assessment and key insights into DO variability characterization in the Deep Water regions of Chesapeake Bay.

2.4.5 Recommendations and Next Steps Emerging from the 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. In order to better document those times and regions for which it is most important that short-duration criteria be assessed, and then to assess short-duration criteria, workshop participants recommended the following activities:

1. Generate a single dataset through a Data Enterprise so that every analyst who participates in the analysis collaboration is using the same version of the available water quality 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 new or updated analytical tools from the Umbrella Criterion assessment, update the segment-by-segment analysis. 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*. This analysis can be used to further target regions of greatest concern for establishing priority high-frequency data collection locations and reducing uncertainty in criteria attainment assessments.

3. With regard to communicating the results of future criteria attainment analyses to managers and decision-makers, quantify and clearly communicate the risk of erroneously classifying segment-DUs impairment status ("false positives" or "false negatives"). In

particular, communicate uncertainties in criteria assessments associated with current calculations of the "30-day mean" (using only the long-term fixed station datasets). 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 approach to communication would be a useful way to visualize comparative levels of protectiveness provided by the 30-day mean DO criterion to shorter duration criteria (see Figure 8).

4. In order to properly assess the instantaneous minimum DO criterion and to compare its violation rates to longer duration criteria, workshop participants recommended reviewing and revising the current definition for the instantaneous minimum DO criterion. The present definition is based on the use of long term, biweekly to monthly water quality measurements. This definition issue becomes paramount when one begins working with high-frequency datasets, for which the measurement intervals are now just seconds to minutes. While a consensus was not reached regarding the most appropriate application of the instantaneous minimum criterion, two recommendations were discussed. First, there was extensive discussion regarding the importance of duration with regard to an instantaneous minimum DO measure. 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. Such measures of violation will be affected by the sampling frequency.

5. Convene an expert panel to review the adequacy of the spectral casting method for assessing short-duration criteria. Through the Umbrella Criterion Assessment Team effort we have already gained insight into uncertainties due to spectral casting based on Chesapeake Bay long term water quality monitoring program data in the USEPA-approved water quality criteria assessments framework. 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 DO criteria assessment programs as necessary to conduct a full regulatory assessment of higher-frequency, shorter duration criteria for Open Water and Deep Water designated uses where appropriate. Consider 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 DO 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 Chesapeake Bay ambient water quality criteria technical 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 USEPA (2003).

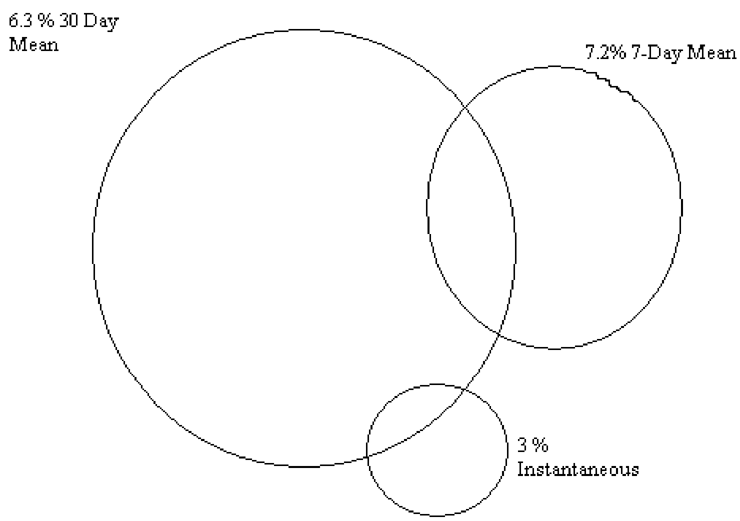


Figure 8. Hypothetical example of usage of a Venn diagram for communicating relative protectiveness of each criterion to managers

CBP STAC Workshop participants emphasized in general that the dissolved oxygen criteria would benefit from further research and conceptual development. Chapter 3 below describes long-term, developmental 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 water quality measurements 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 changing water quality 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.

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

The Umbrella Criteria Assessment Team and STAC workshop participants 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 water volumes for assessment of the DO standards in the Open Water designated use;

- Criteria definitions and implications: The potential importance of event duration associated with hypoxic events (now measurable with high frequency data) on the underlying intention of living resource protection with the criteria, and further, understanding how duration of hypoxic events varies with pollutant loads.

- Questions regarding the potential for patterns of DO concentration measurement variances to change along a trophic status gradient.

3.1 Combining vs. Separating 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 Bay waters may be disproportionate to relative volume of the habitat. 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).

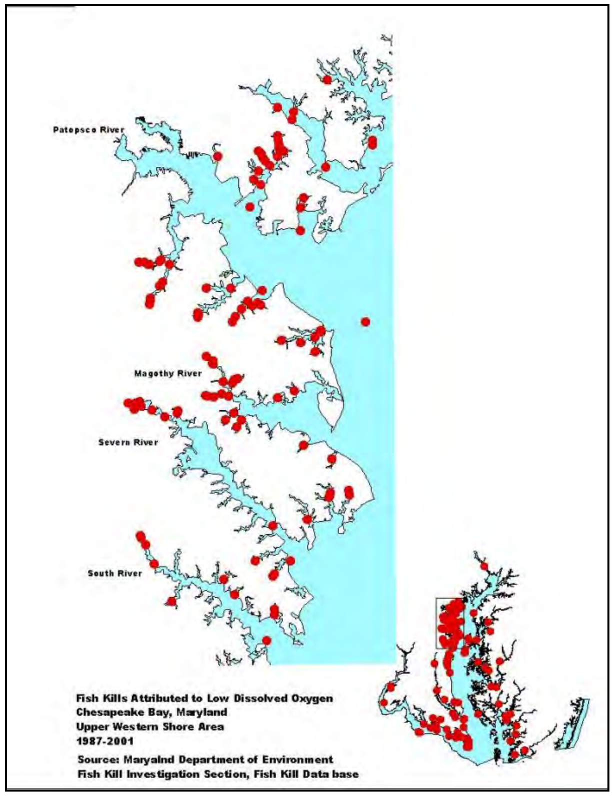


Figure 9. Distribution of fish kills attributed to low dissolved oxygen, Chesapeake Bay, MD, 1987-2001, Source: Maryland Department of the Environment

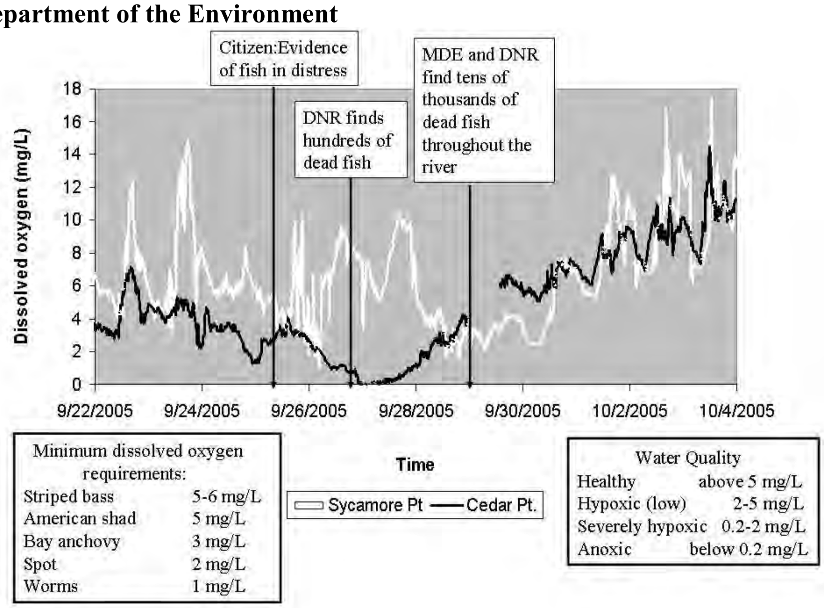


Figure 10. Chronology of a fish kill and associated water quality in Corsica River, MD, 2005. Graphic from Mark Trice, MDDNR originally released in P. Tango, 2005 Waterman's Gazette.

The Umbrella Criterion Assessment Team analyses have also shown that shallow waters may be more likely than open waters to exhibit large 24-hour (diel) fluctuations in DO concentrations. 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 in shallow water habitats 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 may more adequately represent the impact of DO criteria violation in these biologically active regions. The implications of such a partitioning for assessment and management will require further research and analysis.

3.2 The Importance of Hypoxic Event Duration Patterns to Protecting Living Resources with Existing DO Criteria Attainment Protocols

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 DO criteria levels are thresholds of significance to aquatic biota, analysis of *duration below criterion,* now possible with our high-frequency water quality measurements in the Chesapeake Bay Program monitoring program, becomes a new, potentially valuable additional measure of habitat suitability.

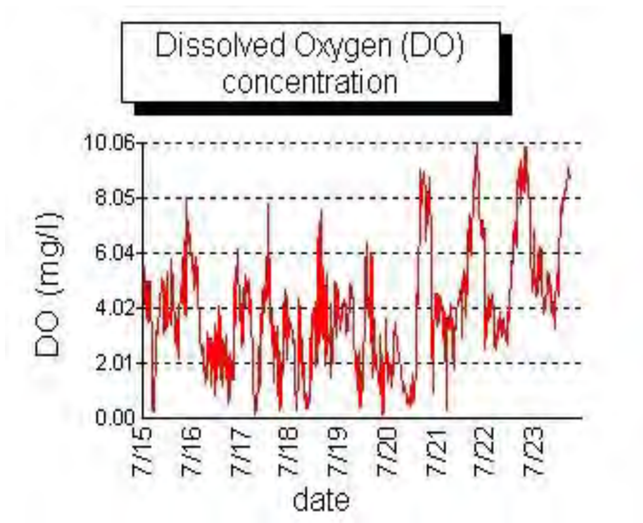


Figure 11. Example of diel hypoxia in shallow water at Ben Oaks, Severn River, MD. Data collected every 15 minutes. Graphic attributed to Maryland Department of Natural Resources.

At a 2011 CBP Tidal Monitoring and Analysis Workgroup meeting, Matt Hall (MD DNR) presented analyses to the Umbrella Criteria Assessment Team demonstrating the use of summer season DO CONMON data, the maximum duration of DO criteria violation varied across shallow water in the Potomac River. **Maximum duration below most DO criteria generally followed a gradient of putative eutrophic condition**. **In other words, the most eutrophic stations had the longest maximum durations below DO criteria.** Boyton et al. (Appendix 3) provided further support for this tendency across other Bay locations. In a preliminary analysis, m**aximum duration of violation was further shown to be linearly and positively related to percent violation of a given DO criterion** (P. Tango, Pers. Comm.)**.** Figure 12 below illustrates this concept using Chesapeake Bay CONMON data evaluated against the Open Water DO 30-day mean criterion; a similar relationship was found when using the OW instantaneous DO criterion. The USEPA-supported CFD water quality criteria assessment methodology does not explicitly address low DO event duration. However, if the total violations of DO attainment are correlated to maximum duration of DO violation against a criterion, then the CFD approach may implicitly capture measures of event duration. We can therefore postulate that as restoration progresses toward improving water quality conditions, habitat stress is not only decreased in the seasonal condition, but in sub-season, short duration measures of habitat conditions.

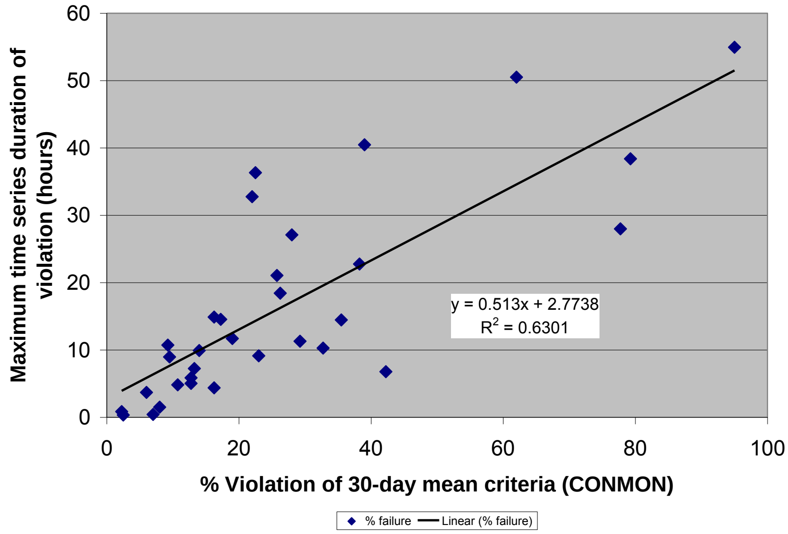
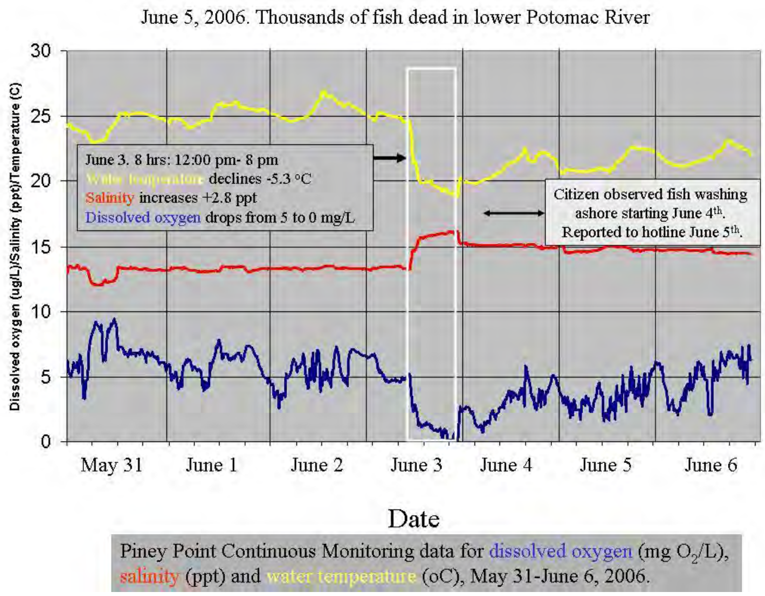


Figure 12. Maryland ConMon assessment of site specific % ConMon criteria violations against the maximum period of duration of criteria violation (hours) in a 30-day period. Data: Matt Hall MD DNR, Graphic: Peter Tango USGS@CBPO.

Results of the USEPA 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 good promise here as an index to the 30-day mean violation rate measured at a point. A more complete development of this index is recommended comparing segment-level DO criteria attainment results and CONMON measures at different locations within the segments.

3.3 Eutrophication Gradients and DO Variability

Suggestions made by the Umbrella Criteria Assessment Team and STAC Workshop participants for further research included better defining patterns and drivers of diel-cycling hypoxia as well as the utility of shallow-water DO concentration variability as a signal of eutrophication status. 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.



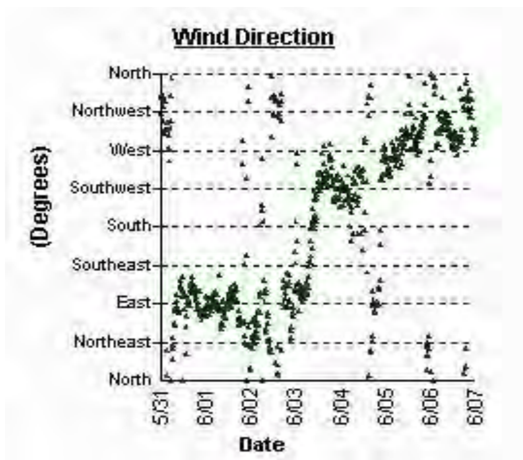


Figure 13. Intrusion of anoxic water from the Bay affecting nearshore dissolved oxygen.

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).

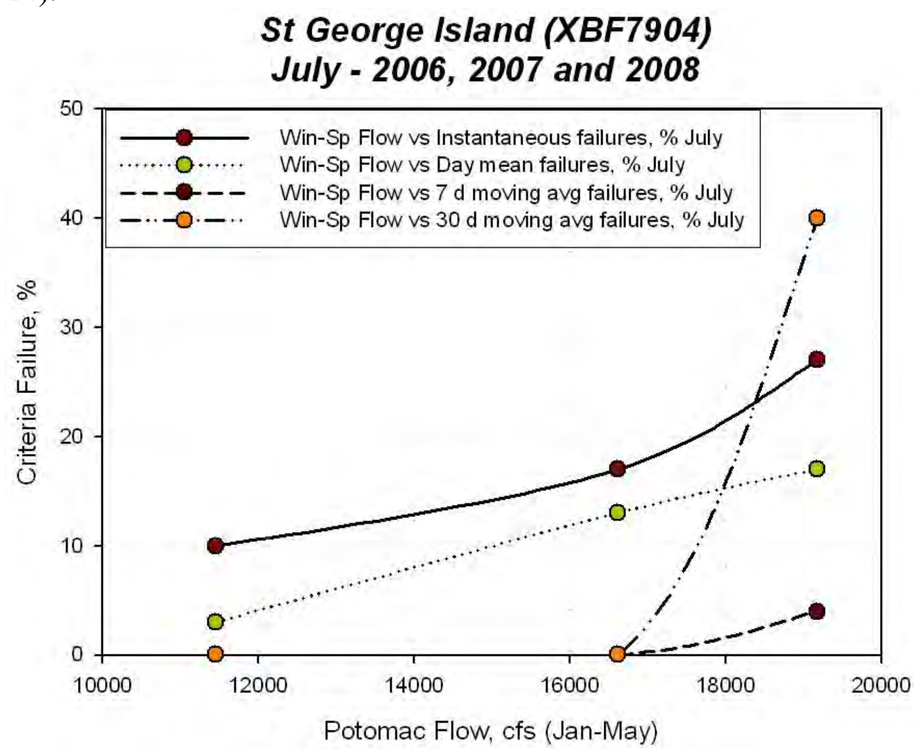


Figure 14. Multiple scatter plot of July dissolved oxygen criteria non-attanment as a function of Potomac River flow.

(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).

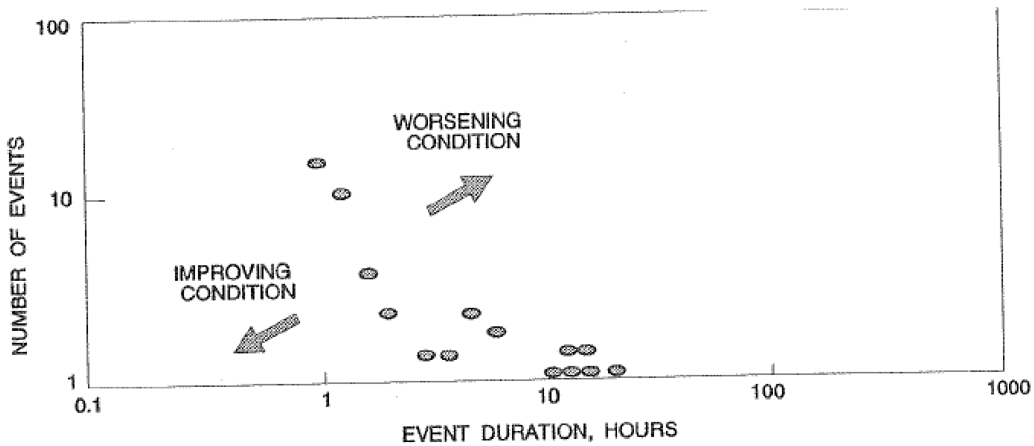


Figure 15. Number of events where DO fell below 3 mg/L in St. Leonard Creek, MD, 1988. Jordan et al. 1992

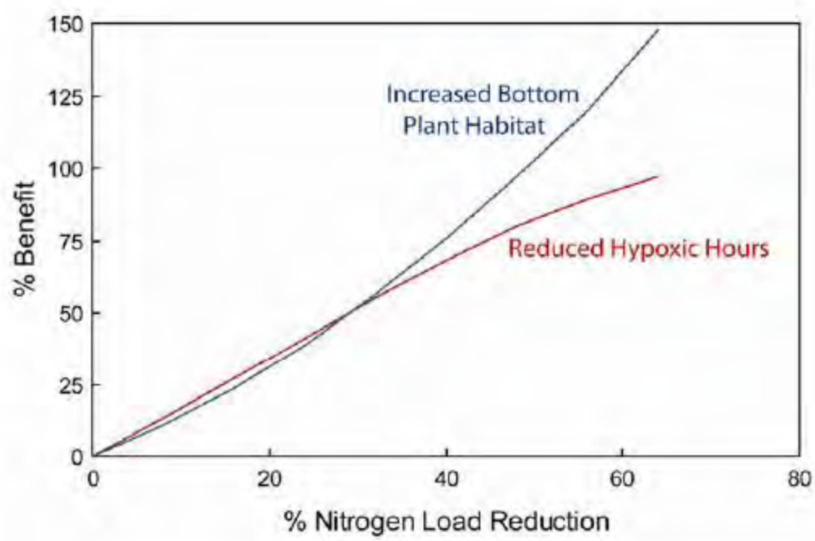


Figure 16. Potential Nitrogen load reductions and simulated % change in summer hypoxic hours and habitat for benthic algae in the Corsica River, Chester River system, MD. Boynton et al. 2009.

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. 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.

APPENDICIES

Appendix 1, Part A of A&B.

A test of the "umbrella criteria" concept in tidal Potomac shallow waters

Claire Buchanan – Interstate Commission on the Potomac River Basin

Dissolved oxygen criteria for migratory fish spawning and nursery uses and open water fish and shellfish uses are presently applied to shallow waters designated for bay grass use on the assumption that DO conditions are not sufficiently different to warrant criteria for shallow waters. High frequency DO data are suggesting this assumption may be faulty and the "umbrella criteria" that perform in open waters may not necessarily work in shallow waters. The validity of the "umbrella criteria" concept in shallow waters was tested with high frequency data collected at 19 nearshore sites in the tidal Potomac River (**Figure 1**). Specifically, does attainment of the 30-day mean DO criteria protect against failure of the 7-day mean DO criteria, and do both of these criteria protect against failure of the instantaneous minimum DO criteria in shallow waters?

Over 1.1 million dissolved oxygen records were collected by Maryland and Virginia at 19 Potomac shallow water stations during spring, summer, and autumn. Most stations were monitored for two or three years between 2004 and 2008; four were monitored all five years. Sondes were positioned at median depths of 0.2 – 3.0 m below the surface and reading made at 15-minute intervals. The 30-day, 7-day, and instantaneous minimum criteria were applied to each site -year-season subset of CMON data to address the "umbrella criteria" question.

Means derived from high frequency (CMON) data are very close to the true mean for a given location while those derived from low frequency, or point sample, data can diverge from the true mean. To evaluate the sensitivity of low frequency data estimates of the 30-day mean to 7-day and instantaneous DO criteria failures, 30-day means were computed both from the available low frequency (calibration) data and high frequency data. Guidelines for computing and applying the 30-day mean, 7-day mean and instantaneous minimum criteria were established to ensure consistency and avoid artifactual results (**Table 1**). The CMON data, the data analysis methods, and some of the results and conclusions are described in more detail in Buchanan (2009).

Figure 2 shows the frequency per month of failing the 7-day mean criteria, plotted against the 30-day mean derived from high frequency (CMON) data and low frequency (calibration) data. **Figure 3** shows the frequency per month of failing the instantaneous minimum criteria, plotted against the 30-day mean derived from high frequency (CMON) data and low frequency (calibration) data. **Figure 4** shows the frequency per 7-day period of failing the instantaneous minimum criteria, plotted against the 7-day mean. Spring results are separated from summer and autumn results and tidal fresh (TF) and oligohaline (OH) results are separated from mesohaline (MH) results because DO criteria differ according to season and salinity zone (see **Table 1**). In each graph, the inverted triangle indicates the criteria applicable to the metric on the x-axis. Points to the right of the inverted triangle are achieving that metric's criteria. **Figure 5** compares monthly means derived from the low and high frequency data. **Figure 6** delineates the threshold for failure of the instantaneous minimum criteria as a function of diel or weekly magnitude of change in DO and daily or weekly mean DO.

Findings:

- *Depending on allowable exceedances*, the 30-day mean criteria applied to the low frequency (calibration) data could be considered protective of the 7-day mean criteria. The 30-day mean criteria applied to the high frequency (CMON) data show similar results and support this finding.
- If the allowable exceedance dictates, for example, that only 1 month of the year can have failures of the 7-day mean criteria, then the 30-day mean criteria is *not* protective of the 7-day mean criteria in Piscataway Creek (2004, 2005), Piney Point (2006), St. Mary's River (2008), and Breton Bay (2008).
- The 30-day mean and 7-day mean criteria are *not* protective of the instantaneous minimum DO criteria, regardless of whether they are derived from low or high frequency data.
- Monthly means derived from low frequency (calibration) data are inaccurate estimates of the true mean, and also appear to be biased in some season-salinity zones.
- Meeting the instantaneous minimum criteria is a largely function of the daily mean DO *and* the diel magnitude of change in DO and the trajectory these parameters take over time. If the diel magnitude is large and the mean is relatively low, the probability of failing the instantaneous minimum criteria is high. If the diel magnitude is small and the mean relatively high and stable, the probability of failing the instantaneous minimum criteria is very low.
- Data points representing weeks with low frequencies (>0% - 1%) of failing the instantaneous minimum criteria provide a linear boundary that separates days or weeks achieving the instantaneous minimum criteria from those failing the criteria. This linear boundary quantitatively describes the relationship between mean DO and magnitude of change and the thresholds of instantaneous minimum criteria failure on daily and weekly scales.
- The 7-day mean DO criterion of 6 mg/liter in spring migratory and spawning reaches is only protective of the instantaneous minimum criteria when the weekly magnitude of change in DO is less than ~6 mg/liter. The 7-day mean DO criterion of 4 mg/liter is only protective of the instantaneous minimum criteria when the weekly magnitude of change is less than ~2 mg/liter, a phenomenon that the tidal Potomac embayments and river flanks never experienced in the spring, summer or fall between 2004 and 2008.
- Daily or weekly DO means of ~10 mg/liter are protective of the instantaneous minimum DO criteria in almost all circumstances. DO means below 10 mg/liter are only protective of the criteria if their magnitudes of cyclic change in DO (diel, weekly) are proportionately smaller, i.e. below the boundary lines indicated in **Figure 5**.

Buchanan, C. 2009. An Analysis of Continuous Monitoring Data Collected in Tidal Potomac Embayments and River Flanks. ICPRB Report 09-3, 56 pgs. Available online at: http://www.potomacriver.org/cms/publicationspdf/ICPRB09-03.pdf

Table 1. Metrics, criteria, and computation guidelines.

Metric and Criteria	CMON data	Calibration data
30-day mean • ≥5.5 mg/liter, TF year-round • ≥5.0 mg/liter, OH & MH year-round	all available data averaged by month (not exactly 30 days); months with less than 20 days of uninterrupted DO recordings not included	samples from 0.5-1.5 m depths only, all available data averaged by day, then averaged by month
7-day mean • ≥6.0 mg/liter, TF&OH Feb1-May 31 • ≥4.0 mg/liter, all salzones Jun1-Jan 31 and MH Feb1-May 31	all available data between midnight on the first day and midnight on the 7th day averaged; means were calculated a) from rolling 7-day averages advanced in 1-day steps or b) from sequential 7-day periods with uninterrupted data records (method noted in figures); weeks with less than 6 days of uninterrupted DO recordings were excluded	n/a
Instantaneous minimum • ≥5 mg/liter, TF & OH Feb1-May 31 • ≥3.2 mg/liter @ ≤29^{o}C and ≥4.3 mg/liter @ >29^{o}C, all salzones Jun1-Jan31 and MH Feb1-May 31	the frequency of observations failing the criteria each day, week, and month; excluded: days with n<95 records, weeks with n<576 records (6 days), months with n<27 days of uninterrupted DO recordings	n/a

Figure 1. Tidal Potomac River "continuous monitoring" (CMON) stations, 2004-2008. Data for 3 of the 22 stations were not included in the analysis: the two stations in the District of Columbia and one in Neabsco Creek. Data for the two District of Columbia stations were available but had not been QA/QC'ed by the provider when the analysis was performed. Data for Neabsco Creek were only collected in the summer of 2006.

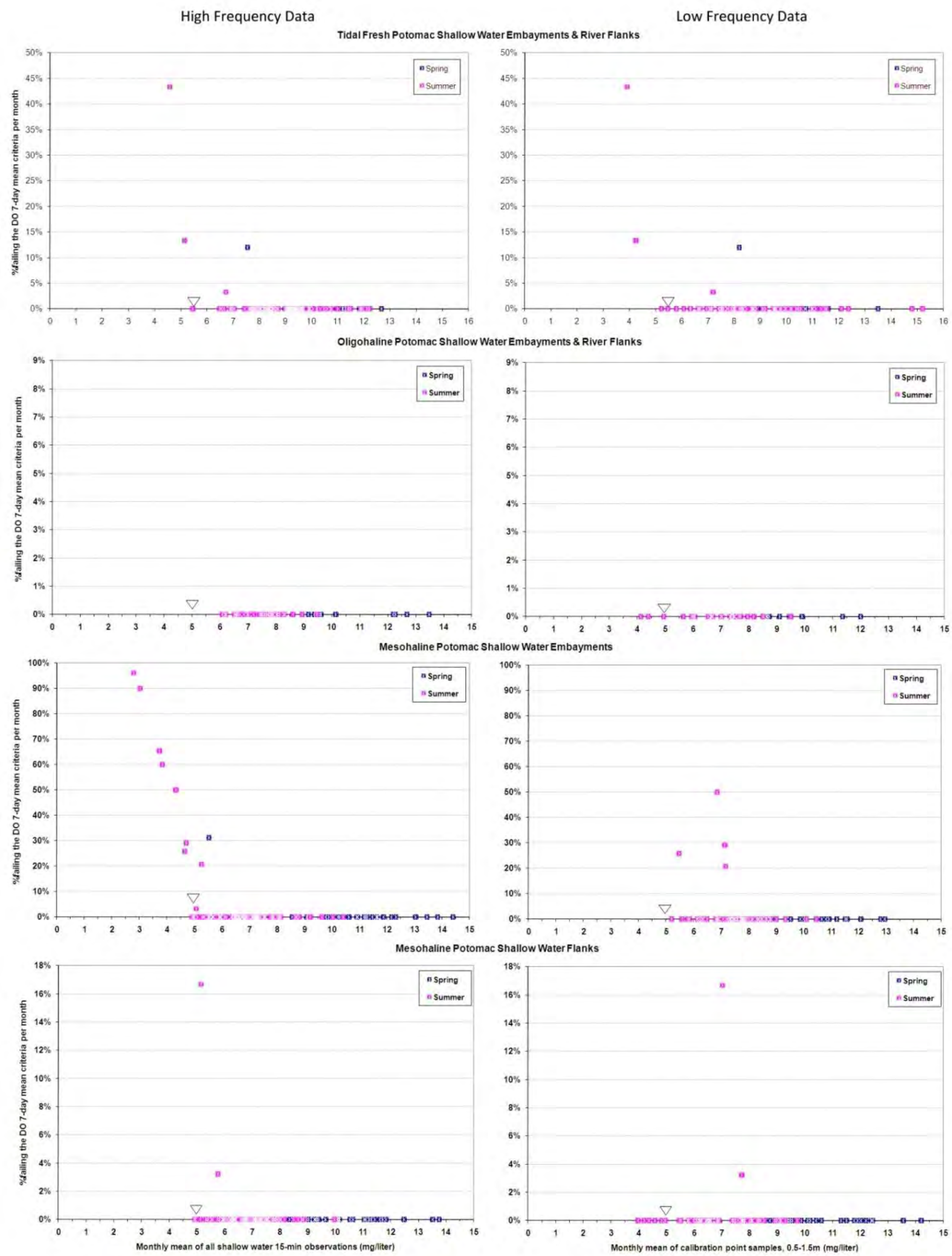


Figure 2. The frequency per month that rolling 7-day periods (1-day step) fail the 7-day mean DO criteria, plotted against the corresponding 30-day mean DO. Overall, 16 of the 415 months (3.86%) represented in the 20 tidal Potomac shallow water stations between 2004 and 2008 had one or more 7-day periods failing the 7-day mean criteria. Approximately half the failures occur in months where the 30-day mean criteria are met. Note: not all of the failing months are apparent in the low frequency data because point samples were not available for all months.

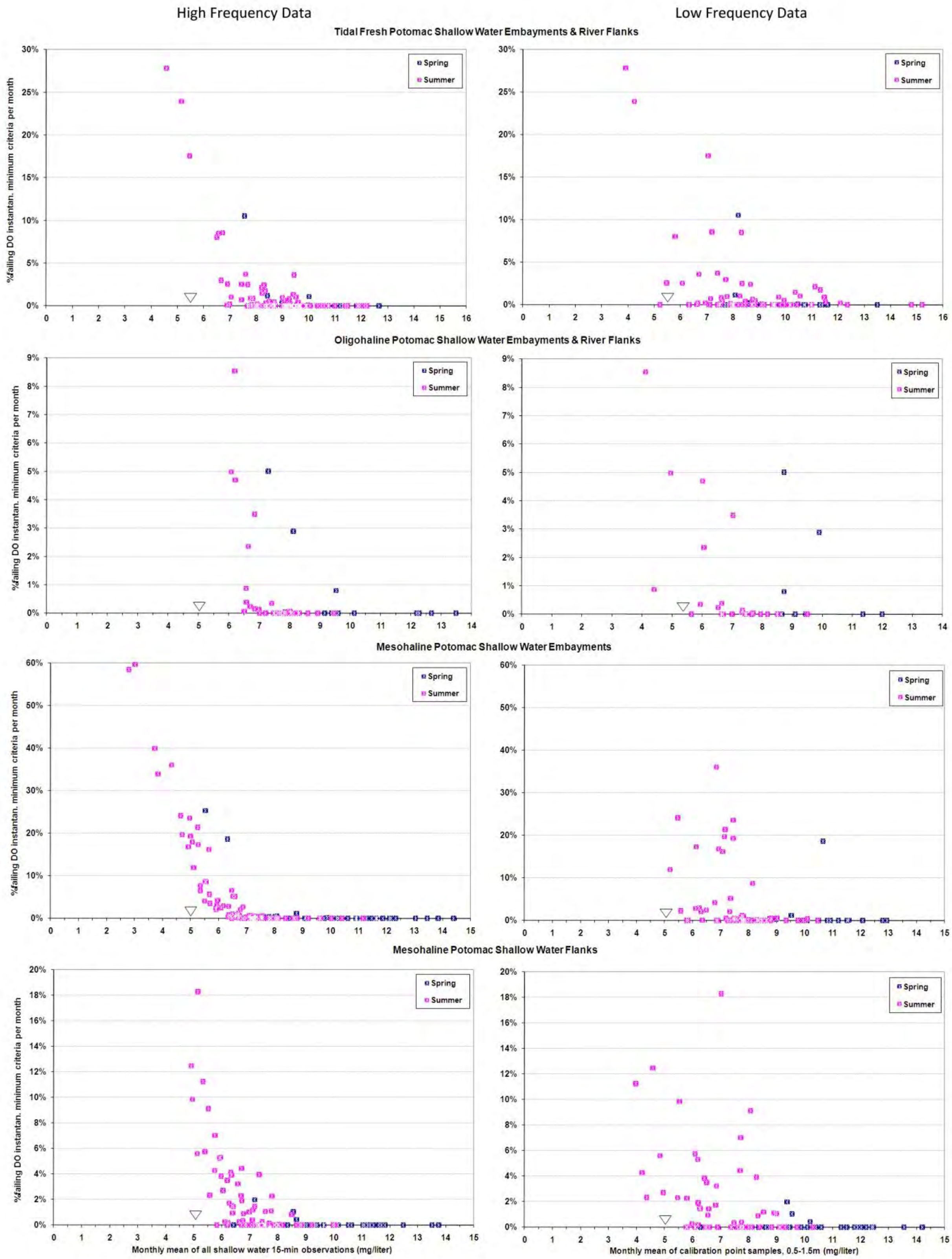


Figure 3. The frequency per month that rolling 7-day periods (1-day step) fail the 7-day mean DO criteria, plotted against the corresponding 30-day mean DO. Overall, 175 of the 415 months (42.2%) represented in the 20 tidal Potomac shallow water stations between 2004 and 2008 had failures of the instantaneous

minimum DO criteria. Most instantaneous minimum criteria failures occurred in months where the 30-day mean criteria are met.

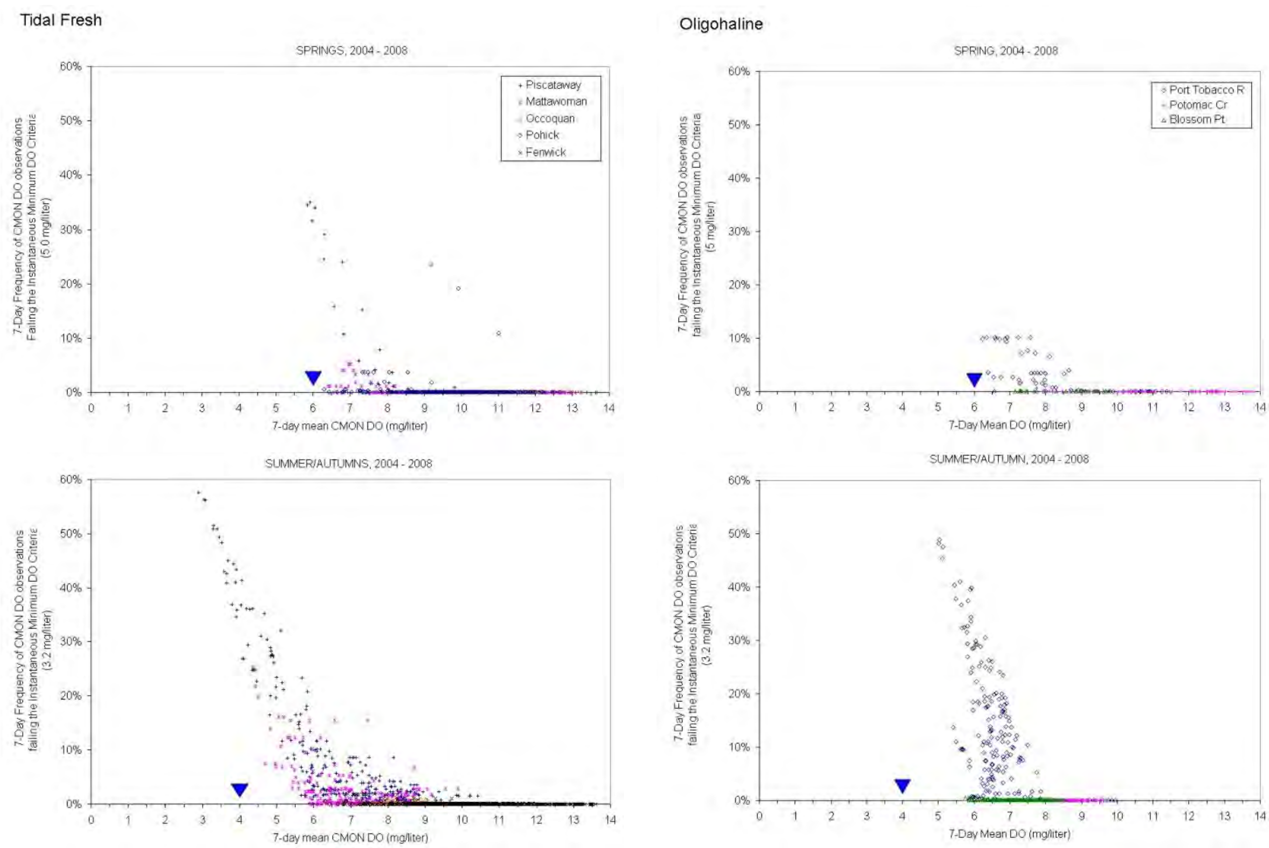


Figure 4. The frequency per 7-day period of failing the instantaneous minimum criteria, plotted against the corresponding 7-day mean in tidal fresh and oligohaline salinities. Frequencies were calculated on rolling 7-day periods (1-day step). Most instantaneous minimum criteria failures occurred in 7-day periods where the 7-day mean criteria are met.

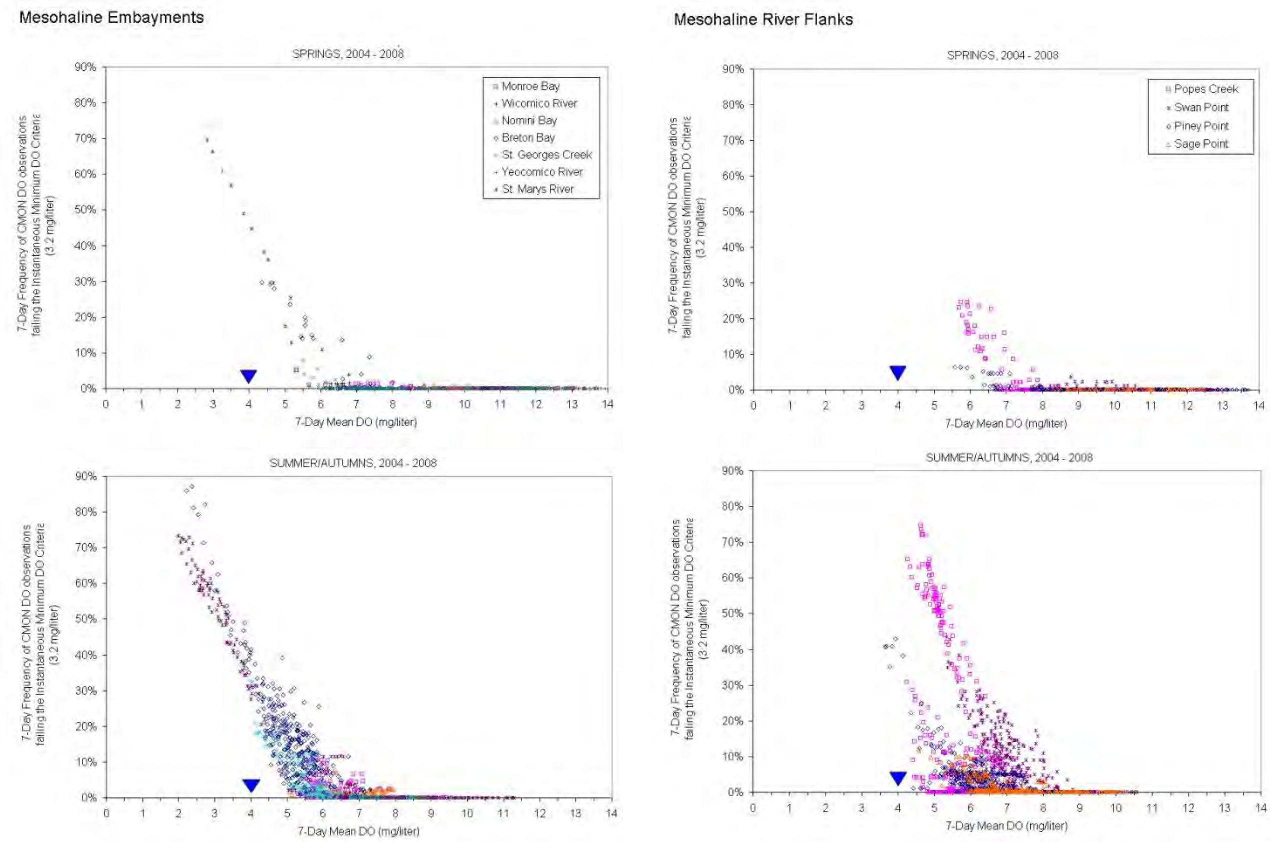


Figure 4 (cont.). The frequency per 7-day period of failing the instantaneous minimum criteria, plotted against the corresponding 7-day mean in mesohaline salinities. Frequencies were calculated on rolling 7-day periods (1-day step). Most instantaneous minimum criteria failures occurred in 7-day periods where the 7-day mean criteria are met.

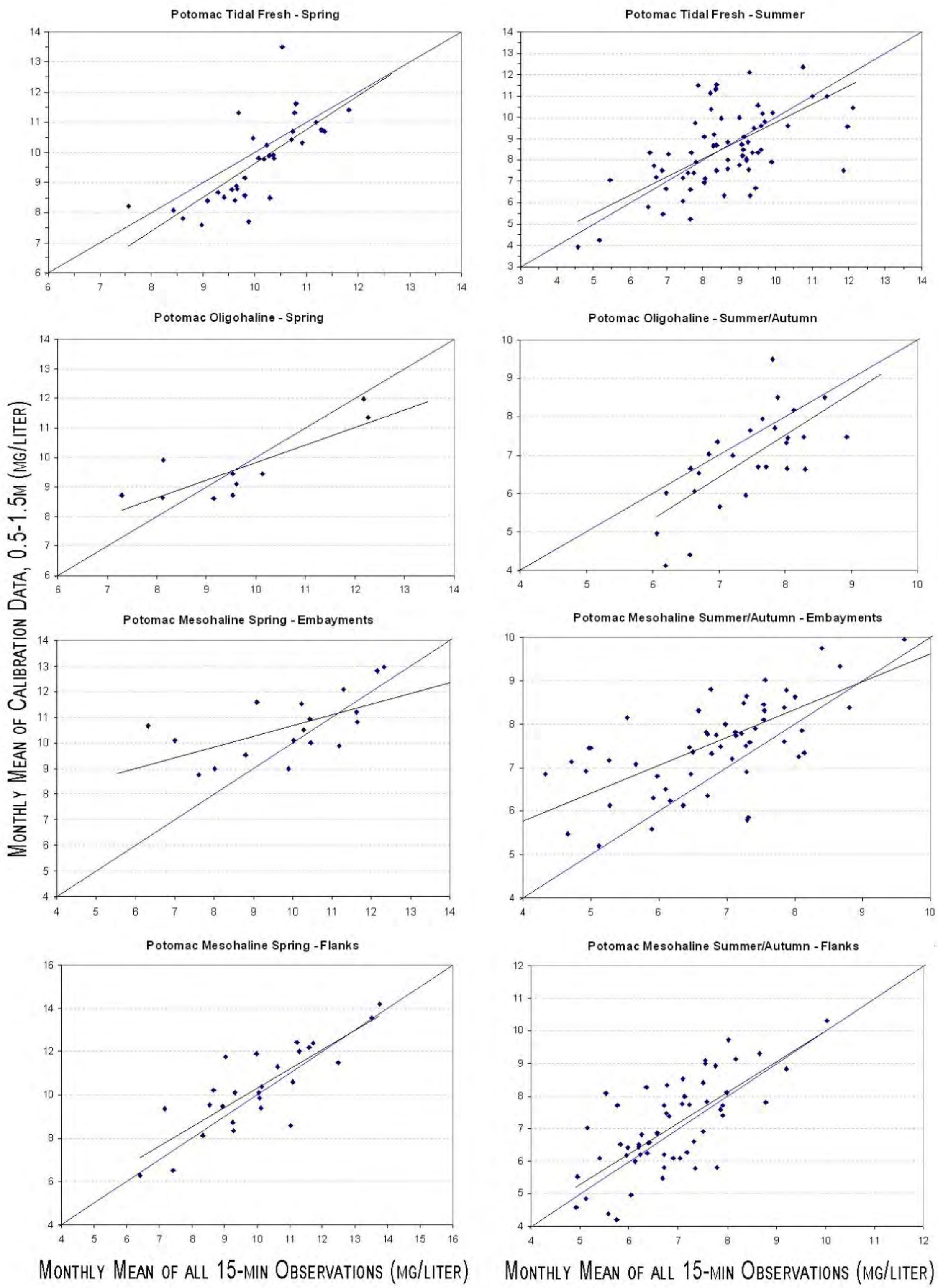


Figure 5. Paired comparison of the 30-day mean DO estimates calculated from low frequency (calibration) and high frequency (CMON) data. Blue line represents the 1:1 relationship, and it is assumed that the means derived from high frequency data (x-axis) are very close to the true mean. There are large differences between the paired 30-day means in all seasons and salinity zones, indicating inaccuracy on the

part of the means derived from the low frequency data. Further, low frequency means appear to be biased downward in TF spring and OH summer/autumn, and biased upward in MH embayments.

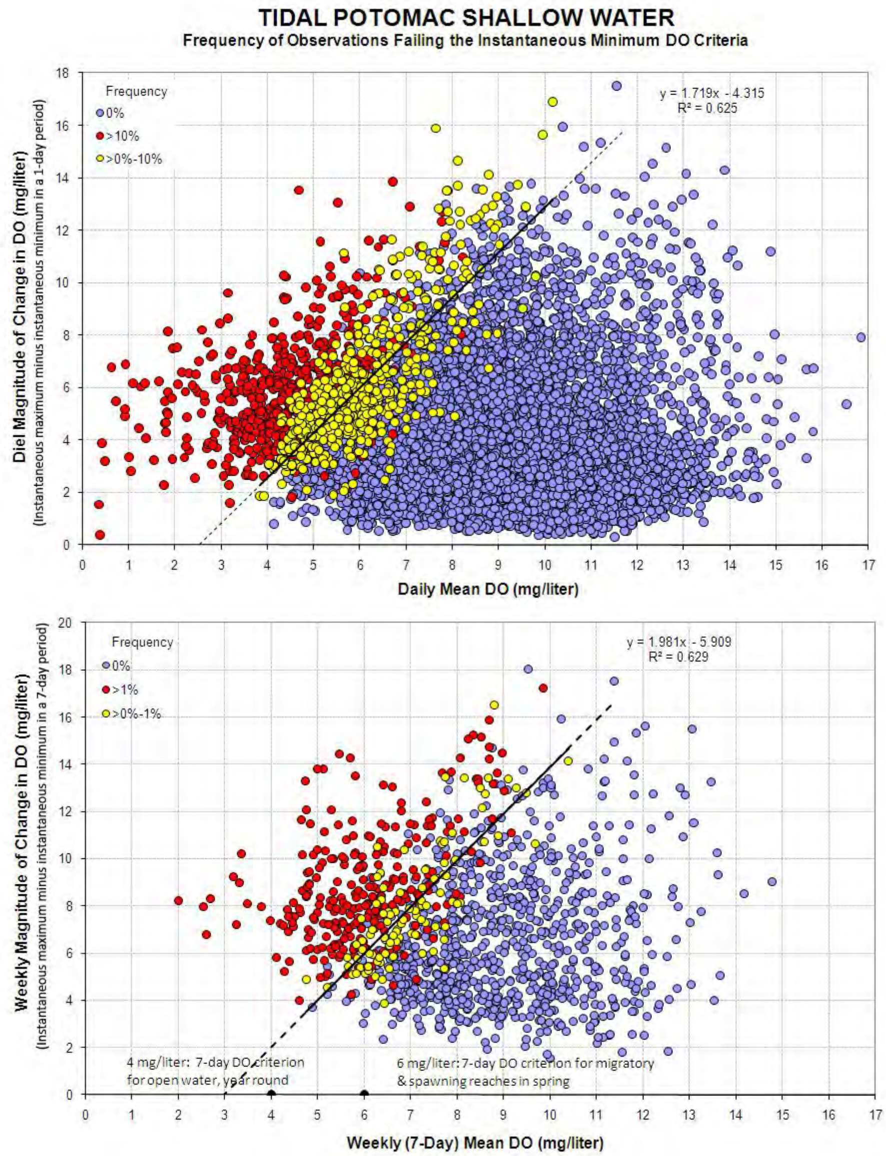


Figure 6. Attainment of the instantaneous minimum DO criteria as a function of daily (top) or weekly (bottom) mean DO and magnitude of change in DO. All tidal Potomac CMON data meeting the guidelines in Table 1 are included in each panel regardless of season or salinity zone. Top, n= 9,879 days; bottom, n= 668 weeks. The 7-day mean DO is calculated on sequential 7-day periods. The 7-day mean DO criteria presently applied to shallow waters are indicated on the x-axis in the bottom panel.

Appendix 1, Part B.

Characterizing shallow waters – C. Buchanan.

Daily mean dissolved oxygen concentrations experienced at individual shallow water sites range broadly over the course of a season (**Figure 1**). Daily means differ significantly from year-to-year and between neighboring sites. Between 2004 and 2008, daily means at the 20 tidal Potomac embayment and river flank stations fell between 1.0 – 16.8 mg/liter in spring, 0.36 – 14.9 mg/liter in summer, and 3.1 – 14.0 mg/liter in autumn. These ranges are broader than those experienced in the surface layer of the Bay's open water environments.

Shallow waters are more likely than open waters to exhibit large 24-hour (diel) fluctuations in DO concentrations. Their close proximity to oxygen-consuming benthic organisms and sediment process can rapidly drive down oxygen levels at night and the presence of up to three plant types (phytoplankton, SAV, benthic algae) can rapidly generate large amounts of oxygen during daylight hours. High frequency data show that the diel magnitudes of change in DO experienced in shallow waters sometimes reach as high as 11.0 mg/liter in spring, 17.52 mg/liter in summer and 10.8 mg/liter in fall (**Figure 1**).

Phytoplankton, expressed as water column chlorophyll a, and SAV are presently monitored in Potomac embayments and their populations can be related to diel magnitudes of change in DO. Sites with abundant SAV or high chlorophyll *a* levels (>50 µg/liter) exhibit some of the largest diel magnitudes of change in DO. Benthic algae are not monitored but their presence can be inferred at sites with low chlorophyll *a* concentrations and no SAV that are coincident with large diel magnitudes of change in DO and pH and moderate-to-high mean DO levels. External factors also control the diel magnitude of change in dissolved oxygen. Magnitude correlates strongly with temperature (**Figure 2A**) and significantly but not as strongly with solar angle (**Figure 2B**). Day-to-day differences in cloud cover and wind introduce additional variability in magnitude.

An evaluation of the high frequency data collected at the 20 tidal Potomac shallow water sites found that 19 of 53 site-year combinations in spring and 46 of 57 site-year combinations in summer/autumn failed the instantaneous minimum DO criteria to some degree (Buchanan 2009). The frequency of near-continuous DO observations failing the instantaneous minimum DO criteria ranged from 0.1% to 16% in spring and from 0.1% to 31.5% in summer and autumn. In comparison, only 10 of the 110 site-season-year combinations had one or more 7-day periods failing the 7-day mean DO criteria. Three occurred in spring and seven in summer/autumn.

- Under what conditions does the umbrella criteria assumption appear to be violated?

The umbrella criteria assumption that the 7-day mean DO criteria are protective of the instantaneous minimum DO criteria does not hold in shallow water conditions where productive communities are forcing large cyclic changes in dissolved oxygen. CMON data are demonstrating that nearshore sites often meet the 7-day mean DO criteria of 4 mg/liter (6 mg/liter in migratory spawning and nursery areas) while failing the instantaneous minimum DO criteria. This occurs because the magnitude of change in DO on both a daily and weekly basis is particularly large at sites with abundant plant communities. The frequency of failing the instantaneous minimum DO criteria in a 24-hour or 7-day period is a function of both the mean DO and the magnitude of change in DO (see **Figure 5, Appendix 1**). Abundant phytoplankton expressed as a high chlorophyll *a* concentration and abundant SAV are both associated with large diel magnitudes of change in DO and high failure rates of the instantaneous minimum criteria. There are also instances in summer where chlorophyll a levels are low and SAV are not present yet diel magnitudes of change in pH and DO are very large and criteria failures occur. This indicates the presence of abundant benthic algae, a plant group that is not presently monitored.

- Define conditions when one criteria is informative about the status of another.

Daily or weekly DO means of ~10 mg/liter are protective of the instantaneous minimum DO criteria in almost all circumstances. DO means below 10 mg/liter are only protective of the criteria if their magnitudes of cyclic change in DO (diel, weekly) are proportionately smaller, i.e. below the boundary lines indicated in **Figure 5, Appendix 1**.

Buchanan, C. 2009. An Analysis of Continuous Monitoring Data Collected in Tidal Potomac Embayments and River Flanks. ICPRB Report 09-3, 56 pgs. Available online at: http://www.potomacriver.org/cms/publicationspdf/ICPRB09-03.pdf

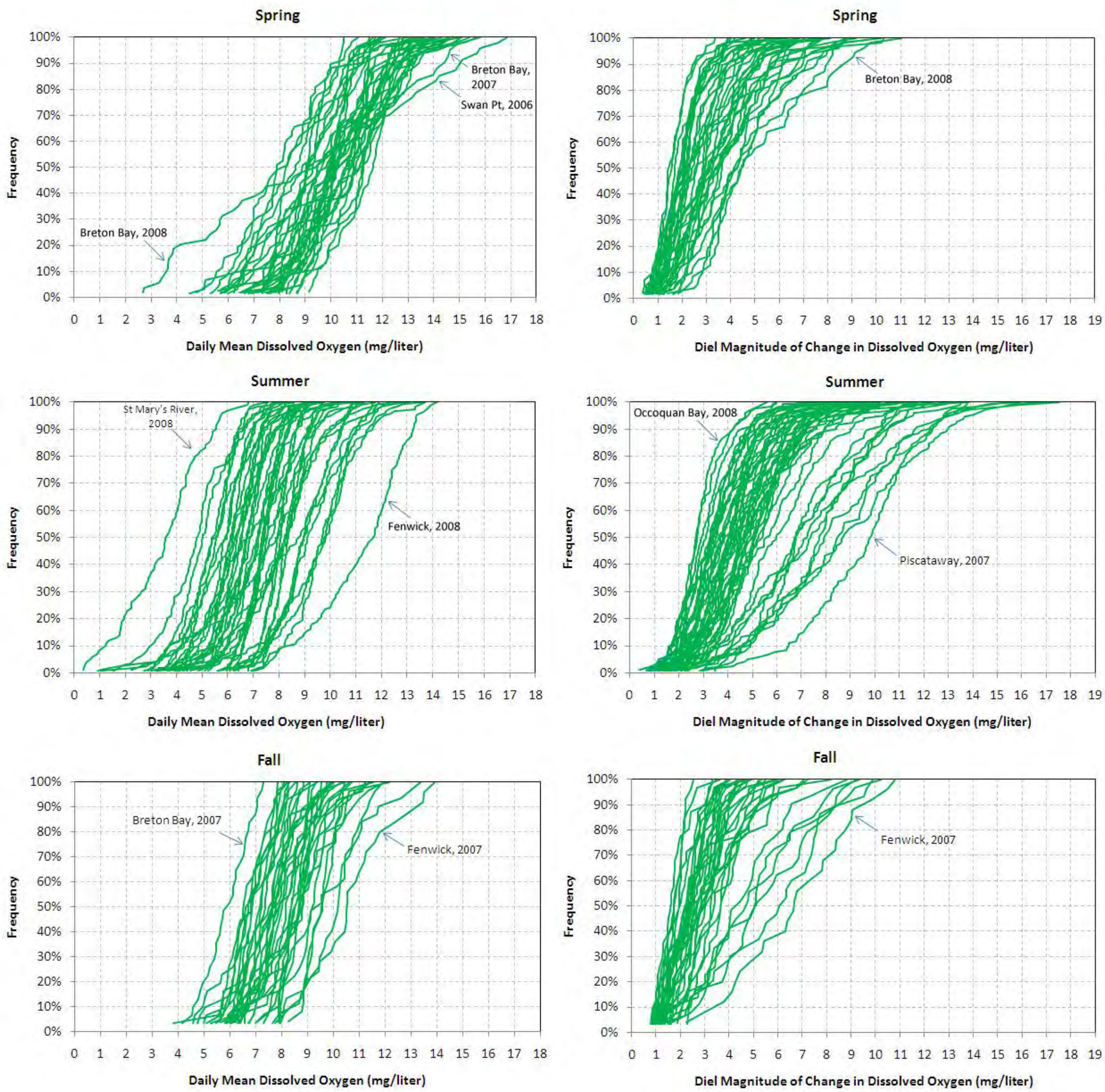
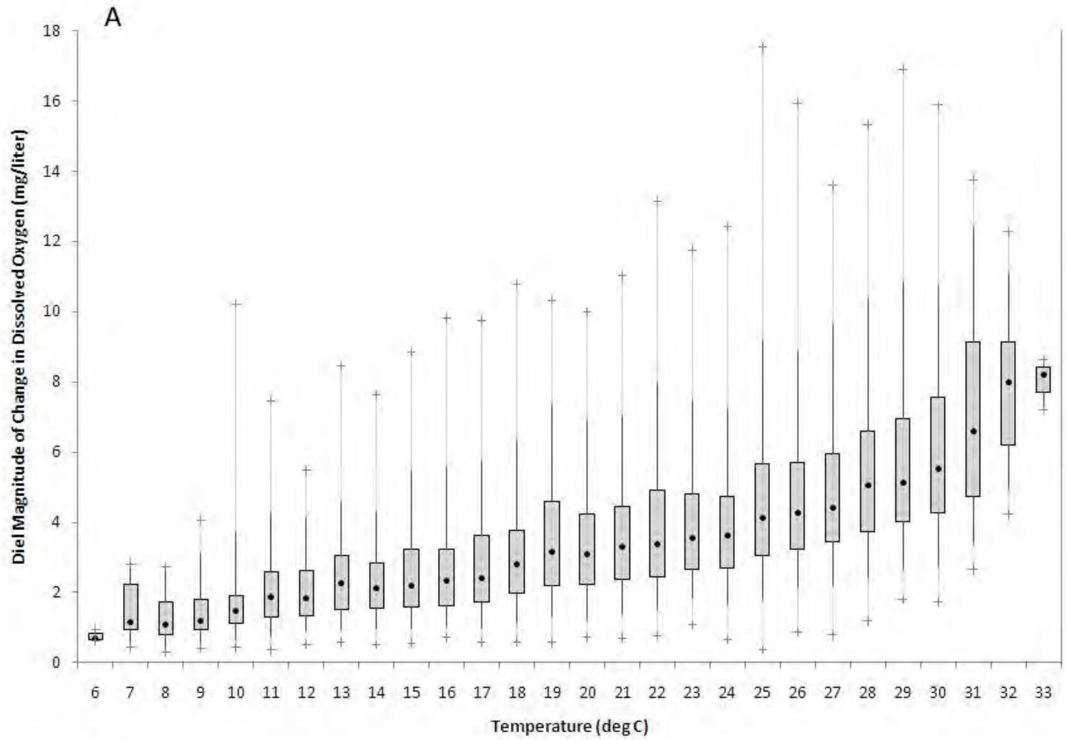


Figure 17. Seasonal cumulative frequencies of daily mean dissolved oxygen (left column) and diel magnitude of change in dissolved oxygen (right column) at shallow water stations in the tidal Potomac. Spring (April-May) n in each station-year >40 days; Summer (June-September) n in each station-year >90 days.



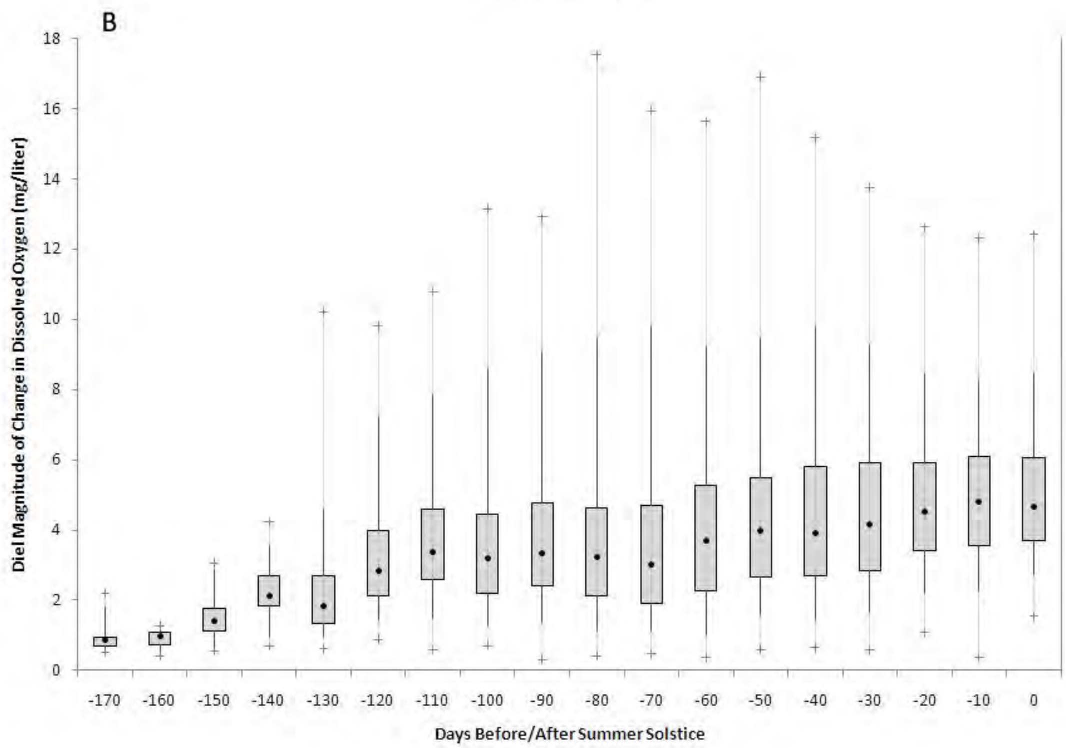


Figure 2. Diel magnitude of change in dissolved oxygen versus temperature (**A**) and solar angle, expressed as days before or after the summer solstice (**B**). The median (•), inter-quartile range (box), 95th%ile and 5th%ile (dark "whiskers") and maximum and minimum (+) are indicated.

Appendix 2
Conditional Probability

30-day mean vs. 7-day mean
Elgin Perry
8-27-2010

This note summarizes some additional analyses I have done with the Potomac ConMon data to address the question of whether the 30-day mean criterion serves as an umbrella for the 7-day mean criterion. The results here seem to confirm that it would be a rare situation where the 30-day mean would be satisfied and the 7-day mean would be violated more than 10% of the time. However, this does not seem to be a broad umbrella in that the margin of protection is not great.

Methods:

The method employed is based on the simple-minded approach (Figure 1.0) that if the variability of the 7-day mean about the 30-day mean has a standard deviation less that 0.7805, then we can expect that the 7-day criterion will be violated less than ten percent of the time if the 30-day criterion is met.

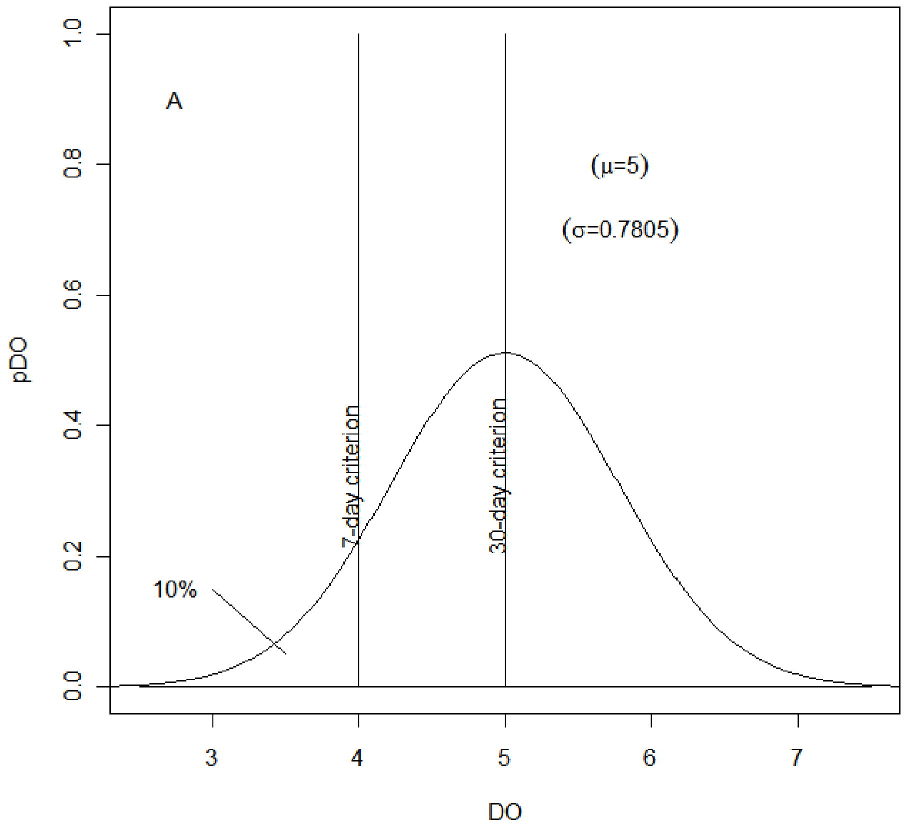


Figure 18. Illustration of the level of variability of the 7-day mean about the 30-day mean that results in up to 10 % violations of the 7-day mean criterion when the 30-day mean criterion is met.

To use this approach, an estimate of the standard deviation of the 7-day mean about the 30-day mean is needed. To estimate this quantity, I used data from the Potomac ConMon locations (Table 1., Figure 2.).

Table 1. Names, locations, and years of Continuous Monitor data used.

location	Latitude	Longitude	years
Occoquan	38.64038	-77.219416	2007-2009
Pohick Creek	38.67591	-77.16641	2007-2009
Potomac Creek	38.3436	-77.30485	2007-2009
Monroe Bay	38.23197	-76.96372	2007-2009
Nomini Bay	38.1316	-76.71759	2007-2009
Yeocomico River	38.02878	-76.55184	2007-2009
Fenwick	38.66993333	-77.11513333	2004-2008
Piscataway Creek	38.70156667	-77.02593333	2004-2008
Mattawoman Creek	38.55925	-77.1887	2004-2008

Figure 19. Locations of Potomac ConMon data collection sites used for this analysis.

Beginning with the first collection day for each year at each location, blocks of 30 days were created to represent months. Partial months at the end of each collection year were counted as a month. Similarly, weeks were created by starting with the first collection day of each year and counting off blocks of 7 days. With these definitions, monthly means were computed as the arithmetic average of DO for each month. Weekly means were computed as the arithmetic average of DO for the intersection of month and week. Thus a week that bridges across two months would have it's data divided by month and a weekly mean computed for each part. Weekly means and Monthly means were merged by month and a residual computed by subtracting the monthly mean from each weekly mean computed within that month. Various analyses were conducted on these residuals to assess the variability of weekly means about the monthly mean.

Graphical analyses were used to assess the uniformity of variation over other factors. Distribution functions and quantile estimation was used to estimate the rate of violation of the 7-day criterion given that the 30-day criterion was satisfied.

Results:

First I report a number of graphical assessments:

The basic distributional assessment of the residuals (Figure 3.) shows that they are reasonably symmetric and centered about zero . The distributions is heavy tailed compared to the normal distribution in the extreme tails suggesting that there are weeks that have a greater deviation (both high and low) than would be expected for a normal distribution. The central part of the distribution seems to follow the normal distribution closely.

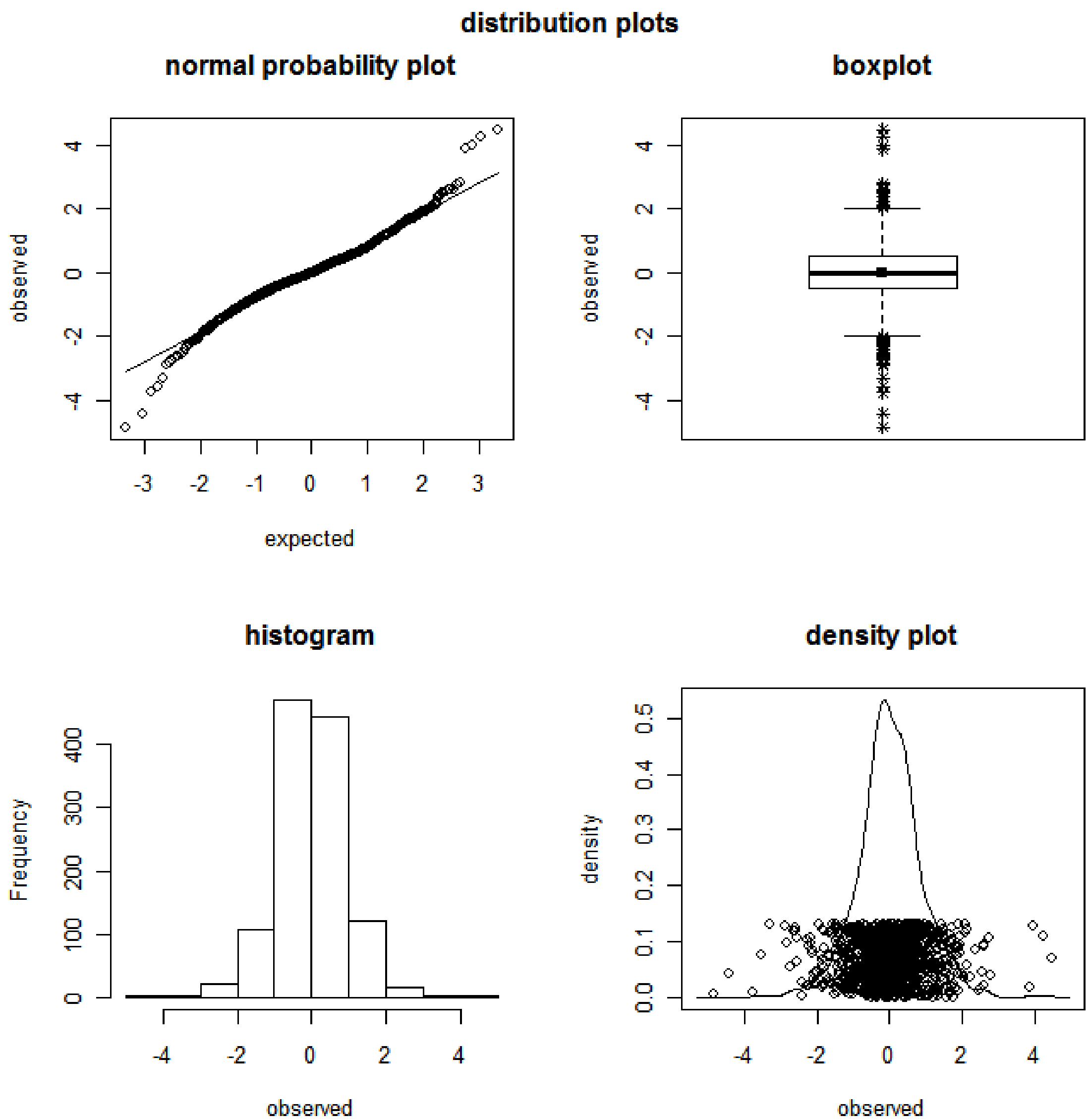


Figure 20. Basic distributional properties of the residuals.

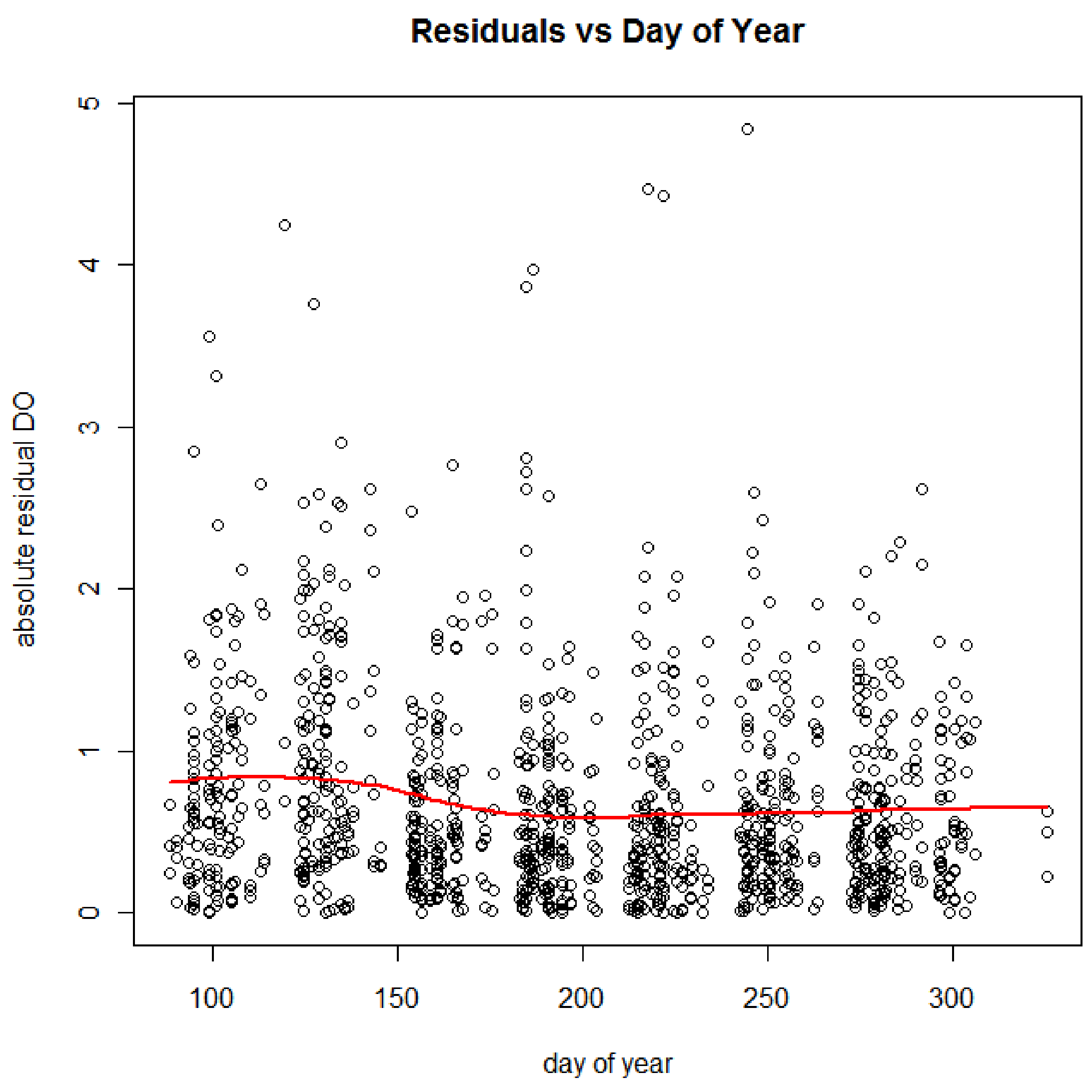


Figure 21. Assessment of seasonal trend in variability.

There is evidence of higher variability in spring than in summer and fall (Figure 4).

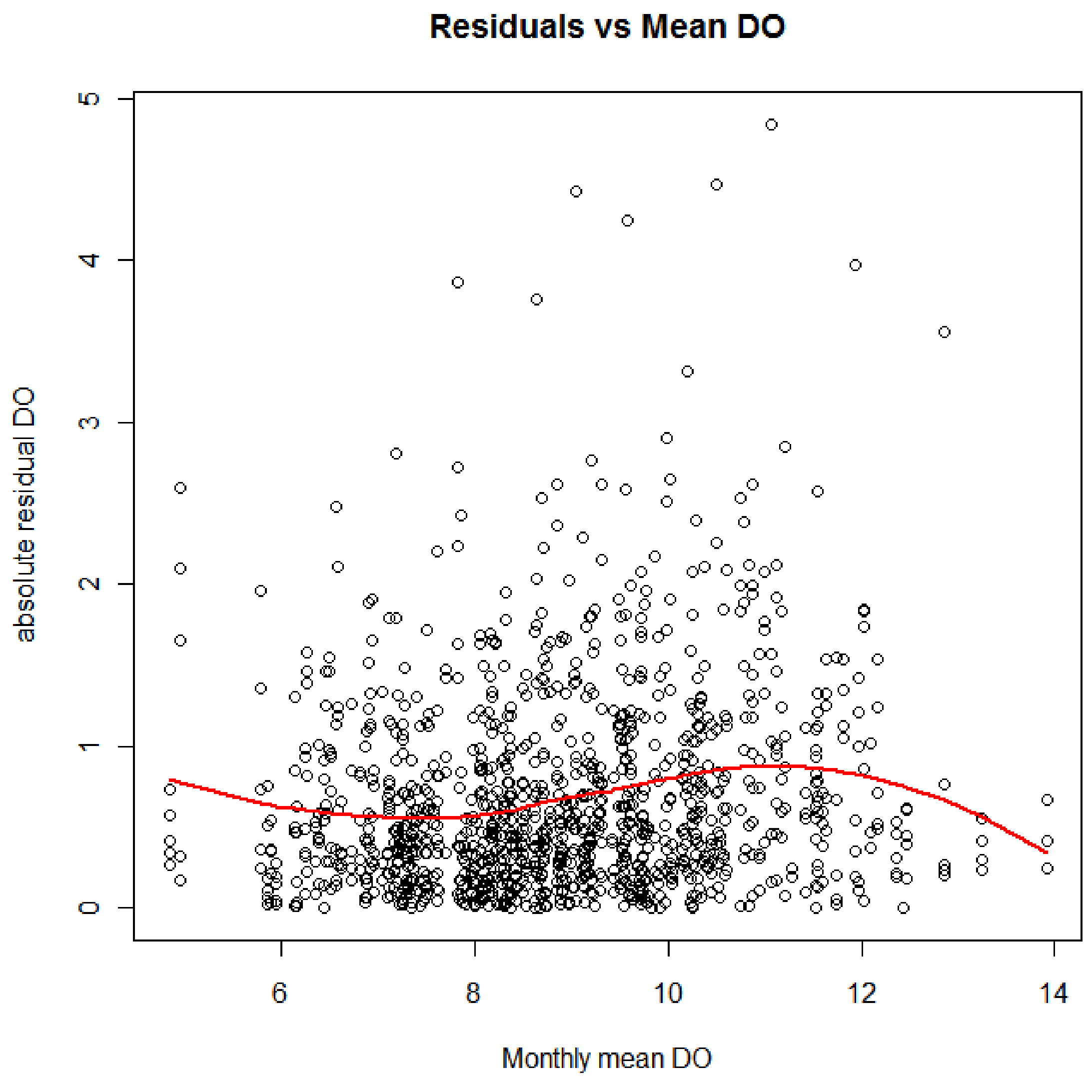


Figure 22. Trend of variability with mean DO.

There is evidence that the variability increases with increasing mean DO (Figure 5.). It is likely that the seasonal trend and the trend with the Mean DO are the same trend because there is a seasonal trend in mean DO.

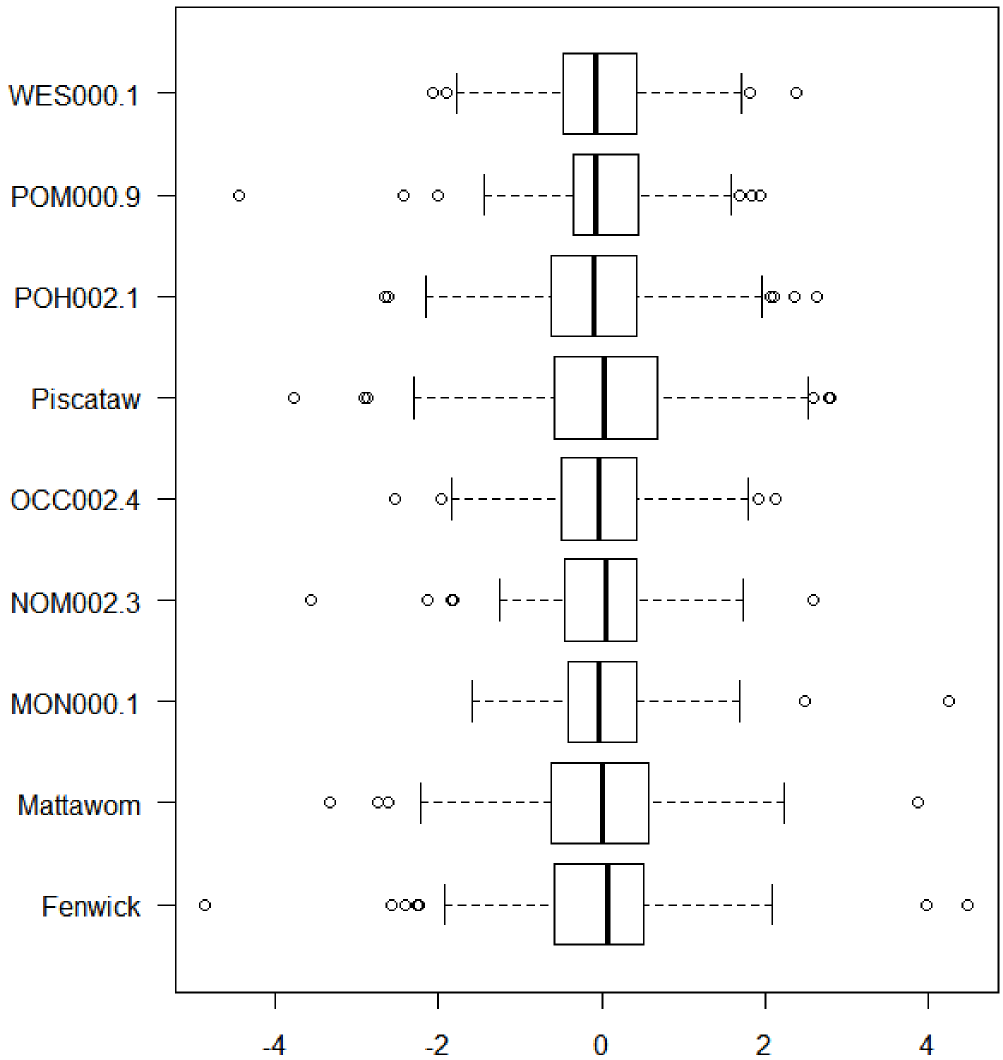


Figure 23. Box and Whisker plots of residuals by sampling location.

There is little evidence of change in variability with location except that Maryland locations appear to have greater variability than Virginia locations.

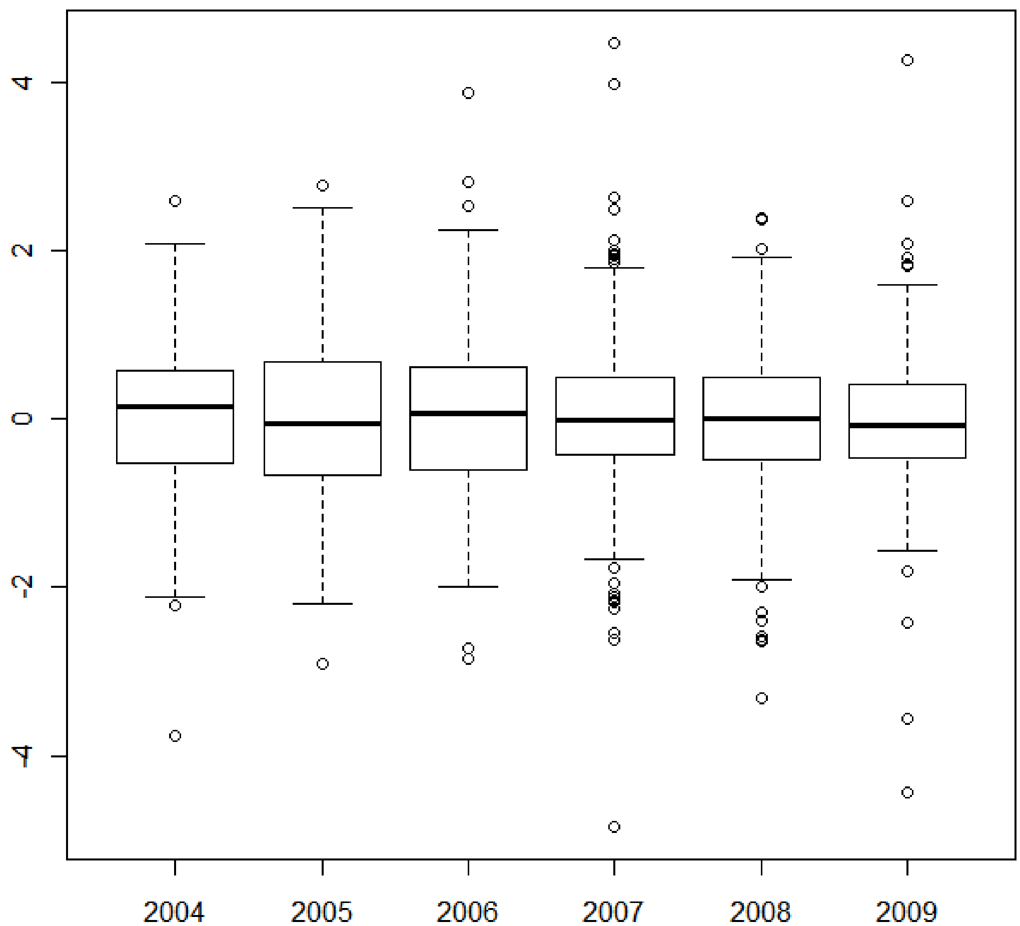


Figure 24. Box and Whisker plots of residuals by year.

There appears to be a time trend in variability (Figure 7.) with a decrease in variability occurring in between 2006 and 2007. However, recall from the collection dates (table 1) that only Maryland collected data prior to 2007 and thus this is the state trend from a different view. The pattern of difference in variability between states persists when the data are subsetted to just 2007-2009 for Mattawoman and Fenwick, but variability at Piscataway is more comparable to Va. locations for this time period (figure not shown).

If the standard deviation of weekly mean about monthly mean is estimated for all data, the value is 0.9648719 which is slightly above the value of 0.7805 which would insure that there would be less than 10% violation of the 7 day criterion is the 30-day criterion were satisfied (Table 2., column 2). However, if the 30-day mean is increased to just 5.3, then we would expect fewer than 10% violation of the 7-day mean. Thus it seems that if the 30-day mean were hovering between 5.0 and 5.3 for an extended period, then there might be greater than 10% violation of the 7-day criterion when the 30-day criterion is satisfied. This circumstance would seem to be a rare event.

Recall that there is evidence that variability increases with the 30-day mean DO (Figure 5.). It is reasonable to exclude the variability associated with high DO because when the 30-day mean DO is high, then it is not hovering in that region close to the criterion which we would expect to also observe violations of the 7-day criterion. To exclude the variability associated with high DO, a subset of the data was created that included only weeks associated the 30-day mean DO of less than 8.0 (the minimum value for the 30-day mean is 4.848). Using this subset of the data, the standard deviation was estimated as 0.8439. This remains slightly larger than the 0.7805 which would insure that the 30-day criterion is an umbrella for the 7-day criterion, but with this, the 30-day mean DO need only be greater than 5.1 to insure fewer than 10% violation of the 7-day criterion (Table 2., column 3).

Note that in low salinity waters where the 30-day criterion is 5.5, then we would expect only 6% or 4% violations of the 7-day criterion for the two estimates of standard deviation (Table 2, row 6). Thus it seems reasonable to conclude that the 30-day mean is an effective umbrella for low salinity. These probability estimates have been made using an assumption that the weekly residuals are approximately normally distributed.

Table 2. Probability of violating 7-day mean criterion as a function of 30-day mean DO for two levels of 7-day mean variability.

30-day mean DO	Prob(|sd=0.9649)	Prob(|sd=0.8439)
5.0	0.1500	0.1180
5.1	0.1271	0.0962
5.2	0.1068	0.0775
5.3	0.0889	0.0617
5.4	0.0734	0.0486
5.5	0.0600	0.0377
5.6	0.0486	0.0290
5.7	0.0390	0.0220
5.8	0.0311	0.0165
5.9	0.0245	0.0122
6.0	0.0191	0.0089

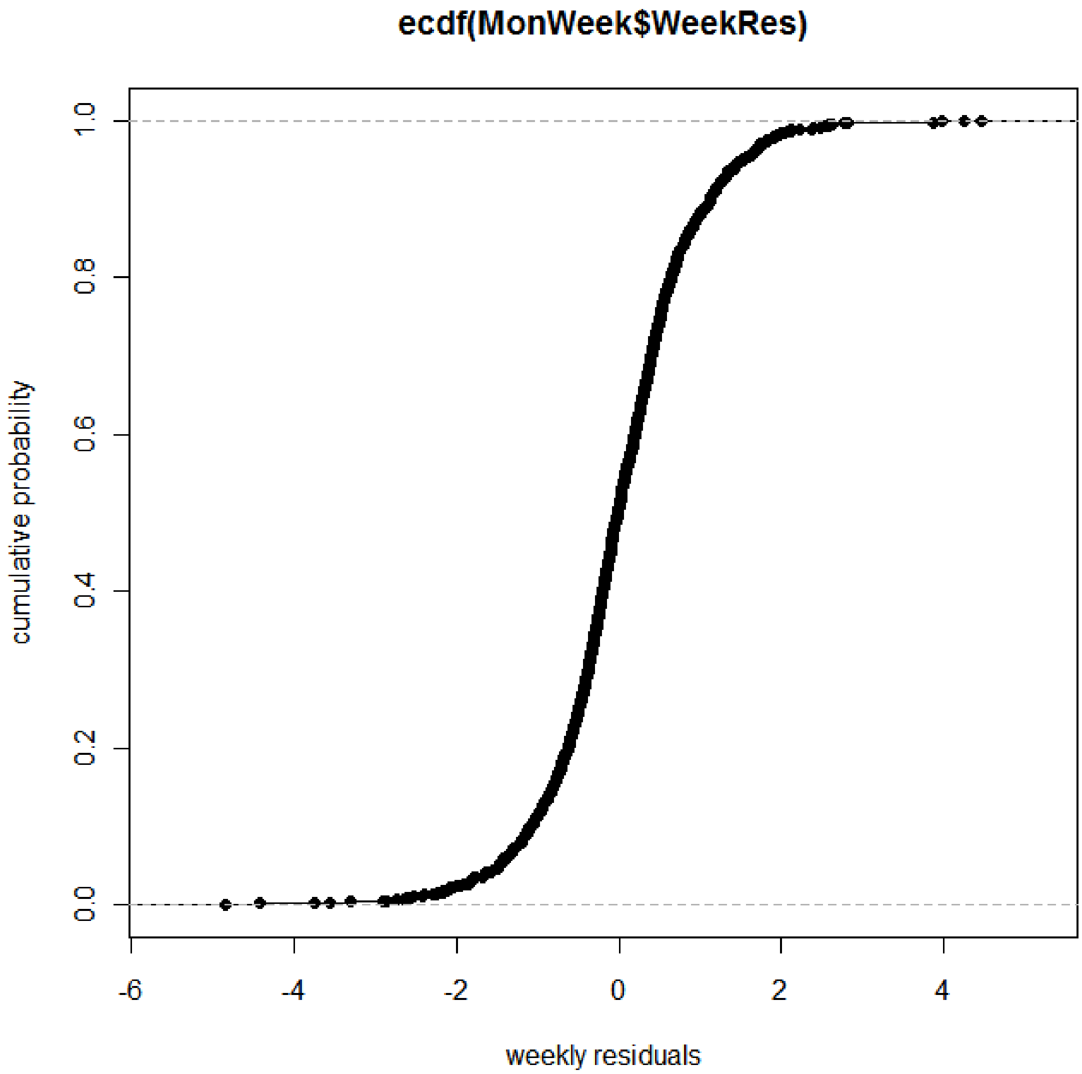


Figure 25. Empirical Cumulative Distribution Function (ECDF) of the set of weekly residuals about the monthly mean.

As a check on the reliability of the normality assumption, we compare the important quantiles of the normal distribution to those of the empirical cumulative distribution function (Figure 8.). The 10th percentile of a normal distribution with mean zero and standard deviation = **0.9649** is
-1.236533. The 10th percentile of the empirical cumulative distribution function is -1.076785. This suggest that the observed deviations at the 10th percentile are less than predicted by the normal distribution which implies that the probabilities in Table 2 are somewhat higher than might be realized. The difference between the 30-day criterion and the 7-day criterion is -1.0 and a quantile of -1.0 corresponds to between 11 and 12 percent (Table 3., column 2) based on the ECDF. This suggests that if the 30-day mean criterion were satisfied exactly, there would be on average 11-12 percent violations of the

7-day criterion. Of course typically the 30-day mean would be satisfied by some margin and if that margin were as little as 0.14 then on average there would be fewer than ten percent violations of the 7-day criterion.

Table 3. Percentiles and corresponding quantiles from the empirical cumulative distribution functions for all weekly residuals (column 2) and weekly residuals given that the 30-day mean is less than 8.0.

percentile	quantile all Weekly residuals	quantile given mean DO < 8
13%	-0.9160	-0.7321
12%	-0.9574	-0.7545
11%	-1.0329	-0.8156
10%	-1.0768	-0.8448
9%	-1.1404	-0.9186
8%	-1.2011	-0.9633
7%	-1.3006	-1.0411
6%	-1.3921	-1.1210
5%	-1.4802	-1.1903

These same calculations are done using the ECDF of weekly residuals given that the 30-day mean is less then 8.0. These results (Table 3, column 3) suggest that if the 30-day criterion is satisfied exactly, there would be on average 7-8 % violations of the 7-day criterion.

Conclusion:

These results suggest that we would only see greater than 10% violations of the 7-day criterion given that they 30-day criterion is met if the 30-day mean were hovering at or just above the 30-day criterion. Because the 30-day mean rarely exhibits this behavior, it seems safe to conclude that in most cases the 30-day criterion acts as an umbrella for the 7-day criterion. However, 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.

Appendix 3
Dissolved Oxygen (DO) Criteria Attainment Analysis for Shallow Water Habitats Using ConMon Data Sets

Boynton, W.R., E.M. Bailey, L.A. Wainger and A.F. Drohan. 2010. Dissolved oxygen (DO) criteria attainment analysis for shallow water habitats using ConMon data sets. In: Ecosystem Processes Component (EPC). Maryland Chesapeake Bay Water Quality Monitoring Program, Level 1 report No. 27. Jul. 1984 - Dec. 2009. Ref. No. [UMCES] CBL 10-098. [UMCES Technical Series No. TS-606-10-CBL].

W.R. Boynton, E.M. Bailey, L.A. Wainger and A.F. Drohan

3.0 Introduction

Until the last decade, water quality monitoring in Chesapeake Bay and tributary rivers was largely based on monthly or bi-monthly sampling at fixed stations located over the deeper (channel) portions of these systems. Such a design had many benefits, especially those related to developing seasonal to inter-annual scale indices of water quality status and trends. However, as in virtually all environmental science activities, a single measurement scheme is not adequate for addressing all questions. Thus, about a decade ago, a new program was initiated, first on a pilot-scale basis, to add measurements of water quality for shallow near-shore habitats. Concern for SAV habitat quality was a prime consideration in developing this program.

The program was named ConMon to indicate the near-Continuous Monitoring feature of this activity. The program used in-situ sensor systems (YSI Sondes) programmed to take measurements of a suite of water quality variables every 15 minutes. Included in the water quality suite was water temperature, salinity, pH, DO, turbidity and chlorophyll-a. In most instances ConMon sites were active from April – October (the SAV growing season) and in most cases sites remained active for three years. In a few cases, sites have remained active for up to 9 years, thus serving as long-term or sentinel sites. To place this sampling intensity in perspective, at a typical main channel site about 16 measurements of water quality variables were collected per year. In contrast, at a ConMon site about 20,500 measurements per year are obtained, an intensity of measurement about three orders of magnitude higher than traditional monitoring.

There have been about 60 sites in the Maryland Bay and Maryland Coastal Bays where ConMon data have been collected. The program is continuing although at somewhat fewer sites than in the recent past. The considerable spatial extent (encompassing sites with water quality varying from quite good to very poor) of these data sets allows for comparative analyses wherein it is likely that relationships between near-shore water quality and management actions can be found.

There are several prime uses of ConMon data sets. First, they have been used as a guide in selecting and monitoring SAV habitat restoration sites. Second, these data have "opened our eyes" to a new scale of hypoxia, namely diel-scale hypoxia wherein DO concentrations can reach critically low levels at night (and especially in the immediate post-dawn hours). Third, these data can be used to make estimates of community production and respiration, both of which are fundamental ecosystem features known to be related to nutrient loading rates. Fourth, these data can be used in DO criteria assessments for shallow open water sites (EPA 2010).

It is the last ConMon use that is the focus of this chapter. In an earlier portion of this report the strategy and details of making DO criteria assessments using ConMon data have been described. In this section we provide examples of DO criteria % non-attainment for three sites in the Bay system. It remains unclear as to which of several approaches best captures meaningful DO non-attainment; we present results of all approached in this section. There is a STAC-sponsored DO criteria workshop scheduled for the fall of 2010 and we will participate in this workshop and hopefully arrive at a consensus approach for computing this metric.

3.1 Methods

Continuous monitoring data was obtained from Maryland Department of Natural Resources Tidewater Ecosystems Assessment division in electronic (.txt file) format (dnr_cmon_sonde_2001-08). This file contained all the collected ConMon data from 2001 to 2008. A SAS® (www.sas.com) program was written to allow selection of dissolved oxygen data by station and year. The program then calculated 6 different methods/averages (Table 3-1) and gave each occurrence of dissolved oxygen (instantaneous or averaged) a score of 1 if lower than the criteria and a score of 0 if equal to or above (based on Chesapeake Bay Program guidelines and discussions with MDDNR and TWMAW). Criteria were chosen prior to selecting specific stations and we selected the higher DO value to make this analysis more "conservative."

Table 3-1. Calculation methods, file names, descriptions and criteria used for DO criteria % non-attainment calculations.

Calculation Method	SAS Filename	Description	DO Criteria
Instantaneous	doyyyST_allcrit	Uses every available data point (~every 15 minutes per annual data set).	4 mg L^{-1}
Daily Mean	doyyyST_daycrit	Takes the mean DO for all measurements over the course of 24 hours. No data point is reused in the calculation.	4 mg L^{-1}
7 Day Moving Average	doyyyyST_wkcrit	Takes the mean DO for a 7 day chunk of data moving forward 15 minutes for each iteration. Measurements are reused.	4 mg L^{-1}
1 Average per 7 Days	doyyyyST_1perwk	Takes a mean DO for all measurements over the course of 7 days. No data point is reused in the calculation.	4 mg L^{-1}
30 Day Moving Average	doyyyyST_moncrit	Takes the mean DO for a 30 day chunk of data moving forward 15 minutes for each iteration. Measurements are reused.	5 mg L^{-1}
1 Average per 30 Days	doyyyyST_1pmo	Takes a mean DO for all measurements over the course of 30 days. No data point is reused in the calculation.	5 mg L^{-1}

Exact criteria values will be refined in FY2011 in consultation with MDDNR for each specific station and month. SAS output files were named DO(dissolved oxygen), yyyy (year), ST (two-letter station code), underscore followed by an identifier for the calculation method used. Percent non-attainment was calculated as: (sum of the non-attainment score)/(total # of observations) * 100. Percent non-attainment was calculated for the entire available annual dataset, June-August and July.

3.2 Results from Selected Sites

Estimates of DO % non-attainment have been developed for three sites in the Bay system. The first site was St George's Island (XBF7904), located in a small embayment of the lower portion of the Potomac River estuary. This site was chosen for initial analysis because water quality at this site is relatively good compared to many other Maryland tributary sites. Water quality here was good enough for this site to be selected for SAV restoration work. ConMon data are available for this site for the years 2006-2008. The second site selected was Sycamore Point (XHH3851), located in the upper portion of the Corsica River estuary. Multi-year monitoring of this site indicates poor to very poor water quality and there are indications from Dataflow mapping that some water quality conditions have been deteriorating further in recent years. Data for the years 2005-2008 were available for this analysis. The third site was the Fort McHenry site (XIE5748) located in the Patapsco River estuary, adjacent to the city of Baltimore, MD. This site was selected because it is an urban site and because there is a considerable ConMon record available from this site.

3.2.1 Low Impact Site (St. George's Island, Lower Potomac River: XBF7904)

Results of DO % non-attainment are summarized for the St George's Island site (2006-2008) in Table 3-1 and Figures 3-1 to 3-3. For each year, 6 different averaging schemes were employed; these have been described earlier in this chapter. Across the top of Table 4-2 a simple average DO concentration was calculated for three time periods, including:

1) the whole year (Jan-Dec); 2) summer period (Jun – Aug) ; and 3) just the month of July. Further to the right in Table 1 DO % non-attainments were calculated for each time period using all 6 averaging schemes. Several patterns are readily evident.

First, % non-attainment consistently increases with smaller time period evaluations. For example, during 2006, the “All Data” computation indicated 4% non-attainment for the whole year evaluation, 8% for the summer evaluation and 10% for the July evaluation. At this site, the July evaluation for all % non-attainment approaches was the highest and this was also true for all three years evaluated. It is interesting to note that hypoxia/anoxia in the mainstem Bay reaches a maximum in July of most years since the monitoring program began in 1985. It may be that the single most critical water quality month is July in most years. Further analysis will clarify this issue.

Second, it is not completely clear which of the averaging techniques provides the most sensitive metric of DO non-attainment. For data collected during 2006 and 2007 it appears that the “All Data” approach detected more non-attainments than any other approach (i.e., it was the most protective). However, during 2008 the same pattern did not emerge. In fact, some counter-intuitive results emerged. The highest July % non-attainment emerged from the 30 day moving average approach, a considerably larger % non-attainment than that obtained from all other approaches, including the “All Data” approach. The fact that the 30 day average had a higher criteria threshold (5 mg/l vs 4 mg/l for other averaging schemes) probably played into this result. Based on results from this single site, it appears that the 7-day moving average and the 1 average per 30 days did not detect DO non-attainment as frequently as did other averaging schemes.

Another way of visualizing these computations is shown as a sequence of three box and whisker plots (Figures 3-1–3; 2006, 2007 and 2008, respectively). In these figures data for the entire annual ConMon data set were included (whole year). What is clear in these diagrams is that the mean of the full data set were always above criteria thresholds (5 and 4 mg L^{-1}). However, instances of non-attainment were most frequently observed using the “all Data”, daily mean and, to a lesser extent, the 7-day moving average approaches. The final three computation methods detected no criteria violations during 2006 (Figure 3-1), only a few during 2007 (Figure 3-2) and a few more during 2008 (Figure 3-3), the year with the poorest water quality.

Table 3-2. A summary of DO % non-attainment estimates from the St George's Island ConMon site for the period 2006-2008. The various methods of computing % DO non-attainment were described in the methods section of this chapter. The "whole year" columns used data for the period April-October. Other calculation periods are as indicated in the table.

St. George's Island (XBF7904)

Year	Method	Available Annual Dataset Mean	June through August Mean	July Mean	Available Annual Dataset % Non-Attainment	June through August % Non-Attainment	July % Non-Attainment
2006	Instantaneous	6.69	5.78	5.68	4	8	10
	Daily Mean				1	2	3
	7 Day Moving Average (15 min. increment)				0	0	0
	1 Average per 7 Days				0	0	0
	30 Day Moving Average (15 min. increment)				0	0	0
	1 Average per 30 Days				0	0	0
2007	Instantaneous	7.05	5.73	5.35	5	9	17
	Daily Mean				2	4	13
	7 Day Moving Average (15 min. increment)				0	0	0
	1 Average per 7 Days				0	0	0
	30 Day Moving Average (15 min. increment)				0	0	0
	1 Average per 30 Days				0	0	0
2008	Instantaneous	7.11	5.33	5.07	10	21	27
	Daily Mean				4	9	17
	7 Day Moving Average (15 min. increment)				1	1	4
	1 Average per 7 Days				4	8	25
	30 Day Moving Average (15 min. increment)				12	25	40
	1 Average per 30 Days				0	0	0

Perhaps the strongest “take-home” messages from analyses at this site is that DO criteria violations occur even at sites with relatively good water quality and that substantial inter-annual variability exists relative to DO non-attainments…some years are clearly better than others. To a large degree this finding is consistent with findings using the historical Cory data set collected from 1964-1969 in the Patuxent River estuary.

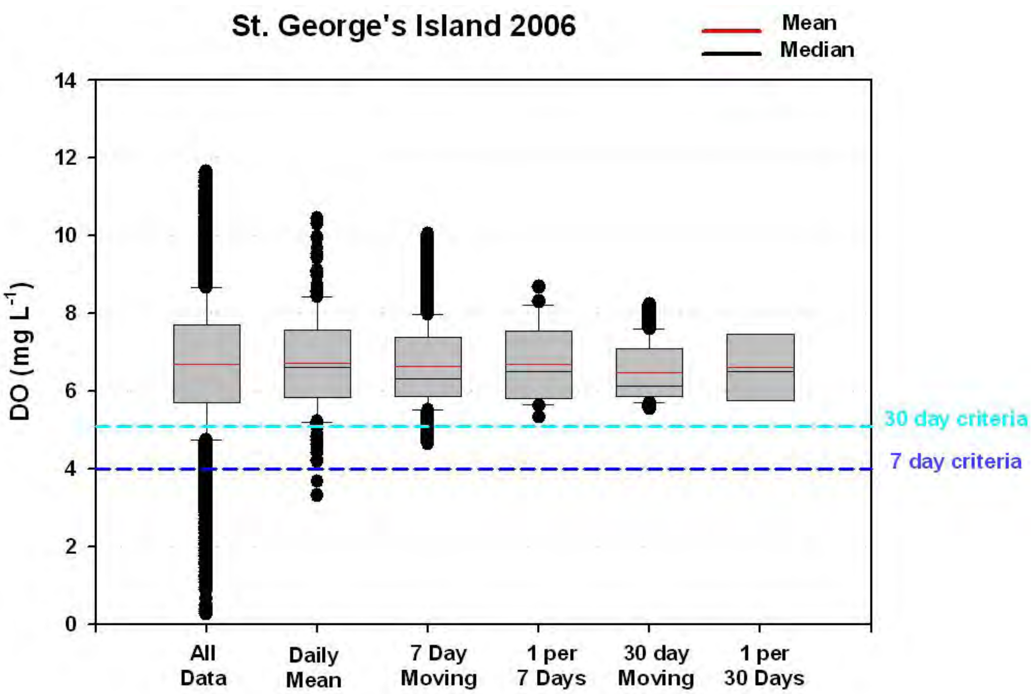


Figure 3-1.

Box and whisker plots of DO concentration based on data collected at the St. George’s Island ConMon site in the lower Potomac River estuary during 2006. The categories indicated on the x-axis were described in the Method section of this chapter. The two horizontal lines indicate DO criteria concentrations for open water sites.

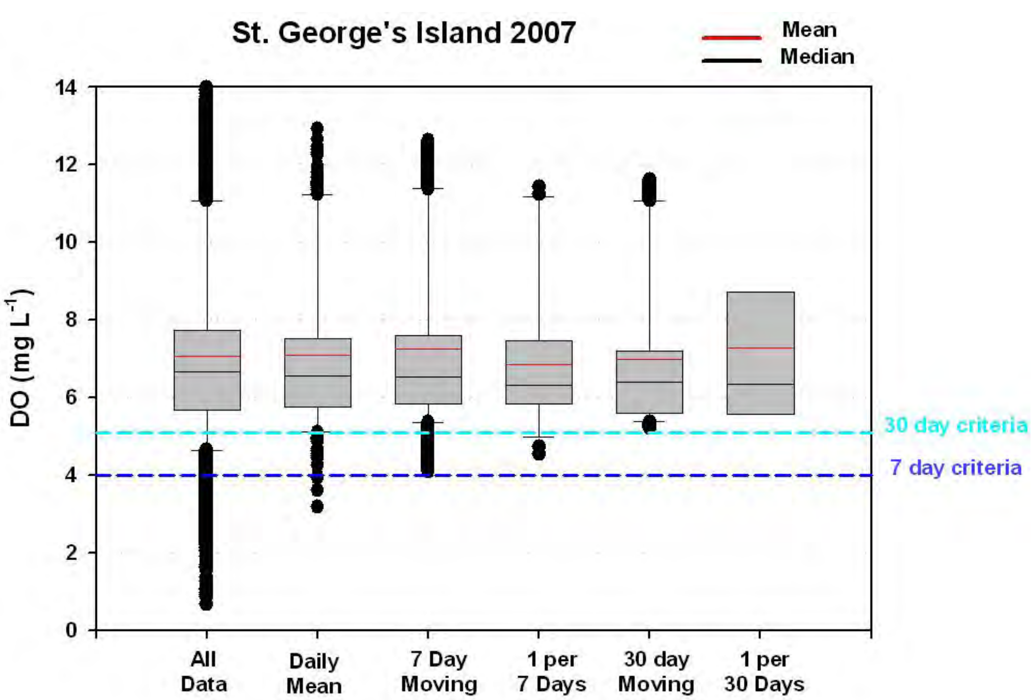


Figure 3-2.

Box and whisker plots of DO concentration based on data collected at the St. George’s Island ConMon site in the lower Potomac River estuary during 2007.

The categories indicated on the x-axis were described in the Method section of this chapter. The two horizontal lines indicate

DO criteria concentrations for open water sites.

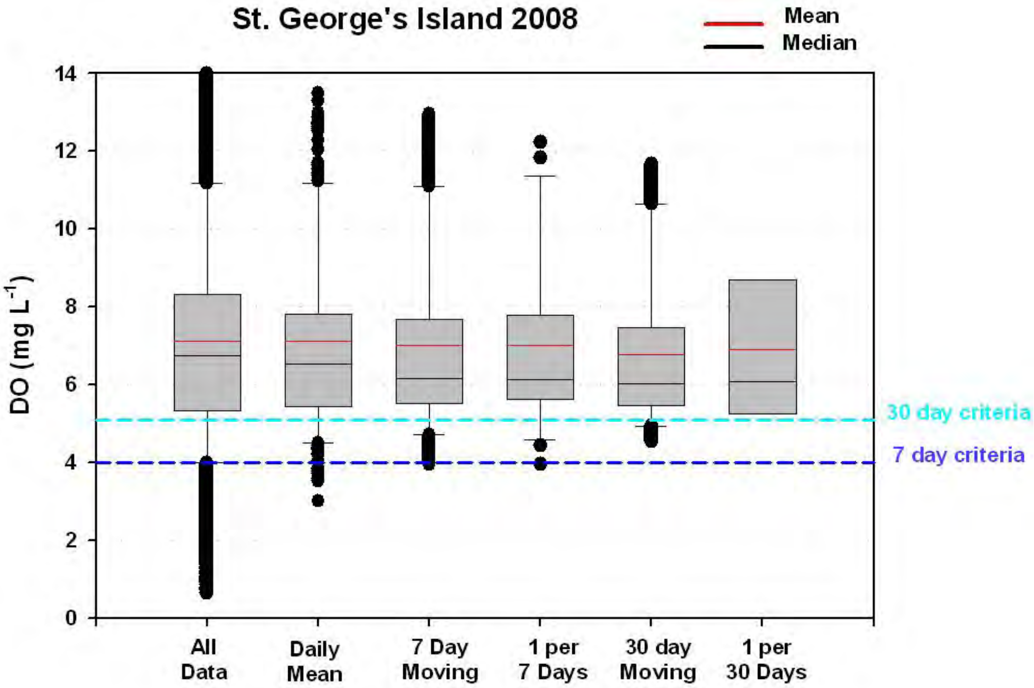


Figure 3-3. Box and whisker plots of DO concentration based on data collected at the St. George's Island ConMon site in the lower Potomac River estuary during 2008. The categories indicated on the x-axis were described in the Method section of this chapter. The two horizontal lines indicate DO criteria concentrations for open water sites.

3.2.2. High Impact Site (Sycamore Point, Upper Corsica River: XHH3851)

The Sycamore Point site in the upper portion of the Corsica River estuary is heavily impacted by nutrient additions, mainly from the agriculturally dominated watershed (Boynton *et al.* 2009). Results from % DO non-attainment for this site are summarized in Table 4-3. Several important points emerge. First, there were far higher % non-attainment rates observed at this site than at the St. George's Island site, as expected. The St. George's Island site is relatively "clean" compared the Sycamore Point site. In addition, the Sycamore Point site has far higher % non-attainment results than found in the historical data from the Cory ConMon site operated in the 1960s. Thus, it appears that there is considerable range in results consistent with our general impressions of water quality.

As at the previous site, there was not a clear result concerning the metric that might be adopted for general use in criteria attainment or non-attainment. For example, the All Data and the Daily Mean approaches tended to detect the highest failure rates. But, this was not always the case. During 2006 both the 30 day moving average and the 1 average per 30 days produced failure rates higher than the previously mentioned metrics. It may well be that the differences in criteria threshold values (4 versus 5 mg O_2 L^{-1}) were that cause of this result. However, data from both 2005 and 2008 do not support this conclusion.

Table 3-3. A summary of DO % non-attainment estimates from the Corsica River, Sycamore Point (XHH3851) ConMon site for the period 2005-2008. The various methods of computing % DO non-attainment were described in the methods section of this chapter. The "whole year" columns used data for the period April-December. Other calculation periods are as indicated in the table.

Sycamore Point (XHH3851)

Year	Method	Available Annual Dataset Mean	June through August Mean	July Mean	Available Annual Dataset % Non-Attainment	June through August % Non-Attainment	July % Non-Attainment
2005	Instantaneous	8.05	5.55	5.51	16	36	39
	Daily Mean				12	25	32
	7 Day Moving Average (15 min. increment)				3	11	16
	1 Average per 7 Days				3	11	0
	30 Day Moving Average (15 min. increment)				3	8	22
	1 Average per 30 Days				0	0	0
2006	Instantaneous	9.10	4.96	5.40	12	37	27
	Daily Mean				8	28	14
	7 Day Moving Average (15 min. increment)				10	36	17
	1 Average per 7 Days				6	29	0
	30 Day Moving Average (15 min. increment)				13	45	29
	1 Average per 30 Days				11	100	ND
2007	Instantaneous	8.57	4.93	5.76	16	47	35
	Daily Mean				13	41	30
	7 Day Moving Average (15 min. increment)				9	29	23
	1 Average per 7 Days				3	9	0
	30 Day Moving Average (15 min. increment)				18	56	6
	1 Average per 30 Days				25	100	ND
2008	Instantaneous	10.03	5.95	5.22	10	29	39
	Daily Mean				6	16	29
	7 Day Moving Average (15 min. increment)				1	4	12
	1 Average per 7 Days				0	0	0
	30 Day Moving Average (15 min. increment)				0	0	0
	1 Average per 30 Days				0	0	0

The time span considered in these evaluations also needs consideration. Without exception, the "Whole Year" computations of % non-attainment were lowest and therefore likely the least protective. When compared to the June-August % non-attainment rates the whole year rates were 2 to 3 times less frequent. However, July alone non-attainment rates were not always higher than those computed from a longer summer period (June – August). We had originally suspected that the July alone computations would yield the highest % non-attainment rates because investigations of hypoxia in deeper waters indicates this month to consistently have the most severe hypoxia. That turns out not to be the case. Of the 24 comparisons that can be made (6 computation schemes for each year and four years of data), 13 times % non-attainment was greater using the June-August data set while on 7 occasions the July only data set yielded higher % non-attainment results (4 cases of zero non-attainment were not included).

3.2.2 Urban Site (Fort McHenry, Patapsco River: XIE5748)

A summary of DO % non-attainment at the urban, Ft. McHenry site is presented in Table 3-4. Here again, results tended to follow many of the patterns seen at the others sites. First, there was substantial inter-annual variability. During 2004 the maximum DO % non-attainment was detected using the instantaneous metric (23%) and four of the remaining five metrics detected no failing DO conditions. During 2007, the instantaneous DO % non-attainment rate was much larger for all time periods (24-39%) and some small failure rates were found with the other metrics. Finally, it is now reasonably clear simple averages (left portion of table; pink background) are not sufficient to detect DO % non-attainment rates. At these relatively shallow sites (<2 m) DO variations on a daily basis can be severe because, in part, the effects of sediment respiration can be large and result in strong DO depressions, especially during the late night and early morning hours. The instantaneous metric appears to capture these events at this site better than any of the other metrics.

Table 3-4. A summary of DO % non-attainment estimates from the Fort McHenry (XIE5748)

Fort Mc Henry (XIE5748)

Year	Method	Available Annual Data Set Mean	June through August Mean	July Mean	Available Annual Data Set % Non-Attainment	June through August % Non-Attainment	July % Non-Attainment
2004	Instantaneous	7.09	6.17	5.65	13	18	23
	Daily Mean				6	10	13
	7 Day Moving Average (15 min. increment)				0	0	0
	1 Average per 7 Days				0	0	0
	30 Day Moving Average (15 min. increment)				0	0	0
	1 Average per 30 Days				0	0	0
2007	Instantaneous	6.85	5.44	5.52	24	39	34
	Daily Mean				18	30	29
	7 Day Moving Average (15 min. increment)				7	7	0
	1 Average per 7 Days				10	9	0
	30 Day Moving Average (15 min. increment)				1	2	0
	1 Average per 30 Days				0	0	0

ConMon site in the Patapsco River for the period 2004 and 2007. The various methods of computing % DO non-attainment were described in the methods section of this chapter. The "whole year" columns used data for the period April-November. Other calculation periods are as indicated in the table.

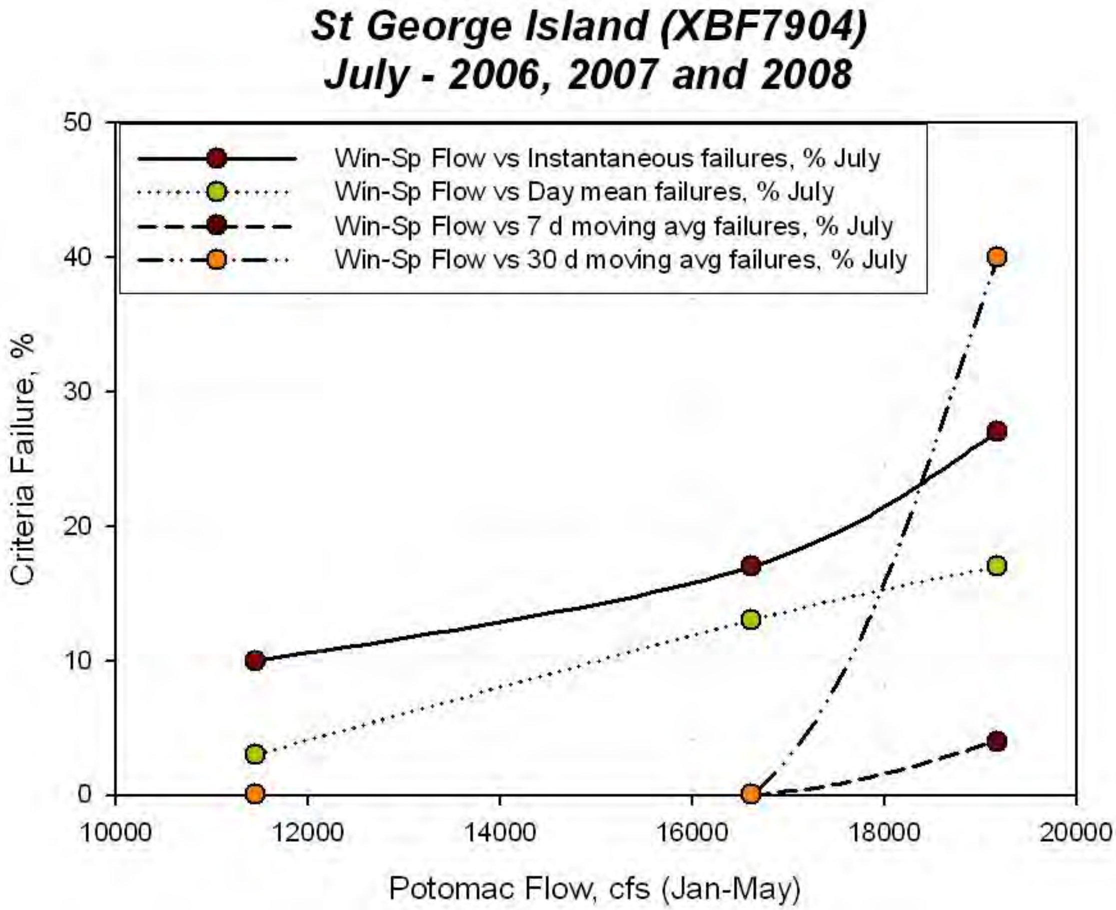
3.3 Relating DO Criteria % Non-Attainment to Other Water Quality Variables

One major goal of this work is to simply compute rates of % DO criteria non-attainment for shallow areas of the open water zone. As with many ecological issues, this one turns out to be not so simple. There are a variety of ways to compute this metric and it remains to be seen which might be the most appropriate method. There is also the issue of merging the DO criteria assessment associated with ConMon based data sets collected in shallow waters relative to open water assessments made with the traditional, low frequency monitoring data. It remains unclear as to just how this will be accomplished.

Finally, since there are not ConMon sites at all locations in the Bay and tributary rivers it would be useful to have some simple water quality variable(s) that could be used as a surrogate for data collected at a ConMon site. It would also be useful to link, in some quantitative fashion, % DO non-attainment results to other ecosystem features to explain the apparent large degree of inter-annual variability observed at some stations.

We are at early stages of this effort. However, data collected at the St George's Island ConMon site can serve as an example of future, and more thorough, efforts to link criteria results with management actions and general understanding. The % DO non-attainment results (developed using 4 different approaches) computed from 2006-2008 ConMon data were plotted as a function of Potomac River flow (Figure 3-4). In this analysis, two metrics of % DO non-attainment increased in a near-linear fashion as a function of river flow. Two other DO % non-attainment metrics remained very low until river flow was quite high at which point one increased slightly while the other exhibited a very large increase, threshold-like in nature. In this simple case the conceptual model supporting this analysis is based on the fact that river flow adds both freshwater (and buoyancy) as well as sediments and nutrients to these systems. Nutrients, in turn, tend to support higher rates of primary production. Organic matter resulting from this nutrient-stimulated production can cause increased respiration rates (utilization of DO) by the heterotrophic community. The net result, in this example, would be higher DO% non-attainment rates. We expect to continue this effort using a variety of water quality variables in addition to freshwater flow and nutrient loading rates. Variables such as TN, TP and chlorophyll-a concentration will be considered in an effort to better understand and predict levels of inter-annual variability of DO % non-attainment rates.

Figure 3-4. 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.



3.4 Current and Future Plans and Activities

During the next few months we will develop DO % non-attainment criteria metrics for other Bay ConMon sites with particular emphasis on sites representing a range of eutrophication intensity. In addition, we will examine some sites because there is a long ConMon record (up to 9 years) in an effort to better understand inter-annual variability. We will also consider water quality data (e.g., nutrient and chlorophyll-a concentrations) collected as part of the calibration activities at ConMon sites. We will consider the use of nutrient loading rates as an explanatory variable but issues remain relative to the most effective way to compute these loading rates for a variety of locations. Finally, there is the need to focus on a smaller selection of methods for computing DO % non-attainment metrics. We expect to have some clarification of this issue following the STAC-sponsored workshop on criteria attainment methods scheduled for the early fall of 2010.

3.5 References

Boynton, W.R., J.M. Testa and W.M. Kemp. 2009. An Ecological Assessment of the Corsica River Estuary and Watershed: Scientific Advice for Future Water Quality Management: Final Report to Maryland Department of Natural Resources. Ref. No. [UMCES]CBL 09-117. [UMCES Technical Series No. TS-587-09-CBL].

USEPA. 2010. Ambient water quality criteria for dissolved oxygen, water clarity and chlorophyll-a for the Chesapeake Bay and its tidal tributaries: 2010 technical support for criteria assessment protocols addendum. EPA 903-R-10_002 CBP/TRS 301-10. May 2010.

Appendix 4

Shallow Water, High Frequency Measurements and the 30 Day Mean Umbrella Approach: Two Preliminary Computations

W. R. Boynton, E. M. Bailey, M. Hall and E. Perry

Background: The traditional water quality monitoring program, in place since 1984, obtains water quality measurements once or twice a month during daylight periods, generally between 0800 and 1500 hours, at many stations in the mainstem Bay and tributary rivers. These stations, for many good reasons, are generally located over the deeper channel areas of the Bay and tributary rivers. Over a decade ago, pilot studies were conducted to see if these main channel (or off-shore) measurements accurately represented water quality conditions in the shallow waters (< 2 m depth) of the Bay and tributary rivers. The answers to this question ranged, in some instances, from generally yes to, in other instances, generally no. Other evaluations concluded that deep versus shallow water quality conditions were inconsistent and little in the way of firm generalities could be developed. In addition, there was a continuing focus in the Bay Program on SAV restoration and these communities were, of course, centered in these shallow water habitats. With these questions and goals in mind, two programs were added to the monitoring program, one focused on obtaining a much finer spatial scale data set of water quality conditions with particular attention paid to actual or potential shallow water SAV habitat (dataflow program) and the other to obtaining both long-term (3 years or more) and temporally detailed (15 minute intervals) water quality measurements in shallow waters (ConMon program). This section deals with ConMon data.

At the present time a very large data base of ConMon measurements has been generated. Some 98 different sites have had ConMon measurements (for at least 3 years) in the Maryland portion of the Bay and tributary rivers and others have been made in Virginia. One of the central findings to emerge from this data set is 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 often indicate a diel-scale (24 hour period) of hypoxia, severe at some locations, wherein dissolved oxygen concentrations drop to low levels during the hours of darkness and sometimes reach dangerously low concentrations at and just after sunrise. Qualitative inspection of these data indicate that the most severe diel-scale hypoxia is observed at sites experiencing severe eutrophication. More quantitative analyses of diel-scale hypoxia related to nutrient conditions are in progress.

Given the above observations it became apparent that ConMon data would be especially useful in at least two ways: 1) these data could be used in a variety of ways to assess trends in water quality conditions in shallow waters and SAV habitat and 2) time high frequency nature of these measurements could be used to directly evaluate surface water DO criteria attainment or failure. The latter of these items is addressed in this section with two different approaches.

ConMon Measurements and the 30-day mean: How protective is it?: The general arguments concerning application of the 30 day DO mean as a protective DO standard have been fully discussed earlier. Here we provide some sample analyses wherein for a variety of shallow water ConMon sites the summertime (Jun-Aug) 30 day mean is directly compared to the rate of DO criteria (instantaneous criteria) failure (Figure 1). In this example ConMon data were assembled from nine different locations, ranging from those having severe water quality issues to those having relatively good water quality conditions. The procedure for computing the DO means, percent failure rates and criteria values are provided in Table 1. In all there were 104 months of data included in this analysis. Several issues are apparent. First, when the 30 day mean is below the 30 day criteria value (DO < 5.0 mg/l) the rate of instantaneous DO criteria (<3.2 mg/l) failure rare is often quite high (>25%). In this case, both results signal a DO criteria failure. However, there were approximately 22 months (of 104) where the 30 day mean DO criteria was satisfied but the instantaneous criteria was not satisfied. Similar analyses have been conducted by C. Buchanan (see section X of this report) focused on the ConMon sites along the Potomac River estuary and similar results were obtained. We also conducted this same type of analysis but used the 7 day failure rate and in that case the 30 day mean was more protective but not completely protective. As a part of the Maryland Chesapeake Bay Biomonitoring Program we will continue to use ConMon data and make these computations for additional sites in the Maryland Bay. At this point it seems safe to tentatively conclude that for shallow water areas the 30 day mean is not protective of short-term DO criteria in many instances during summer periods (Jun – Sep).

The Issue of Duration of Low DO Conditions: For the formal DO criteria analysis there are both temporal and spatial considerations. In this analysis, using ConMon data, we are only considering the temporal aspect. However, in the formal analysis there is recognition in both the temporal and spatial domains that there needs to be some degree of "forgivness" of criteria violations and this seems appropriate given the very dynamic nature of estuarine systems. In our analysis of DO conditions in the Patuxent estuary during the 1960s, a period before this system underwent severe eutrophication, there were times (not very frequent) when surface DO criteria were violated. Thus, if a single violation was all it took to fail DO criteria, we would likely always have DO failures in most places for most time periods. That being the case, a 10% failure buffer has been adopted. However, this buffer needs to be considered in the light of just how the 10% acceptable violation rate is distributed in time.

Consider for a minute the breathing rate of a human as an analog of this problem. If we inhale once every 6 seconds we take 10 breaths per minute, 600 hundred breaths per hour and 14,400 per day. If we were to skip 10% of those breaths at a rate of 1 in every 10 breaths we would be fine…maybe a bit inconvenienced, but basically fine. However, if we were to skip all 10% at one time we would be dead…quite the difference.
We have examined the issue of DO criteria violation rate duration at a selection of Maryland ConMon sites and will continue to examine additional sites for the next several months. The data used for dissolved oxygen criteria failure duration calculations was extracted from the 2001 to 2008 Maryland Department of Natural Resources Continuous

Monitoring database (www.eyesonthebay.net) provided by Ben Cole (MDDNR). The file was in .txt format and imported into SAS® 9.2 (http://www.sas.com/). Data found to have error codes (http://mddnr.chesapeakebay.net/eyesonthebay/documents/SWM_QAPP_2010_2011_FINALDraft1.pdf) were removed prior to analysis. For this exercise the duration of time a measurement of dissolved oxygen was found below 3.2 mg L^{-1} (instantaneous criteria) and separately for 5.0 mg L^{-1} (30 day mean criteria) was calculated for the period of record at a Con Mon station. Con Mon measurements are made up of a dissolved oxygen reading taken every 15 minutes so each increment of duration of failure is 15 minutes. A duration sequence of failure was calculated as a series of continuous 15 minute intervals where the measured dissolved oxygen value was below the chosen criteria. If a measurement time stamp exceeded 40 minutes (to allow some variance in time stamp intervals due to data sonde set up) or changed dates (data sonde was removed or unavailable for some period of time) the duration sequence was reset to start again. Total duration of dissolved oxygen failure for a sequence was the sum of the 15 minute intervals. In this early version of the duration calculator, failures are terminated at the end of each 24 hour period. We know that in some cases the failure duration continues into the next day. The calculator needs to be up-graded to address this issue as well as several other problems. So, at present the calculator provides a minimum estimate of DO failure duration.

Examples of DO criteria failure duration for two criteria levels are provided in Table 2. We selected sites exposed to very severe eutrophication (Bishopville Prong in the MD Coastal Bays), reasonably good water quality conditions (St. George Island), a tidal freshwater site in an enriched estuary (Jug Bay; Patuxent River) and a mesohaline site exposed to open waters (Pin Oak; Patuxent River). As expected, at the site with severe eutrophication there were many criteria failures and criteria failure durations ranged from 12 to 24 hours (likely longer than this). At the less impacted sites, DO criteria failures were of shorter durations, especially for the instantaneous criteria (< 3.2 mg/l). At the higher DO criteria (< 5.0 mg/l) duration of failures remained long, often up to 24 hours. This evaluation is in early stages and some refinements have already been suggested. The point we make here is that it does not appear that DO failure rates are evenly distributed in time and this needs t be further evaluated to be certain that DO criteria values are as protective as they were intended to be.

What's Next?: There appear to be many avenues worth exploring relative to DO criteria issues. Several are listed below:

1. How can we generate a spatial dimension for the ConMon data set? Are there ways to convincingly link ConMon data to the spatially intensive Dataflow data?

2. We need to up-grade the duration calculator to capture DO criteria failures for periods longer than 24 hours. This effort is underway and we see no large problems with getting this completed.

3. How can we “link” data from the shallow water ConMon sites with the broad expanses of shallow “open waters” present in many segments of the Bay?

4. While there are many ConMon sites, there remains the problem of extending ConMon data to areas of the Bay and tributaries not having ConMon sites. Can we develop a statistical model of diel DO behavior based on data (temperature, PAR, salinity, chlorophyll-a concentrations) that will allow for an acceptable evaluation of DO criteria attainment at sites not monitored with ConMon technology?

5. We have seen considerable inter-annual differences in DO criteria failure rates at ConMon sites. For example, there was a steady improvement in failure rates at the Coastal Bay site. There were strong inter-annual changes in failure rates at the St George Island site in the mesohaline Potomac, possibly linked to high and low river flow years. How do we deal with this variability?

Appendix 5

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.

Mark J. Brush, Iris C. Anderson, Howard I. Kator, Virginia Institute of Marine Science College of William and Mary, Gloucester Point, VA. (Presented in MRAT 2009, Appendix 9)

Background: In the following analysis, we focus on the discussion of dissolved oxygen (DO) criteria in the section of the MRAT optimization report entitled, "*Considerations for Modifications of the Chesapeake Bay Program Partnership's Long-term Tidal Water Quality Monitoring Program While Addressing Funding Realignment*". As noted in the report, the EPA has established DO criteria at four timescales (instantaneous concentrations, 1-day mean, 7-day mean, and 30-day mean) for various designated uses in the Chesapeake and at various times of the year. However, there are limited data available with which to assess these criteria in the sub-surface open water, deep water, and deep channel designated uses. Currently the monthly Bay Program fixed station monitoring program is the primary data source for assessing these criteria, but the data can only be used to assess the 30-day mean criterion which has been proposed as potentially protective of the other criteria. However, one value per month may not be enough to assess a 30-day mean value given the multiple scales of temporal variability in the bay and its tributaries. This may particularly be the case in Virginia tributaries in which cycles of stratification and resulting bottom water hypoxia followed by mixing and re-aeration are to a large degree controlled by the spring-neap tidal cycle.

Given the interest in using the 30-day mean as protective of the other criteria, and the general lack of available data with which to evaluate the full suite of criteria in sub-surface waters, we used our spatially- and temporally-intensive monitoring data from the 2007-09 VIMS Chesapeake Bay Initiative – Open/Deep Water Component to assess York River hypoxic volume and compute DO concentrations on all four criteria timescales. Data are used from a vertical profiler in the polyhaline region of the river, a DO record from the very bottom of this segment, and 3D surveys with a towed, undulating ACROBAT instrument platform in the polyhaline and mesohaline segments. Similar to other analyses in the MRAT optimization report, the following work is preliminary and ongoing and requires additional data for verification of the findings to date. Nevertheless, the observations highlight a major difference between the mainstem model of Chesapeake hypoxia/anoxia and the occurrence of hypoxia in at least some tributaries, which could complicate attempts to evaluate DO criteria and make de-listing decisions in these segments with the existing monitoring program.

VIMS Chesapeake Bay Initiative: In 2007, the VA Department of Environmental Quality funded the VIMS Chesapeake Bay Initiative, which was designed to take advantage of VIMS' state of the art high spatial and temporal resolution monitoring capabilities and provide much greater monitoring coverage of the main VA tributaries for the purposes of criteria assessment for chlorophyll-*a*, water clarity, and DO. For the sub-surface open water, deep water, and deep channel designated uses, our monitoring was focused on the York River to demonstrate its capability, and included a fixed station vertical profiler in the polyhaline zone of the river and bimonthly, 3D monitoring of the mesohaline and polyhaline zones using our ACROBAT towed instrument package (Fig. 1).

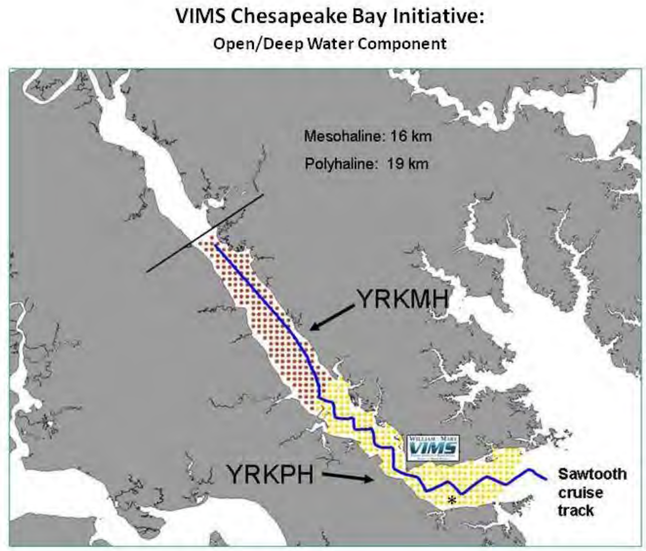


Fig. 1. Typical cruise track of the VIMS ACROBAT sampling in 2007-08 (blue line). Location of the VIMS vertical profiler is shown with the asterisk (*).

The vertical profiler is mounted on the Yorktown Coast Guard Training Center pier, the deepest shore-based point in the river, and provides hourly water column profiles with a YSI datasonde (temperature, salinity, depth, pH, DO, chlorophyll-*a*, and turbidity) at 1-meter intervals from the surface to a depth of approximately 9 m, which is within the typical pycnocline in the lower river and therefore just captures the signal from the deep water designated use. The profiler was run from June through September in 2007-08, and a third year of data from 2009 is currently being collected.

The ACROBAT collects data (temperature, salinity, depth, DO, chlorophyll-*a*, and turbidity) four times a second with a horizontal resolution of 6-8 m and a vertical resolution of 5-10 cm, for a total of 40,000-50,000 data scans per cruise. This rate of collection results in approximately 10 complete water column profiles per kilometer in a 10 m water column. Surveys in 2007 were conducted along a transect that covered the full polyhaline zone and the portion of the mesohaline zone over which hypoxia has historically occurred (Fig. 1). Two types of surveys were conducted: (1) bimonthly 3D surveys from June through September along the zig-zag track in Figure 1 to compute hypoxic volume, once following a spring tide and once following a neap tide each month, and (2) high frequency 2D cruises along a straight-line version of the track in Figure 1 approximately every other day through a neap-spring-neap cycle in June and again in August. 3D ACROBAT surveys were also conducted once per month from June through September in 2008 but data are not presented here.

Profiler data from 2007-08 demonstrate two key patterns: (1) data do not vary consistently throughout the water column, and (2) near-pycnocline data at the bottom of the profiler record display a periodic variation between normoxic and hypoxic conditions (Fig. 2). Occurrence of hypoxic waters typically corresponds to the presence of high density water at the bottom and stratification of the water column. The data suggest that at least for this tributary, surface ConMon data are not sufficient for assessing DO in deeper waters, and DO values vary on a time scale significantly less than the 30-day criterion, which is being proposed as protective of the other criteria, and which would not be captured by the current monthly mid-channel monitoring. As expected from previous studies in the VA tributaries, this variation in stratification and hypoxia appears strongly

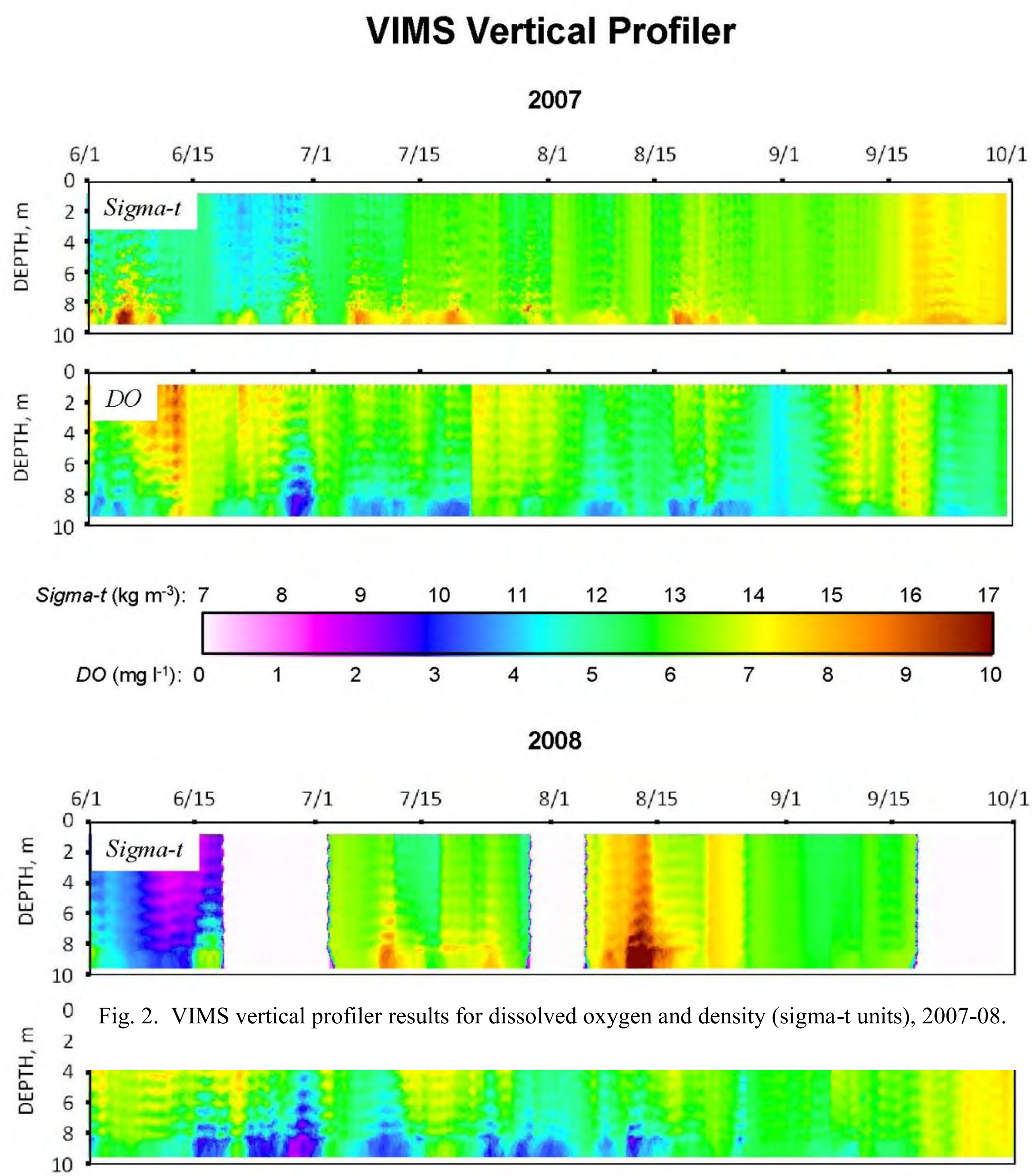


Fig. 2. VIMS vertical profiler results for dissolved oxygen and density (sigma-t units), 2007-08.

related to the spring-neap tidal cycle (Fig. 3).

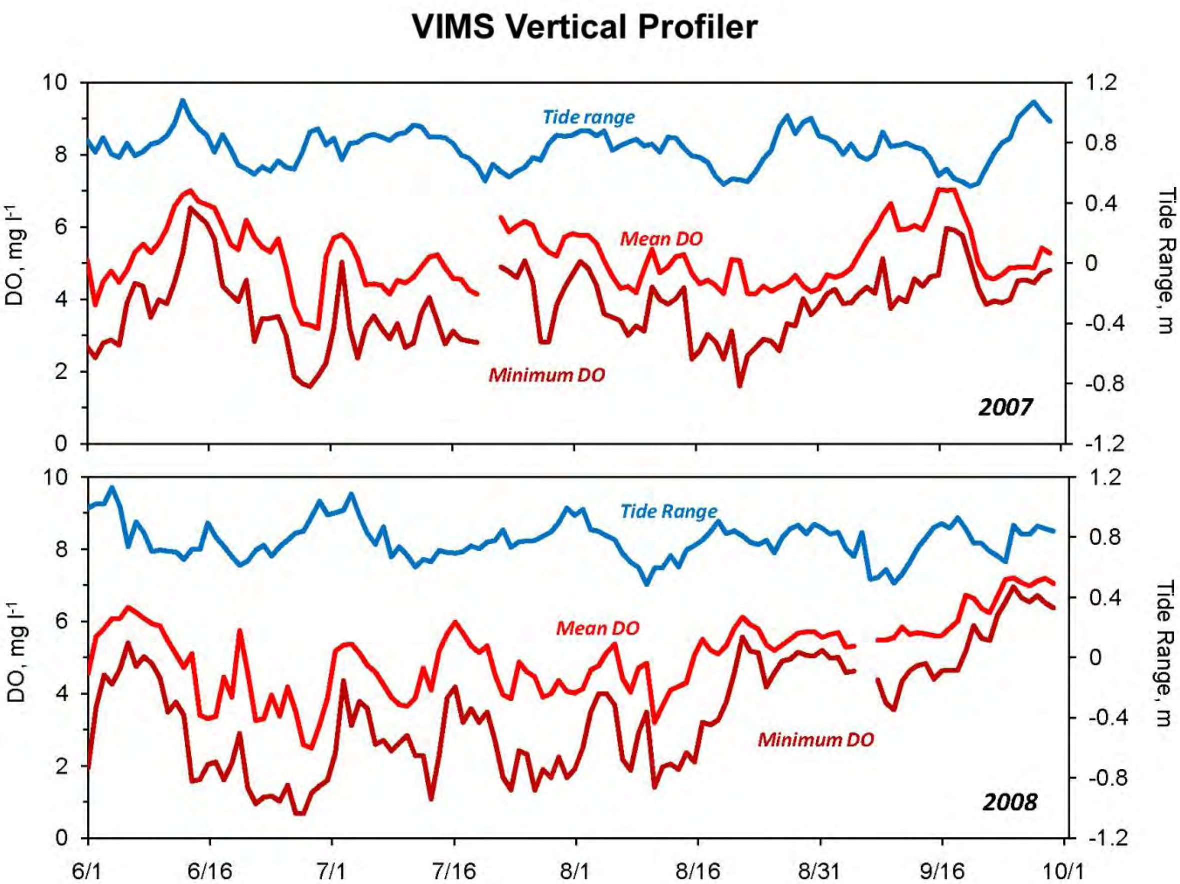


Fig. 3. Daily mean and minimum bottom DO readings from the VIMS vertical profiler in 2007 (top) and 2008 (bottom) and the corresponding daily tide range at the Yorktown Coast Guard Training Center (data from NOAA).

Profiler data from both years were used to compute 1-day, 7-day, and 30-day means for analysis of criteria attainment. Since the profiler samples from the surface to the region of the pycnocline, the analysis was performed using an average across the water column to represent a depth-integrated assessment of the open water use zone, and using only the bottom values as an indication of the deep water use zone. Results for average (i.e. open water) DO suggested that the 30-day mean criterion may indeed be protective of the other criteria, but instantaneous bottom (i.e. deep water) DO was more frequently in violation of this criterion while the 1-day and 30-day criteria were met, suggesting the 30-day criterion may not be protective in deep water (Fig. 4).

The problem with this analysis is that the VIMS profiler only reaches to the pycnocline and does not include more hypoxic water throughout the deep water zone. To address this, we analyzed a record of bottom DO from a datasonde placed offshore from the vertical profiler at the very bottom of the York channel this summer by our graduate student, Mr. Sam Lake (Fig. 5). The record is incomplete but consists of approximately a month in June and a second month in late-July to August. In both periods, the 30-day mean DO was very close to the 30-day criterion while the 1-day and instantaneous criteria were violated. Because of the spring-neap induced cyclic nature of bottom water DO in this system, the 30-day criterion does not appear to always be protective of deep water DO, and this spring-neap cycling can only be captured with more frequent monitoring than the current monthly mid-channel cruises.

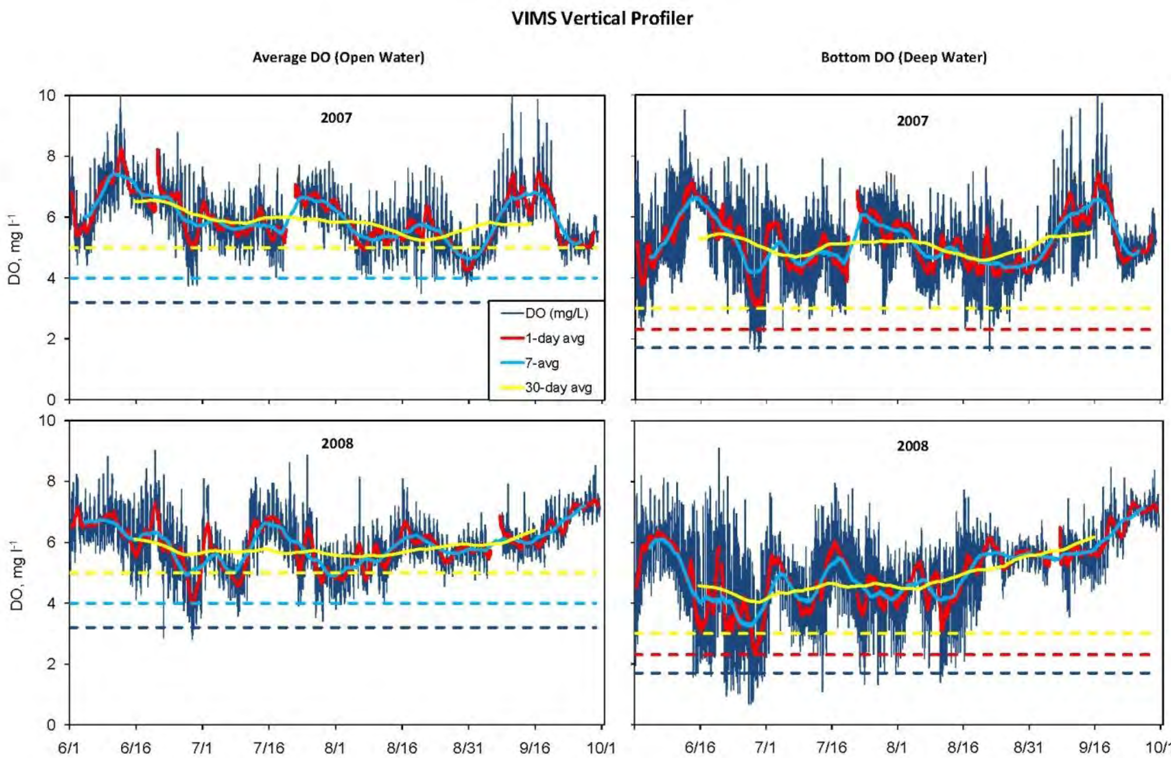


Fig. 4. VIMS vertical profiler DO record from 2007 (top) and 2008 (bottom) with running 1-day, 7-day, and 30-day mean values. Records were computed for the open water use zone by taking an average across each profile (left) and using only the bottom reading for the deep water zone (right). DO criteria are indicated with dashed lines of the corresponding color.

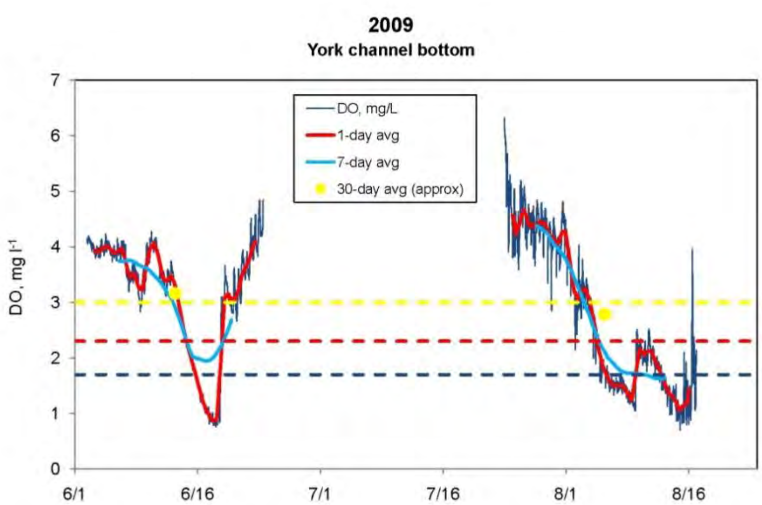


Fig. 5. DO record from the bottom of the York River, summer 2009, collected by VIMS graduate student Mr. Sam Lake, with running 1-day, 7-day, and 30-day mean values. DO criteria are indicated with dashed lines of the corresponding color.

VIMS ACROBAT data provide another clear picture of the spring-neap variation in stratification and hypoxia in the York. The high frequency 2D cruises through a neap-spring-neap tidal cycle in June and August 2007 demonstrated the rapidity with which the York oscillates between stratified, hypoxic conditions following a neap tide to well-mixed, normoxic conditions following a spring tide, and back again (Fig. 6). The data also illustrate the spatial patchiness of York hypoxia, with low DO water generally being confined to the deepest regions and appearing to develop in the lowermost portion of the river and spread up-river over a few days. Given this spatial and temporal patchiness, traditional fixed monitoring has the potential to miss important dynamics in the system.

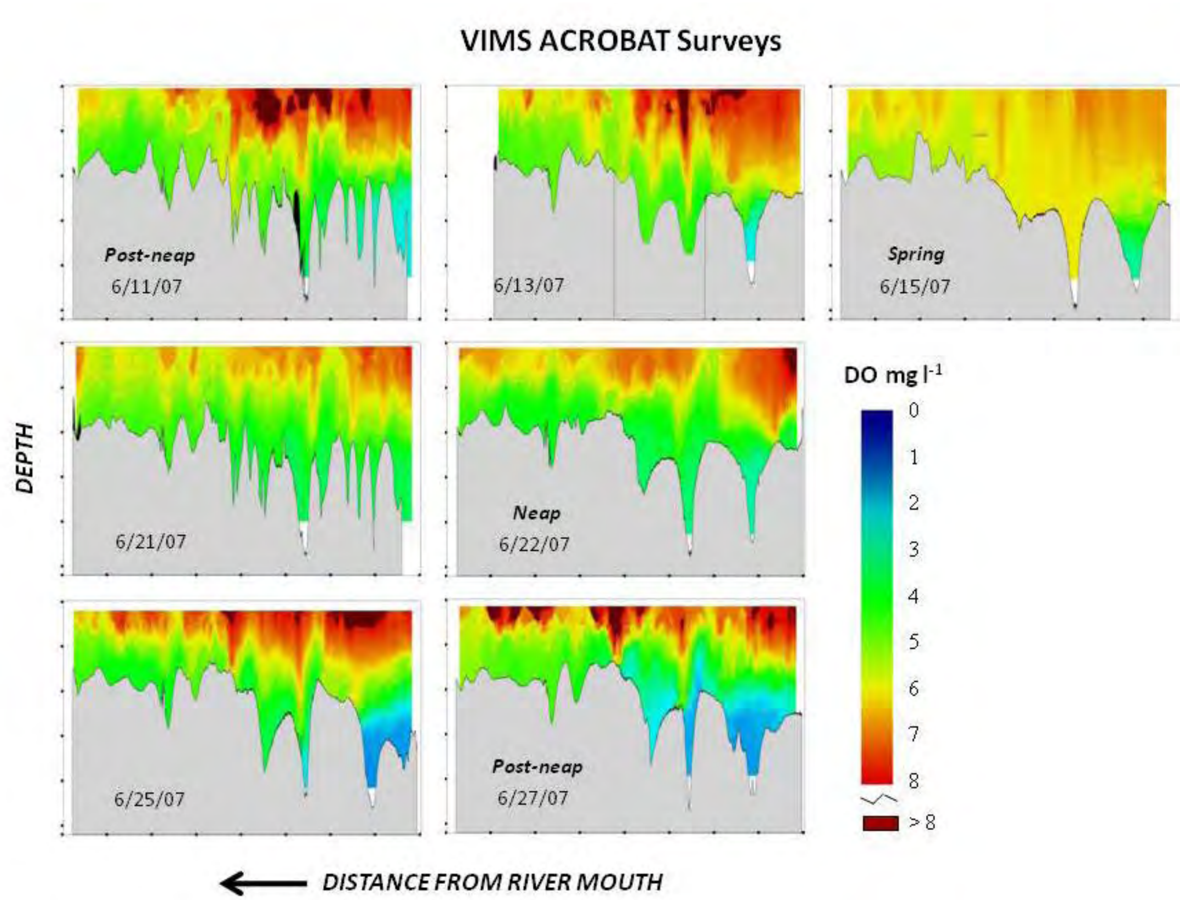


Fig. 6. 2-D cross-sections of DO in the York River from ACROBAT surveys over a two week neap-spring-neap tidal cycle in June 2007.

ACROBAT data were volumetrically interpolated using the NOAA Chesapeake Bay and Tidal Tributary Interpolator (http://chesapeakebay.noaa.gov/interpolator.aspx) to determine the volume of hypoxic water in the mesohaline and polyhaline zones below the 30-day mean DO criterion, both in open water (30-day criterion = 5 mg l^{-1}) and deep water (30-day criterion = 3 mg l^{-1}), using our average observed pycnocline depth of 9 m. Hypoxic volumes varied throughout the summer in the open water zone, largely showing higher hypoxic volumes during neap tides as opposed to spring tides (Fig. 7 top).

Hypoxic volumes were much smaller in the deep water owing to the smaller volumes of this zone (almost no volume in the mesohaline region), and were only well developed in early summer (Fig. 7 bottom).

For comparison, we also computed hypoxic volumes with the NOAA interpolator using the monthly Bay Program fixed monitoring station data (Fig. 7 dashed lines). In some cases the fixed station volumes tracked the ACROBAT volumes well, although they could not capture the spring-neap variation in hypoxic volume. In other cases, the Bay Program data appear to greatly overestimate hypoxic volume, which we believe is due to the apparent patchy distribution of hypoxia as indicated by the ACROBAT surveys, which the fixed station data cannot capture (Fig. 8). In this case, the ACROBAT data provide a more accurate estimate of the distribution of hypoxic water in the York, which is less than what the existing monitoring program suggests.

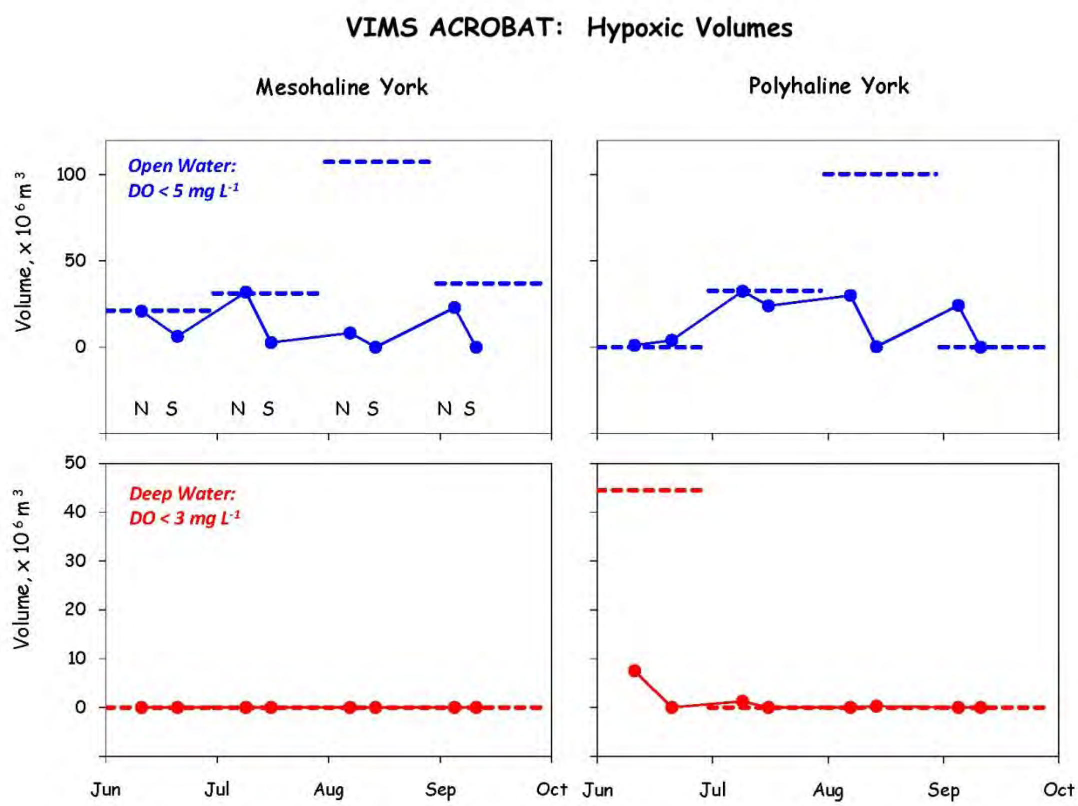


Fig. 7. Computed volume of water in the mesohaline (left) and polyhaline (right) York River in violation of the open water (top) and deep water (bottom) 30-day mean DO criterion based on 2007 ACROBAT surveys (solid lines) and Bay Program monitoring station data (broken lines).

Summary: Given the dominance of the spring-neap cycle in driving York River stratification and hypoxia, the existing fixed-station and ConMon monitoring data may not be adequate to assess the distribution of *sub-surface* DO in this system, and the 30-day mean criterion may not necessarily be protective of the other criteria for *sub-surface* waters. In some cases the 30-day mean may indeed be protective, but in other cases high frequency monitoring appears necessary to assess the criteria on higher-frequency timescales, especially in deeper waters. The fixed stations appear to usually overestimate hypoxic volume so in a sense these data may also be protective, but this may come at the expense of accuracy and may set an unrealistically high target for reducing hypoxia. Our assessments have been based on a single tributary, and more data from additional systems are certainly needed to make a better assessment, as well as York River data from deeper than our profiler can reach. However, given the established dominance of the spring-neap cycle in driving stratification throughout the VA tributaries (Fig. 9), our preliminary findings for the York likely apply in other tributaries. As noted above, these observations highlight an important difference between the mainstem conceptual model of Chesapeake hypoxia/anoxia and the occurrence of hypoxia in some tributaries, which could complicate attempts to evaluate DO criteria and make de-listing decisions in these segments with the existing monitoring program.

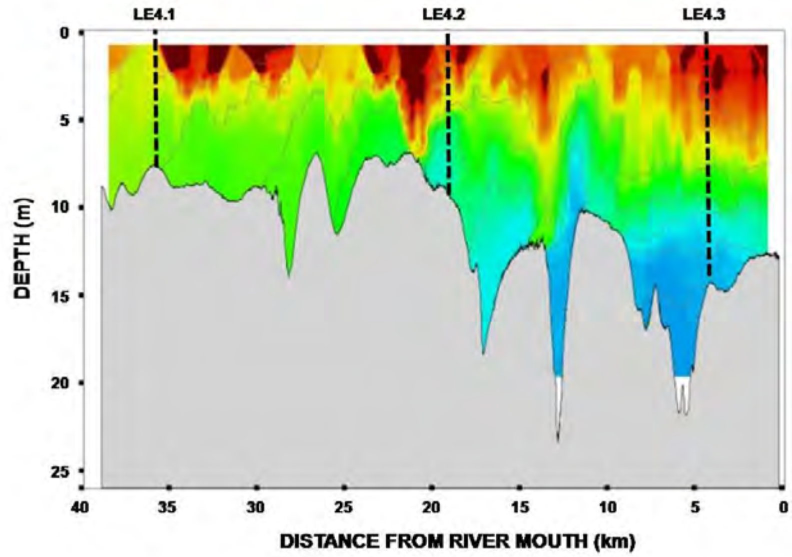


Fig. 8. Location of the Chesapeake Bay Program mid-channel fixed monitoring stations in the lower York River relative to DO concentrations on 6/11/07 measured by the ACROBAT. Color scale is the same as in Figure 6.

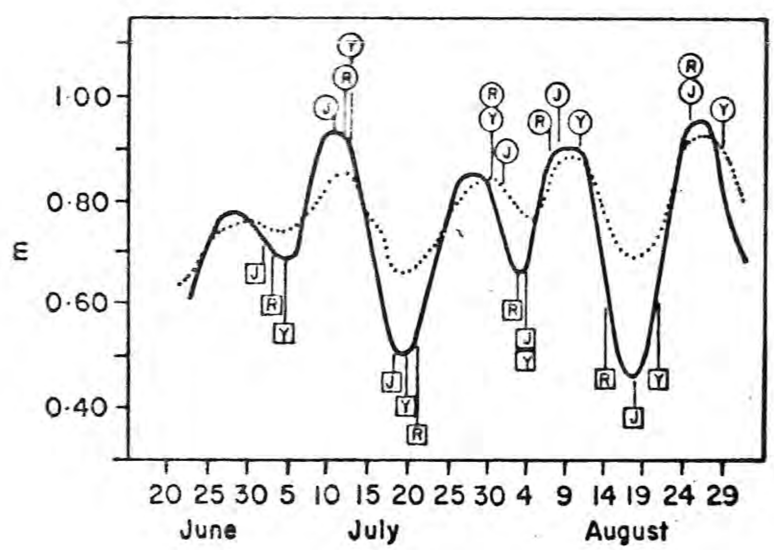


Fig 9. Figure 7 reproduced from Haas (1977, Est Coast Mar Sci 5:485-496). Predicted tidal range (solid line) and high tide height (broken line) at Hampton Roads, 1972. Circles and squares indicate the times of maximum observed homogeneity and stratification, respectively, in the James (J), York (Y), and Rappahannock (R) Rivers.

Appendix 6

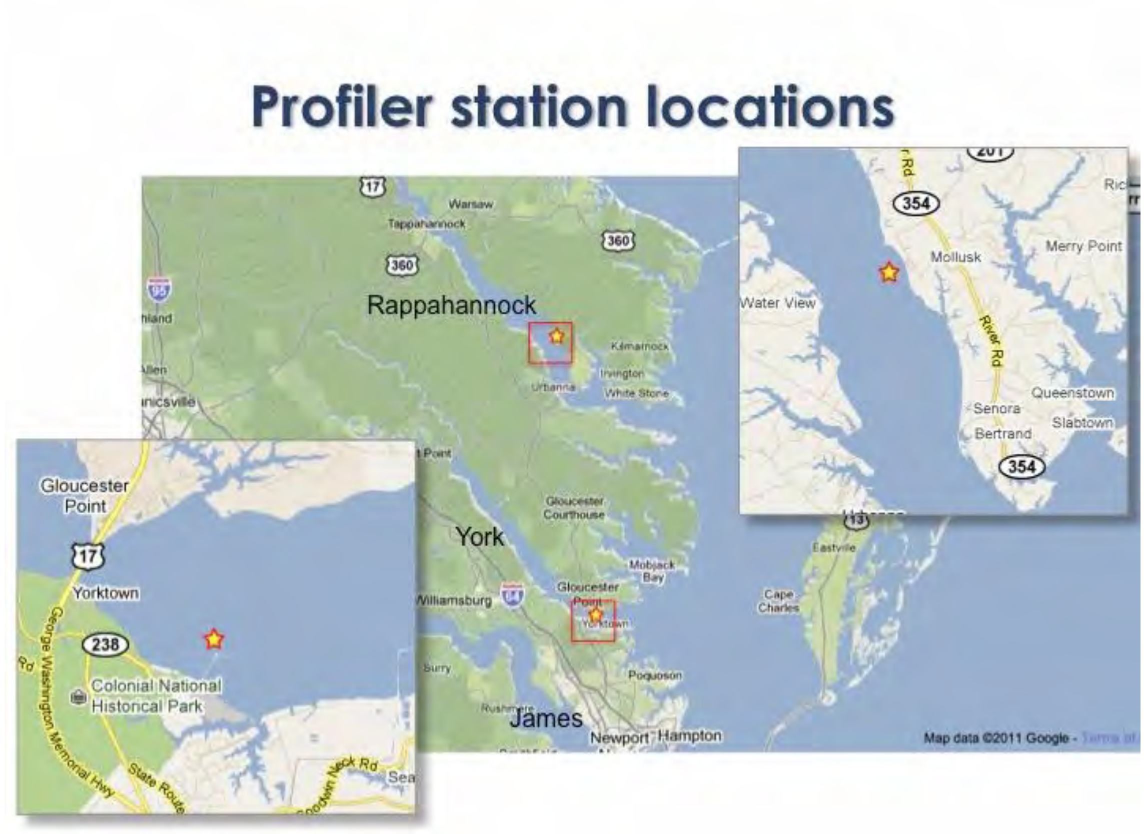
YORK AND RAPPAHANNOCK RIVER PROFILER DATA EVALUATION

D.M. Bilkovic, D. Rudders, Virginia Institute of Marine Science
D. Jasinski, Chesapeake Environmental Communications

MARCH 2011

DATA DESCRIPTION & MANIPULATION

Water quality vertical profiler data are collected at high frequency from a fixed location. Vertical profiles are made every hour at approximately 1 meter intervals during summer months. Profiles measure temperature, salinity, conductivity, dissolved oxygen, chlorophyll and turbidity using YSI 6600EDS V2 sondes. High frequency vertical profile data from two fixed stations on the York and Rappahannock rivers were examined to evaluate the protectiveness of DO criteria at varying temporal scales.



<u>YORK RIVER</u>

- There were a total of 61,248 observations during 2007-2009 used in the analysis
- Data cover summer months (June 1 – Sept 30) for 3 years 2007-2009
- Previous data manipulation (from Mike Lane, Oct 21, 2010) included
    - Removal of values classified as QC problems (e.g. general probe failure)

        - 20June2007-30June2007
        - 20July2007-31July2007
        - 30Sep2007
        - 01Sept2008-08Sept2008
        - 24June2009-30June2009
        - 01July2009-08July2009
    - The original time stamp has been converted to the hour of collection so that there is one value per depth for each hour rather than values at various time intervals within a given hour (these were averaged).
    - All values below 8 meters were dropped since these values occurred infrequently.
    - There were cases for which data were not collected at exactly one meter intervals. This occurred with pretty high frequency. This problem was fixed by substituting the mean of the values at the depth above and the depth below the missing interval.
    - Other problematic observations were substituted with an appropriate mean value. Typically this amounted to the substitution of one days worth of data with the mean hourly values of the previous and the next day.

<u>RAPPAHANNOCK RIVER</u>

- Rappahannock River Profiler Location: Sta: RPP021.36; 37.7202, -76.56688
- Dates of record: 16 June – 5 October 2009 (*28162 raw data records; 21888 QAQC data records*)
- Data were collected hourly and once per depth. Depths ranged from 0 to 13 m.
- Rappahannock River Profiler 2009 data manipulation included
    - Removal of values classified as QC problems (e.g. general probe failure)
        - 29June2009 -1July2009
        - 5July 2009
        - 18July 19July 2009
        - 12Aug – 24Aug 2009
        - 7Sept – 21Sept 2009
        - 27Sept2009
        - 5Oct 2009
        - Negative depth measures (28 removed)
    - The original time stamp has been converted to the hour of collection so that there is one value per depth for each hour rather than values at various time intervals within a given hour.
    - All values below 11 meters were dropped since these values occurred infrequently.
    - There were cases (~1500) for which data were not collected at exactly one meter intervals. This problem was fixed by substituting the mean of the values at the depth above and the depth below the missing interval.
    - Other problematic observations (e.g. shortened profiles) were substituted with an appropriate mean value. Typically this amounted to the substitution of data for several depths during a given hour of a day with the associated mean hourly values for the missing depth of the previous and the next day (1819 records interpolated in this manner).

APPROACHES

York and Rappahannock Profiler Data:

1) To distinguish deep water from open water (and deep water from deep channel on the Rappahannock) the upper and lower boundaries of the pycnocline (if present) were estimated using the standardized method for calculating using measurements of water temperature and salinity (see **Appendix A**).
2) Moving 7-day averages (rolling means), and 1-day and 30-day means were calculated. 7-day mean: begins on Day 1 of each month; evaluate first 4 weeks of month and ignore trailing days. Data were excluded if there were not 7 consecutive days for which calculate a moving average. 1-day means represent the average for each 24-hr period. 30-d mean: begins on Day 1 of each month; trailing days were ignored.
3) The percentage of instantaneous measures that failed the DO criteria was calculated separately for deep water (≤ 1.7 mg/L) and open water (≤ 3.2 mg/L).
4) To further assess the protective nature of the 30-day mean, comparisons were made between
    a. The % non-attainment of the 7-day mean criterion (4 mg/L) for open water in relation to the 30-day mean.
    b. The % non-attainment of the 1-day mean criterion (2.3 mg/L) for deep water in relation to the 30-day mean
5) To examine the scenario in which the 7-day mean criterion is just barely satisfied (in the 4-5 mg/L interval) in relation to instantaneous assessments in greater detail, the 7-day mean and the 10^{th} percentile of instantaneous observations that make up the 7-day mean were compared. If the instantaneous minimums associated with the 7-day means near the criterion are generally above the instantaneous criterion then even if the 7-day criterion for open water is barely satisfied, there are likely fewer than 10% violations of the instanteneous minimum criterion.
6) In a similar manner as above, the 1-day mean criterion (2.3 mg/L) was related to the 10^{th} percentile of instantaneous observations that make up the 1-day mean for deep water.

RESULTS

I. York River

York River Profiler Observations

- The location (and presence) of the pycnocline is highly variable on a daily basis with a reduced presence of a pycnocline in the fall (September).
- Overall, deep water accounted for 20% of the readings (n=12,379), open water = 80% (48,869). This pattern was fairly consistent among years with slightly lower percentage of deep water in 2007 (15%).
- When the pycnocline was present, there was no consistent trend in the depth of occurrence (i.e. similar distribution among depths).
- For open water, there were 7 instantaneous violations in 2007, 36 in 2008 and 10 in 2009. For deep water, there were only instantaneous violations in 2008 (all but one in June) numbering 15.
- There were 792 instances when temperature was > 29ºC (most occurrences were in 2007 (758). For all these instances, instantaneous DO was ≥ 4.3 mg/L.
- The 1-day, 7-day and 30-day means always met the designated DO criteria (e.g. the 7-day mean for open water was always ≥ 4 mg/L).
- In the York River profiler location, the data have few observations where the 7-day mean criterion is just barely satisfied (in the 4-5 mg/L interval), which is when it is more likely that the 7-day mean assessment and instantaneous minimum assessment would be inconsistent.

30 day mean in relation to 7-day mean violations (Open Water)

- There were no open water 7-day mean violations in a given 30-day period observed in 2007-2009. In the York River, the 30-day mean may be protective of the 7-day mean (**Fig 1**).

30 day mean in relation to 1-day mean violations (Deep Water)

- There were no deep water 1-day mean violations in a given 30-day period observed in 2007-2009. The 30-day mean DO criterion for deep water during summer months is 3.0 mg/L (**Fig 2**).

30-day mean in relation to instantaneous violations (Open and Deep Water)

- Open water DO instantaneous violations in a given 30-day were observed in all years and most months ranging in frequency from 0.04 – 0.4% (**Fig 3**).
- Deep water DO instantaneous violations in a given 30-day were observed in 2008 (June and July) and ranged in frequency from 0.2 – 0.4% (**Fig 4**).
- For the 30-day mean, 25% of open water (3 of 12 data points) and 83% of deep water (10 of 12 data points) had NO instantaneous violations.

7-day mean in relation to instantaneous violations (Open Water)

- Percentage of open water instantaneous DO violations were less than 1.2% for all years. In general, the 7-day mean appears to be protective of the instantaneous minimum DO criterion for open water in the Lower York River (**Fig 5**).
- For the moving 7-day mean, 60% of open water (n=247) observations had NO instantaneous violations.

<u>10th percentile of instantaneous observations in relation to 7-day mean</u> (Open Water)

- To assess the liklihood that the 7-day mean is protective of the instantaneous minimum for open water, the values of the 10^{th} percentile of the instantaneous minimums were fit with the running 7-day mean. Although there were not any observed values for the 7-day mean that were < 5.0 mg/L, the regression line crosses near 3.2 mg/L (instantaneous) at 4 mg/L 7-day mean. This may be further support for the 7-day mean being protective of the instantaneous minimum (in the York River) (i.e. if the 7-day criterion is barely satisfied, there is likely fewer than 10% violations of the instantaneous minimum criterion) (**Fig 6**).

<u>1-day mean in relation to instantaneous violations</u> (Deep Water)

- For the 1-day mean, 98% of deep water (224 out of 229 daily data points) had no instantaneous violations.
- Deep water DO instantaneous violations in a given 1-day period were infrequent (6 days in 2008) and range in frequency from 0.6 - 7.8%. Overall, the 1-day mean appears to be protective of the instantaneous minimum DO criterion for deep water in the Lower York River. Failures were observed only during 2008 (**Fig 7**).

<u>10th percentile of instantaneous observations in relation to 1-day mean</u> (Deep Water)

- The values of the 10^{th} percentile of the instantaneous minimums were fit with the 1-day mean for deep water. Although there are not any observed values for the 1-day mean that below the criterion (2.3 mg/L), the regression line crosses near 1.5 mg/L (instantaneous) at a 3 mg/L 1-day mean. This suggests that for deep water the 1-day mean may not be protective of instantaneous minimum when it is near the criterion (i.e. if the 1-day criterion is barely satisfied, there are likely greater than 10% violations of the instantaneous minimum criterion) (**Fig 8**).

YORK RIVER GRAPHICS

30 day mean in relation to 7-day mean violations (Open Water)

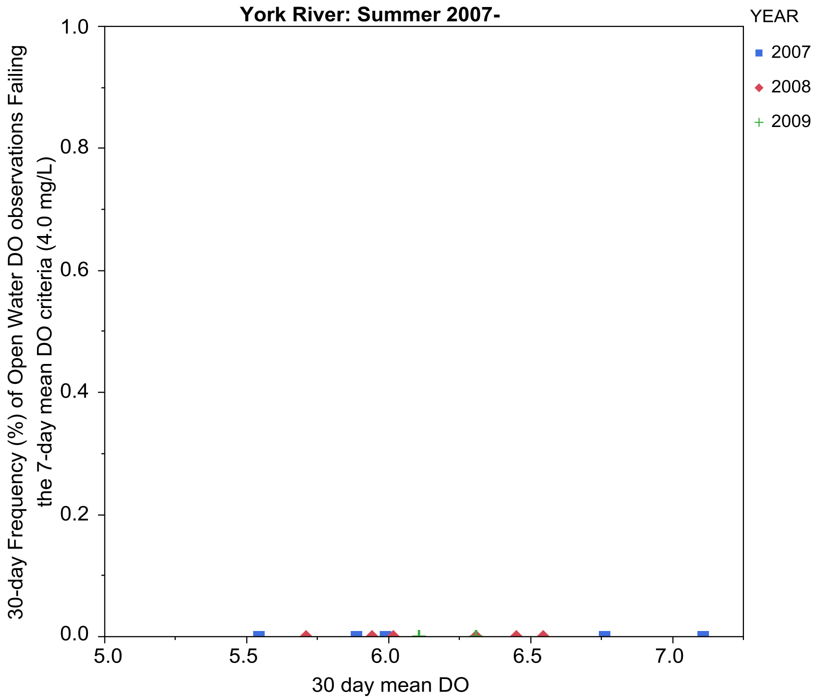


Figure 1. There were no open water 7-day mean violations in a given 30-day period observed in 2007-2009. The 30-day mean DO criterion for open water during summer months is 5.0 mg/L. However, the 30 day mean is probably close to the true 30-day mean since temporal high frequency measurements were used. Additional analysis is needed to satisfy the question of whether the fixed station data with limited observations per month show similar patterns

30 day mean in relation to 1-day mean violations (Deep Water)

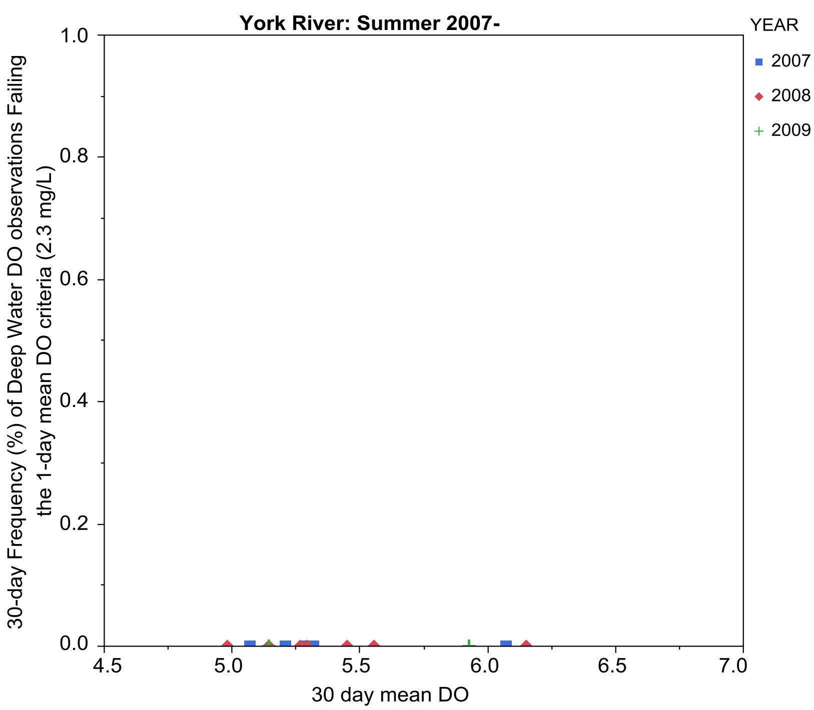


Figure 2. There were no deep water 1-day mean violations in a given 30-day period observed in 2007-2009. The 30-day mean DO criterion for deep water during summer months is 3.0 mg/L. However, the 30 day mean is probably close to the true 30-day mean since temporal high frequency measurements were used. Additional analysis is needed to satisfy the question of whether the fixed station data with limited observations per month show similar patterns

30-day mean in relation to

instantaneous violations (Open and Deep Water)

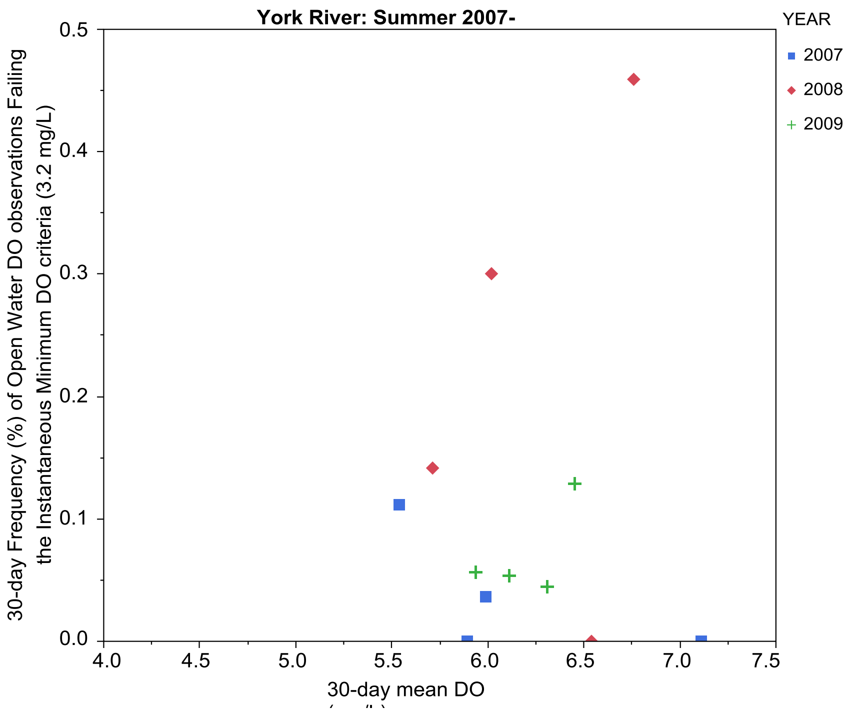


Figure 3. Open water DO instantaneous violations in a given 30-day timeframe were observed in every month and year and ranged from 0.04 – 0.4%. There were instantaneous violations associated with 75% of the monthly means (9 of 12 data points (3 years x 4 months)). The 30-day mean DO criterion for open water during summer months is 5.0 mg/L.

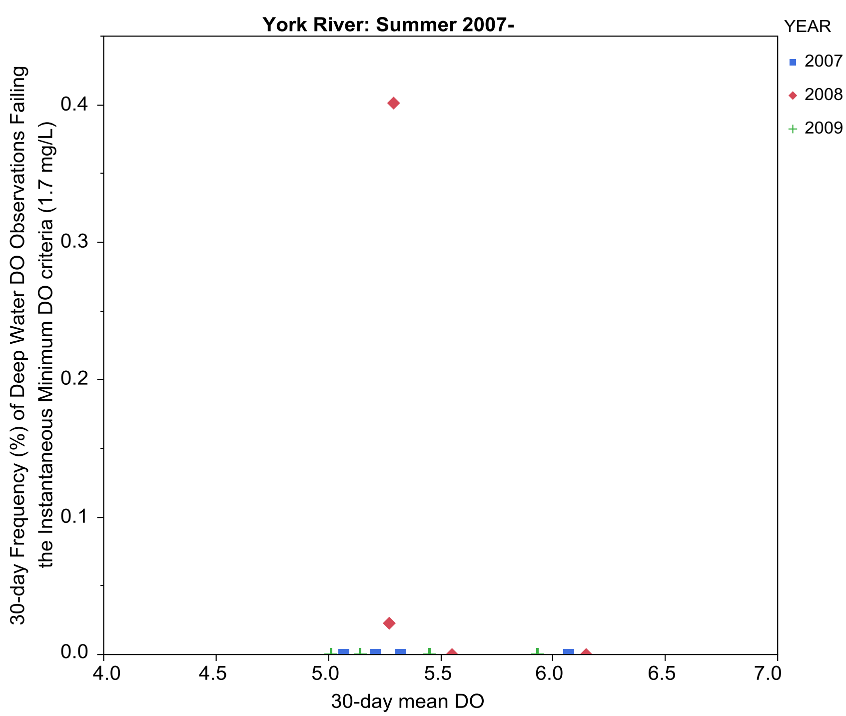


Figure 4. Deep water DO instantaneous violations in a given 30-day timeframe were observed in 2008 (June and July) and ranged from 0.2 – 0.4%. There were instantaneous violations associated with 17% of the monthly means (2 of 12 data points (3 years x 4 months)). The 30-day mean DO criterion for deep water during summer months is 3.0 mg/L.

7-day mean in relation to instantaneous violations (Open Water)

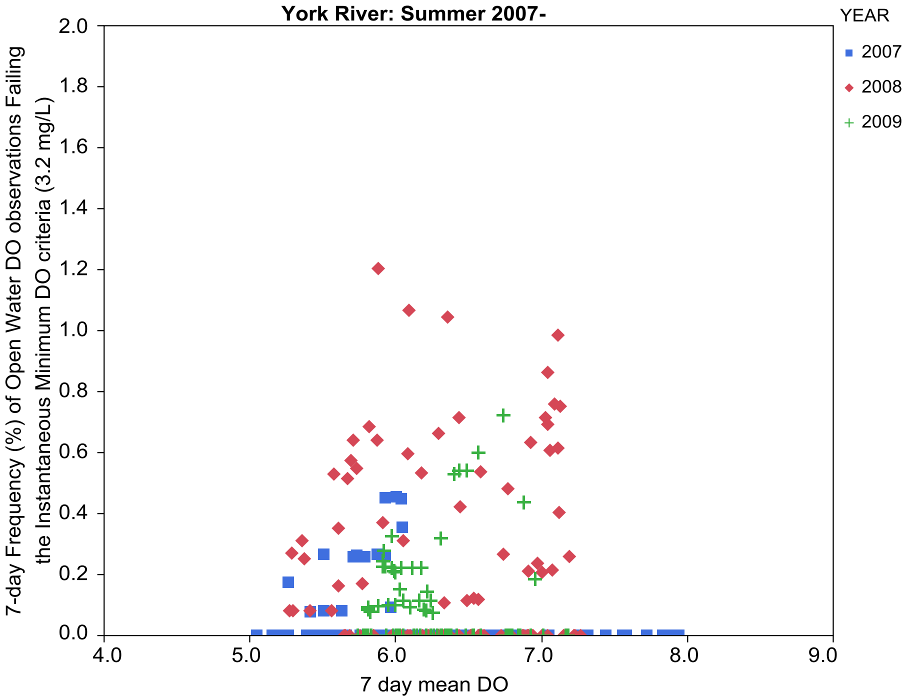


Figure 5. Percentage of open water DO instantaneous violations in a given 7-day period are less than 1.2% for all years. Overall, the 7-day mean appears to be protective of the instantaneous minimum DO criterion for open water in the Lower York River (violations < 1.2%). The 7-day mean DO criterion for open water during summer months is 4 mg/L. 7-day means represent running mean values.

10th percentile of instantaneous observations in relation to 7-day mean (Open Water)

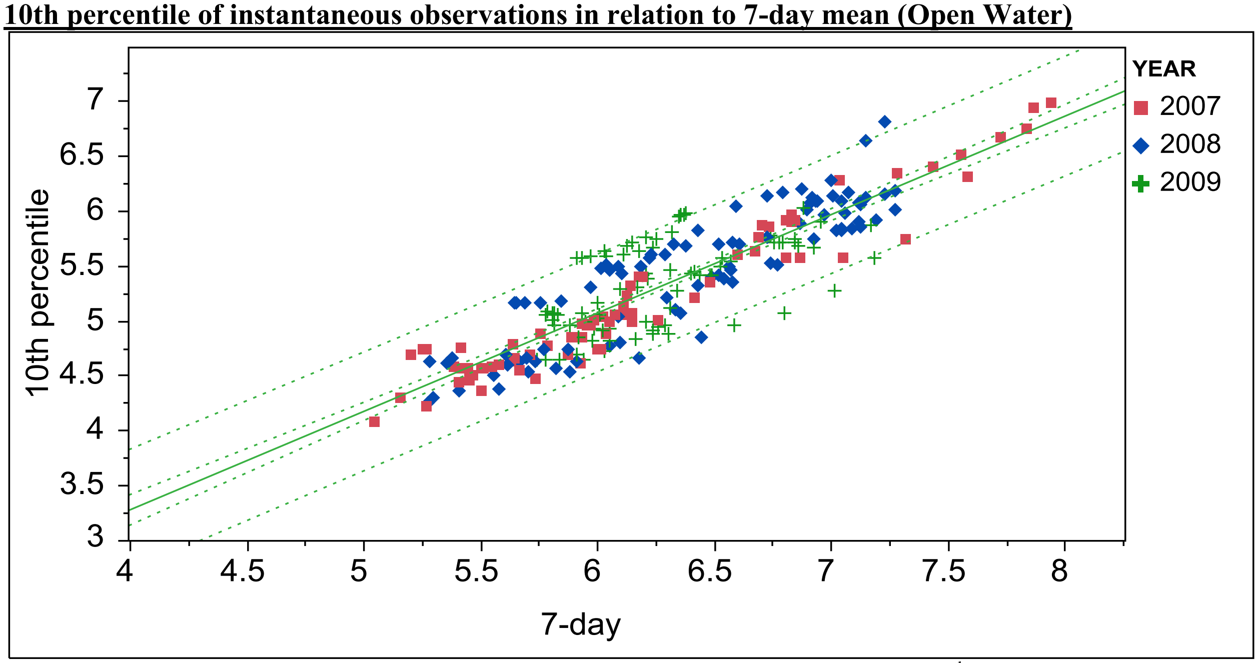


Figure 6. To further examine the relationship between actual values, the values of the 10th percentile of the instantaneous minimums fit with the running 7-day mean with 95% confidence intervals about the fitted line and predicted observations. Although there are not any observed values for the 7-day mean near the criterion (4.0 mg/L), the regression line crosses near 3.2 mg/*L* (instantaneous) at 4 mg/L 7-day mean. This may be further support for the 7-day mean being protective of the instantaneous minimum for open water (in the York River) (i.e. if the 7-day criterion is barely satisfied, there is likely fewer than 10% violations of the instantaneous minimum criterion).

Bivariate Fit of 10th percentile of instantaneous observations by 7-day mean (Fig 6b)

The years varied slightly with non-random residuals observed in 2008 and 2009 near late summer, but in these instances the 10^{th} percentile instantaneous minimum was actually closer to the 7-day mean (i.e. difference was < 0.8).

Linear Fit

10th percentile = -0.300083 + 0.8958658*7-day mean

Summary of Fit

RSquare	0.785504
RSquare Adj	0.784628
Root Mean Square Error	0.271338
Mean of Response	5.328229
Observations (or Sum Wgts)	247

Analysis of Variance

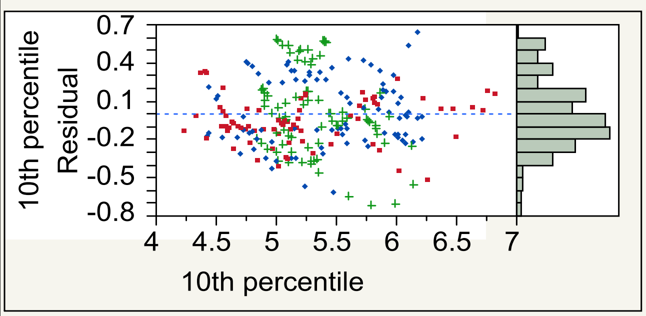
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Ratio**
Model	1	66.056743	66.0567	897.2107
Error	245	18.038016	0.0736	**Prob > F**
C. Total	246	84.094759		<.0001*

Parameter Estimates

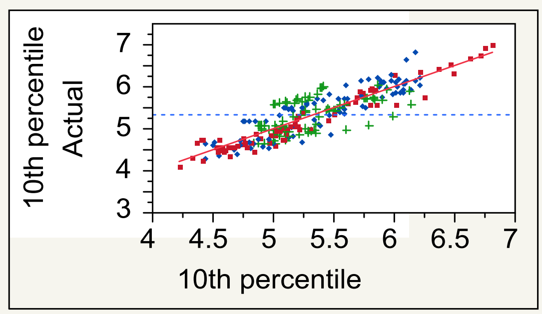
Term	**Estimate**	**Std Error**	**t Ratio**	**Prob>|t|**
Intercept	-0.300083	0.188693	-1.59	0.1131
7-day mean	0.8958658	0.029909	29.95	<.0001*

Diagnostics Plots

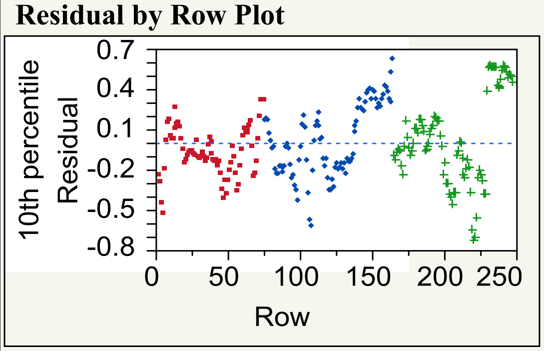
Residual by Predicted Plot



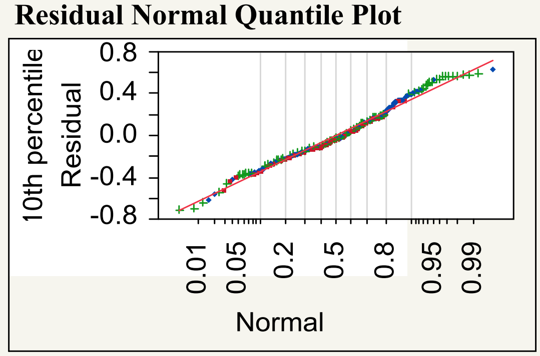
Actual by Predicted Plot



Residual by Row Plot



Residual Normal Quantile Plot



1-day mean in relation to instantaneous violations (Deep Water)

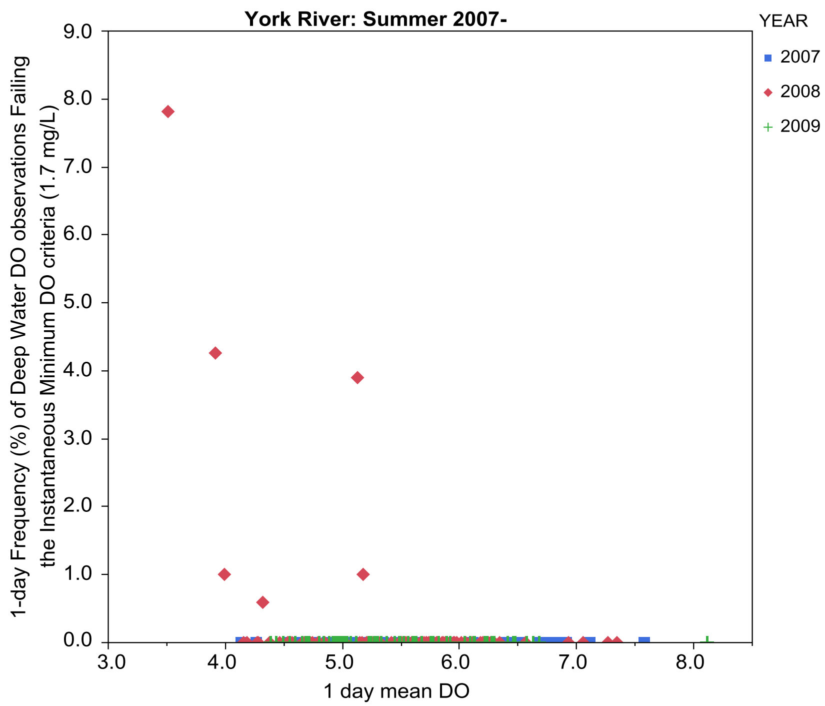
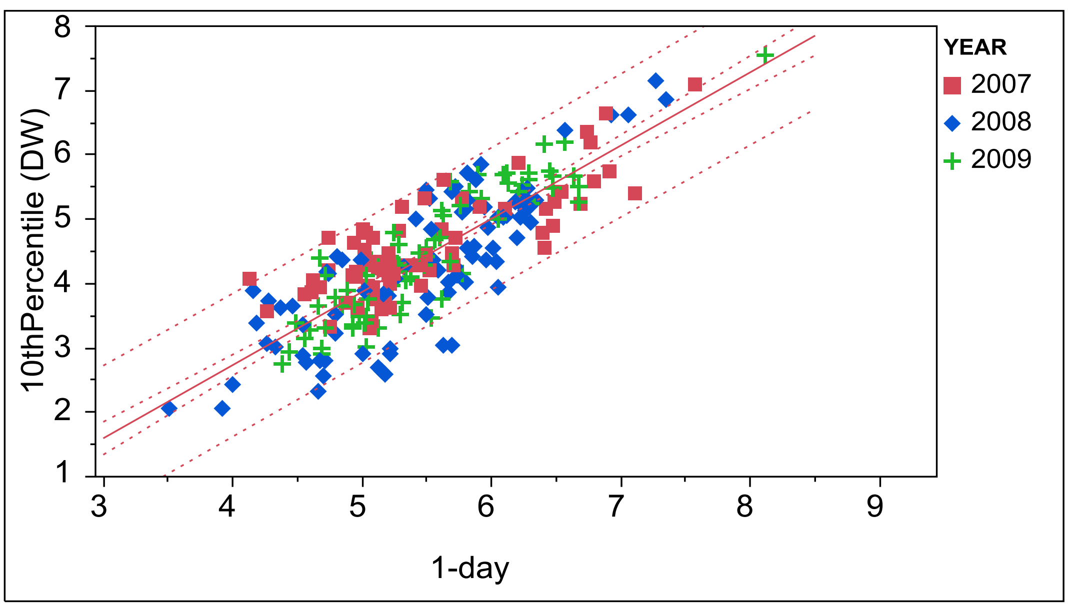


Figure 7. Deep water DO instantaneous violations in a given 1-day period were infrequent (6 days in 2008) and range from 0.6 - 7.8%. Overall, the 1-day mean appears to be protective of the instantaneous minimum DO criterion for deep water in the Lower York River. Failures were observed only during 2008. The 1-day Mean DO criterion for deep water during summer months is 2.3 mg/L.



10th percentile of instantaneous observations in relation to 1-day mean

Figure 8. The values of the 10th percentile of the instantaneous minimums fit with the 1-day mean for deep water with 95% confidence intervals about the fitted line and predicted observations. Although there are not any observed values for the 1-day mean that below the criterion (2.3 mg/L), the regression line crosses near 1.5 mg/L (instantaneous) at 3 mg/L 1-day mean. This suggests that for deep water the 1-day mean may not be protective of instantaneous minimum when it is near the criterion. (i.e. if the 1-day criterion is barely satisfied, there are likely greater than 10% violations of the instantaneous minimum criterion)

Bivariate Fit of 10th percentile of instantaneous observations by 1-day mean (Deep Water) (Fig. 8b)

Linear Fit

10thPercentile_DW = -1.816817 + 1.1376489*Mean daily DO (DW)

Summary of Fit

RSquare	0.695015
RSquare Adj	0.693672
Root Mean Square Error	0.558112
Mean of Response	4.420229
Observations (or Sum Wgts)	229

Analysis of Variance

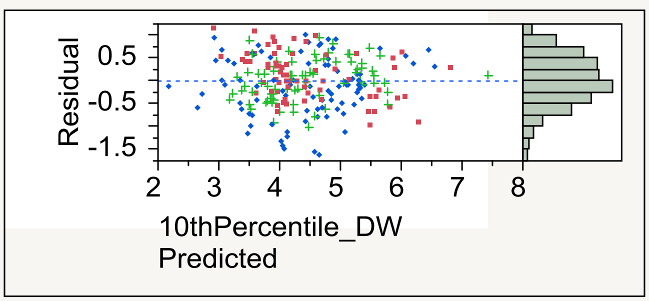
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Ratio**
Model	1	161.13303	161.133	517.2994
Error	227	70.70798	0.311	**Prob > F**
C. Total	228	231.84101		<.0001*

Parameter Estimates

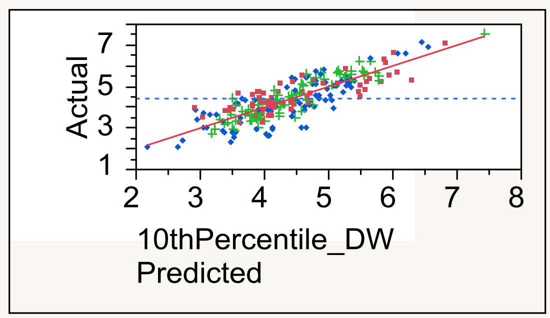
Term	**Estimate**	**Std Error**	**t Ratio**	**Prob>|t|**
Intercept	-1.816817	0.276695	-6.57	<.0001*
Mean Daily DO_DW	1.1376489	0.050019	22.74	<.0001*

Diagnostics Plots

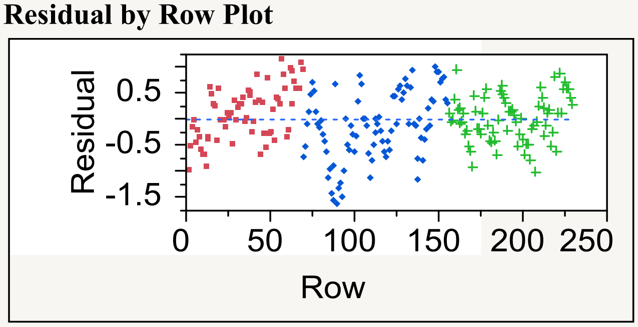
Residual by Predicted Plot



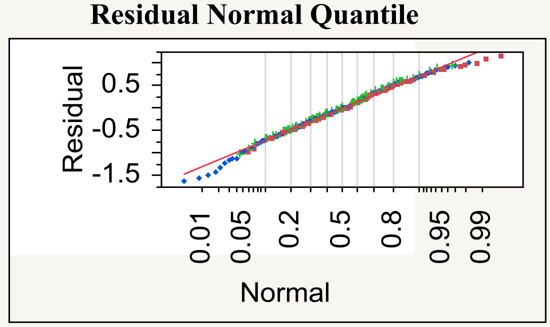
Actual by Predicted Plot



Residual by Row Plot



Residual Normal Quantile



Plot

II. Rappahannock River

Rappahannock River Profiler Observations

- A pycnocline was predominantly present at the Rappahannock River profiler station location, with a reduced presence in the fall (September).
- Overall, deep channel accounted for 27% (n=5,884), deep water accounted for 45% (n=9,746), and open water accounted for 29% (6,258) of the readings (**Fig 9**).
- For open water, there were 96 instantaneous violations (1.5% of the observations) in 2009 which occurred in June -early September (3rd). For deep water, there were 1,207 instantaneous violations (~12% of readings were in violation) in 2009 which occurred in June - early September (2nd). For deep channel, there were 1,976 violations (34% of the observations) of the instantaneous minimum criterion (1 mg/L).
- The 1-day, 7-day and 30-day means always met the designated DO criterion for open and deep water.

30 day mean in relation to 7-day mean violations (Open Water)

- There were no open water 7-day mean violations in a given 30-day period (month) in 2009. The 30-day mean DO criterion for open water during summer months is 5.0 mg/L (**Fig 10**).

30 day mean in relation to 1-day mean violations (Deep Water)

- There were no deep water 1-day mean violations in a given 30-day period in 2009 (**Fig 11**).

30-day mean in relation to instantaneous violations (Open and Deep Water)

- Open water DO instantaneous violations in a given 30-day period were < 5% in all months (range 0-3.3%) (**Fig 12**).
- Deep water DO instantaneous violations in a given 30-day timeframe ranged from 0-27%. The 30-day mean is not protective of the instantaneous minimum for deep water at the Rappahannock location. Violations exceeded 10% in June and August (**Fig 13**).

7-day mean in relation to instantaneous violations (Open Water)

- Rappahannock open water DO instantaneous violations in a given 7-day period are less than 7% (**Fig 14**).

10th percentile of instantaneous observations in relation to 7-day mean (Open Water)

- To assess the liklihood that the 7-day mean is protective of the instantaneous minimum for open water, the values of the 10th percentile of the instantaneous minimums were fit with the running 7-day mean. There were not any observed values for the 7-day mean that did not meet the open water criterion (5.0 mg/L). There was significant variability among observations and the regression suggests that if the 7-day mean nears 5 mg/L the instantaneous minimum violations will likely exceed 10%. However, additional data, particularly weekly means <6mg/L, are necessary to improve the fit (**Fig 15**).

<u>1-day mean in relation to instantaneous violations</u> (Deep Water)

- Rappahannock River deep water DO instantaneous violations in a given 1-day period were frequent (occuring 48 of the 74 data days in 2009) and range from 0.8 – 51.6% (overall mean = 12%). The 1-day mean was not protective of the instantaneous minimum DO criterion for deep water in the Rappahannock River. The 1-day mean DO criterion for deep water during summer months is 2.3 mg/L (**Fig. 16**).

<u>10th percentile of instantaneous observations in relation to 1-day mean (Deep Water)</u>

- The values of the 10^{th} percentile of the instantaneous minimum in relation to the 1-day mean for deep water with 95% confidence intervals about the fitted line and predicted observations. For observations near the 1-day criterion (2.3 mg/L), the 10^{th} percentile of observed instantaneous values was often < 1.7 mg/L). For deep water, as the 1-day mean nears the criterion, the instantaneous minimum violations exceed 10% (**Fig 17**).

RAPPAHANNOCK RIVER GRAPHICS

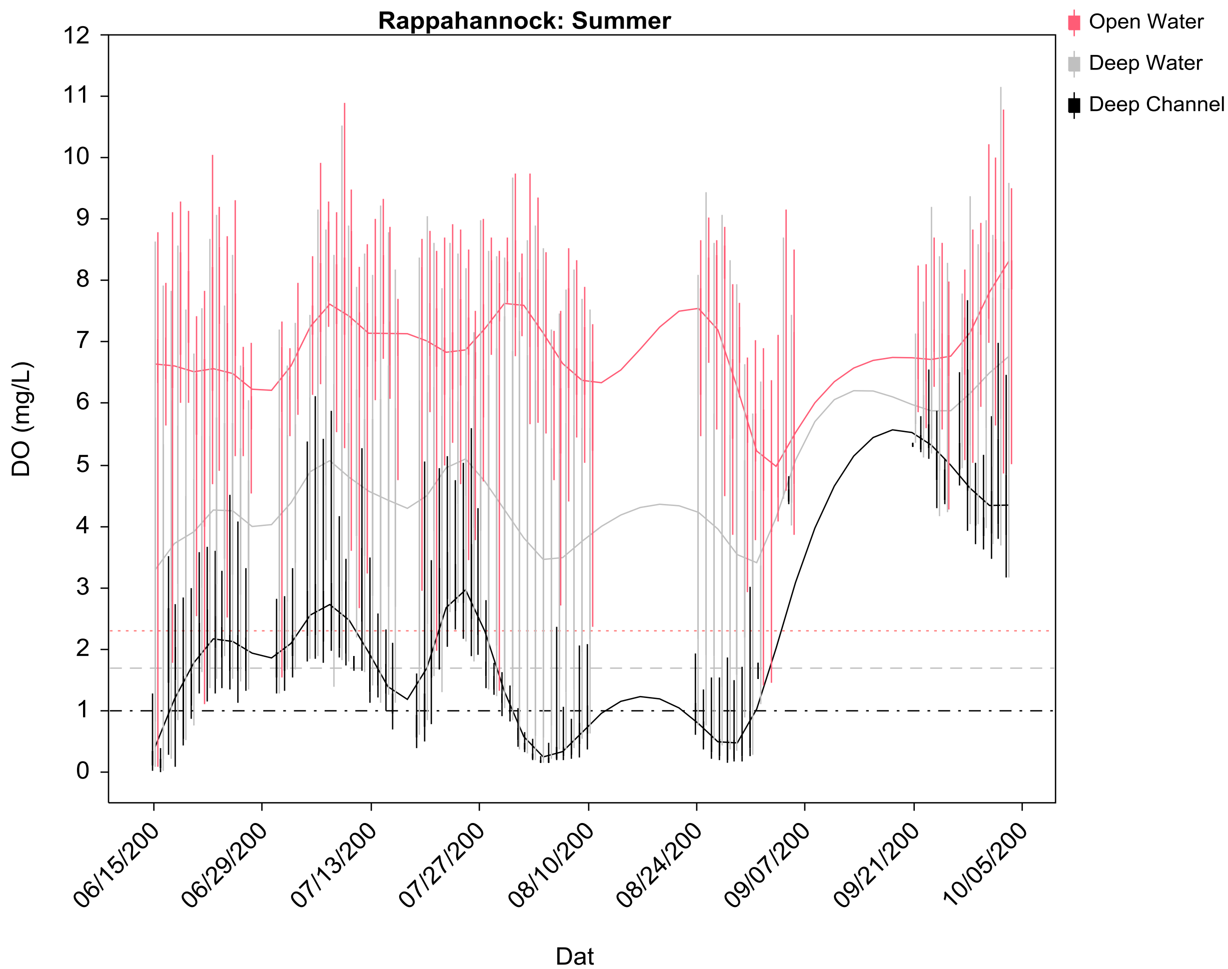


Figure 9. Rappahannock River Profiler Data – 2009. Quantile box plots of DO (mg/L) during 2009. Large gaps in dataset are due to profiler failures (e.g. 12 Aug – 24 Aug; 7 Sep – 21 Sept). Instantaneous minimum criteria are noted (3.2 mg/L for open water, 1.7 mg/L for deep water, 1.0 mg/L for deep channel). DO observations were highly variable on a daily basis.

30 day mean in relation to 7-day mean violations (Open Water)

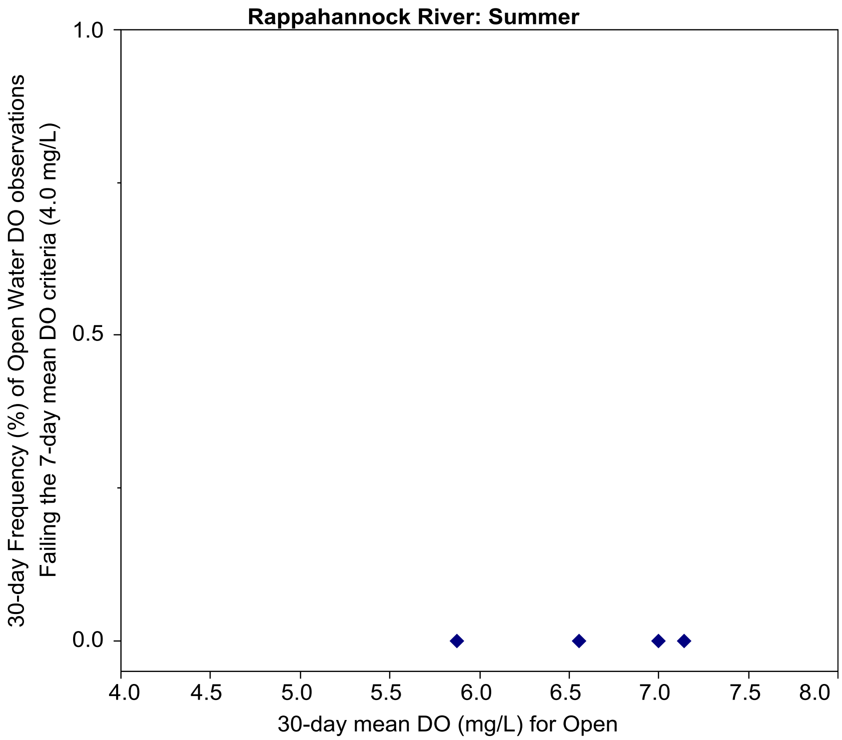


Figure 10. There were no open water 7-day mean violations in a given 30-day period (month) in 2009. The 30-day mean DO criterion for open water during summer months is 5.0 mg/L. However, the 30-day mean is probably close to the true 30-day mean since temporal high frequency measurements were used. Additional analysis is needed to satisfy the question of whether the fixed station data with limited observations per month show similar patterns.

30 day mean in relation to 1-day mean violations (Deep Water)

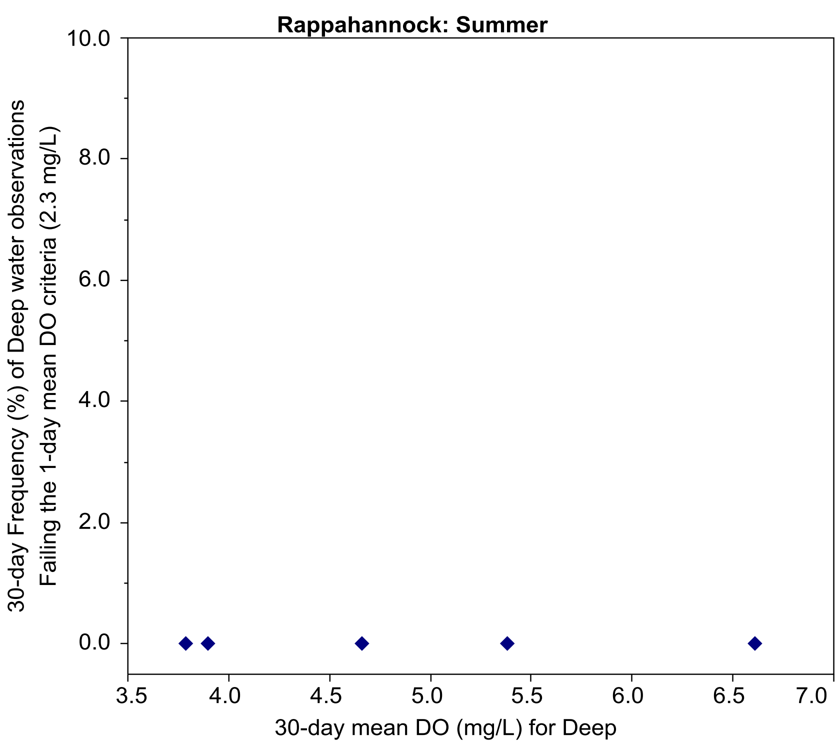


Figure 11. There were no deep water 1-day mean violations in a given 30-day period in 2009.

30-day mean in relation to instantaneous violations (Open and Deep Water)

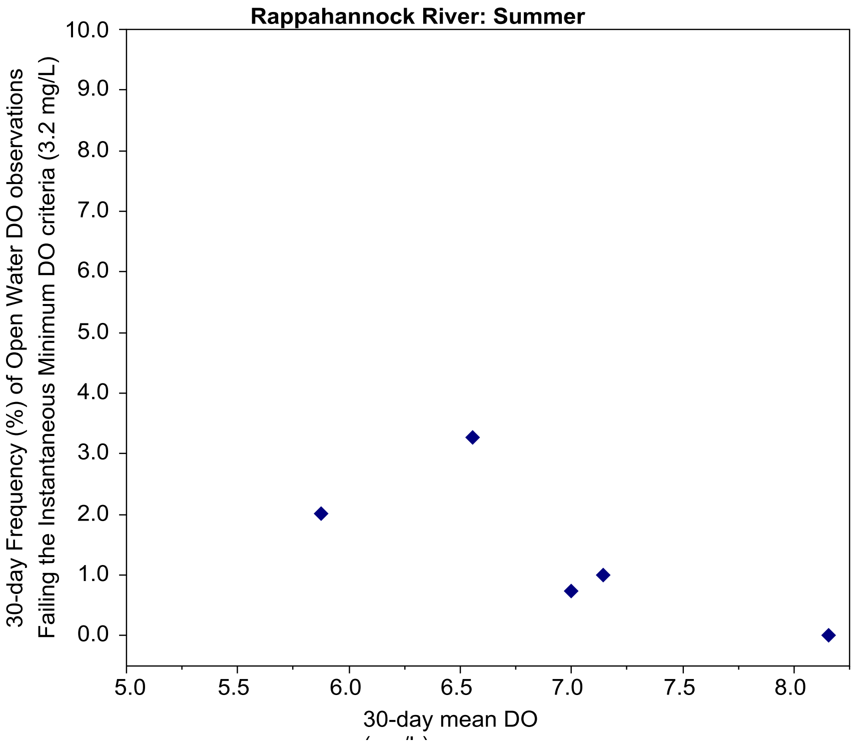


Figure 12. Open water DO instantaneous violations in a given 30-day period were < 5% in all months (range 0-3.3%).

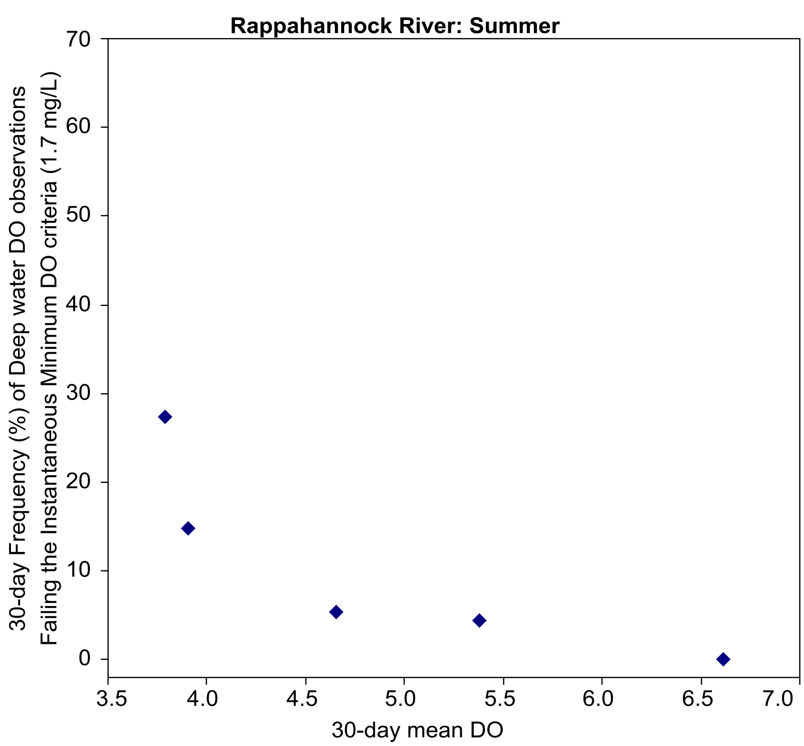


Figure 13. Deep water DO instantaneous violations in a given 30-day timeframe ranged from 0-27%. The 30-day mean is not protective of the instantaneous minimum for deep water at the Rappahannock location. Violations exceeded 10% in June and August.

7-day mean in relation to instantaneous violations (Open Water)

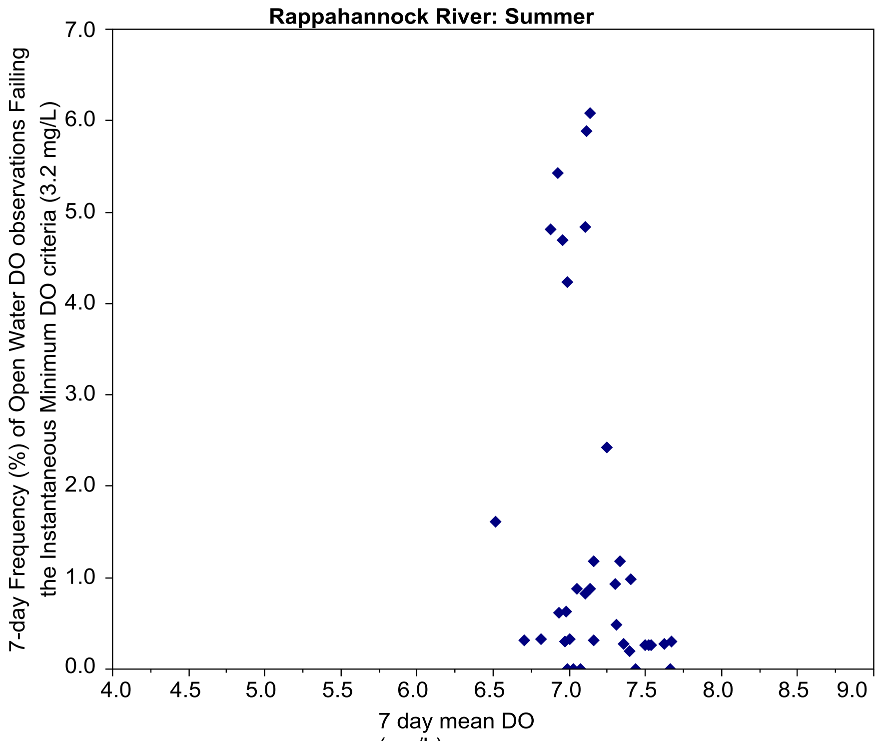


Figure 14. Rappahannock River open water DO instantaneous violations in a given 7-day period are less than 7%.

10th percentile of instantaneous observations in relation to 7-day mean (Open Water)

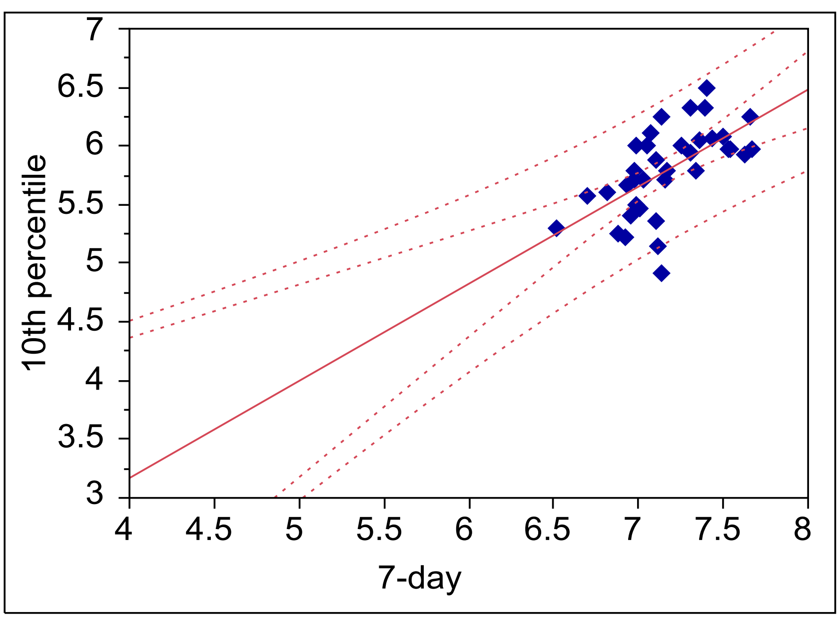


Figure 15. The values of the 10^{th} percentile of the instantaneous minimum in relation to the running 7-day mean for open water with 95% confidence intervals about the fitted line and predicted observations. There were not any observed values for the 7-day mean that did not meet the criterion (4.0 mg/L) and regression fit was not strong. The regression suggests that if the 7-day nears 4 mg/L the instantaneous minimum violations (< 3.2 mg/L) may exceed 10%. Additional data, particularly weekly means <6mg/L, should improve the fit.

Bivariate Fit of 10th percentile of instantaneous observations by 7-day mean (Fig 15b)

Linear Fit

10th percentile Instantaneous = -0.143134 + 0.8281057*7-day mean

Summary of Fit

RSquare	0.370674
RSquare Adj	0.352164
Root Mean Square Error	0.298216
Mean of Response	5.792431
Observations (or Sum Wgts)	36

Analysis of Variance

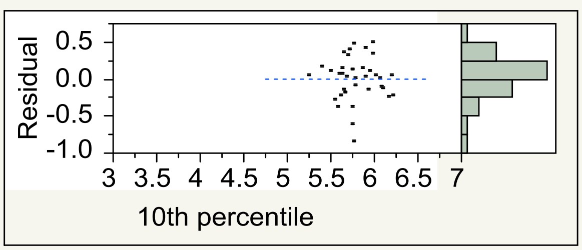
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Ratio**
Model	1	1.7809683	1.78097	20.0260
Error	34	3.0237128	0.08893	**Prob > F**
C. Total	35	4.8046811		<.0001*

Parameter Estimates

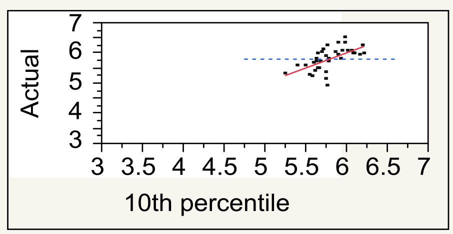
Term	**Estimate**	**Std Error**	**t Ratio**	**Prob>|t|**
Intercept	-0.143134	1.327301	-0.11	0.9148
sevendaymo	0.8281057	0.18505	4.48	<.0001*

Diagnostics Plots

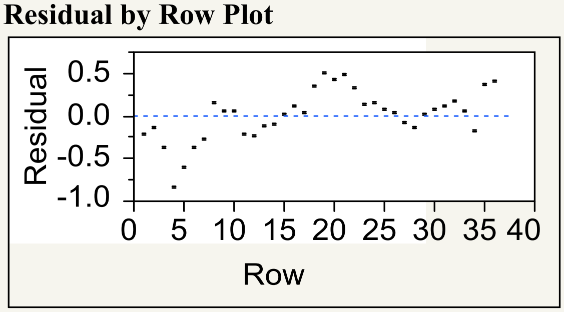
Residual by Predicted Plot



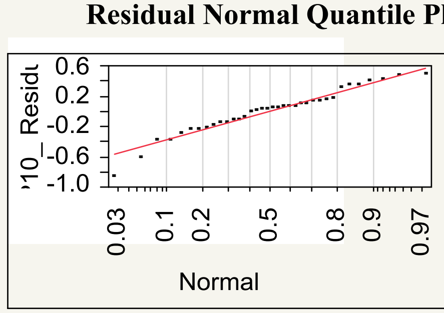
Actual by Predicted Plot



Residual by Row Plot



Residual Normal Quantile Plot



1-day mean in relation to instantaneous violations (Deep Water)

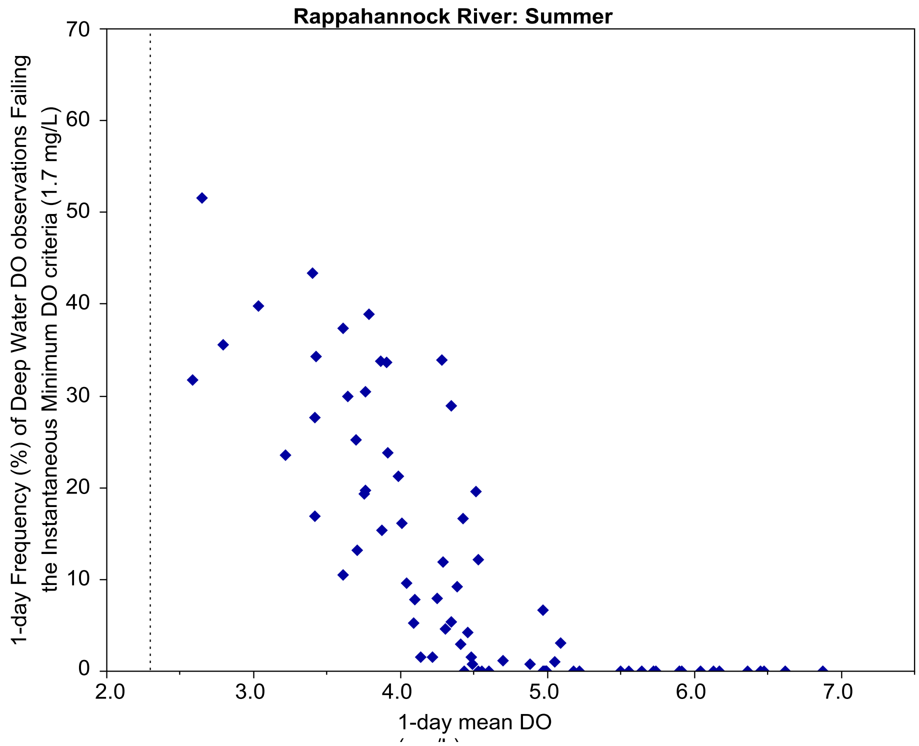


Figure 16. Rappahannock River deep water DO instantaneous violations in a given 1-day period were frequent (occuring 48 of the 74 data days in 2009) and range from 0.8 – 51.6% (overall mean = 12%). The 1-day mean was not protective of the instantaneous minimum DO criterion for deep water in the Rappahannock River. The 1-day mean DO criterion for deep water during summer months is 2.3 mg/L.

10th percentile of instantaneous observations in relation to 1-day mean (Deep Water)

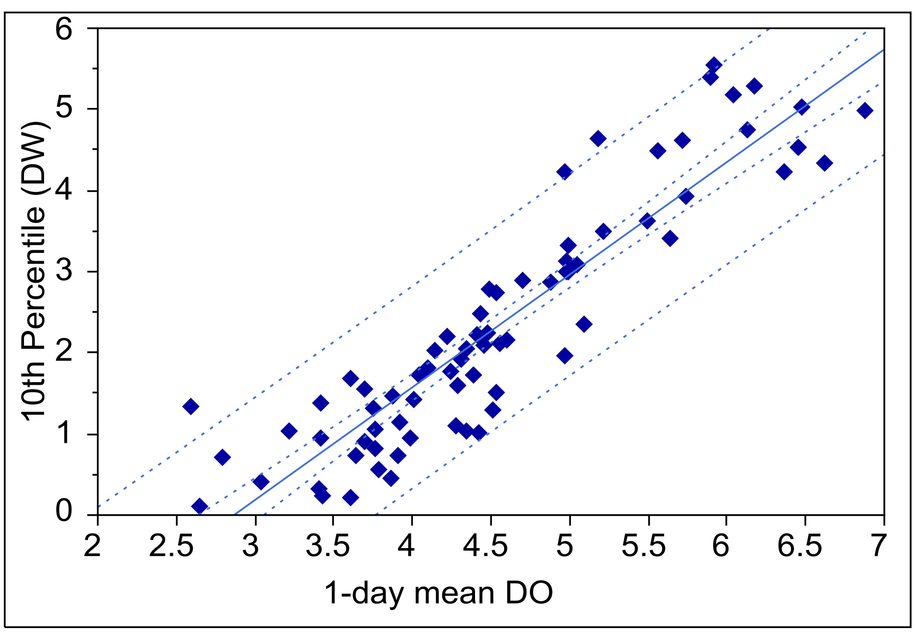


Figure 17. The values of the 10th percentile of the instantaneous minimum in relation to the 1-day mean for deep water with 95% confidence intervals about the fitted line and predicted observations. For observations near the 1-day criterion (2.3 mg/L), the 10th percentile of observed instantaneous values was often < 1.7 mg/L). For deep water, as the 1-day mean nears the criterion, the instantaneous minimum violations exceed 10%.

Bivariate Fit of 10th percentile of instantaneous observations by 1-day mean (DeepWater) (Fig. 17b)

Linear Fit

10thPercentile_DW = -3.978322 + 1.3884194*meandailyDO_DW

Summary of Fit

RSquare	0.829863
RSquare Adj	0.8275
Root Mean Square Error	0.619568
Mean of Response	2.328615
Observations (or Sum Wgts)	74

Analysis of Variance

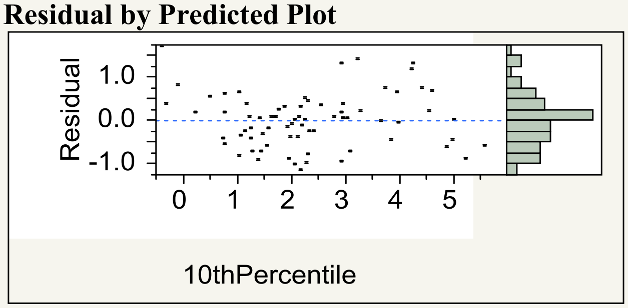
Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	134.80836	134.808	351.1878
Error	72	27.63821	0.384	**Prob > F**
C. Total	73	162.44657		<.0001*

Parameter Estimates

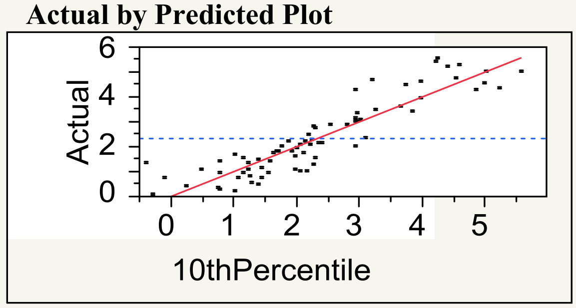
Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	-3.978322	0.34417	-11.56	<.0001*
Mean daily DO_DW	1.3884194	0.074089	18.74	<.0001*

Diagnostics Plots

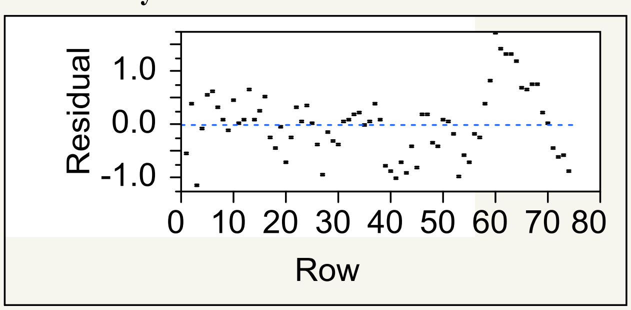
Residual by Predicted Plot



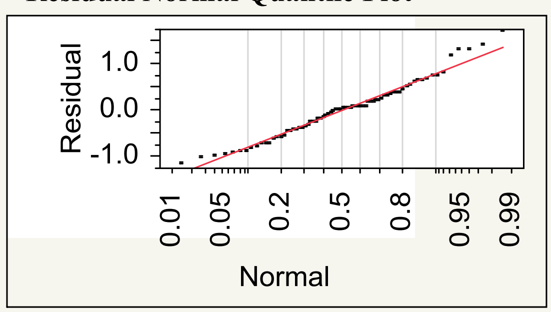
Actual by Predicted Plot



Residual by Row Plot



Residual Normal Quantile Plot



SUMMARY

While definitive conclusions for an entire system cannot be determined based on a single location, high frequency datasets can provide insight into spatial and temporal patterns and trends of dissolved oxygen. There is general support for the 30-day mean being protective over the 7 and 1-day mean in the York and Rappahannock rivers. However, there is questionable support for the 30-day mean being protective over the instantaneous minimum, particularly for deep water. One caveat is that the 30-day mean as calculated from the profiler data is probably close to the true 30-day mean since temporal high frequency measurements were used. Additional analyses are necessary to satisfy the question of whether the fixed station data with limited observations per month show similar patterns. For 7-day and 1-day mean observations near the criterion (4 and 2.3 mg/L, respectively), regression analysis suggests that these criteria may not be completely protective of the instantaneous minimum. While a single year of data was analyzed in the Rappahannock River, the Rappahannock River location appears to be more variable than the York River with higher percentages of violations. Additional years of data should be examined to determine if these patterns are consistent.

APPENDIX A in Appendix 6.

Standardized method for calculating upper and lower boundaries of the pycnocline using measurements of water temperature and salinity

Requires a vertical profile of salinity & water temperature measurements at multiple depths.

1. Sort the vertical profile of data from the surface downwards.
2. For each depth at which there are measurements, calculate a water density value as **σT**, or "sigma T", using water temperature and salinity measurements for that depth. Use the following method and equations:
   - $\sigma T = a(T) + b(T)*S$, where
   - T = temperature (°C); S = salinity; a & b are polynomial functions of T:
   - $a(T) = -9.22x10^{-3} + 5.59x10^{-2} * T – 7.88x10^{-3} * T^2 + 4.18x10^{-5} * T^3$
   - $b(T) = 8.04x10^{-1} – 2.92x10^{-3} * T + 3.12x10^{-5} * T^2$
3. Look down through the profile. Wherever the difference between sequential depth measurements is < 0.19 meters, average the 2 depth measurements and their corresponding salinity and density measurements.
4. Look down through the profile again. If there are still any depths (depth, salinity, temp and density measurements) < 0.19 meters apart, then average them again. Continue until there are no depths < 0.19 meters apart.
5. Starting at the surface and continuing until the deepest measurement in the profile, calculate the change in salinity and density between each sampling depth. For example, for 2 density values at 1 meter depth (y_1) and 2 meters depth (y_2) respectively, change in density, or $\Delta\sigma T = y_2 – y_1$. Likewise, for salinity measurements $\Delta S = y_2 – y_1$.
6. Assign a depth measurement to each pair of Δ values (ΔS, ΔσT) equal to the average of 2 depths used to calculate the Δ values. Thus for the 2 measurements y_2 and y_1, calculate accompanying depth as $(x_1 + x_2)/2$. You should now have a vertical profile of ΔS and ΔσT values with an accompanying depth.
7. To find the upper boundary of the pycnocline, look at the vertical profile of ΔσT, beginning with the 2nd value (from the surface) and excluding the 2 deepest values:
   a. IF ΔσT > 0.1,
   b. AND IF ΔσT for the *next* depth is greater than zero,
   c. AND IF ΔS > 0.1,
   d. Then this depth represents the upper boundary of the pycnocline.
8. Identify whether there is a lower mixed layer: use the same vertical profile but examine it from the 2nd deepest value upward (exclude deepest value):
   a. IF change in density (ΔσT) at the 2nd deepest depth < 0.2
   b. OR IF ΔσT at the next depth (moving upwards, i.e. shallower) < 0.2
   c. THEN a lower mixed layer (i.e. a layer at depth where the density is not changing) below the pycnocline exists.
9. If a lower mixed layer exists, then look for the lower boundary of the pycnocline. Beginning at the 2nd deepest value, and stepping up to the depth immediately below the upper pycnocline boundary, for ΔS and ΔσT values at each depth:
   a. IF ΔσT > 0.2,
   b. AND IF ΔS > 0.1,
   c. Then this depth is the lower pycnocline boundary.
10. If a pycnocline exists, then the upper and lower (if present) boundaries of the pycnocline have now been identified.

Appendix 7

Chapter 5 of the USEPA 2004
Olson et al.

Guidance for Attainment Assessment of Instantaneous Minimum and 7-day Mean Dissolved Oxygen Criteria

Available as a separate PDF on the Umbrella Criteria Workshop webpage

http://www.chesapeake.org/OldStac/umbrellacriteria/Appendix7-Chapter5ofUSEPA2004.pdf

Appendix 8

Evaluation of spectral analysis for dissolved oxygen criteria assessment

Matthew Hall
Maryland Department of Natural Resources

Introduction

As part of the goal of the 1987 Chesapeake Bay Agreement, the signatories agreed "to provide for the restoration and protection of living resources, their habitats and ecological relationships." Further, the Chesapeake Executive Council (CEC) committed to "develop and adopt guidelines for the protection of water quality and habitat conditions necessary to support the living resources found in the Chesapeake Bay system, and to use these guidelines in the implementation of water quality and habitat protection programs." A document was produced by the Chesapeake Bay Program outlining dissolved oxygen thresholds for various living resources (Jordan et al. 1992). The State of Maryland adopted these dissolved oxygen thresholds as standards in 1995 (COMAR 1995). Since the individual states within the Chesapeake Bay watershed are responsible for developing water quality standards under Section 303c of the Clean Water Act, the US Environmental Protection Agency (USEPA) later developed a guidance document outlining water quality thresholds for dissolved oxygen, water clarity and chlorophyll *a* (USEPA 2003). The recommended dissolved oxygen thresholds were derived from previous work, including the Jordan et al. report, and were based on designated uses of various Chesapeake Bay habitats (Table 1 in main document). Dissolved oxygen criteria were established for several different durations, ranging from instantaneous concentrations to 30-day means. Due to the lack of monitoring on short duration dissolved oxygen criteria, the authors asserted that "the assessment of attainment for some geographic regions and for some short-term criteria elements must be waived for the time being or must be based on statistical methods that estimate probable attainment." Proposed models included logistic regression of routine fixed-station data, which suffered from relying on once- or twice-a-month sampling at varying times within the natural diurnal cycle of dissolved oxygen concentration, and the creation of synthetic datasets from fixed-station data and short-duration continuous monitoring data, which suffered from a lack of short-duration data.

In or around 2001, the Maryland Department of Natural Resources (DNR) and the Virginia Department of Environmental Quality (DEQ) began deploying continuous monitors in shallow-water (<3 m) habitats within selected Chesapeake Bay tributaries. In Maryland, this led to the formation of the Shallow Water Quality Monitoring Program, which consisted not only of continuous monitoring on small temporal scales, but also intensive water quality mapping (DATAFLOW) on broad spatial scales. Over the succeeding years through the present, DNR, along with multiple partners, performed three-year assessments of Chesapeake Bay segments within Maryland's jurisdiction. DEQ and its partners continue to perform the same assessments in Virginia tidal segments.

DEQ subsequently performed a pilot study (Robertson 2009) using spectral analysis based on the work of Neerchal et al. (1994). This study showed the potential utility of combining short- and long-term datasets for assessment purposes for short duration criteria assessment (down to 7-days). In an effort to test the same method in Maryland waters, DNR conducted a similar case study in the lower Potomac River where shallow water monitoring was conducted between 2006 and 2008. The results were used to test the idea of the "Umbrella Criteria", i.e., that a single criterion (the 30-day mean) was protective of shorter duration criteria (7-day mean and instantaneous minimum).

Three questions arose from the VADEQ and DNR studies:

1. How does modeling long-term open water monitoring station data synthetically with short term continuous monitoring data match up with actual short term data from open water stations?
2. Is it appropriate to use surface continuous monitoring data to synthetically model deeper open water profile data?
3. What does short term open water data tell us about the umbrella criterion?

When the results of the DNR Lower Potomac spectral analysis study were presented to the Tidal Monitoring and Analysis Workgroup (TMAW) of the Chesapeake Bay Program, several members suggested using data from a continuous water quality profiler deployed in the deep water of the lower Potomac during the same time period to compare to modeled datasets. This analysis would yield a comparison of actual monitored data against modeled data in the same area.

This report describes the two studies discussed above. Part I describes the initial spectral analysis yielding synthetic data used to evaluate umbrella criteria, and Part II describes the subsequent deep-water vertical profiler sampling comparative analysis.

Part I: Evaluation of umbrella dissolved oxygen criteria using synthetic datasets

Methods

For this case study, the methods established by VADEQ (Robertson 2009), based on earlier work by Neerchal et al. (1994), were used to evaluate the umbrella criterion. Use of the same methodology insured comparability between the states and decreased the time necessary to perform the analyses. The author is indebted to Tish Robertson and Elgin Perry for guidance and troubleshooting of the methodology. For a full description of the statistical method, please see Robertson (2009), Neerchal et al. (1994), or earlier descriptions in this report.

This case study utilized two fixed monthly stations paired with nearby continuous monitors in the mesohaline segment of the Potomac River (POTMH; Figure 1). The first was fixed station LE2.2 paired with continuous monitor XBE8396 (Piney Point). The second was fixed station RET2.4 paired with continuous monitor XDC3807 (Pope's Creek). Within each year of the three-year assessment (2006-2008), the data were trimmed to the putative summer season (June through September) both to limit the analysis to a single dissolved oxygen criterion (see Table 1 in main document) and to decrease computing time. Only surface data were used in the models since the continuous monitors were fixed at the surface.

The purpose of this case study was to evaluate the only dissolved oxygen criterion currently used by the Chesapeake Bay Program (30-day mean) as an umbrella criterion for lower frequency criteria (7-day mean or instantaneous minimum). The synthetic datasets were evaluated by comparing the frequency of failing the higher frequency criterion versus the frequency of failing the lower frequency (umbrella) criterion. In this case, the 30-day mean criterion was tested as an umbrella criterion for the 7-day mean, and the 7-day mean was tested as an umbrella criterion for the instantaneous minimum. These evaluations were based on those performed by Buchanan on the combined DNR and VADEQ Potomac continuous monitor data (Buchanan 2009).

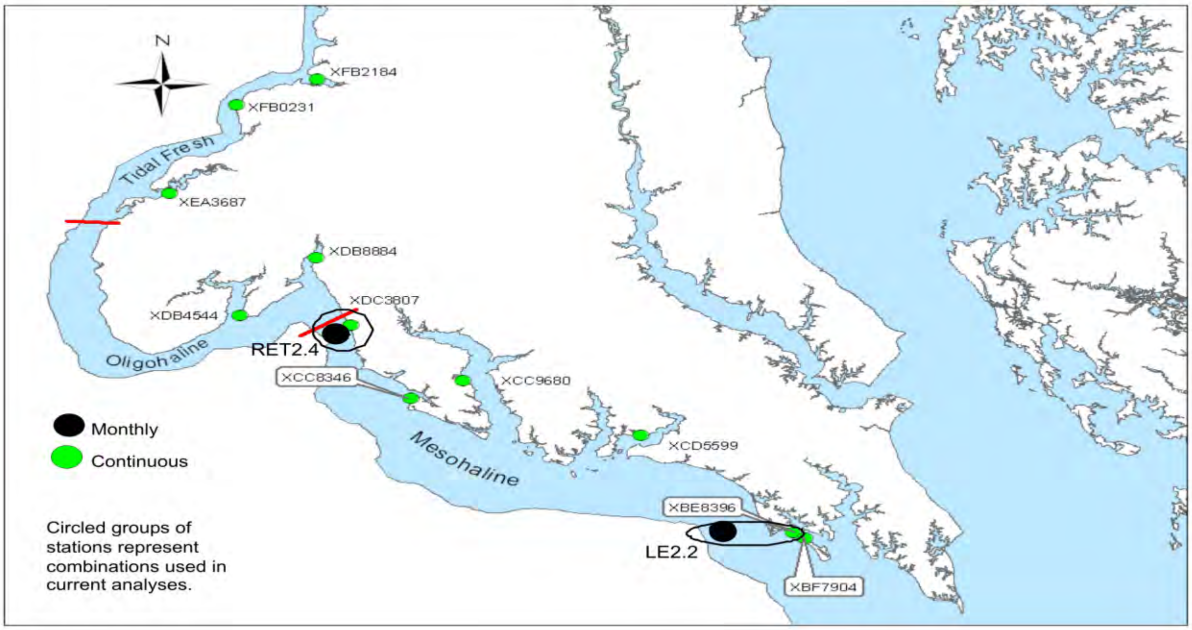


Figure 1: Map showing grouped stations used in the analysis, as well as all continuous monitors deployed by DNR in the tidal Potomac River between 2006 and 2008.

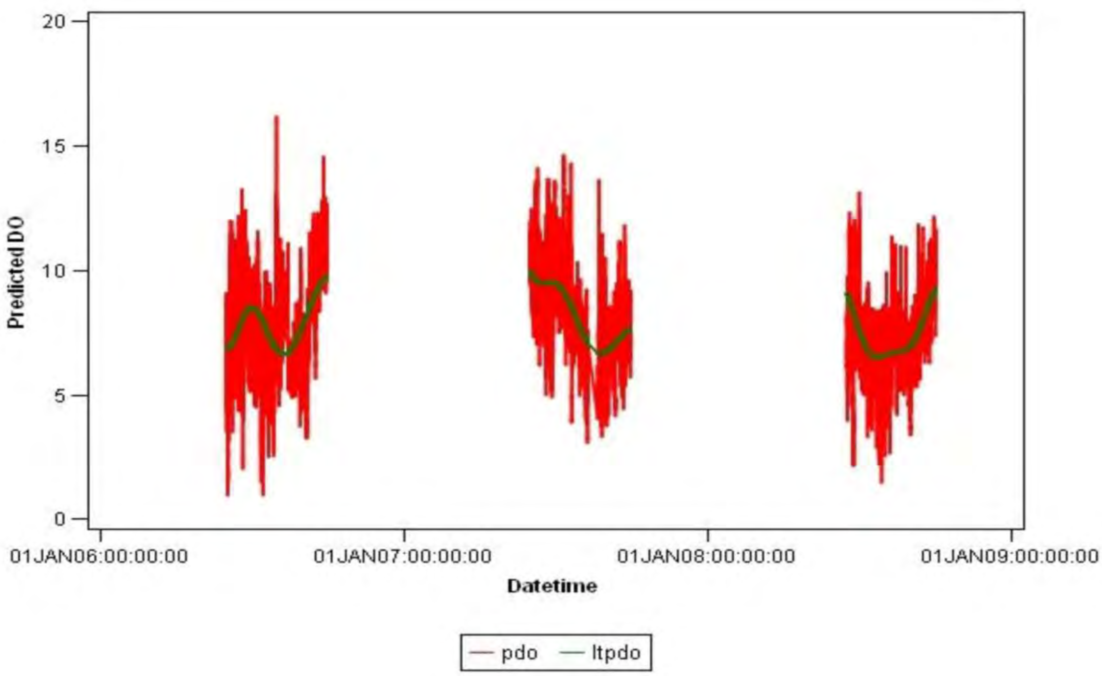


Figure 2: Modeled synthetic data from station LE2.2-Piney Point. The predicted dissolved oxygen (pdo) is overlaid with the long-term predicted dissolved oxygen (ltpdo) for comparison.

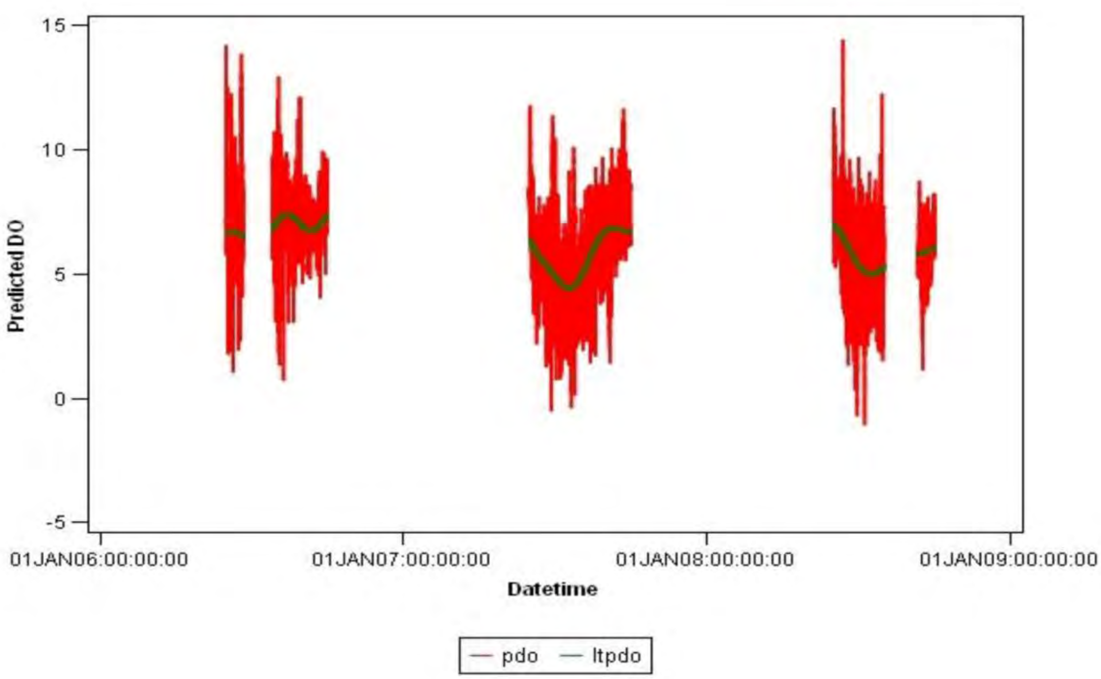


Figure 3: Modeled synthetic data from station RET2.4-Pope's Creek. The predicted dissolved oxygen (pdo) is overlaid with the long-term predicted dissolved oxygen (ltpdo) for comparison.

Results and Discussion

Modeled data for both station combinations are shown in Figures 2 and 3. These synthetic data were then tested to determine whether the 30-day mean criterion (5 mg/L) was protective of the 7-day mean (4 mg/L) and whether the 7-day mean criterion was protective of the instantaneous minimum (3.2 mg/L). Within each year, the period of record was divided into contiguous 30-day and 7-day segments. Periods less than 30 or 7-days at the end of each year were counted as a 30-day or 7-day period. To test the hypothesis that the 30-day mean criterion acts as an umbrella for the 7-day mean criterion, the frequency of violations of the 7-day mean within each 30-day period was plotted against the 30-day mean for each station (Figures 4 and 5). Similarly, the frequencies of failing the instantaneous minimum within each 7-day period were plotted against the 7-day means to determine how well the 7-day mean acted as an umbrella criterion for the instantaneous minimum (Figures 6 and 7).

At the two Potomac monthly monitoring stations during the summer season (June-September), no evidence of the 30-day mean not protecting the seven-day mean was present (Figures 4 and 5). The only time that the 7-day mean was violated, the 30-day mean was also violated (Figure 4). However, this study provided evidence that the 7-day mean criterion does not protect the instantaneous minimum all of the time at either station (Figures 6 and 7). The data suggest that the instantaneous minimum was violated during many 7-day periods where the 7-day mean was not violated. Buchanan (2009) also found that the 7-day mean criterion did not protect the instantaneous minimum in a survey of continuous monitoring data throughout the Potomac River and tributaries. Future analyses should evaluate a greater number of continuous monitoring/fixed monthly monitoring station combinations and utilize the Chesapeake Bay Program methodology for spatially evaluating Bay segments for compliance with dissolved oxygen criteria.

Conclusions

- The 30-day mean criterion is an effective umbrella for the 7-day mean criterion.
- The 7-day mean criterion is not an effective umbrella for the instantaneous minimum criterion.
- Caution is urged since only two stations were considered, and the synthetic datasets were not subjected to the spatially-interpolated criteria assessment used by the Chesapeake Bay Program.

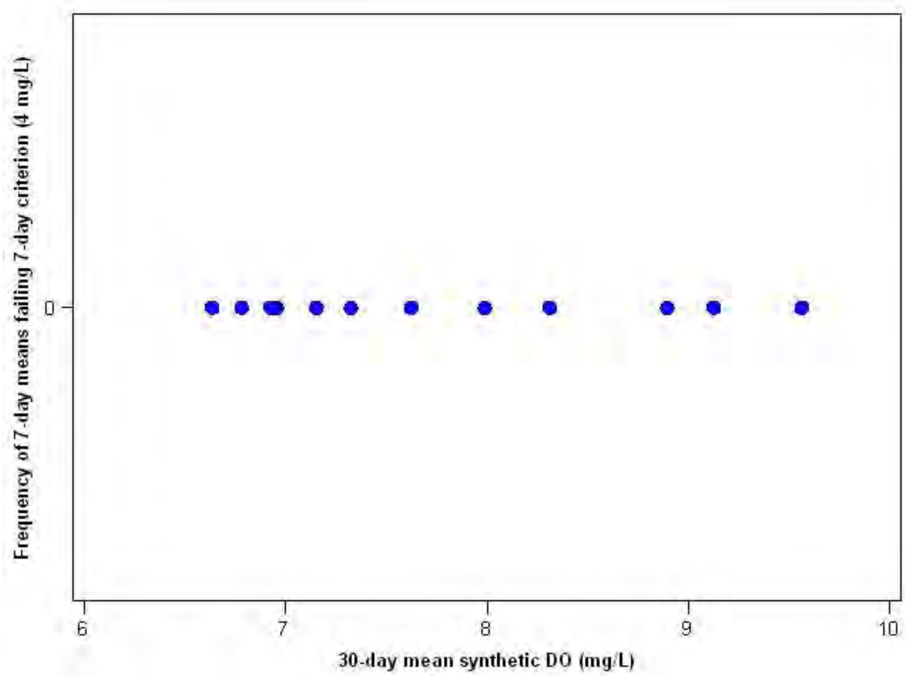


Figure 4: Dissolved oxygen seven-day mean failure rates for successive seven-day periods within each month (June-September 2006-2008) versus the 30-day mean for the same months using the LE2.2-Piney Point synthetic data. The 30-day mean criterion (5 mg/L) was never violated, nor was the 7-day mean criterion (4 mg/L). The 30-day mean criterion protected the 7-day mean criterion at this site grouping.

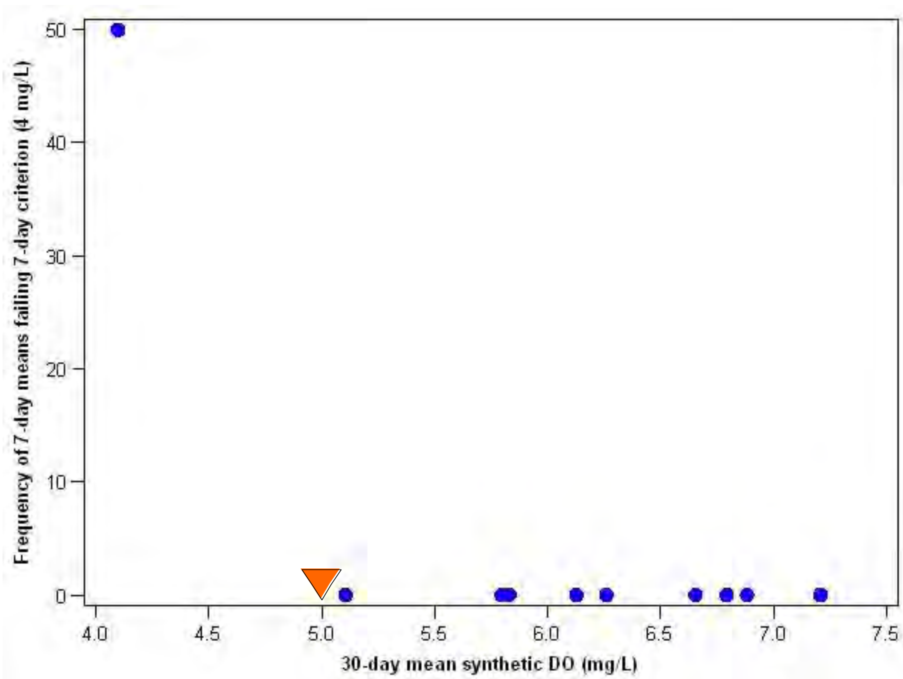


Figure 5: Dissolved oxygen seven-day mean failure rates for successive seven-day periods within each month (June-September 2006-2008) versus the 30-day mean for the same months using the RET2.4-Pope's Creek synthetic data. The orange triangle marks the 30-day mean criterion (5 mg/L). One mean violated this criterion, but half of the 7-day means also violated the 7-day mean criterion within that 30-day period. The 30-day mean criterion protected the 7-day mean criterion at this site grouping.

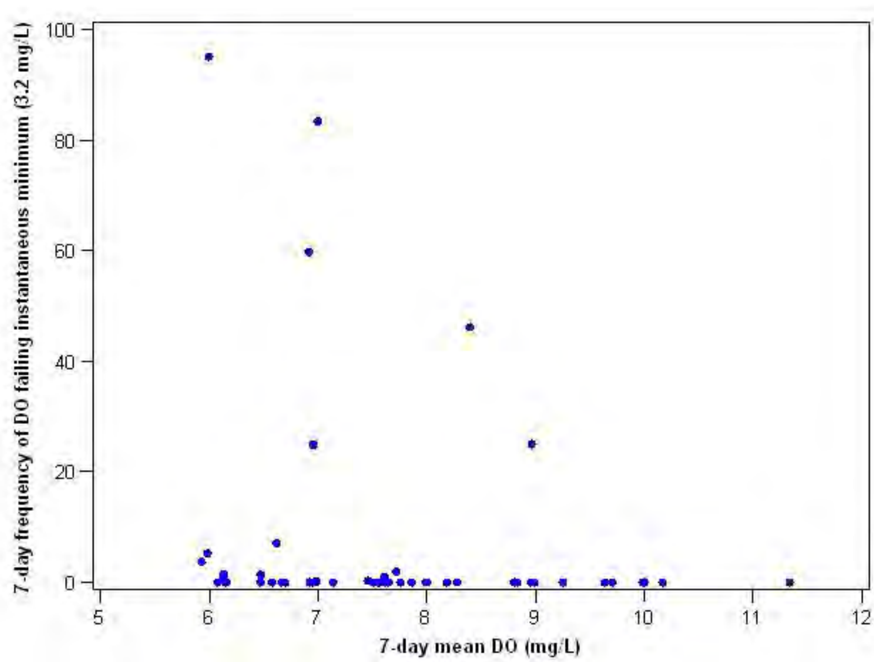


Figure 6: Dissolved oxygen instantaneous minimum failure rates for successive seven-day periods versus mean dissolved oxygen for the same seven-day periods using the LE2.2-Piney Point synthetic data. All of the 7-day periods passed the 7-day mean criterion, but many did not pass the instantaneous minimum criterion. The 7-day mean criterion did not protect the instantaneous minimum criterion at this site during 2006-2008.

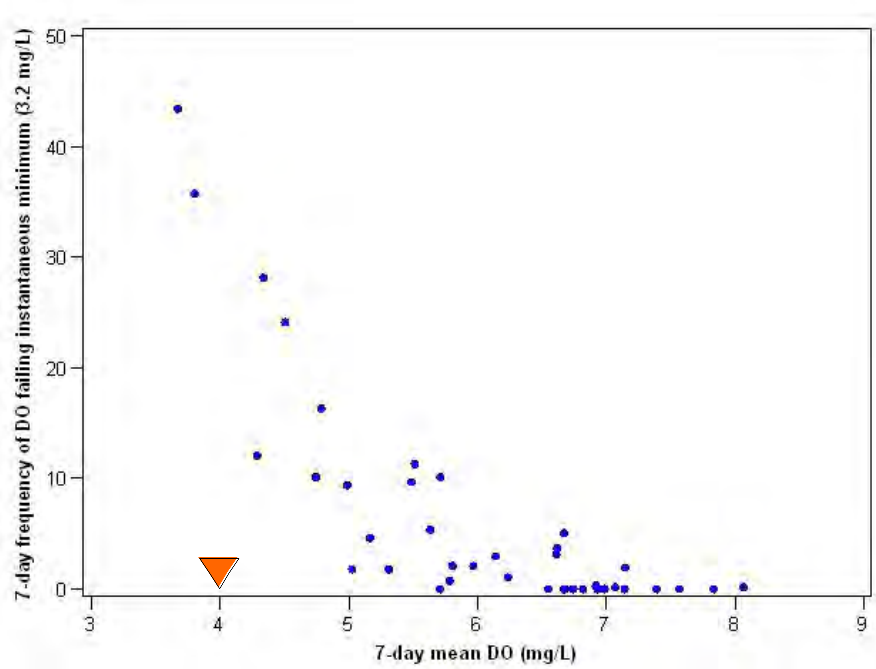


Figure 7: Dissolved oxygen instantaneous minimum failure rates for successive seven-day periods versus mean dissolved oxygen for the same seven-day periods using the RET2.4-Pope's Creek synthetic data. . The orange triangle marks the 7-day mean criterion (4 mg/L). Many 7-day periods that pass the 7-day mean criterion contain failures of the instantaneous minimum. The 7-day mean criterion did not protect the instantaneous minimum at this site during 2006-2008.

Part II: Evaluation of spectral analysis for dissolved oxygen criteria assessment using open water vertical profiler data

Methods

In order to test the suitability of spectrally modeling shallow water continuous monitoring data with mid-channel deep water data, data from a vertical profiler deployed in the lower Potomac River in 2009 were used. The profiler was deployed near Sandy Point in the lower Potomac River (Figure 8) in approximately 15 meters of water. Water quality samples (dissolved oxygen, pH, temperature, salinity, turbidity, chlorophyll) were collected at 1, 3, 5, 7, 9, and 11 meters every three hours between June and November. 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. Examination of the resulting datasets revealed long periods (days to weeks) of missing data after August 16th. Data between June 22 and August 16 were used for all comparisons to eliminate these data gaps. 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. In this case, 30-day and 7-day means were calculated as moving averages over the entire period of record (June 22nd – August 16th). Moving averages were used because TMAW was evaluating the use of moving averages versus fixed averages during the time that these analyses were performed. The resulting frequencies were then compared to the failure frequencies for the vertical profiler data at each depth.

The Chesapeake Bay Program dissolved oxygen criteria differ for surface and deep water, defined as below an established pycnocline. The presence and depth of a pycnocline in the vertical profiler data were determined using the Chesapeake Bay Program Calculated Threshold Value equation (Maryland Department of Natural Resources 2009). A pycnocline was found in approximately 80 percent of the profiler samples. The median upper pycnocline was found to be 6.0615 meters, and the median lower pycnocline was found to be 7.9855 meters. Therefore, deep water criteria were used to evaluate depths below 6 meters, with the knowledge that the 7 meter depth would fall within the pycnocline a large portion of the time.

In a number of cases, synthetic datasets yielded no significant violations of the criterion in question. In other cases, all of the synthetic data violated the criterion in question. To gain further insight into the criteria levels where violations (or no violations) would begin to occur, the synthetic datasets were tested for violations when criteria were incrementally raised or lowered. Increments of 0.5 mg/L DO were used, except in the case of surface instantaneous minimum (3.2 mg/L) where the level was lowered to 3.0 mg/L or raised to 3.5 mg/L.

.

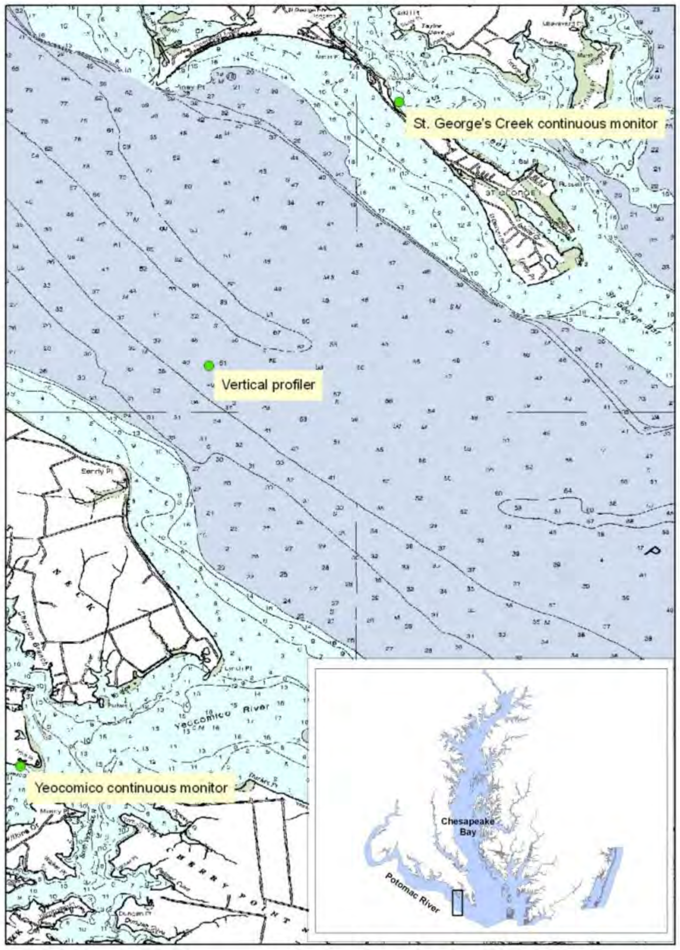


Figure 8: Map representing the lower Potomac River, two shallow water continuous monitor stations, and the 2009 location of the vertical profiler. The inset shows the site (rectangle) within the context of Maryland portion of the Chesapeake Bay.

Results and Discussion

Synthetic datasets modeled the simulated long-term data sampled from each depth at the vertical profiler with the variation from the short-term continuous monitors in Maryland and Virginia. Figure 9 shows an example of a synthetic dataset. Note the large section of missing data in August and September that necessitated cutting the datasets at August 16th.

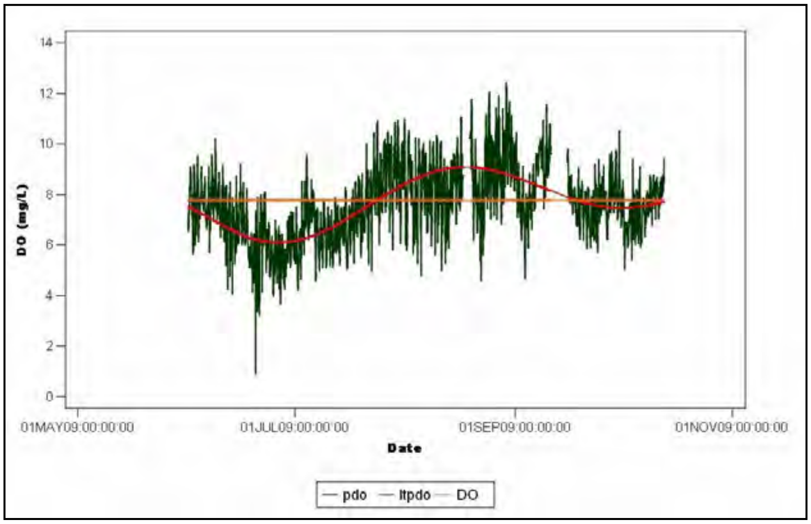


Figure 9: Synthetic data produced using spectral analysis for the St. George's Creek/ sampled vertical profiler at 1 meter depth. Legend abbreviations are as follows: pdo=predicted dissolved oxygen (from the synthetic dataset), ltpdo=long-term predicted dissolved oxygen (from the sampled vertical profiler modeled data), DO=mean dissolved oxygen (from the sampled vertical profiler data).

The synthetic data for the St. George's Creek/modeled vertical profiler station did not violate the 30-day mean open water or deep water criteria (Table 1). However, the real data from the vertical profiler did not violate the 30-day mean criteria either.

Depth	Synthetic data %	Profiler continuous data %
1 m	0	0
3m	0	0
5m	0	0
7m	100	100
9m	100	100
11m	100	100

Table 1: St. George's Creek/Sandy Point Profiler Synthetic Dataset percent failures of 30-day mean. Grey background=open water criterion 5 mg/L, blue background=deep water criterion 3 mg/L.

Comparison of the 7-day (open water) and 1 day (deep water) mean criteria produced similar results (Table 2). The only discrepancy occurred at the 7 meter depth where the percentage failures differed greatly. This may be attributable to the 7 meter depth being within the pycnocline where greater dissolved oxygen variability might overly influence the modeled data under the 1 day mean criteria.

Depth	Synthetic data %	Profiler continuous data %
1 m	0	0
3m	0	0
5m	0	0
7m	96.67	3.03
9m	100	100
11m	100	100

Table 2: St. George's Creek/Sandy Point Profiler Synthetic Dataset percent failures of 7-day average (open water) or daily average (deep water). Grey background=open water criterion 4 mg/L, blue background=deep water criterion 2.3 mg/L.

Under the instantaneous minimum criteria, many discrepancies between the modeled and real data occurred (Table 3). 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.

Depth	Synthetic data %	Profiler continuous data %
1 m	0	2.68
3m	0	3.13
5m	0	6.5
7m	60.57	41.87
9m	69.49	88.57
11m	92.11	96.9

Table 3: St. George's Creek/Sandy Point Profiler Synthetic Dataset percent failures of instantaneous minima. Grey background=open water criterion 3.2 mg/L, blue background=deep water criterion 1.7 mg/L.

Tables 4 through 6 show percent failure of criteria for the Yeocomico Creek/ Profiler synthetic data.

Depth	Synthetic data %	Profiler continuous data %
1 m	0	0
3m	0	0
5m	0	0
7m	100	100
9m	100	100
11m	100	100

Table 4: Yeocomico Creek/Sandy Point Profiler Synthetic Dataset percent failures of 30-day average. Grey background=open water criterion 5 mg/L, blue background=deep water criterion 3 mg/L.

Depth	Synthetic data %	Profiler continuous data %
1 m	0	0
3m	0	0
5m	0	0
7m	93.94	3.03
9m	96.97	100
11m	100	100

Table 5: Yeocomico Creek/Sandy Point Profiler Synthetic Dataset percent failures of 7-day average (open water) or daily average (deep water). Grey background=open water criterion 4 mg/L, blue background=deep water criterion 2.3 mg/L.

Depth	Synthetic data %	Profiler continuous data %
1 m	0.38	2.68
3m	0.44	3.13
5m	0.87	6.5
7m	56.57	41.87
9m	67.11	88.57
11m	87.89	96.9

Table 6: Yeocomico Creek/Sandy Point Profiler Synthetic Dataset percent failures of instantaneous minima. Grey background=open water criterion 3.2 mg/L, blue background=deep water criterion 1.7 mg/L.

Both synthetic datasets (St. George's/profiler and Yeocomico/profiler) showed identical failures for the 30-day mean criteria and nearly identical failure percentages for 7-day mean and instantaneous minima criteria. The Yeocomico monitor was significantly closer to the profiler, while the St. George's Island monitor was not only a greater distance away, but also located in an isolated area (Figure 8). Greater differences between the two synthetic datasets were expected, although the results suggest that the mid-channel vertical profiler data had greater influence on the model. Direct comparison of the St. George's Creek and Yeocomico Creek continuous monitoring datasets may also yield insight as to why the modeled datasets exhibited such similar criteria failure percentages.

Evaluation of the lower frequency (30 and 7-day mean) criteria as umbrella criteria for higher frequency data yielded mixed results. Above the pycnocline, no violations of the 7-day or 30-day mean occurred (Table 1 compared to Table 2 and Table 4 compared to Table 5). This may indicate that the 30-day mean criteria protect the 7-day mean criteria, but this is not certain since there were no violations. Below pycnocline, differences in failure frequency between the 30-day and daily mean criteria are evident (Table 1 compared to Table 2 and Table 4 compared to Table 5). Since the 30-day criteria were violated 100 percent of the time and the 1-day mean was violated generally less than that, the 30-day mean may be protective of the 1-day mean. There is less evidence that the instantaneous minimum criteria are protected by either the 30-day or 7-day mean criteria, especially when considering the actual vertical profiler data (Figures 3 and 5 compared to Figures 1-2 and 4-5 respectively). Except for the St. George's Island synthetic data, where no violations of the instantaneous minimum above pycnocline were found (Figure 3), violations of instantaneous minima occurred when no violations of 30 or 7-day mean criteria occurred. Below pycnocline, the 30-day and 1-day means tended to be violated 100 percent of the time, inferring that they protected the instantaneous minimum. An exception to this occurred at the 7 meter depth; however, this depth was generally located within the pycnocline where greater variability would be expected.

The following two charts (Tables 7 and 8) show percent failures for increasing criteria values for above pycnocline Yeocomico/profiler data and for decreasing criteria values below pycnocline. Once 100 percent failure was reached for above pycnocline and once 0 percent failure was reached for below pycnocline data, the analysis was stopped.

Above pycnocline, increasing criterion levels (from 5 mg/L), % failure							
Depth	**5.5**	**6**	**6.5**	**7**	**7.5**	**8**	**8.5**
1	0 (0)	0 (0)	0 (0)	29.6 (1.9)	55.5 (1.9)	74.1 (36.5)	100 (71.1)
3	0 (0)	0 (0)	0 (0)	22.2 (2.8)	51.9 (20.2)	81.5 (77.9)	100 (100)
5	0 (0)	0 (9.6))	22.2 (79.8)	55.5 (93.2)	100 (100)	100 (100)	100 (100)

Below pycnocline, decreasing criterion levels (from 3 mg/L), % failure						
Depth	**2.5**	**2**	**1.5**	**1**	**0.5**	**0**
7	100 (9.6)	96.3 (0)	81.5 (0)	59.3 (0)	0 (0)	0 (0)
9	100 (71.1)	100 (65.4)	81.5 (54.8)	63 (44.2)	51.9 (19.2)	33.33 (0)
11	100 (86.5)	100 (80.8)	100 (75)	100 (65.4)	74.1 (48.1)	37.04 (0)

Table 7: 30-day mean percent failures for increasing criteria levels (top table) and decreasing criteria levels (bottom table) in the profiler/Yeocomico synthetic data. Corresponding percent failures for the actual profiler data are shown in parentheses.

Above pycnocline, increasing criterion levels (from 4 mg/L), % failure												
Depth	**4.5**	**5**	**5.5**	**6**	**6.5**	**7**	**7.5**	**8**	**8.5**	**9**	**9.5**	**10**
1	0 (0)	0 (0)	6 (0)	16 (0)	22 (0)	26 (5)	52 (23)	74 (44)	78 (70)	86 (89)	94 (95)	100 (100)
3	0 (0)	0 (0)	2 (0)	12 (0)	20 (1)	24 (15)	52 (39)	78 (71)	84 (89)	94 (100)	100 (100)	100 (100)
5	0 (0)	0 (5)	8 (14)	20 (28)	24 (50)	50 (82)	78 (92)	88 (100)	96 (100)	100 (100)	100 (100)	100 (100)

Below pycnocline, decreasing criterion levels (from 2 mg/L; crierion is daily mean of 2.3), % failure					
Depth	**2**	**1.5**	**1**	**0.5**	**0**
7	87.9 (19.3)	75.8 (11.6)	60.6 (6.3)	15.2 (0.9)	0 (0)
9	87.9 (64.9)	69.7 (63.9)	57.6 (56.5)	48.5 (40)	42.4 (0)
11	100 (74)	100 (69.6)	100 (40)	54.6 (0)	42.4 (0)

Table 8: 7-day mean percent failures for increasing criteria levels (top table) and daily average decreasing criteria levels (bottom table) in the profiler/Yeocomico synthetic data. Corresponding percent failures for the actual profiler data are shown in parentheses.

Some of the below pycnocline failure percentages in the synthetic data are greater than zero when the criterion reaches 0 mg/L dissolved oxygen. This is an artifact of spectral modeling where values around zero caused negative model values. This occurred at greater depths where more values around or at 0 mg/L dissolved oxygen would be expected.

For both 30-day and 7-day mean criteria, ranges between set criterion concentrations and concentrations of totality (100 percent failure above pycnocline and 0 percent failure below pycnocline) are large (generally several mg/L DO). This suggests that slight changes in criteria levels, which are based on biologically relevant thresholds, will not result in greater violation rates. Differences in failure percentages for incrementally increasing criteria between modeled data and actual data collected by the profiler were correlated overall ($r_{spearman}=0.92$, $p<0.0001$ above pycnocline; $r_{spearman}=0.75$, $p<0.0001$ below pycnocline). This indicates that the modeled data and actual data generally had similar percentage failure rates over all tested criteria levels, even though paired values at a few criteria levels differed greatly.

Conclusions

- For 30-day and 7-day mean criteria above pycnocline, the synthetic data tracks well with continuous profiler data at all depths.
- The agreement is less for instantaneous minima.
- Use of synthetic datasets to evaluate deep water criteria may not be appropriate, except perhaps for the 30-day mean criteria. This is especially true when only shallow water continuous monitors are available for the high frequency data component.
- Validates methodology (at this site, at least). Synthetic data captures variability of open water continuous data overall, though a few paired comparisons differ greatly.
- The umbrella criterion (30-day mean) may be protective of the 7-day mean criterion above pycnocline. Data that may have come from within the pycnocline obscure this determination below pycnocline. The instantaneous minimum may not be protected. The data from this report provide weak evidence of umbrella protection.

References

Buchanan, C. 2009. An analysis of continuous monitoring data collected in tidal Potomac embayments and river flanks. Interstate Commission on the Potomac River Basin, Report 09-3. Rockville, Maryland.

COMAR (Code of Maryland Regulations). 1995. Code of Maryland Regulations: 26.08.02.03 – Water Quality Criteria Specific to Designated Uses. Maryland Department of the Environment. Baltimore, Maryland.

Jordan, S., C. Stegner, M. Olson, R. Batiuk, and K. Mountford. 1992. Chesapeake Bay dissolved oxygen goal for restoration of living resources habitats. Chesapeake Bay Program, Reevaluation Report #7c. CBP/TRS88/93. Annapolis, Maryland.

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.

Maryland Department of Natural Resources. 2009. Quality assurance project plan for the Maryland Department of Natural Resources Chesapeake Bay water quality monitoring program – chemical and physical properties component for the period July 1, 2009 through June 30, 2010. Accessed January 12, 2011 from website: http://mddnr.chesapeakebay.net/eyesonthebay/documents/MD_DNR_MTQAPP09_final.pdf.

Robertson, T. 2009. Spectral analysis: a potential tool for short-term criteria assessment. Unpublished report. Virginia Department of Environmental Quality. Richmond, Virginia.

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.

Appendix 9, (Part A of A & B).

EPA Criteria Assessment-Based Approach to Spectral Casting: Evaluation of the 30-day mean "umbrella" hypothesis against the 7-day mean

Tish Robertson
Virginia Department of Environmental Quality
Michael Lane
Old Dominion University

Introduction

For more than a decade, EPA Bay Program and its partners have developed and refined dissolved oxygen (DO) standards designed to protect the integrity of the Bay's aquatic life designated use. A state-of-the-art methodology was established to assess these criteria. According to the July 2007 Addendum of the *Ambient Water Quality Criteria for Dissolved Oxygen, Water Clarity and Chlorophyll* a *for the Chesapeake Bay and Is Tidal Tributaries,* it was the goal of the Bay Program that this methodology be applied consistently to all water quality criteria, allowing for assessment of data at multiple scales, and be capable of incorporating as much data possible into its framework. In addition, the methodology should provide a clear basis for making decisions based on assessment results and assess conditions not only on a temporal scale (frequency of violations), but also on a spatial scale (volumetric or areal units of violation). The methodology designed for the Bay Program allows for nuances characteristic of natural conditions, through both time and space, providing for a more robust assessment than what is provided by very localized, temporally-insensitive assessments traditionally employed for smaller systems.

The assessment method used by the Bay Program for DO assessments is described in detail in section 2.1.1 of the main document. With respect to the Open Water subuse (see Table 1), this methodology has, to date, been used in the assessment of one criterion—the 30-day mean—while the applicable 7-day and instantaneous minimum criteria remain unassessed. This is primarily because the data currently used in assessments are collected at monthly or semi-monthly frequencies. While data collected at higher frequencies have been available for some time through a network of continuous shallow water monitors in Maryland and Virginia's tributaries, these data have been deemed inappropriate for assessment purposes.

However, the 30-day "umbrella" question has presented Virginia and Maryland with an opportunity to delve into these high-frequency datasets, using the spectral analysis approach initially introduced to the Program by Neerchal et al. (1994). To fully test the hypothesis that the 30-day mean criteria are more protective of the Open Water designated use than the 7-day mean, it is not enough to simply generate predictive time series for a few long-term stations, but to apply the complete assessment methodology used by the Bay Program on predicted short-term data.

This was achieved using models of the of high-frequency variability from the continuous monitors, "casted" onto the low-frequency data generated at a particular mid-channel "grab" station.) This complete assessment methodology requires that data from a station be spatially interpolated for each time interval (30-day or 7-day), the volumetric proportion of violations of each interpolated "snap shot" be calculated and ranked, and a cumulative frequency diagram be generated to determine excessive violation rates in space-time. Multiple stations and segments should be assessed in this way for a robust comparison of 30-day and 7-day exceedence rates.

Virginia is fortunate to have access to multiple years' worth of data collected from over twenty continuous monitoring stations, some having records going back as far as 2002. Most of these stations are maintained by Kenneth Moore's team at the Virginia Institute of Marine Science. The placement of these monitors is on a rotating schedule, so that for any given period, some segments are sampled while others are not. For the analysis be presented here, we analyzed data from two sets of monitors: those sampled for the years 2006 thru 2008 and those sampled from 2007 to 2009 (see Figure A6-1). Not all Virginia tributary segments were analyzed, however. Potomac embayment segments were not analyzed because it was felt that the continuous monitoring "sending" stations situated on the shoreline were too far away from their closest "receiving" mid-channel stations. In addition , the 7- day mean criteria are not applicable to the Pamunkey and Mattaponi segments due to the presence of naturally occurring hypoxia, so these segments were not analyzed. The lower Rappahannock and York Rivers were also excluded due to the complexities involved with separating the multiple designated uses (Open Water, Deep Water, and/or Deep Channel) that occur in these segments. See Table A6-1 for a full list of stations used in our analyses.

Figure A6-1. Map of the water quality monitoring stations in Virginia's tributaries used in comparing the results of the 30-day and the 7-day mean assessments for two temporal windows—2006-2008 and 2007-2009.

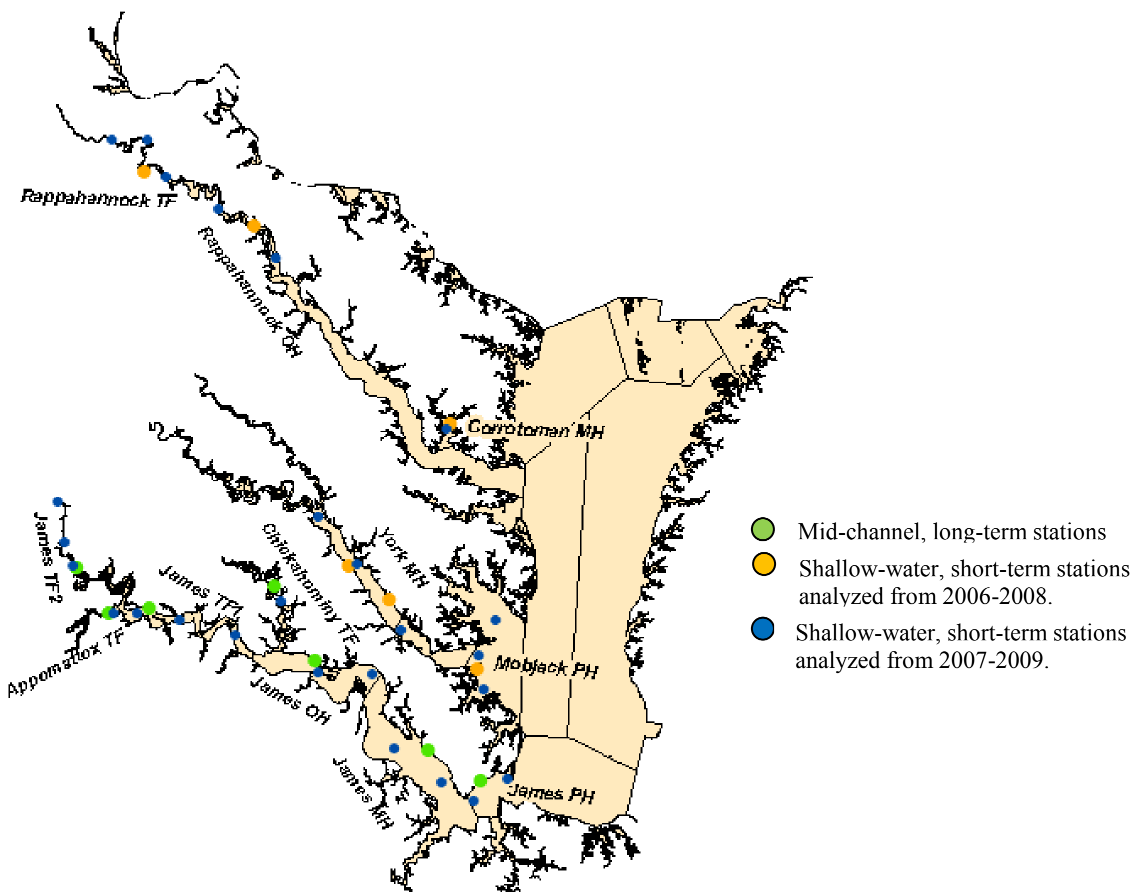


Table A6-1. List of mid-channel (long-term) and shallow-water, continuous monitoring (short-term) stations, by Bay segment. TF = tidal fresh, OH = oligohaline, MH = mesohaline, and PH = polyhaline

Segments/Years Analyzed	**Long-term ("receiving") stations**	**Short-term ("sending") stations**
Appomattox TF/2006-2008	TF5.4	APP001.83
Chickahominy OH/2006-2008	RET5.1A	CHK015.12
James TF2/2006-2008	TF5.2	JMS009.00
	TF5.2A	JMS009.00
	TF5.3	JMS009.00
James TF1/2006-2008	TF5.5	JMS073.37
	TF5.5A	JMS073.37
James OH/2006-2008	TF5.6	JMS043.78
	RET5.2	JMS043.78
	LE5.1	JMS043.78
James MH/2006-2008	LE5.2	JMS018.23
	LE5.3	JMS018.23
James PH/2006-2008	LE5.4	JMS002.55
	LE5.5-W	JMS002.55
York MH/2007-2009	RET4.3	TSK000.23
	LE4.1	TSK000.23
Mobjack Bay PH/2007-2009	WE4.1	CHE019.38
	WE4.2	CHE019.38
	WE4.3	CHE019.38
Corrotomon MH/2007-2009	LE3.3	CRR004.02
Rappahannock TF/2007-2009	TF3.2	RPP084.32
	TF3.1E	RPP084.32
	TF3.1B	RPP084.32
Rappahannock OH/2007-2009	TF3.2A	RPP057.00
	TF3.3	RPP057.00

We addressed three questions with our analyses:

1. Is the 30-day mean criteria more protective than the 7-day mean criteria for segments from a variety of locations and salinity regimes, during different time periods, using the EPA-criteria based assessment methodology?
2. Do the results of the above change when a rolling, rather than a sequential 7-day period, is used in the analysis?
3. Do the results depend on how the predictive data are generated (i.e., Fourier analysis versus spline interpolation)?

Part I. Protectiveness of the 30-day mean versus the 7-day mean criteria

Methodology

Generating the predictive DO daily time-series from "sending" to "receiving" stations:

Dissolved oxygen data collected during the summer months at both long-term, mid-channel stations and short-term, shallow-water stations were obtained from the Chesapeake Bay Program Water Quality Database and VIMS's Virginia Estuarine and Coastal Observation System (VECOS), respectively. The three most recent years of short-term data collected in each segment dictated the three-year window of long-term data that were analyzed. Sending and receiving stations were selected based on their proximity within the same Bay segment. Some sending stations were used to "cast" signals onto multiple receiving stations (see Table A6-1).

Synthetic data sets were created by conducting two separate spectral analyses, one each for the sending and receiving data sets, using the SAS (v.9.2) SPECTRA procedure. In spectral analysis, time series variation is decomposed into a set of N/2 cyclic components or frequencies that are equally spaced in time series. Sine and cosine coefficients are calculated to fit a sinusoid for each of these frequencies. The resulting coefficients can then be used to create separate predicted data sets for long term signal and short term signals which are combined to create a synthetic data set for the receiving stations. These synthetic data sets used subsequently for the interpolation and assessment analyses. Long-term signals were created from monthly means of mid-channel stations for the three year period that corresponded to the same three year period for which shallow water data were available. In order to provide for equally spaced data points in the long-term data sets, missing values were replaced with the monthly mean value for the three year period. Predicted values for short-term data sets were generated on a monthly basis each station and year. Synthetic data sets were not generated for any assessment periods with less than a total of 5 days of continuous data. Data sets with 24 hours or more of missing data were divided and the longest of the two resultant series was used as the assessment window and used to generate synthetic data. For some data sets, short periods (< 6 hours) of missing data were replaced with mean of values collected at corresponding hours of the previous and next day in the data set.

EPA criteria-based approach to 7-day mean assessment, using the predictive DO data:

- The three-year window (summer 2006-2008 or summer 2007-2009) was partitioned into sequential 7-day intervals for each segment. If the predictive dataset was not complete (i.e., it had breaks due to missing data in the short-term dataset), then a minimum of 5 consecutive days was treated as a sufficient interval. The maximum number of data points in a predictive dataset with no gaps was 35,136 (representing a summertime DO value occurring every 15 minutes).

- For each 24-hour period, at each receiving station, the predictive dataset was averaged, thereby generating a 1-day mean. These days were then averaged together for a 7-day mean. "Week 1" was started on June 1 of the first year of the assessment window and continued to "Week 51", ending September 27th at end of the last year of the assessment window. September 28th to 30th were always excluded in the averaging. Two predictive time series were generated for each receiving station—one for the shallow layer and one for the bottom layer.
- The predicted 7-day means for both shallow and bottom depths at receiving stations were spatially interpolated using the Bay Interpolator, which uses Inverse Distance Weighting to generate estimated values through vertical and horizontal space. Predictions based on 2006-2008 data were interpolated separately from predictions based on 2007-2009 data. Each 7-day interval had its own interpolated three-dimensional "snap-shot."
- Each grid was run through an assessment binary (≥4 mg/l = PASS, <4 mg/l = FAIL) and each segment in Table A6-1 was assessed. The number of volumetric estimates varies based on the size of the segment and the number of gridded Interpolator cells assigned an estimate by the Interpolator which depended on the spatial extent of available data for any given 7-day interval.
- For each segment and 7-day interval, the spatial violation rate was calculated by counting the number of "FAILS" and dividing by the number of assessed Interpolator cells.
- These violation rates were ranked as described in Section 2.1.1 and a CFD was generated for each segment. Because bioreference curves are not available for the 7-day mean criterion, assessment curves were compared to a default 10% reference curve Assessment curves for any segment that crossed the reference curve were deemed as non-attaining of the 7-day mean criterion.

EPA-criteria based approach to 30-day mean assessment, using empirical data gathered from long-term, mid-channel stations:

The methodology for assessing the 30-day mean Open Water criteria was similar to the approach used by the Bay Program for 303(d) listings of impaired waters (described in Section 2.1.1 and EPA(2007)). The only differences were that: (1) a 10% reference curve was used rather than the established bioreference curve, because recent investigations have shown this to be more appropriate (USAEPA, 2010) and (2) only DO values from shallow and bottom waters were analyzed. For a complete 303(d) assessment, the full vertical profile is analyzed and assessed. This is not difficult to do for the 30-day mean assessment because empirical data is used but to for 7-day assessments, predictive data sets would be needed for all depth intervals. At the deepest receiving stations, as many as 13 or 14 models would have to generated and then aggregated into 7-day intervals. As a result, assessments were limited to only shallow and bottom depth data in order to streamline the analyses and allow for comparability between the 30-day mean and 7-day mean assessments.

Part I. Protectiveness of the 30-day mean versus the 7-day mean criteria

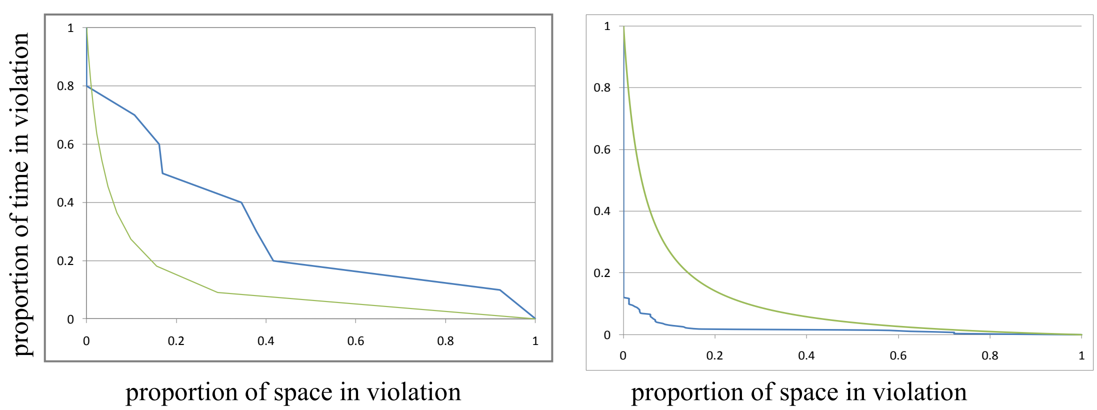
Results and Discussion

Table A6-2. Excessive violation rates for assessments of Open Water 30-day and 7-day mean criteria, by Bay segment.

Segments/Years Analyzed	30-day mean Excessive violation rate%	7-day mean Excessive violation rate%
Appomattox TF/2006-2008	4.4	0.0
Chickahominy OH/2006-2008	0.0	0.0
James TF2/2006-2008	1.3	0.0
James TF1/2006-2008	6.4	0.0
James OH/2006-2008	0.0	0.0
James MH/2006-2008	0.0	0.0
James PH/2006-2008	0.0	0.0
York MH/2007-2009	18.5	0.0
Mobjack Bay PH/2007-2009	<<0.1	0.0
Corrotomon MH/2007-2009	10.7	0.0
Rappahannock TF/2007-2009	3.8	0.0
Rappahannock OH/2007-2009	0.0	0.0

As Table A6-2 shows, none of the twelve segments assessed for the 7-day mean criterion showed non-attainment. The mesohaline portion of the Corrotomon River came the closest to non-attainment. It also failed the 30-day mean criterion, along with six other segments. The results of our analysis support the hyothesis that the 30-day mean assessment is more protective (or conservative) than the 7-day mean assessment, since we did not find any instance of a segment failing the 7-day mean that passed th*e* 30-day mean. Attainment of the 7-day mean was found even in segments with relatively high excessive violation rates for the 30-day mean criteria—such as seen in York MH (see Figure A6-2).

Figure A6-2. Comparison of 30-day (left) and 7-day (right) assessment CFDs for York River MH. The green curve represents the 10% reference curve.



<u>Part 2. Rolling versus Sequential 7-Day Mean Assessments</u>
<u>Methodology</u>

The convention for defining the 30-day period is to use calender months, or sequential 30-day intervals. Sequential 7-day means, starting at June 1 of each summer-year, were therefore initially used to be consistent with this procedure. However, it was decided at the June 10th, 2010 Tidal Monitoring and Analysis Workgroup meeting that it makes more sense biogically to use rolling means because periods of hyopoxia could be missed by partioning the assessment window into what could be considered arbitrary blocks of time.

Thus, the analysis detailed above was repeated using rolling means, to determine if more 7-day criteria violations were revealed than when sequential 7-day means were assessed. For any assessment window of (at most) 51 sequential 7-day periods, 348 rolling 7-day periods—and therefore the same number of interpolated grids—were used to assess a segment.

Part 2. Rolling versus Sequential 7-Day Mean Assessments

Results and Discussion

The absolute violation rates of the assessment curves determined that the assessments of rolling means were slightly higher than those from sequential mean assessments, but overall results were no different than that shown in Table A6-2.

Part 3. Spline Interpolation versus Fourier Analysis

Methodology

At the June 10th, 2010 TMAW meeting, Elgin Perry suggested that Fourier analysis might not be the best method for modeling and synthesizing both long-term and short-term variability into a predictive dataset and introduced spline interpolation as an alternative. We investigated to see if the two procedures produced different results for the 7-day mean assessment. The SAS code used to generate estimates via spline interpolation is provided at the end of this Appendix:.

Prior to conducting interpolations, paired comparisons of simulated data sets produced by the Fourier and spline procedures were made using Student's *t* test for each assessment window and station combinations (a total of 572 comparisons). For these combinations, over 80% showed a statistically significant difference between data sets generated using the Fourier and spline procedures. However the mean absolute difference across all comparisons was less than 0.40 mg/L and less than 25% of assessments had absolute differences greater than 0.50 mg/L between the paired synthetic data sets. Less than 1% had absolute differences greater than 1 mg/L. These results suggested that the differences in synthetic data sets were not likely to result in substantial changes in the assessment results for the two simulation methods. Nonetheless assessments were performed using synthetic data sets created using the two procedures.

Part 3. Spline Interpolation versus Fourier Analysis

Results and Discussion

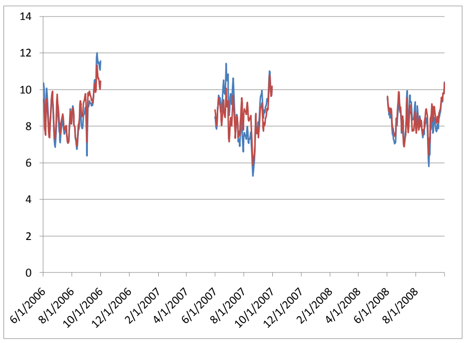
Spline interpolation and Fourier analysis were shown to generate slightly different predictions, with the former generally producing lower violation rates. Thus, using spline interpolation did not produce a different assessment result from the Fourier approach used since this work was initiated.

Figure A6-3 shows both spline and Fourier-generated time series for stations within segments where 7-day violations were observed. The cumulative frequency diagram illustrates the violation rate in space-time for these segments.

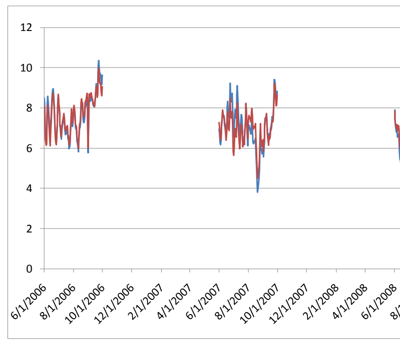
Figure A6-3. Predicted DO time-series for surface (a) and bottom(b) waters at a single station and c) the CFD for the segment. . Station time-series are grouped together by their respective segments. The blue line represents predictive values produced through Fourier analysis while the spline interpolation is shown in red. The green line on the CFD plot represents the 10% reference curve.

James Tidal Fresh Upper

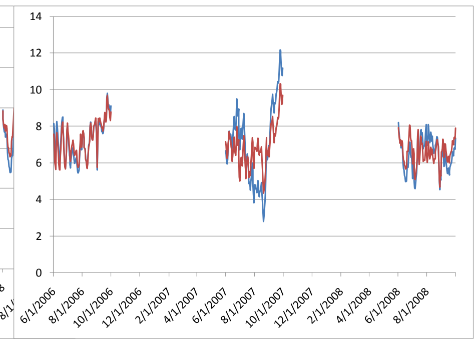
a) TF 5.2 Surface Surface



TF 5.2A Surface



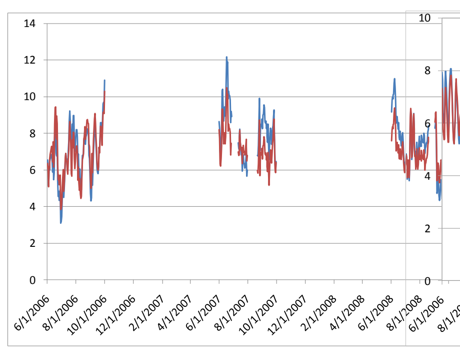
TF 5.3



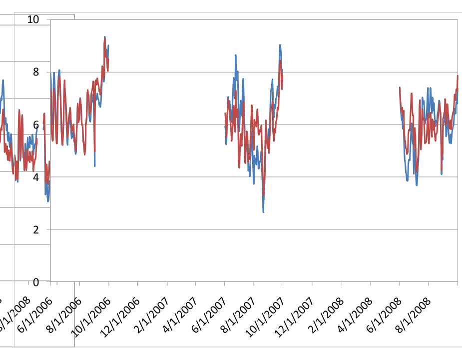
b)

No bottom waters at this station

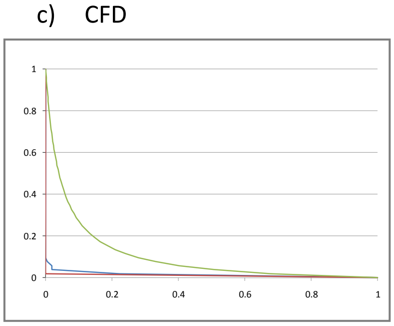
TF 5.2A Bottom



TF 5.3 Bottom



c) CFD

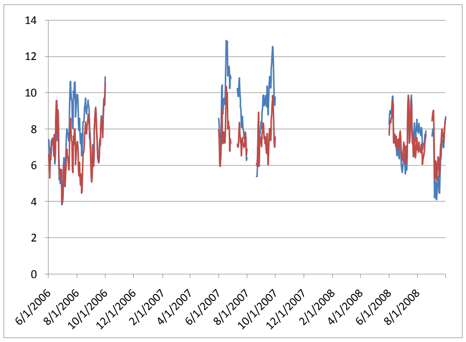


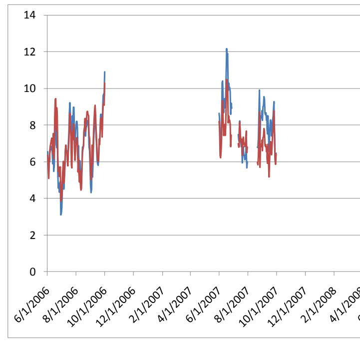
James River Tidal Fresh Lower

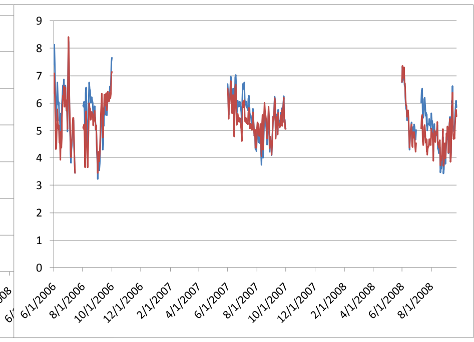
a) TF 5.5 Surface

TF 5.5A Surface

TF 5.6 Surface



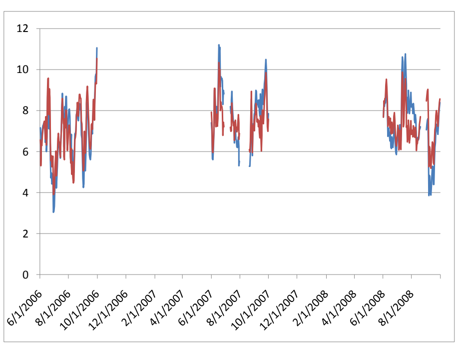


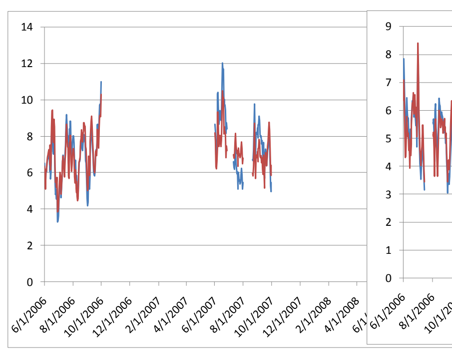


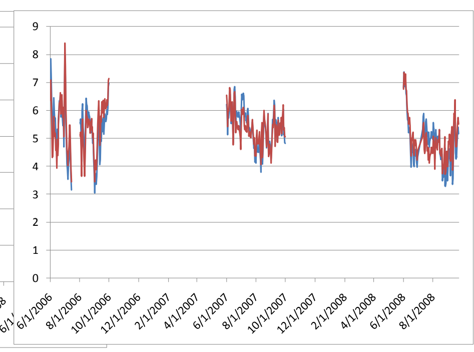
b) TF 5.5 Bottom

TF 5.5A Bottom

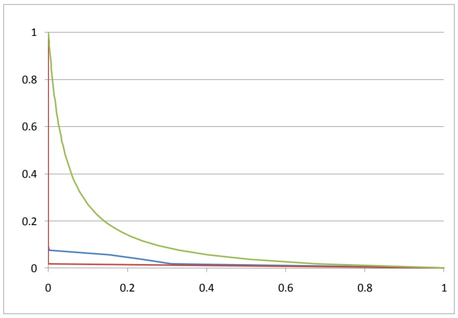
TF 5.6 Bottom





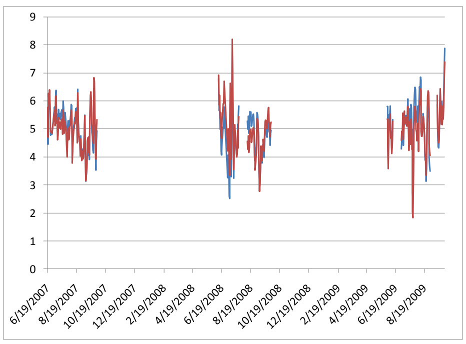


c) CFD for segment

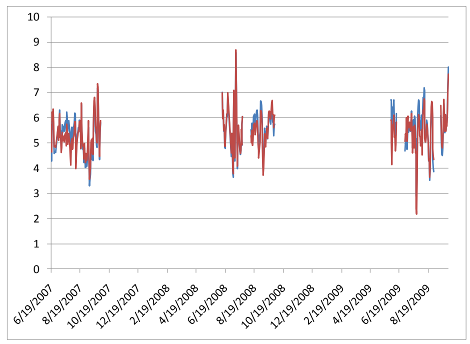


York River Mesoholine

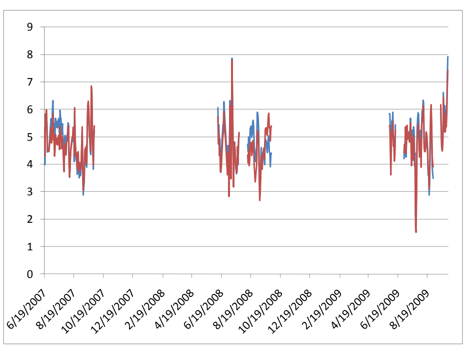
a) RET 4.3 Surface



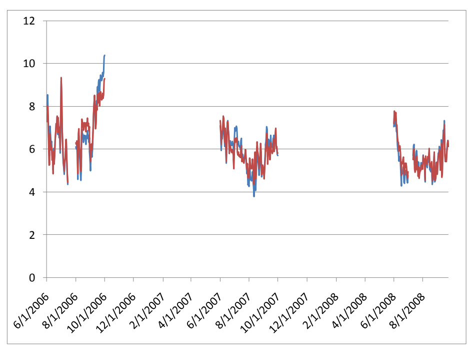
LE 4.1 Surface



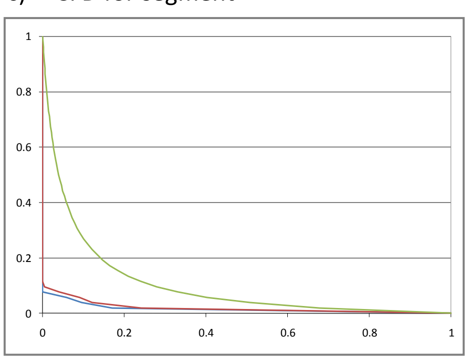
b) RET 4.3 Bottom



LE 4.1 Bottom

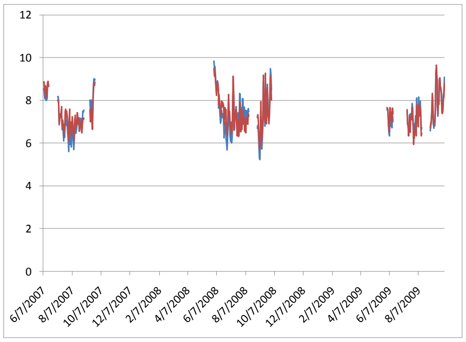


c) CFD for segment

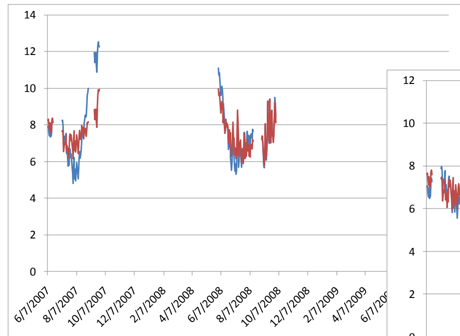


Mobjack Bay Polyhaline

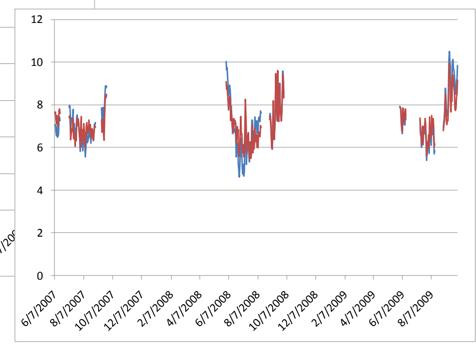
a) WE 4.1 Surface



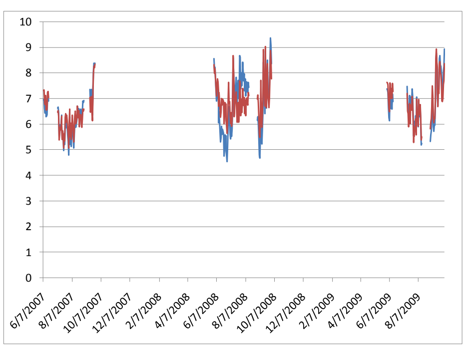
WE 4.2 Surface



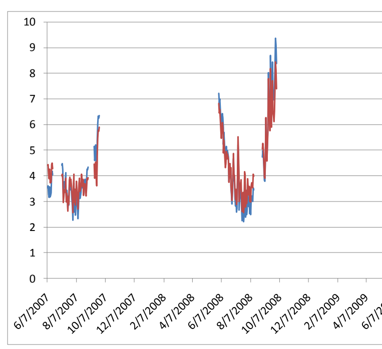
WE 4. 3 Surface

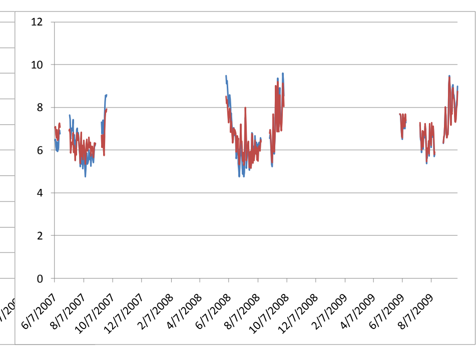


b) WE 4.1 Bottom



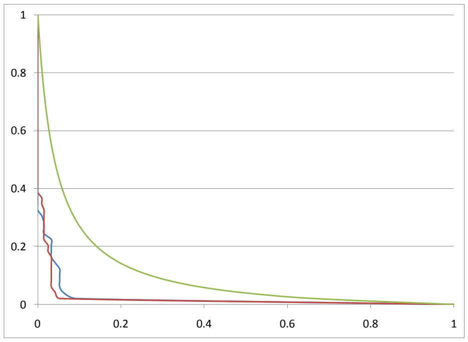
WE 4.2 Bottom





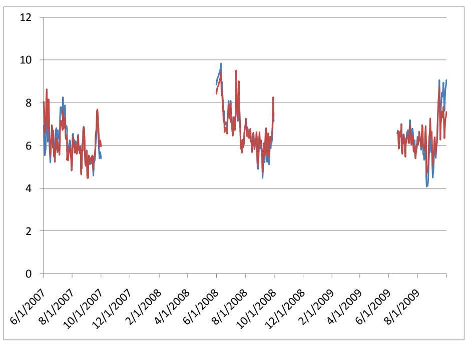
WE 4.3 Bottom

c) CFD for segment

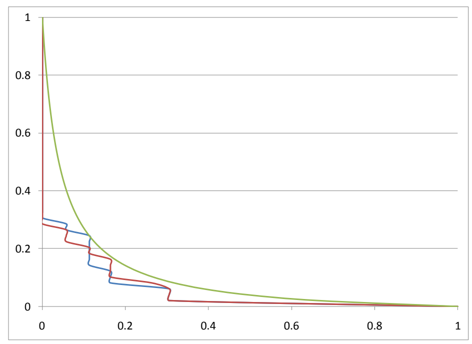


Corrotoman River Mesohaline

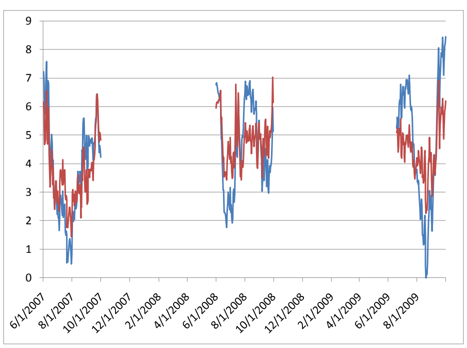
a) LE 3.3 Surface



c) CFD for segment



b) LE 3.3 Bottom



SAS Program used for generating Simulated Low Frequency Data using Fourier Analysis

The version provided below produces simulated data for multiple stations in the James River. See the macro select statements at the end of the program for examples.

```
**********************************************************************;
file:          C:\Projects\CBP\DO_spectral\ret24.sas
function:      spectral analysis of ret24 data
programmer:    Elgin S. Perry, Ph. D.
date:          10/7/2009
```

```
address:        2000 Kings Landing Rd.
                Huntingtown,  Md.  20639
voice phone:    (410)535-2949
email:          EPERRY@chesapeake.net
**********************************************************************;

********************************************************************;
*DATA STEP 1: INPUT AND MODIFY LONG-TERM (SENDING) AND SHORT-TERM  ;
*RECEIVING DATA SETS                                                ;
********************************************************************;

OPTIONS LS=120 PS=55 REPLACE NOCENTER MPRINT;
*VERTICAL PROFILE DATA SET;
LIBNAME sas7bdat "K:\sci\sci chesapeake bay program projects\CBP_DATABASE\CBPWATER\";
DATA ONE; SET sas7bdat.cbpwq_profile85_09;

FORMAT DATE DATE9.;

DATA ONE; SET ONE;
IF LAYER='S';
YEAR=YEAR(DATE);
MONTH=MONTH(DATE);
DAY=DAY(DATE);

IF STATION="LE5.5W"  THEN STATION="LE5.5";
IF STATION="LE5.5-W" THEN STATION="LE5.5";

DATA ONE;  SET ONE;

PROC SORT; BY STATION YEAR MONTH;
RUN;

*CONTINUOUS MONTIORING DATA SET;
LIBNAME SAS7BDAT "K:\sci\sci chesapeake bay program projects\CURRENT\CRITERIA\DATA\";
DATA CONMON; SET Sas7bdat.conmon2006_2009new_mod_b;
RUN;

%MACRO
SELECT(NUM,STAT,OUTSTAT,CONSTAT,CONSTAT2,MON1,MON2,MON3,ED,ST,YR,YR2,VAR,TIMESPAN,TIMESPAN3,YLABEL
,YAXISLAB);

*********************************************************************;
*SELECT SENDING SET DATA BASED ON ASSESSMENT WINDOW, AVERAGE BY DATE
AND ADD ;
*********************************************************************;

DATA &OUTSTAT;  SET ONE;
IF STATION="&STAT";
IF LAYER="S";
MONTH=MONTH(DATE);
YEAR=YEAR(DATE);

DATA &OUTSTAT; SET &OUTSTAT;
DATE=MDY(MONTH,DAY,YEAR);
FORMAT DATE DATE9.;

DATA &OUTSTAT&VAR&MON2&YR; SET &OUTSTAT;
IF &TIMESPAN;

PROC MEANS NOPRINT DATA=&OUTSTAT&VAR&MON2&YR;
BY STATION DATE;
VAR DO;
ID YEAR MONTH DAY;
OUTPUT OUT=&OUTSTAT&VAR&MON2&YR(DROP=_TYPE_ _FREQ_)  MEAN=DO SALINITY WTEMP;

DATA &OUTSTAT&VAR&MON2&YR;  SET &OUTSTAT&VAR&MON2&YR;
TL=_N_-1;  *(create variable t as a sequential index for time;
FORMAT DATE DATE9.;
```

```
*******************************************************************;
*ADD MISSING MONTH PLACE HOLDER TO LONG-TERM SENDING DATA SET
*******************************************************************;

PROC IMPORT OUT=WORK.MONTHS
            DATAFILE= "K:\sci\sci chesapeake bay program
projects\\CURRENT\CRITERIA\Missing_Months.xls"
            DBMS=EXCEL REPLACE;
SHEET="Sheet1$";
GETNAMES=YES;
MIXED=NO;
SCANTEXT=YES;
USEDATE=YES;
SCANTIME=YES;

DATA MONTHS; SET MONTHS;
STATION="&STAT";
DATE=MDY(MONTH,15,YEAR);
IF &TIMESPAN;

PROC SORT; BY STATION YEAR MONTH;

DATA &OUTSTAT&VAR&MON2&YR; MERGE &OUTSTAT&VAR&MON2&YR MONTHS; BY STATION YEAR MONTH;
YEAR=YEAR(DATE);

*****************************************************************;
*PROC GPLOT FOR PLOT OF ORIGINAL TIME SERIES               ;
*****************************************************************;

ODS DOCUMENT NAME=&OUTSTAT&VAR&MON2&YR;
GOPTIONS NODISPLAY NOBORDER;
PROC GPLOT DATA=&OUTSTAT&VAR&MON2&YR;
AXIS1 VALUE=(H=1.5 FONT=ARIAL COLOR=BLACK)
         LABEL=(A=90 H=1.5 FONT=ARIAL COLOR=BLACK "&YLABEL")
         MAJOR=(H=1.5 COLOR=BLACK)
         MINOR=NONE
         ORDER=(&YAXISLAB);

AXIS2 VALUE=(H=1.5 FONT=ARIAL COLOR=BLACK)
         LABEL=(H=1.5 FONT=ARIAL COLOR=BLACK "Date")
         MAJOR=(H=1.5 COLOR=BLACK)
         MINOR=(H=0.5 COLOR=BLACK NUMBER=10);

SYMBOL1 COLOR=BLUE  I=J   V=CIRCLE L=1; *OBSERVED DATA;
TITLE1 H=1.5 F=ARIAL JUSTIFY=CENTER C=BLACK "Observed Bottom &YLABEL at Station" ;
TITLE2 H=1.5 F=ARIAL JUSTIFY=CENTER C=BLACK "&STAT. from &TIMESPAN3";
PLOT &VAR*DATE=1 / VAXIS=AXIS1 HAXIS=AXIS2 NOLEGEND NAME="LTDAT&NUM";
RUN;
QUIT;
ODS DOCUMENT CLOSE;


*******************************************************************;
*CALCULATE OVERALL MONTHLY MEAN AND SUBSTITUTE FOR MISSING VALUES
*******************************************************************;

PROC SORT DATA=&OUTSTAT; BY MONTH;
PROC MEANS DATA=&OUTSTAT NOPRINT;
VAR &VAR;
BY MONTH;
OUTPUT OUT=MON_MEANS (DROP=_TYPE_ _FREQ_) MEAN=M_&VAR;

PROC SORT DATA=&OUTSTAT&VAR&MON2&YR; BY MONTH;
DATA &OUTSTAT&VAR&MON2&YR; MERGE &OUTSTAT&VAR&MON2&YR MON_MEANS; BY MONTH;

IF &VAR=. THEN &VAR=M_&VAR;

PROC SORT; BY STATION YEAR MONTH;
```

```
DATA &OUTSTAT&VAR&MON2&YR; SET &OUTSTAT&VAR&MON2&YR;
TL=_N_-1;
KEEP STATION DATE YEAR MONTH DO TL;
PROC SORT; BY STATION DATE;

DATA LT&OUTSTAT&VAR&MON2&YR;
DO DATE='01JAN2006'D TO '31DEC2008'D;
OUTPUT;
END;
FORMAT DATE DATE9.;
RUN;

DATA LT&OUTSTAT&VAR&MON2&YR; SET LT&OUTSTAT&VAR&MON2&YR;
STATION="&STAT";
LAYER="S";
PROC SORT; BY STATION DATE;
RUN;

*********************************************************************;
*PROC GAM FOR GENERATING SENDING DATA SET PROC PLOTS FOR OBSERVED AND
*PREDICTEDS AND ODS DOCUMENT GENEARATION
*********************************************************************;

PROC GAM DATA=&OUTSTAT&VAR&MON2&YR;
MODEL DO=SPLINE(DATE,DF=18);
*use score state to get predicted DO values added to the synthetic data set;
SCORE DATA=LT&OUTSTAT&VAR&MON2&YR OUT=LT&OUTSTAT&VAR&MON2&YR;
*( plot data as a check;
DATA LT&OUTSTAT&VAR&MON2&YR; MERGE LT&OUTSTAT&VAR&MON2&YR &OUTSTAT&VAR&MON2&YR;   BY STATION DATE;

ODS DOCUMENT NAME=&OUTSTAT&VAR&MON2&YR;
GOPTIONS NODISPLAY NOBORDER;
PROC GPLOT DATA=LT&OUTSTAT&VAR&MON2&YR;
AXIS1 VALUE=(H=1.5 FONT=ARIAL COLOR=BLACK)
         LABEL=(A=90 H=1.5 FONT=ARIAL COLOR=BLACK "&YLABEL")
         MAJOR=(H=1.5 COLOR=BLACK)
         MINOR=NONE
         ORDER=(&YAXISLAB);

AXIS2 VALUE=(H=1.5 FONT=ARIAL COLOR=BLACK)
         LABEL=(H=1.5 FONT=ARIAL COLOR=BLACK "Date")
         MAJOR=(H=1.5 COLOR=BLACK)
         MINOR=(H=0.5 COLOR=BLACK NUMBER=10);

SYMBOL1 COLOR=BLUE  I=J   V=CIRCLE L=1; *OBSERVED DATA;
SYMBOL2 COLOR=RED   I=J   V=NONE   L=1; *PREDICTED DATA;

TITLE1 H=1.5 F=ARIAL C=BLACK JUSTIFY=CENTER "Long Term Observed and GAM Predicted Data at Station
&STAT." ;
PLOT &VAR*DATE=1 P_&VAR*DATE=2 /OVERLAY VAXIS=AXIS1 HAXIS=AXIS2 NAME="LTPDAT&NUM";
RUN;
ODS DOCUMENT CLOSE;
RUN;
QUIT;

GOPTIONS DISPLAY NOBORDER;
ODS DOCUMENT NAME=&OUTSTAT&VAR&MON2&YR;
PROC GREPLAY IGOUT=GSEG TC=SASHELP.TEMPLT TEMPLATE=V2S NOFS;
TREPLAY 1:LTDAT&NUM
        2:LTPDAT&NUM
        DES="";
RUN;
ODS DOCUMENT CLOSE;
RUN;
QUIT;

*********************************************************************;
*CALCULATE SENDING DATA SET GRAND MEAN AND SENDING DATA PREDICTED
*VALUE DATA SETS FOR MERGE WITH SHORT-TERM PREDICTEDS
```

```
**********************************************************************;

PROC SORT DATA=LT&OUTSTAT&VAR&MON2&YR; BY STATION;
DATA LT&OUTSTAT&VAR&MON2&YR; SET LT&OUTSTAT&VAR&MON2&YR;
PROC MEANS DATA=&OUTSTAT&VAR&MON2&YR NOPRINT;
VAR &VAR;
BY STATION;
OUTPUT OUT=LTMEAN&VAR MEAN=LTMN&VAR;

DATA LT&OUTSTAT&VAR&MON2&YR; MERGE LT&OUTSTAT&VAR&MON2&YR LTMEAN&VAR; BY STATION;
LTP&VAR=P_&VAR;
MONTH=MONTH(DATE);
DAY=DAY(DATE);
YEAR=YEAR(DATE);
IF P_DATE=. THEN DELETE;
DROP TL _TYPE_ _FREQ_;
RUN;
QUIT;

*SPECTRAL ANALYSIS SECTION;

**********************************************************************;
*SELECT CONMON ASSESSMENT PERIOD DATA SET
**********************************************************************;

DATA &CONSTAT2&VAR; SET CONMON;
IF STATION="&CONSTAT";

YEAR = YEAR(DATE);
MONTH=MONTH(DATE);
DAY = DAY(DATE);

IF "01&MON2.&YR"D <= DATE <= "&ED.&MON2.&YR"D;
DATETIME = DHMS(DATE,HOUR(TIME),MINUTE(TIME),00);
CDATE=DATE - "15&MON2.&YR"D;
CDATESQ = CDATE*CDATE;
FORMAT DATETIME DATETIME15.;

DATA &CONSTAT2&VAR; SET &CONSTAT2&VAR;
TS=_N_-1;

**********************************************************************;
*SHORT TERM MEAN VALUE
**********************************************************************;

PROC MEANS DATA=&CONSTAT2&VAR NOPRINT;
VAR &VAR;
OUTPUT OUT=STMEAN&VAR MEAN=STMN&VAR;

**********************************************************************;
*SPECTRAL ANALYSIS OF SHORT TERM DATA SET
**********************************************************************;

*Spectral analysis for shorterm do data;
PROC SPECTRA DATA=&CONSTAT2&VAR OUT=CM&OUTSTAT&VAR&MON2&YR P S COEF CENTER;
VAR &VAR;
WEIGHTS 1 2 3 4 3 2 1;
RUN;

*Rescale the period and frequency variables to enhance interpretation;
DATA CM&OUTSTAT&VAR&MON2&YR; SET CM&OUTSTAT&VAR&MON2&YR;
PI=3.14159265;
PER_HR=PERIOD/4;  *( convert period into hours;
CYC_DAY=FREQ*96/(2*PI);  *( convert frequency into cycles per day;

**********************************************************************;
*GPLOT PERIODGRAM OUTPUT
**********************************************************************;
```

```
GOPTIONS NODISPLAY NOBORDER;
TITLE1  H=1.5 F=ARIAL JUSTIFY=CENTER C=BLACK "Station &CONSTAT. ConMon &MON1. &YR";
SYMBOL1 I=JOIN V=NONE C=BLUE;
AXIS1 VALUE=(H=1.5 FONT=ARIAL COLOR=BLACK)
         LABEL=(A=90 H=1.5 FONT=ARIAL COLOR=BLACK "Periodogram")
         MAJOR=(H=1 COLOR=BLACK)
         MINOR=NONE;
AXIS2 VALUE=(H=1.5 FONT=ARIAL COLOR=BLACK)
         LABEL=(H=1.5 FONT=ARIAL COLOR=BLACK "Cycles per day")
         MAJOR=(H=1.5 COLOR=BLACK) MINOR=NONE;

GOPTIONS NODISPLAY NOBORDER;
PROC GPLOT DATA=CM&OUTSTAT&VAR&MON2&YR;
PLOT P_01 * CYC_DAY / VAXIS=AXIS1 HAXIS=AXIS2 NOLEGEND NAME="PGCYDY&NUM";
RUN;

TITLE1  H=1.5 F=ARIAL JUSTIFY=CENTER C=BLACK "Station &CONSTAT. ConMon &MON1. &YR";
SYMBOL1 I=JOIN V=NONE C=BLUE;
AXIS1 VALUE=(H=1.5 FONT=ARIAL COLOR=BLACK)
         LABEL=(A=90 H=1.5 FONT=ARIAL COLOR=BLACK "Spectral Density")
         MAJOR=(H=1 COLOR=BLACK)
         MINOR=NONE;
AXIS2 VALUE=(H=1.5 FONT=ARIAL COLOR=BLACK)
         LABEL=(H=1.5 FONT=ARIAL COLOR=BLACK "Cycles per year")
         MAJOR=(H=1.5 COLOR=BLACK) MINOR=NONE;
PROC GPLOT DATA=CM&OUTSTAT&VAR&MON2&YR;
PLOT S_01 * CYC_DAY / VAXIS=AXIS1 HAXIS=AXIS2 NOLEGEND NAME="SDCYDY&NUM";
RUN;

TITLE1  H=1.5 F=ARIAL JUSTIFY=CENTER C=BLACK "Station &CONSTAT. ConMon &MON1. &YR";
SYMBOL1 I=JOIN V=NONE C=BLUE;
AXIS1 VALUE=(H=1.5 FONT=ARIAL COLOR=BLACK)
         LABEL=(A=90 H=1.5 FONT=ARIAL COLOR=BLACK "Periodogram")
         MAJOR=(H=1 COLOR=BLACK)
         MINOR=NONE;
AXIS2 VALUE=(H=1.5 FONT=ARIAL COLOR=BLACK)
         LABEL=(H=1.5 FONT=ARIAL COLOR=BLACK "Periods per hour")
         MAJOR=(H=1.5 COLOR=BLACK) MINOR=NONE;
PROC GPLOT DATA=CM&OUTSTAT&VAR&MON2&YR;
PLOT P_01 * PER_HR / VAXIS=AXIS1 HAXIS=AXIS2 NOLEGEND NAME="PGPRHR&NUM";
RUN;

TITLE1  H=1.5 F=ARIAL JUSTIFY=CENTER C=BLACK "Station &CONSTAT. ConMon &MON1. &YR";
SYMBOL1 I=JOIN V=NONE C=BLUE;
AXIS1 VALUE=(H=1.5 FONT=ARIAL COLOR=BLACK)
         LABEL=(A=90 H=1.5 FONT=ARIAL COLOR=BLACK "Spectral Density")
         MAJOR=(H=1 COLOR=BLACK)
         MINOR=NONE;
AXIS2 VALUE=(H=1.5 FONT=ARIAL COLOR=BLACK)
         LABEL=(H=1.5 FONT=ARIAL COLOR=BLACK "Periods per hour")
         MAJOR=(H=1.5 COLOR=BLACK) MINOR=NONE;
PROC GPLOT DATA=CM&OUTSTAT&VAR&MON2&YR;
PLOT S_01 * PER_HR / VAXIS=AXIS1 HAXIS=AXIS2 NOLEGEND NAME="SDPRHR&NUM";
RUN;

*********************************************************************;
*OUTPUT GPLOTS TO ODS DOCUMENT
*********************************************************************;

GOPTIONS DISPLAY;
ODS DOCUMENT NAME=&OUTSTAT&VAR&MON2&YR;
PROC GREPLAY IGOUT=GSEG TC=SASHELP.TEMPLT TEMPLATE=L2R2S NOFS;
TREPLAY 1:PGCYDY&NUM
        2:SDCYDY&NUM
        3:PGPRHR&NUM
        4:SDPRHR&NUM
        DES="";
RUN;
ODS DOCUMENT CLOSE;
```

```
RUN;

************************************************************************;
*TRANSPOSE SIN AND COS COEFFICIENTS TO CREATE A VARIABLE FOR EACH SIN/COS
*COEFFICIENT NUMBER OF VARIABLES FOR SIN IS N/2 WHERE N=NUMBER OF DAYS IN THE
*RECEIVING DATA SET.  SAME NUMBER FOR COS COEFFICIENTS
************************************************************************;

*Transpose sin and cos coefficients to create a appear on one row of data;
%LET MS=&ST;
DATA STCF&OUTSTAT&VAR&MON2&YR; SET CM&OUTSTAT&VAR&MON2&YR end=last;
ARRAY COS [&MS] SC1-SC&MS;
ARRAY SIN [&MS] SS1-SS&MS;
IF _N_ = 1 THEN SET STMEAN&VAR;
COS[_N_] = COS_01;
SIN[_N_] = SIN_01;

RETAIN SC1-SC&MS SS1-SS&MS STMN&VAR;
DROP FREQ PERIOD COS_01 SIN_01 P_01 S_01 Per_yr Cyc_yr;

IF LAST THEN DO;
EVEN = 1-MOD(&MS-1,2);
IF EVEN THEN DO;
SIN[_N_]=SIN[_N_]/2;
COS[_N_]=COS[_N_]/2;
END;
OUTPUT;
END;

*FIRST STEP OF CONMON PREDICTION IS TO CREATE A DATETIME VARIABLE STARTING AT ZERO;
DATA STPR&OUTSTAT&VAR&MON2&YR (KEEP = TS DATETIME SDATETIME PDATETIME); SET &CONSTAT2&VAR;
IF _N_ = 1 THEN SDATETIME=DATETIME;
PDATETIME=DATETIME-SDATETIME;
RETAIN SDATETIME;
FORMAT SDATETIME PDATETIME DATETIME15.;


*COMBINE PREDICTION DATES WITH COEFFICIENTS AND COMPUTE A PREDICTED DO FOR EVERY 15MIN;
DATA STPR&OUTSTAT&VAR&MON2&YR; SET STPR&OUTSTAT&VAR&MON2&YR;
IF _N_=1 THEN SET STCF&OUTSTAT&VAR&MON2&YR;
RETAIN SC1-SC&MS SS1-SS&MS STMN&VAR;
ARRAY A [&MS] SC1-SC&MS;
ARRAY B [&MS] SS1-SS&MS;
ARRAY W [&MS] W1-W&MS;

DO K=1 TO &MS-1;
W[K] = 2*PI*K/(2*&MS);
END;
RETAIN W1-W&MS;
P&VAR=STMN&VAR;

*START SUM WITH DO MEAN AND SUM THE FOURIER TERMS TO GET PREDICTED DO;
DO K = 1 TO &MS-1;
P&VAR = P&VAR + A[K+1]*COS(W[K]*TS)+ B[K+1]*SIN(W[K]*TS);
END;
*merge monthly DO and predicted DO to plot on one plot;
DATA CMST&OUTSTAT&VAR&MON2&YR; MERGE &CONSTAT2&VAR STPR&OUTSTAT&VAR&MON2&YR;
BY TS;
PROC SORT; BY DATETIME;

************************************************************************;
*GPLOT SHORT-TERM PREDICTED DATA
************************************************************************;

GOPTIONS NODISPLAY NOBORDER;
TITLE1  H=1.5 F=ARIAL JUSTIFY=CENTER C=BLACK "Station &CONSTAT. ConMon &MON1. &YR ";
PROC GPLOT DATA=CMST&OUTSTAT&VAR&MON2&YR;
AXIS1 VALUE=(H=1.5 FONT=ARIAL COLOR=BLACK)
         LABEL=(A=90 H=1.5 FONT=ARIAL COLOR=BLACK "&YLABEL")
```

```
         MAJOR=(H=1.5 COLOR=BLACK)
         MINOR=NONE
         ORDER=(&YAXISLAB);

AXIS2 VALUE=(H=1.5 FONT=ARIAL COLOR=BLACK)
         LABEL=(H=1.5 FONT=ARIAL COLOR=BLACK "Date")
         MAJOR=(H=1.5 COLOR=BLACK);

SYMBOL1 COLOR=BLUE  I=J   V=CIRCLE L=1; *OBSERVED DATA;
SYMBOL2 COLOR=RED   I=J   V=NONE   L=1; *PREDICTED DATA;

PLOT &VAR*DATETIME P&VAR*DATETIME / OVERLAY VAXIS=AXIS1 HAXIS=AXIS2 NAME="STPRDT&NUM";
RUN;

***************************************************************************;
*COMBINE LONG AND SHORT-TERM SIGNALS
***************************************************************************;

DATA LTSTPR&OUTSTAT&VAR&MON2&YR;
FORMAT START DATETIME16.;
START = "01&MON2.&YR.:00:00"DT;
DATETIME=START;
DATE=DATEPART(DATETIME);
TS=0;
OUTPUT;
DO TS = 0 to 2*&MS;
DATETIME=DATETIME+15*60;
DATE=DATEPART(DATETIME);
OUTPUT;
END;
RETAIN DATETIME;
FORMAT START DATETIME DATETIME16. DATE DATE9.;

DATA LTDATES (KEEP = DATE TL LTP&VAR); SET LT&OUTSTAT&VAR&MON2&YR;
DATA LTSTPR&OUTSTAT&VAR&MON2&YR; MERGE LTSTPR&OUTSTAT&VAR&MON2&YR (IN=INLTST) LTDATES;
BY DATE;
IF INLTST;

DATA LTSTPR&OUTSTAT&VAR&MON2&YR; SET LTSTPR&OUTSTAT&VAR&MON2&YR;
ARRAY SA [&MS] SC1-SC&MS;
ARRAY SB [&MS] SS1-SS&MS;
ARRAY SW [&MS] SW1-SW&MS;

IF _N_=1 THEN DO;
SET STCF&OUTSTAT&VAR&MON2&YR;
END;
RETAIN SC1-SC&MS SS1-SS&MS SW1-SW&MS;

DO K=1 TO &MS-1;
SW[K] = 2*PI*K/(2*&MS-1);
END;
P&VAR = LTP&VAR;

DO K=1 TO &MS-1;
P&VAR=P&VAR+SA[K+1]*COS(SW[K]*TS)+SB[K+1]*SIN(SW[K]*TS);
END;
RUN;

PROC SORT DATA=LTSTPR&OUTSTAT&VAR&MON2&YR;
BY DATETIME;

DATA LTSTPR&OUTSTAT&VAR&MON2&YR; SET LTSTPR&OUTSTAT&VAR&MON2&YR;
IF DATETIME > "&ED.&MON2.&YR2.:23:45"DT THEN DELETE;

***************************************************************************;
*GPLOT OF SYNTHETIC DATA AND OUTPUT TO ODS DOCUMENT
***************************************************************************;
```

```
TITLE1 H=1.5 F=ARIAL JUSTIFY=CENTER C=BLACK "Synthetic data for Station &STAT. &MON1. &YR
(Bottom)";
PROC GPLOT DATA=LTSTPR&OUTSTAT&VAR&MON2&YR;
AXIS1 VALUE=(H=1.5 FONT=ARIAL COLOR=BLACK)
         LABEL=(A=90 H=1.5 FONT=ARIAL COLOR=BLACK "&YLABEL")
         MAJOR=(H=1.5 COLOR=BLACK)
         MINOR=NONE
         ORDER=(&YAXISLAB);

AXIS2 VALUE=(H=1.5 FONT=ARIAL COLOR=BLACK)
         LABEL=(H=1.5 FONT=ARIAL COLOR=BLACK "Date")
         MAJOR=(H=1.5 COLOR=BLACK);
         *ORDER=("01&MON2.&YR2.:00:00"DT to "&ED.&MON2.&YR2.:23:45"DT);
SYMBOL1 COLOR=BLUE  I=J   V=CIRCLE L=1; *OBSERVED DATA;
SYMBOL2 COLOR=RED   I=J   V=NONE   L=1; *PREDICTED DATA;
PLOT P&VAR*DATETIME LTP&VAR*DATETIME / OVERLAY VAXIS=AXIS1 HAXIS=AXIS2 NAME="SYNDT&NUM";
RUN;

GOPTIONS DISPLAY;
ODS DOCUMENT NAME=&OUTSTAT&VAR&MON2&YR;
PROC GREPLAY IGOUT=GSEG TC=SASHELP.TEMPLT TEMPLATE=V2S NOFS;
TREPLAY 1:STPRDT&NUM
        2:SYNDT&NUM
        DES="";
RUN;
ODS DOCUMENT CLOSE;
RUN;

***************************************************************************;
*OUTPUT DATA TO PERMANENT SAS DATA SETS AND ODS DOCUMENT TO PDF FILE
***************************************************************************;

DATA LTSTPR&OUTSTAT&VAR&MON2&YR; SET LTSTPR&OUTSTAT&VAR&MON2&YR;
STATION="&STAT";
CONMONSTAT="&CONSTAT";
RUN;

DATA LTSTPR&OUTSTAT&VAR&MON2&YR;
KEEP STATION CONMONSTAT DATETIME DATE LTPDO PDO;
RETAIN STATION STATION CONMONSTAT DATETIME DATE LTPDO PDO;
SET LTSTPR&OUTSTAT&VAR&MON2&YR;
RUN;

LIBNAME sas7bdat "K:\sci\sci chesapeake bay program
projects\CURRENT\CRITERIA\DATA\Output_Data\James_River\Bottom\";
DATA sas7bdat.LTSTGAM&OUTSTAT&VAR&MON2&YR; SET LTSTPR&OUTSTAT&VAR&MON2&YR;
RUN;

ODS PDF FILE="K:\sci\sci chesapeake bay program
projects\CURRENT\CRITERIA\Graphics\James_River\Bottom\
GAM&OUTSTAT&VAR&MON2&YR..pdf";
PROC DOCUMENT NAME=&OUTSTAT&VAR&MON2&YR;
REPLAY / LEVELS=ALL;
RUN;
ODS PDF CLOSE;
RUN;

PROC DATASETS;
DELETE &CONSTAT2&VAR CMST&OUTSTAT&VAR&MON2&YR CM&OUTSTAT&VAR&MON2&YR LTCF&OUTSTAT&VAR&MON2&YR
LTCP&OUTSTAT&VAR&MON2&YR LTDATES
LTMEAN&VAR LTPR&OUTSTAT&VAR&MON2&YR LTSTPR&OUTSTAT&VAR&MON2&YR LT&OUTSTAT&VAR&MON2&YR MON_MEANS
STCF&OUTSTAT&VAR&MON2&YR
STMEAN&VAR STPR&OUTSTAT&VAR&MON2&YR &OUTSTAT &OUTSTAT&VAR&MON2&YR;
RUN;
QUIT;

%MEND;
```

```
%SELECT(I8,LE5.5,LE55,JMS002.55,JMS002,September,SEP,9,30,1441,2006,06,
DO,2006<=YEAR<=2008,2006-2008 ,Dissolved Oxygen (mg/L),0 TO 16 BY 2);

%SELECT(I9,LE5.5,LE55,JMS002.55,JMS002,September,SEP,9,30,1441,2007,07,
DO,2006<=YEAR<=2008,2006-2008, Dissolved Oxygen (mg/L),0 TO 16 BY 2);

RUN;
```

Appendix 9, (Part B).

Procedure for Spectral Analysis

Tish Robertson & Mike Lane

Note: There are a number of ways to do this, some no doubt better than others. The following is the procedure that I used to generate the models presented in this document. I make no claim about its strengths or weaknesses relative to other methods.

1. Create a model representing seasonal, long-term variation using log-transformed midchannel data. The model is represented by the equation:

$$y = x + \sum_{j=1}^{J} [a_j \cos(2\pi jT) + b_j \sin(2\pi jT)]$$

where:

x is the y-intercept
J is the total number of monthly data points minus one ÷ 2 (ie.,[36-1]/2)
j is the fourier frequency
a and b are regression coefficients
T is period (scaled in months)

To obtain regression coefficients, model the linear relationship between y (dissolved oxygen values) and cosine and sine functions at increasing fourier frequencies and months. This means that for a three-year time frame, you will have potentially 17 independent cosine variables and 17 independent sine variables. To reduce the group down to the most important variables, I use the exhaustive search method of the function *regsubsets* included in the R package *leaps*. I have found that most variation is sufficiently captured with just six or seven variables.

Each depth should be modeled separately. So for a station where six depths are routinely sampled, you would create six different models.

2. Repeat step 1 for the continuous monitoring data using weekly averages. The model is built the same way except that the total number of fourier frequencies will be larger and of course weekly periods (e.g., T=t/156) should be used instead of monthly (e.g., T=t/36). For continuous monitoring data that are not actually continuous (i.e., no data collected in the winter), the time-series for each year should be modeled separately.

3. Create a synthetic dataset by combining the long-term and short-term equations, setting *T* to the scale of the short-term model (e.g., week), using the y-intercept of the long-term model, and using coefficients from the long-term model in the case of frequency redundancies. Using the resulting equation, you should be able to generate a weekly DO estimate. To err on the side of caution, you should only generate estimates for those periods when conmon data were available.

Long-term stations should be "paired" with the closest continuous monitoring station unless professional judgment indicates otherwise.

$$y_{syn} = x_{lt} + \Sigma\,[a_j\cos(2\pi jT) + b_j\sin(2\pi jT)]_{lt} + \Sigma\,[a_j\cos(2\pi jT) + b_j\sin(2\pi jT)]_{st}$$

4. Interpolate each weekly estimate. For summer assessment of the 7-day mean, depending on how much data were available, you would produce about 52 interpolation grids. Each cell in the interpolation grid should be assessed against the 7-day mean criterion for open water (4.0 mg/l), with non-attainment percentages used to produce a CFD evaluated against a 10% reference curve.

Appendix 9, Part C

Spectral Analysis: A Potential Tool for Short-Term Criteria Assessment

Tish Robertson

Virginia Department of Environmental Quality

The Chesapeake Bay Program Office has established multiple criteria for the assessment of dissolved oxygen. These criteria are specific to designated use (e.g., Open Water, Migratory Fish Spawning and Nursery, Deep Water, and Deep Channel) and are applied to different time intervals (e.g., 1-day, 7-day mean, 30-day mean and instantaneous minimum). At present, only the Deep Water use has been assessed based on the instantaneous minimum—the sole criteria for this use. Assessment has been restricted to

30-day mean criteria in Open Water and Deep Water designated uses. Such criteria are appropriately applied to the data generated by the long-term, midchannel, fixed station water quality monitoring conducted in the Bay and tidal tributaries. But because these data are collected at only monthly/semi-monthly intervals, they do not allow for the application of short-term criteria. Fortunately, this is the biggest advantage of the data generated by continuous monitoring stations, which both VA and MD have employed for their shallow water monitoring programs.

Appendix 1 in the April 2003 *Ambient Water Quality Criteria for Dissolved Oxygen, Water Clarity and Chlorophyll a for the Chesapeake Bay and its Tidal Tributaries* presents different analytical approaches to short-term DO assessment. The most promising is the spectral analysis approach, which combines mid-channel monthly data from long-term, fixed stations with shallow water, temporally-intensive data gathered at continuous monitoring stations. Using models created from these different datasets, a “synthetic” dataset is produced that simultaneously reflects seasonal (long-term) patterns and tidal (short-term) fluctuations of DO. Figures 1-3 illustrate this approach using fixed station TF5.6 and a nearby continuous monitoring station in JMSTF1.

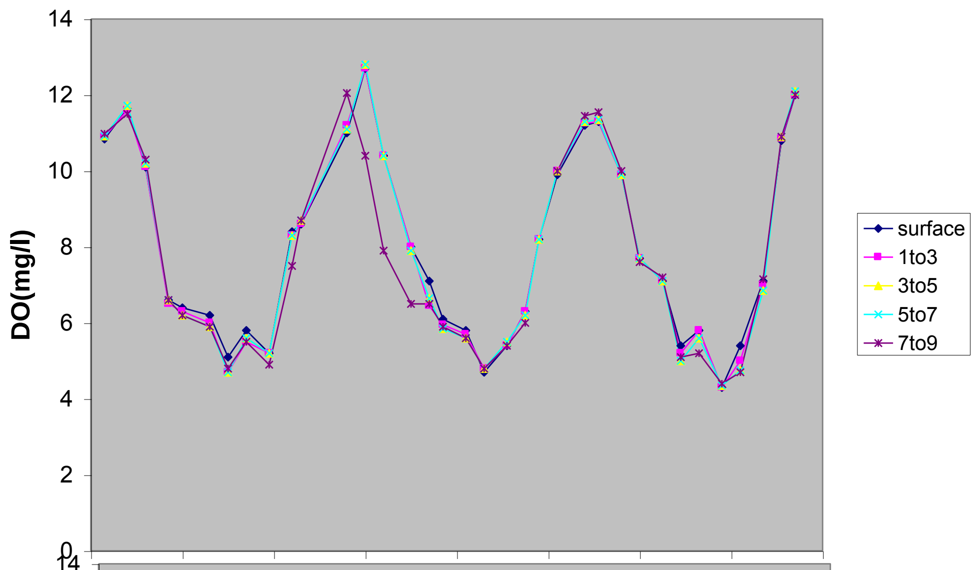


Figure 1. Three-year monthly time series (Jan-2006 to Dec-2008) of surface dissolved oxygen values from station TF5.6 in JMSTF1.

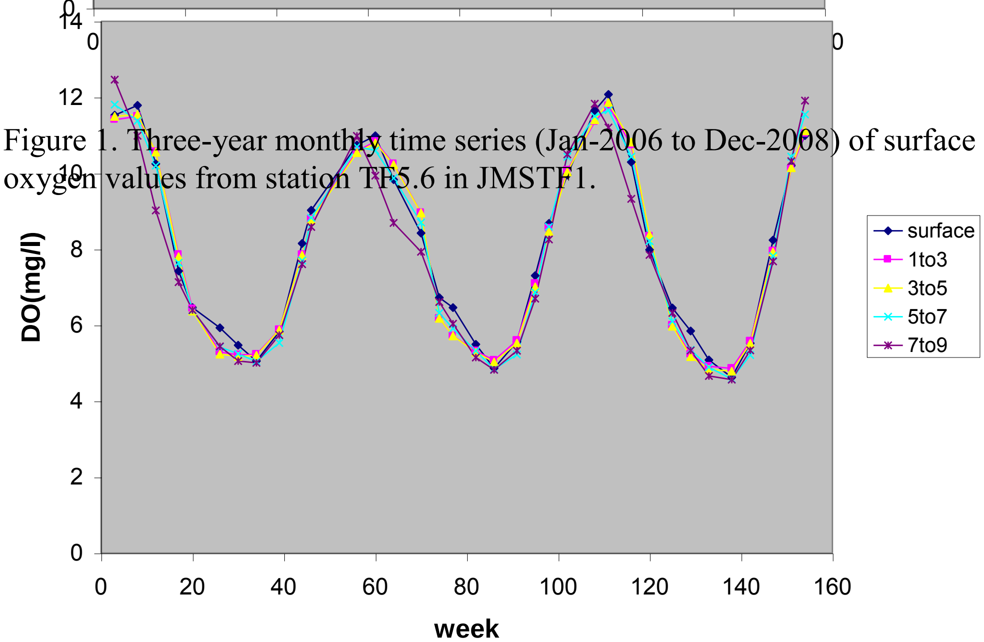


Figure 2. Modeled time-series of DO from station TF5.6. The average coefficient of determination is 0.94.

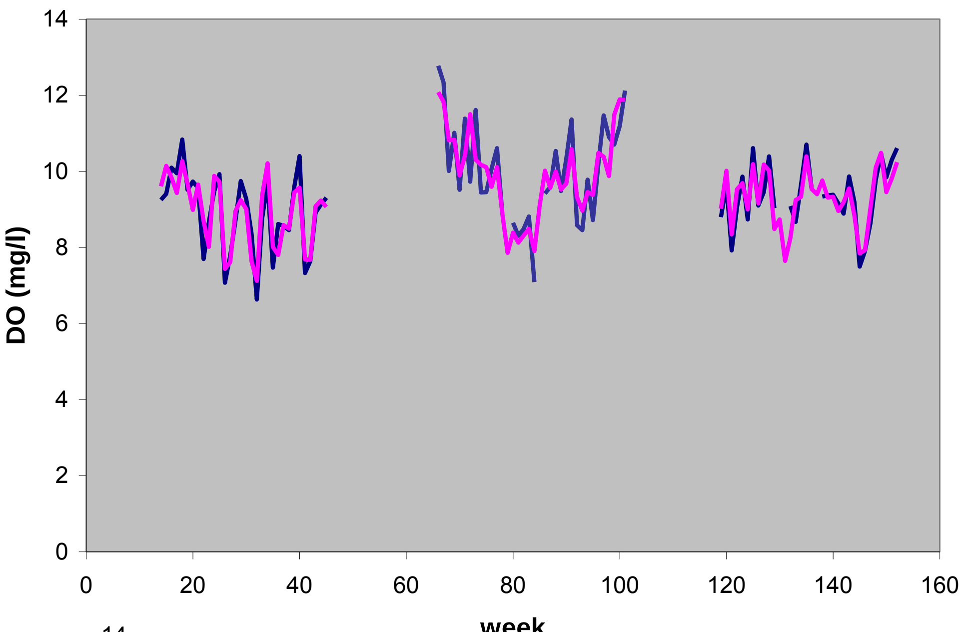


Figure 3. The empirical (blue) and modeled (pink) time-series of DO from a continuous monitoring station in JMSTF1. The average coefficient of determination is 0.71 for the three models (one for each year).

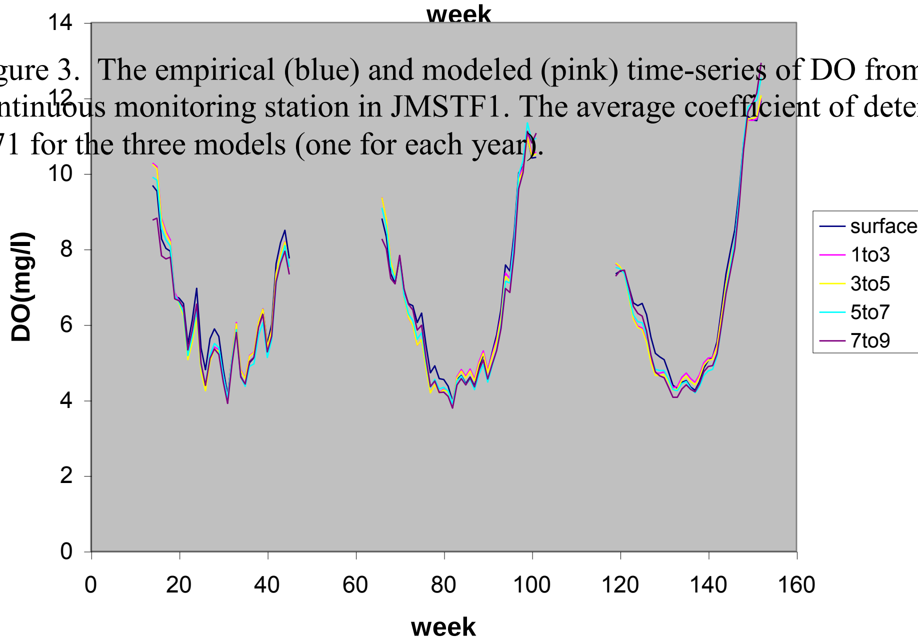


Figure 4. The “synthetic” dataset created by combining the long-term model shown in Fig. 2 with the short-term models shown in Fig 3.

Ideally, monthly time series generated from all long-term stations within close proximity (~ 3 km) of a continuous monitoring station would be “transformed” so that daily or weekly averages could be estimated, interpolated spatially, and then assessed using the CFD approach.

The approach is not without its considerations, however. Most importantly, its application rests of the assumption that short-term fluctuations in shallow water are similar to fluctuations in the midchannel. As illustrated in Figure 5a, this is not necessarily the case. Daily averages of DO values from a midchannel YSI Vertical Profiler are statistically different, in terms of both variability and overall trend, to daily averages from a shallow-water continuous monitoring station approximately 3 km away. However, when weekly averages are used instead, the variability of the two datasets is statistically similar (Figure 4b). This example underscores the importance of selecting the appropriate time interval. The spectral analysis approach is probably too risky for 1-day criteria but most likely appropriate for 7-day criteria. We need to analyze additional datasets, like those provided by NOAA’s midchannel buoys, to make a more definitive conclusion.

Vertical homogeneity of short-term patterns is another major assumption that should be tested before employing spectral analysis. Continuous monitoring stations provide information about surface conditions only. If surface conditions are significantly less or more variable than bottom conditions, using the short-term signal generated by continuous monitoring stations may result in an inaccurate assessment of DO. Vertical Profiler data show that on a weekly basis DO varies similarly throughout the water column (see Figure 5b). While these are promising results, more data are needed to thoroughly test this assumption.

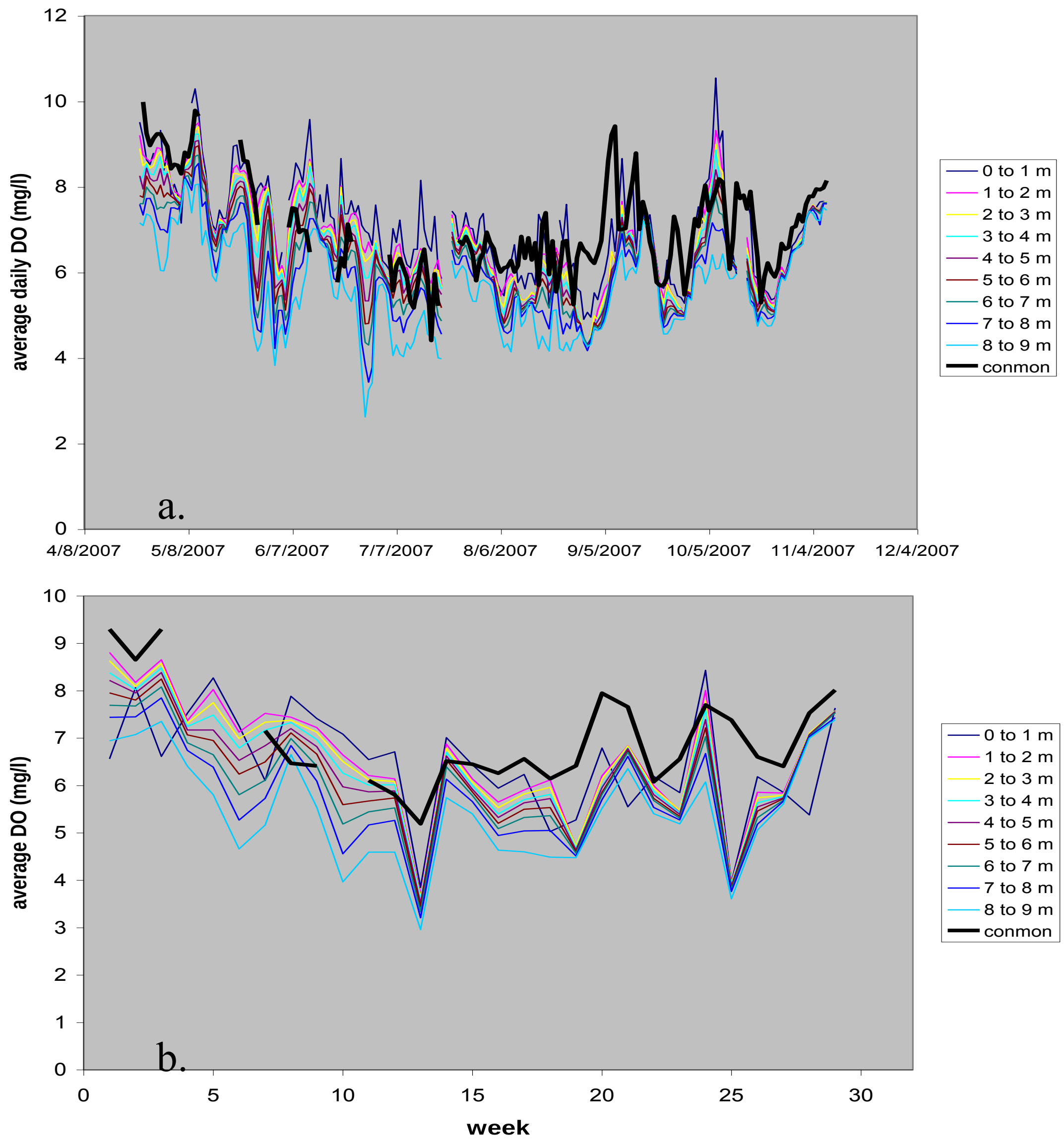


Figure 5. Comparison of DO from a Vertical Profiler and a continuous monitoring station (conmon) in YRKPH. A) DO represented by daily averages. The variance and trend are significantly different between continuous monitor and Profiler time-series. B) DO represented by weekly averages. The variance and 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.

Literature Cited

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

Appendix 10

Spectral Casting validation (Perry): Brief description of using spectral analysis to assess the 'umbrella' criterion concept.

Elgin Perry
9/1/2010

Definitions:

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, 1992) that spectral analysis could be used to create a synthetic high frequency data set at a location with only low frequency data. Because the technique involves transporting the high frequency signals from one location to a nearby location, I propose we call this technique 'Spectral Casting' - an analogy to casting with a spinning rod.

Sending Site: Spectral casting involves combining the high frequency signal from one site with the low frequency signal of a second site. The result is a synthetic high frequency record for the low frequency only site. Because the high frequency signal is being transported from one site to another, I call 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. For example, the 30-day mean criterion can be assessed with low frequency sampling whereas the instantaneous minimum requires high frequency data.

Description of Method:

Spectral Casting uses spectral analysis as an interpolation device to create a synthetic high frequency data set for locations where high frequency data are not available. The sending site is a location with high frequency data (e.g., 5-, 15-, 30-min intervals) which might be taken by an automated DO sensor on a buoy. The receiving site is a location with low frequency data (e.g., 1 or 2 x/month) which might be one of the fixed station monitoring locations of the CBP fixed station network. High Frequency synthetic data for the receiving site are formulated by combining the low frequency signal from the receiving location with the high frequency signal for the sending location. The computation begins by using a Fast Fourier Transform (FFT) to obtain a spectral decomposition at both locations (equations 1 and 2). High frequency terms are trimmed from equation 1 (frequency < 1 month) to leave a smooth function that interpolates the long term means (Figure 1.) for the receiving site. The mean term and low frequency terms (frequency > 1 month) are trimmed from equation 2 to leave just short-term variation about zero (Figure 2) at the sending site. These two trimmed series are summed to form the synthetic data which track the long term mean for receiving site and reflect the short term variability of the sending site (Figure 3). The high frequency criteria for the receiving site are then assessed using the high frequency synthetic data.

$$x_t^r = \bar{x}^r + \sum_{k=1}^{rn/2} \left\{ a_k^{rt} \cos(2\pi f_k^r t) + b_k^r \sin(2\pi f_k^r t) \right\} \qquad \text{Eqn. 1}$$

where:
r is a superscript indicating receiving,

x_t^r are observations in the long term time series, $t = 1,2,\ldots rn$

$\bar{x}^r$ is the mean over the *rn* observations

a_k^r, b_k^r are spectral coefficients estimated by the FFT

f_k^r are Fourier frequencies

$$x_t^s = \bar{x}^s + \sum_{k=1}^{sn/2} \left\{ a_k^s \cos(2\pi f_k^s t) + b_k^s \sin(2\pi f_k^s t) \right\} \qquad \text{Eqn. 2}$$

where terms are defined analogously to Eqn. 1 with *s* indicating sending site.

Umbrella Assessment Project

Various projects are underway to assess the Umbrella concept. As a rule, we are seeking criteria which require only low frequency data that will serve as an umbrella for criteria that require high frequency data. One method of assessment is to use spectral casting to generate synthetic high frequency data for the fixed station monitoring network. These synthetic high frequency data can then be processed through existing computer software to complete a CFD assessment of both the low frequency and high frequency criteria. A comparison of the assessment of the two criteria is a test of the umbrella concept.

Steps for the spectral casting project:

1. Develop a method for assigning sending sites to receiving sites.
2. Use results of Step 1 above to create synthetic data for receiving sites in test. If the sending site is observed at 15 minute intervals then a 3-year synthetic data set has 3x365x24x4= 105,120 observations. This takes 1-2 minutes per site x depth so for the 50+ stations in the main bay fixed station network, this will represent 8-10 hours of computation.

3. After populating the high frequency time series, these data are interpolated for spatial coverage. This will entail 105,120 interpolations which at 15 seconds per interpolation (the reported execution speed of the VB interpolator) will require 438 hours of computation and require an enormous amount of disk storage. The disk storage problem could likely be solved by summarizing for step 4. as we go.

4. The last step (which encapsulates all steps of CFD assessment) entails using the results of step 3 to compute the percent of space and percent of time criteria exceedances for the CFD assessment. A comparison of CFD results for high and low frequency criteria serve to test the umbrella concept.

This is a validation of the spectral casting method. Two near continuous data sets from the buoy data files are selected (Figure 1., Table 1.).

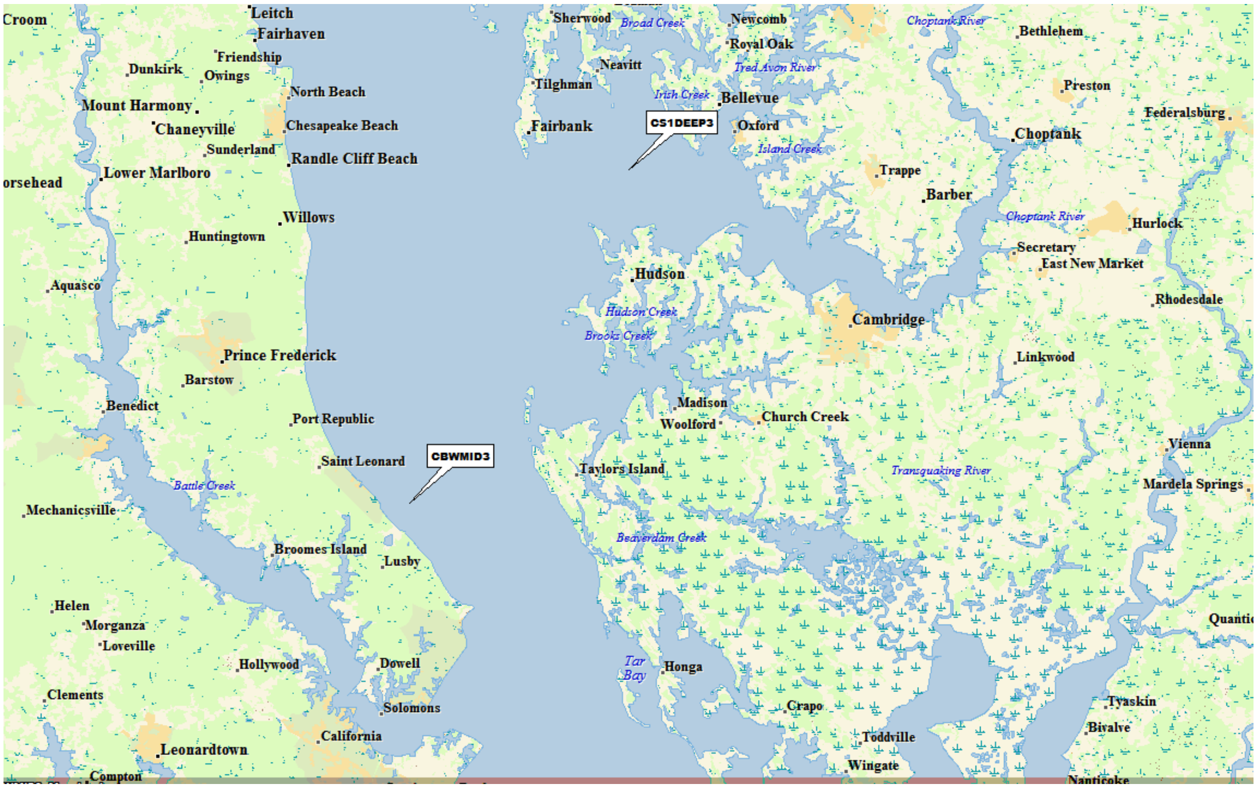


Figure 26. Location of send and receive sites.

Table 1. Properties of send and receive sites used in this test of spectral casting.

DataName	source	Lat	Long	Sensor Depth	Total Depth	Time Interval	Begin Date	End Date
CBWMID3	SANFORD	38.4517	76.4333	6	11	5	12-Aug-87	9-Sep-87
CE1DEEP3	SANFORD	38.5542	76.3967	19	23	15	12-Aug-87	9-Sep-87
CE2DEEP3	SANFORD	38.6458	76.3097	13.1	16	5	12-Aug-87	9-Sep-87
CS1DEEP3	SANFORD	38.6625	76.2567	6	8	12	12-Aug-87	7-Sep-87

CSDEEP3 is the sending site, CBWMID3 is the receiving site.

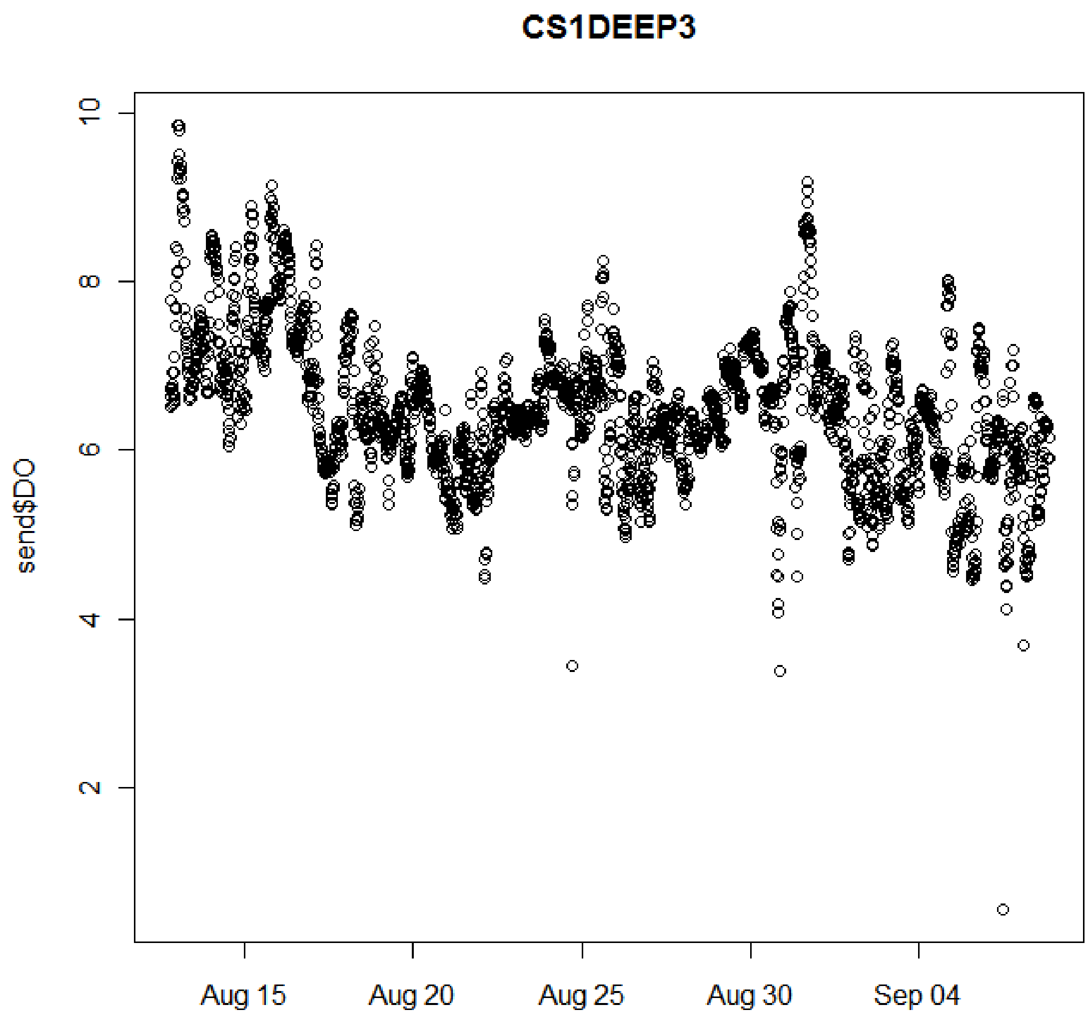


Figure 27. Time series of DO data for sending site.

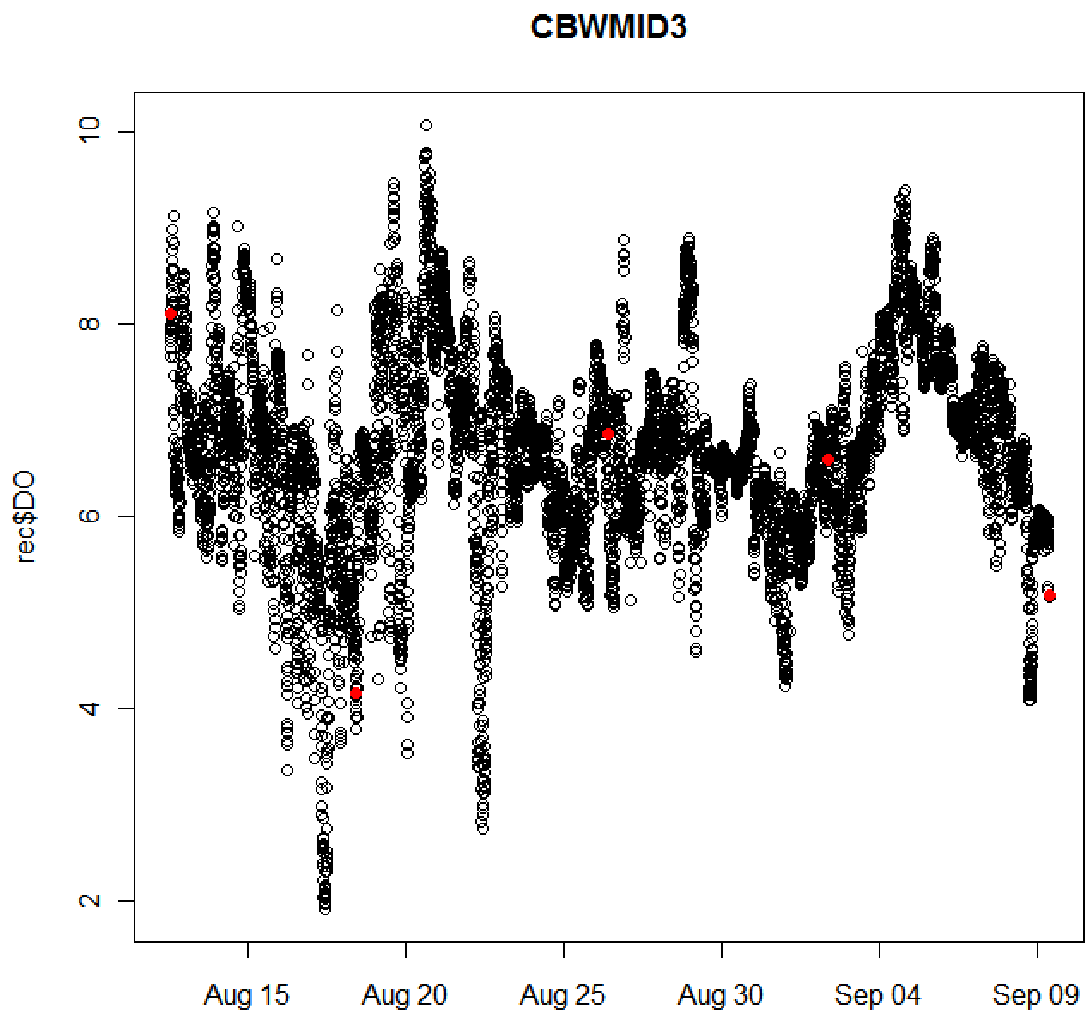


Figure 28. Time series of DO data for receiving site. Red dots are subsample for low frequency interpolation.

The low frequency sub-sample was chosen as the beginning and end + points once a week. The once a week points were chosen as day-time (morning) and about the same time of day.

1987-08-12 13:20:00

1987-08-18 10:20:00

1987-08-26 11:00:00

1987-09-02 09:50:00

1987-09-09 09:20:00

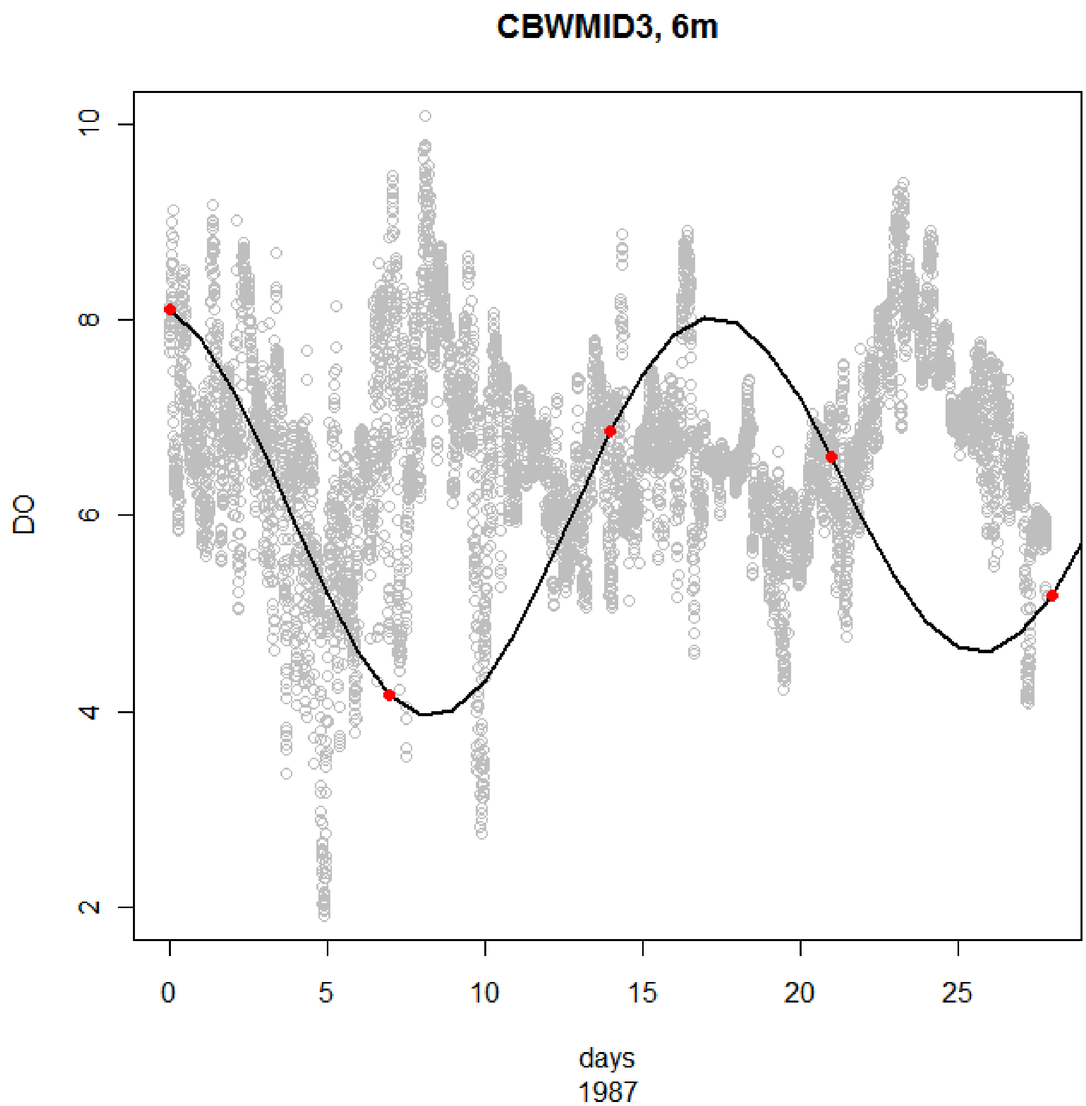


Figure 29. Time series of DO data for receiving site. Red dots are subsample for low frequency interpolation. Black line is interpolating function.

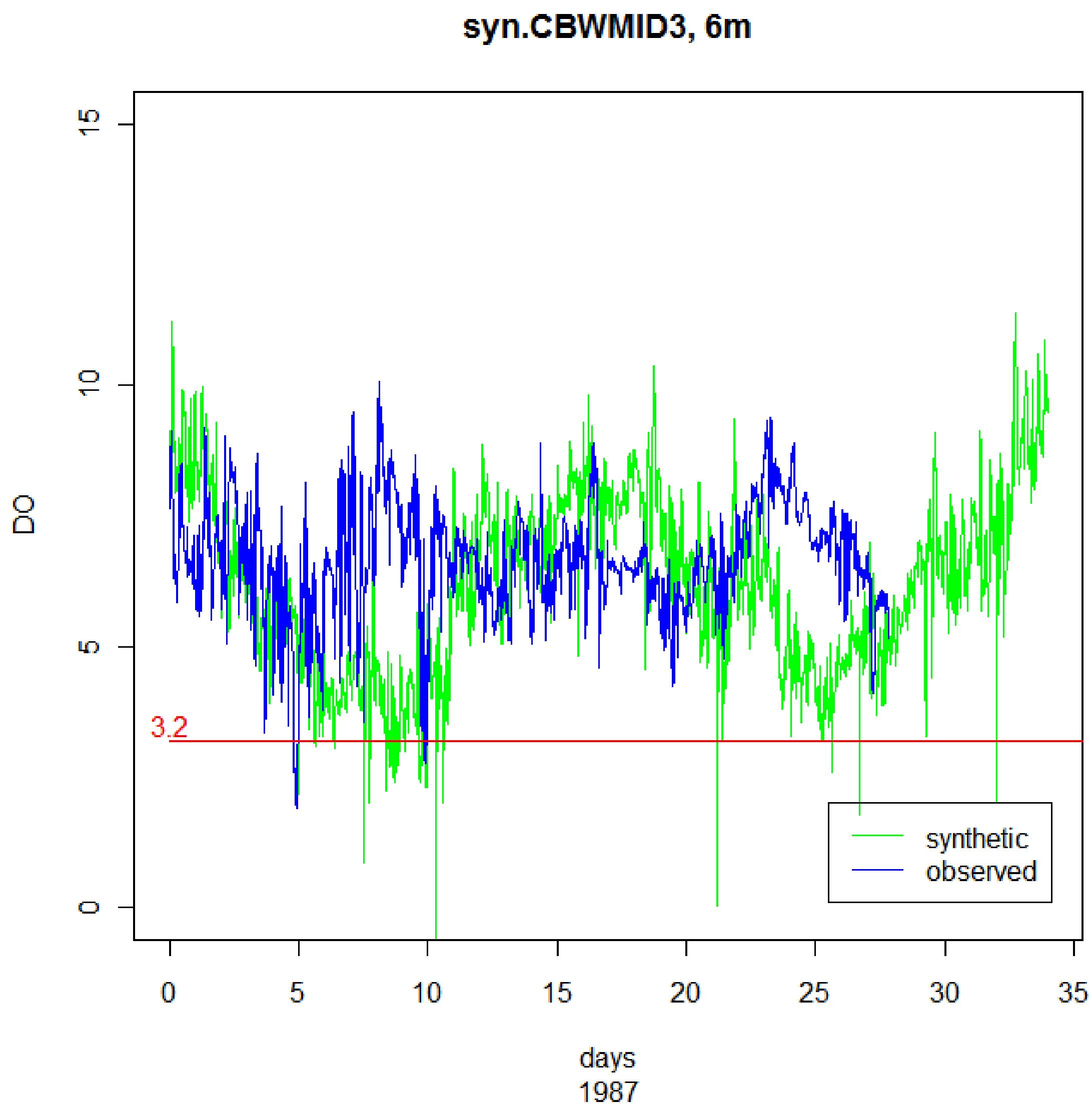


Figure 30. Time series of observed (blue) and synthetic (green) data at the receiving site.

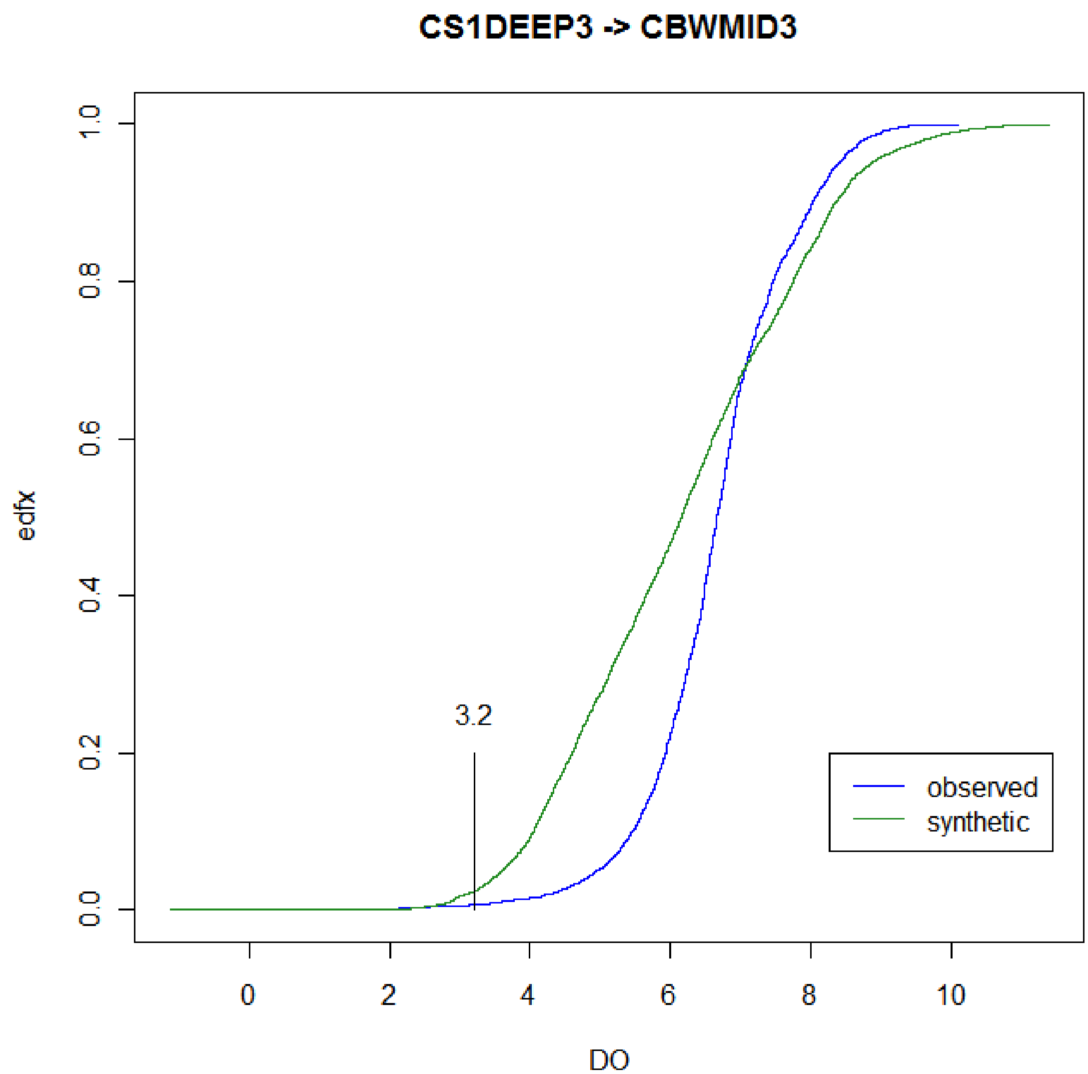


Figure 31. Empirical Cumulative Distribution Function for observed and synthetic data at receiving site.

- **Synthetic data show slightly greater violations of the Instantaneous Minimum criterion (3.2) than observed data.**
- **If the curves were shifted to the left (i.e. lower mean DO) the difference in violation rate could be substantial.**
- **Overall, the synthetic data have greater variability than the observed, but this greater variability did not originate at the sending site.**

var(rec$DO) = 1.089469
var(rec.syn$DO) = 2.787914
var(send$DO) = 0.7394112

sd(rec$DO) = 1.043776
sd(rec.syn$DO) = 1.669705
sd(send$DO) = 0.8598902

APPENDIX 11

Notes on relative noise contribution of low frequency time subsampling vs. high frequency casting, based on Shallow Water Data.

Elgin Perry

2/22/2011

In my last discussion, I reviewed the relative contribution to uncertainty from two sources in the spectral casting process for data from the Southern Bay. 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. Each step of the process will cause the percent of violations in the synthetic data to deviate from the percent of violations in the true DO time series. ***Using a long term - high frequency record near VIMS as a receiving site and casting from 10 different sites with 3 days of high frequency data from around the southern Bay, produced results that showed greater uncertainty due to the selection of the low frequency sample than due to the selection of casting site***.

Here I present more results using this same approach on shallow water data.

Brief recap of the methods.

To assess which step contributes the greatest uncertainty, we perform a validation excercise. In the validation, we subsample a high frequency time series to create a low frequency subsample that is the receiving site of the spectral cast. The low frequency subsample is interpolated to as if it were low frequency time series from the fixed station data. Using Fourier analysis, the high frequency variation from a sending site is cast in and convoluted with the low frequency interpolation to form synthetic data. The validation step is to compare 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 allow us to differentiate uncertainty due to low frequency sampling and uncertainty due to casting.

Step 1 is to examine the variability due to low frequency sample selection. To examine this, two ConMon sites are selected. One will serve as the receiving site and the other as the sending site. For one iteration of the spectral casting process, a low frequency (once every two weeks) sample is selected from the receiving site. This low frequency sample is interpolated using a Fourier series to estimate the central tendency of the synthetic data. Spectral Analysis is used to capture the high frequency signal from a subset of the sending site which is super-imposed on the central tendency to create the synthetic data. The synthetic data are compared to the observed high frequency data at the receiving site. This spectral casting process is repeated iteratively using a different low frequency subsample for each iteration and using the same high frequency signal. The variability among these iterations measures uncertainty due to the selection of the low frequency subsample.

Step 2 is to examine the variability due to choosing different sending data. In this exercise, we hold the low frequency subsample constant and study the variation contributed by different sending data. Unlike the assessment of buoy data in the southern bay where different sending sites were used to assess variability due to casting, for these shallow water data, a single sending site is used and data from this site are broken into two-week segments to create multiple sending data sets. Thus we hold the subset selected as the low frequency sample constant and use multiple temporal subsets from the sending site to created variability due to casting selection.

Results:

The two sites (yellow squares in Figure 1.) used for this exercise are Maryland ConMon Locations 'XBF7904' near St. George's Island and 'XCD5599' in Breton Bay (Figure 1). Both locations were sampled at 15 minute intervals in the years 2006-2009.

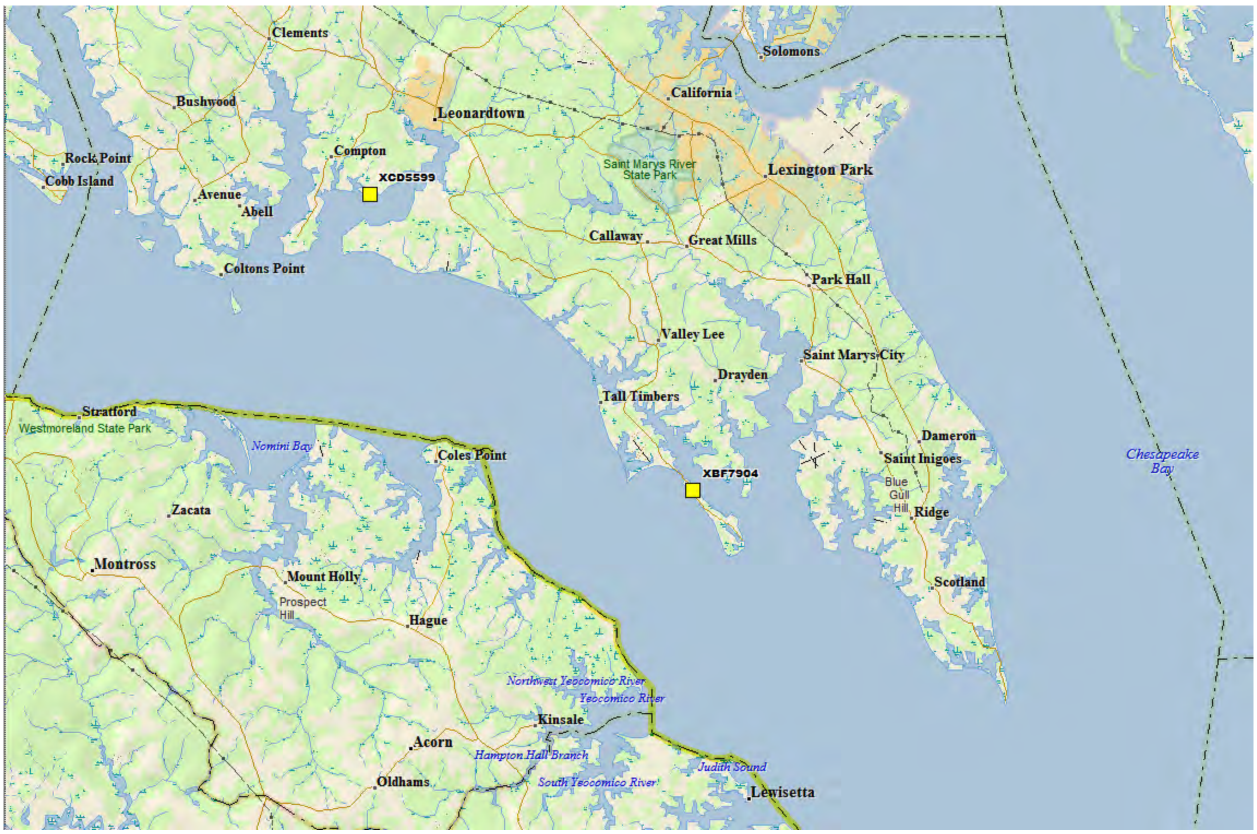


Figure 32. Location of two sites used to compare uncertainty from low frequency sampling vs. uncertainty from casting.

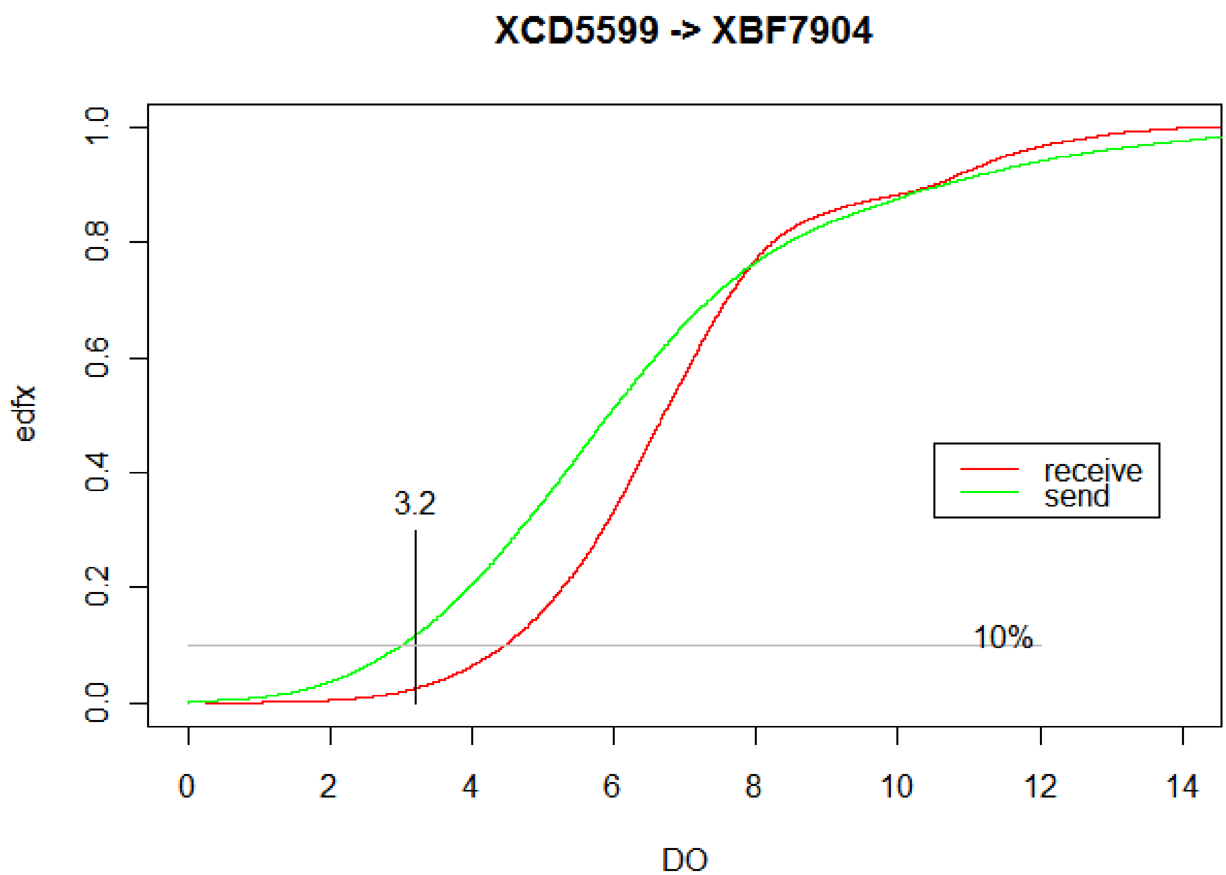


Figure 33. Cumulative empirical distribution functions for sending and receiving sites.

Over the three year period, the sending site has a greater proportion of violations of the instantaneous minimum criterion than the receiving site (Figure 2.). In addition, the sending site show greater variability in DO than does the receiving site. The greater variability at the sending site is mostly in the form of more low values than at the recieving site.

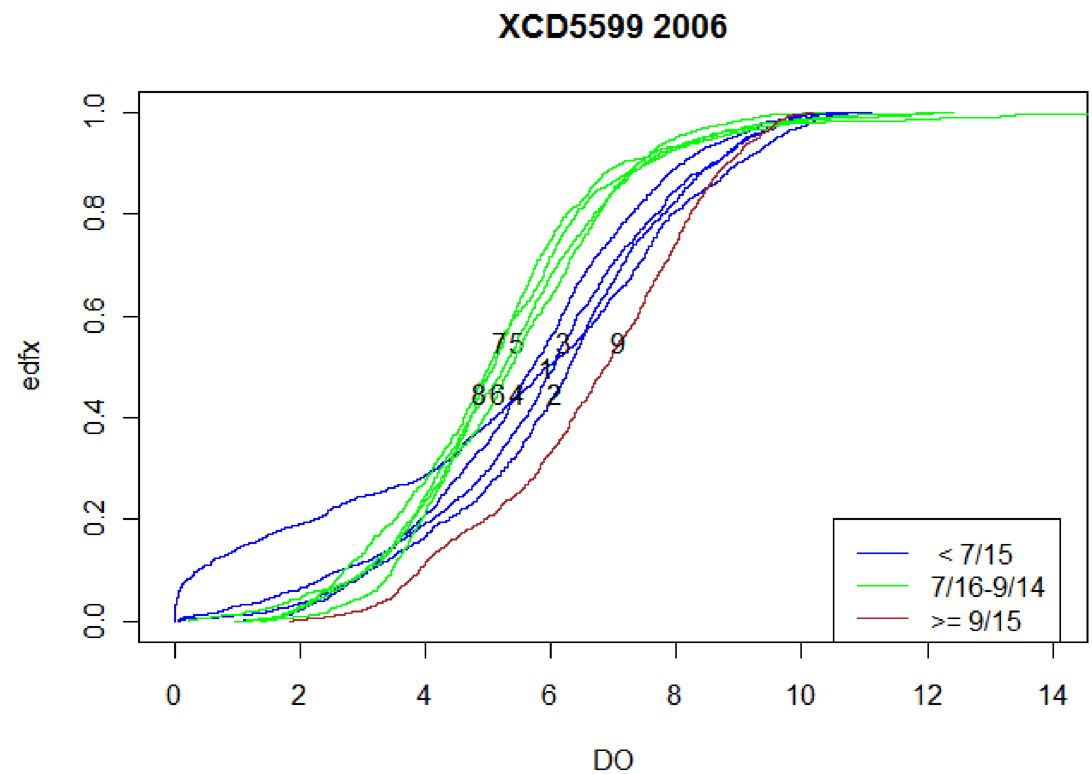


Figure 34. The cumulative empirical distribution function of raw data for each sending data set for 2006.

To create multiple sending data sets, for each of 3 years of data, the summer time series was broken into two week intervals. To get so concept of the variability among these different sending data sets, the ECDF for each one is plotted both in raw form (Figure 3.) and after the long term signal is removed leaving only the high frequency variability (Figure 4.). The 2006 data show a seasonal pattern of DO being relatively high through mid July (blue curves), low in late July and August (green curves) and returning to relatively high in fall (brown curve).

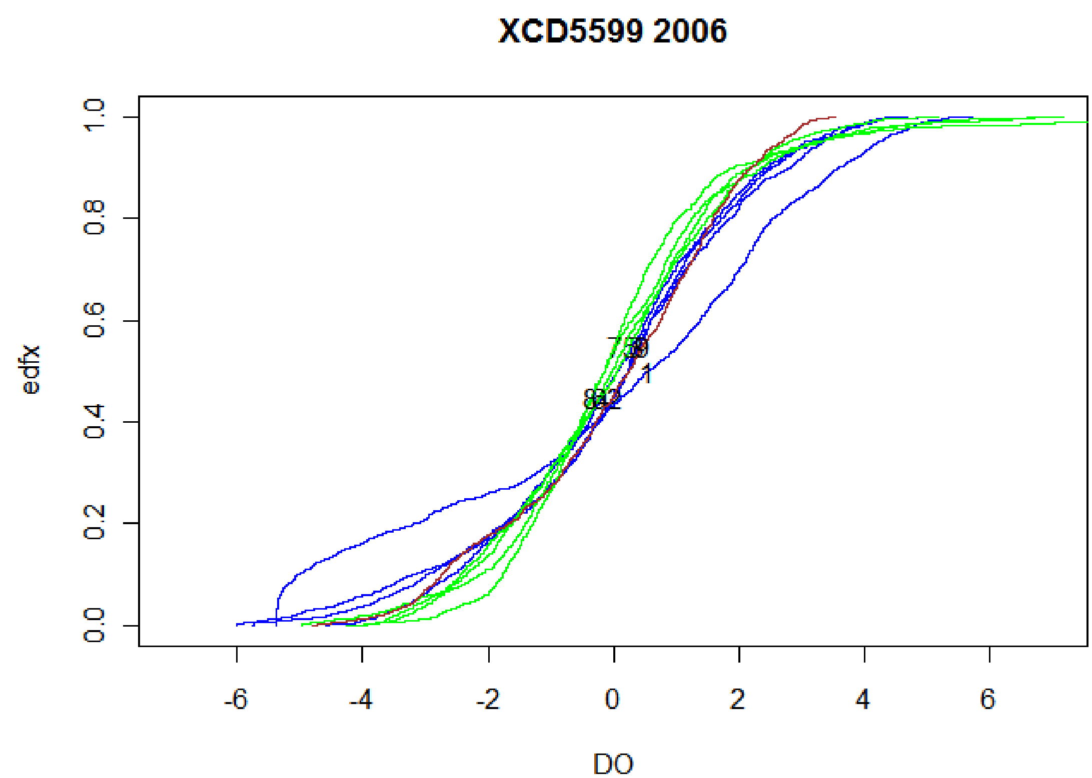


Figure 35. The cumulative empirical distribution function of data adjusted for central tendency leaving just high frequency variability for each sending data set for 2006.

The short-term variability for each two-week period (Figure 4.) shows a much tighter cluster than the raw data. One curve (labeled 1 in Figures 3 and 4) show a marked deviation from the remainder. It is the first 2-week period of the 2006 time series and has a mix of very high and very low DO values. Looking at a time series plot (Figure 5.) for this two-week period reveals a period low DO that occurred early in the season. While the data are unusual, it does appear to be a valid data record.

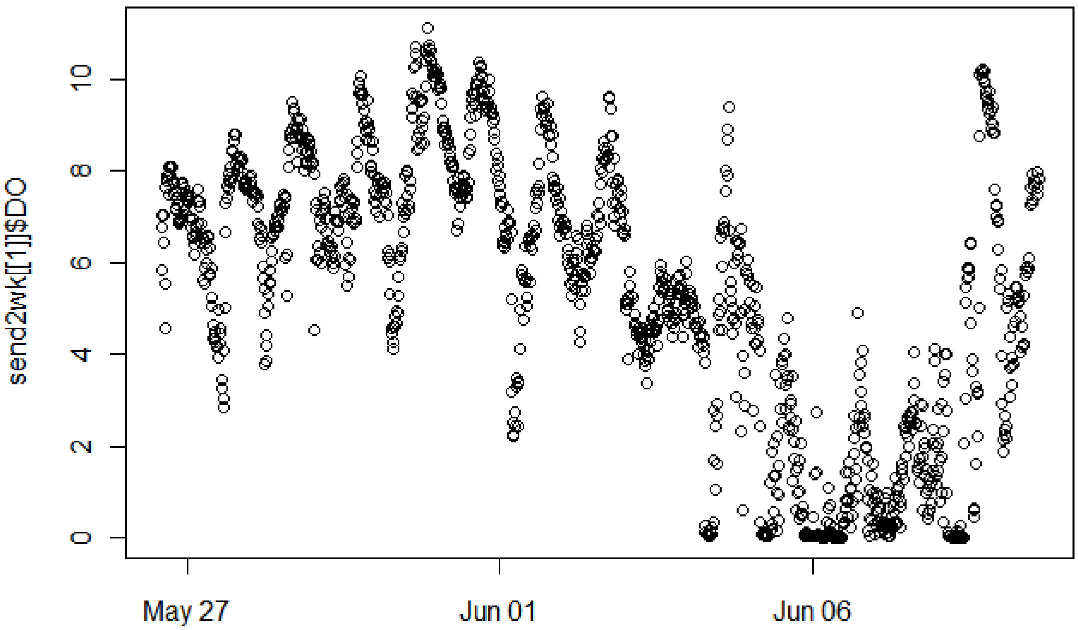


Figure 36. Time series plot of DO from the 1st two-week period in 2006 at the sending site.

This exercise of comparing the variability among cedf's for raw data and cedf's for centered data is repeated for the years 2007 (Figure 6.) and 2008 (Figure 7.).

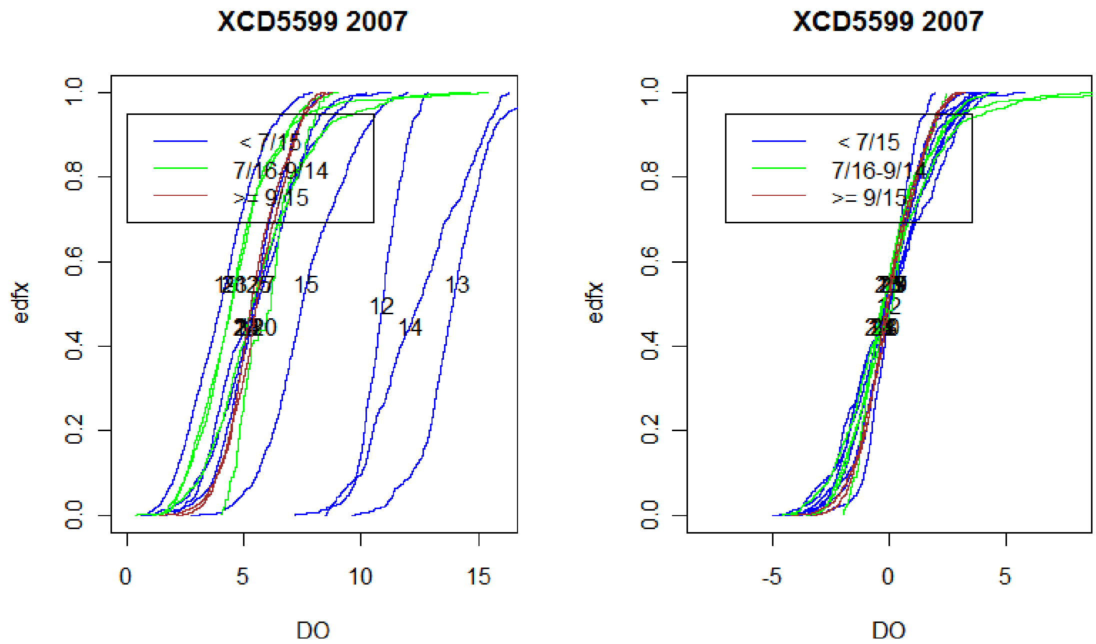


Figure 37. Comparison raw data (left) and centered data (right) for 2007.

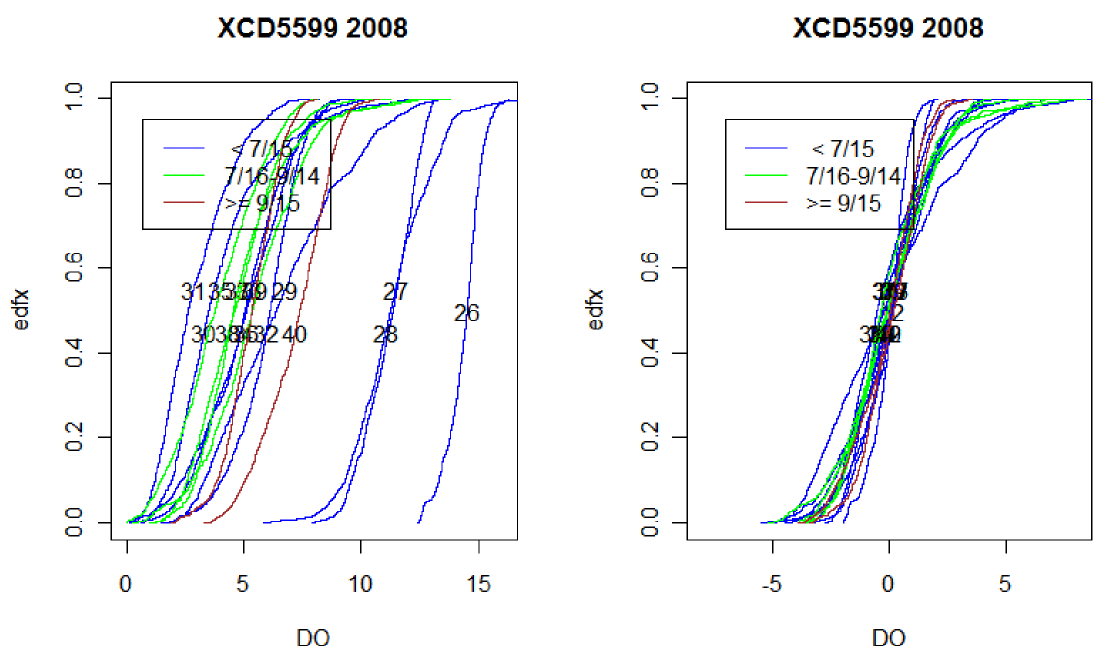


Figure 38. Comparison raw data (left) and centered data (right) for 2008.

The seasonal pattern in the raw data differs among the three years, but the centered data shows a fairly tight cluster in each case.

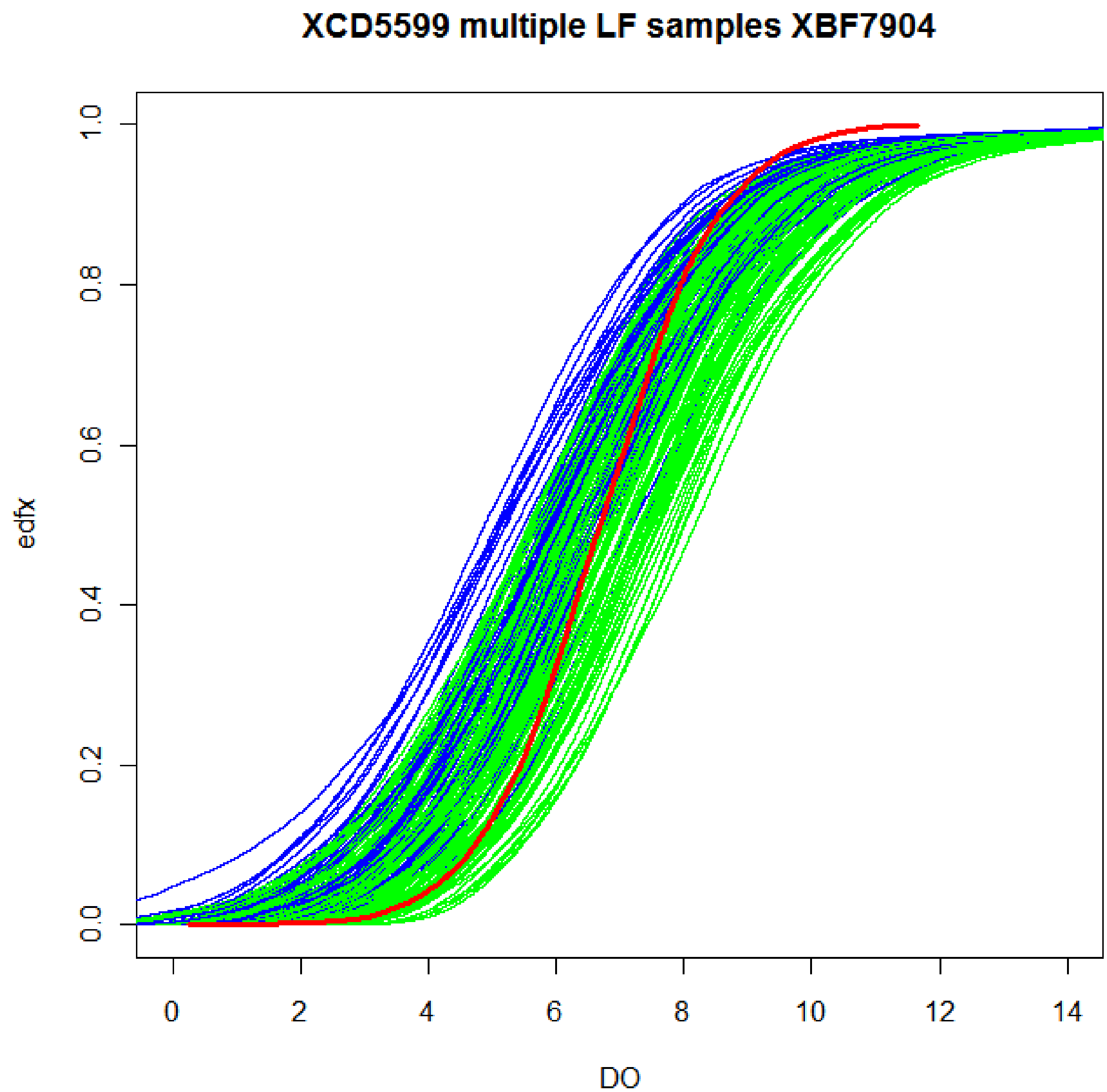


Figure 39. Variation due to multiple low-frequency samples from the receiving site with Fourier Series interpolation. The sending data set is held constant at one two-week interval. Blue curves synthetic data based on a series of night samples. Green curves are from a series of day samples. The red curve is the receiving site high frequency data.

The results of step 1, holding the sending data constant and varying the low frequency subsample (Figure 8.) show considerable variability. In general the synthetic data show greater failure of the instantaneous minimum relative to the observed data. However, it is clear that by collecting low frequency data only during the day is possible to obtain synthetic data that are anti-conservative.

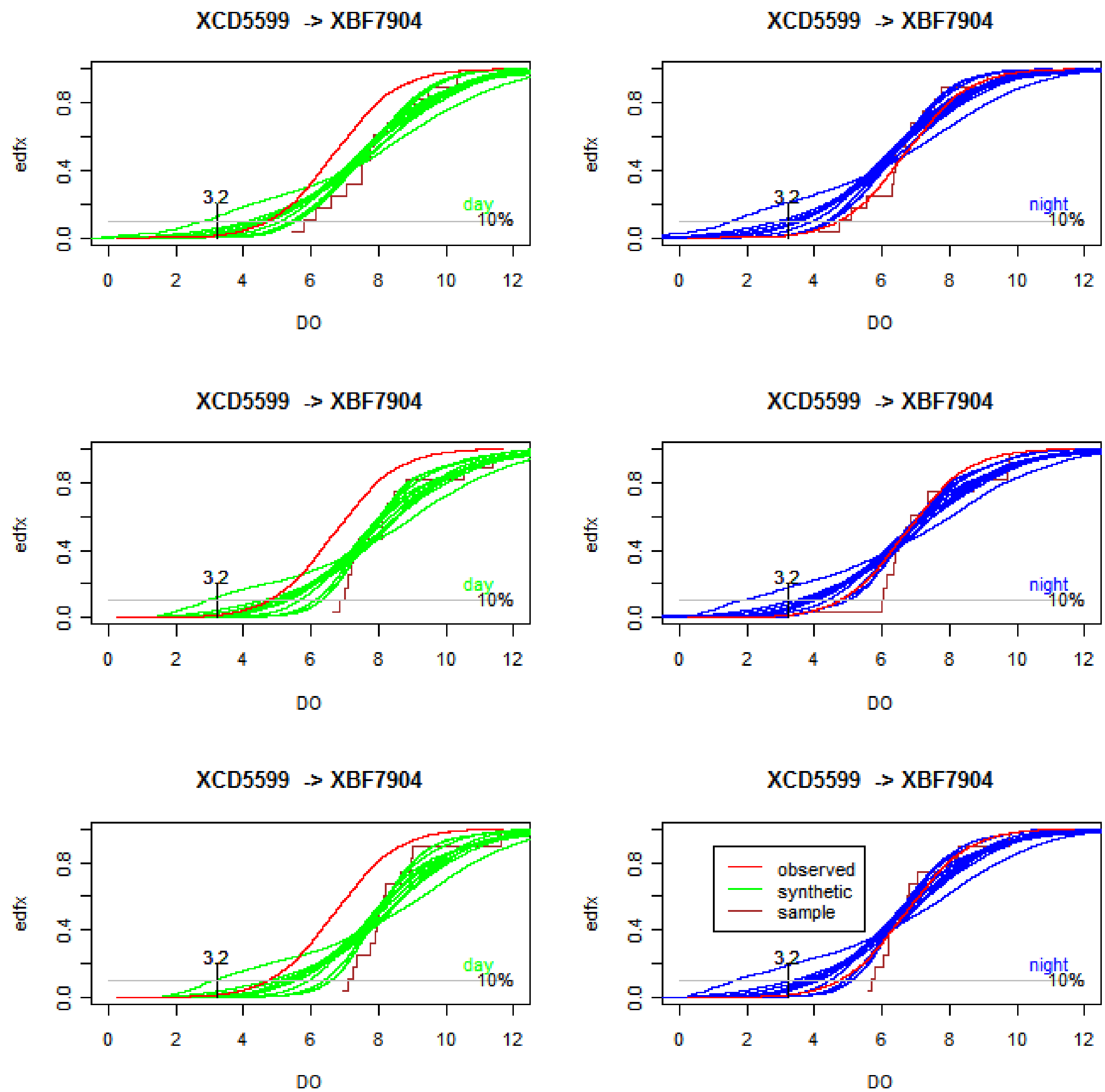


Figure 40. In each panel are the original receiving high frequency data (red) , the low-frequency sample (brown) and 12 synthetic data sets (blue or green) base on 12 sending data sets. Green curves use low-frequency data collected during the day. Blue curves use low frequency data collected at night. Each panel is for a different set of low-frequency data.

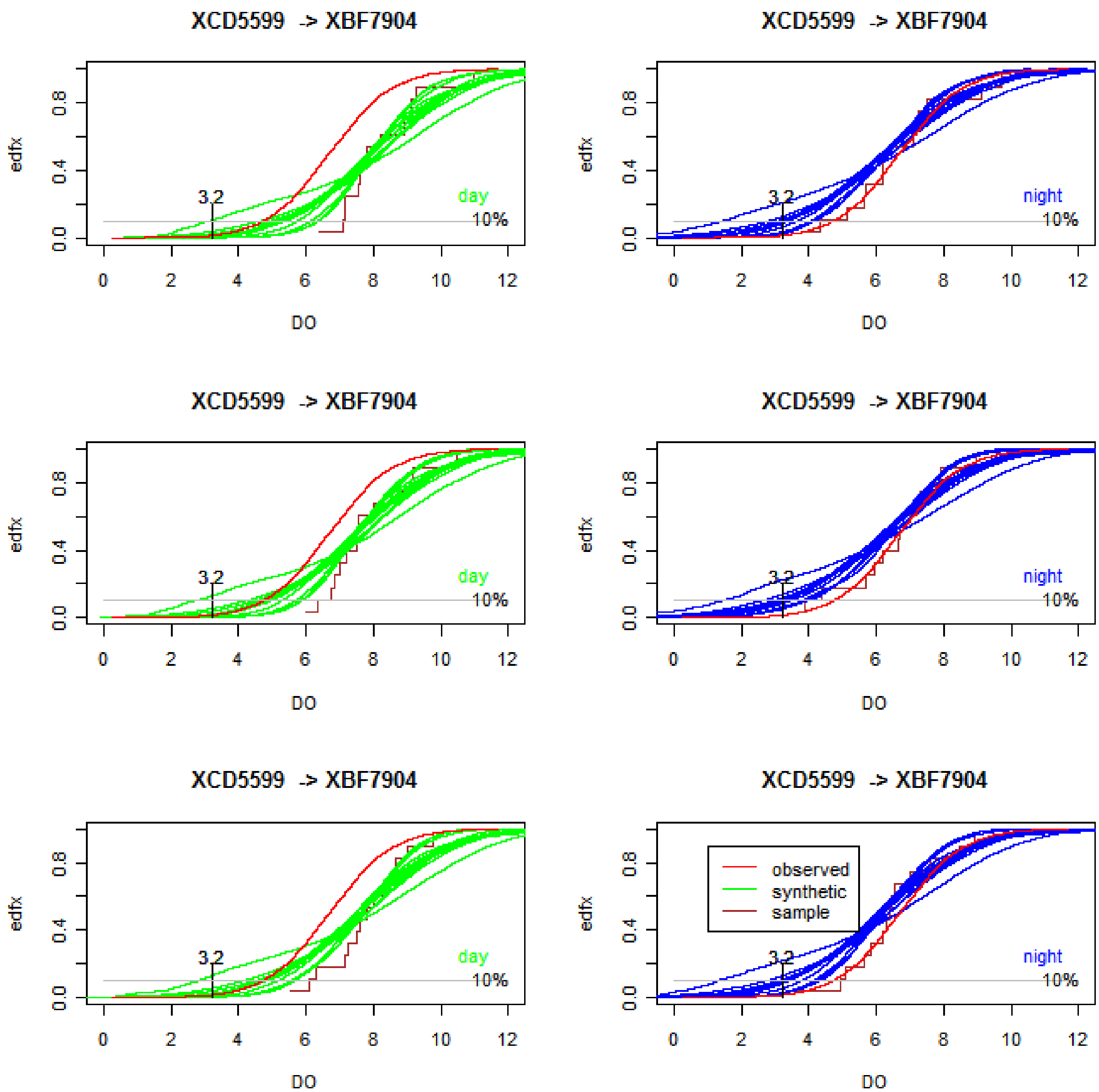


Figure 41. This is a continuation of results in Figure 9.

Each panel of this figure shows 12 synthetic data sets (green or blue curves) based on 12 sending data sets superimposed on 1 low frequency sample (brown step curve) as compared to the original high frequency receiving data (red). Moving from panel to panel changes the low frequency sample (brown) that was drawn from the high frequency data (red). It is clear that daytime sampling leads to a positive bias in the estimated status of DO while night sampling results in negative bias. It is also clear that deviations among the 12 Castings tend to be less than the deviations of the low frequency sample from the true CDF.

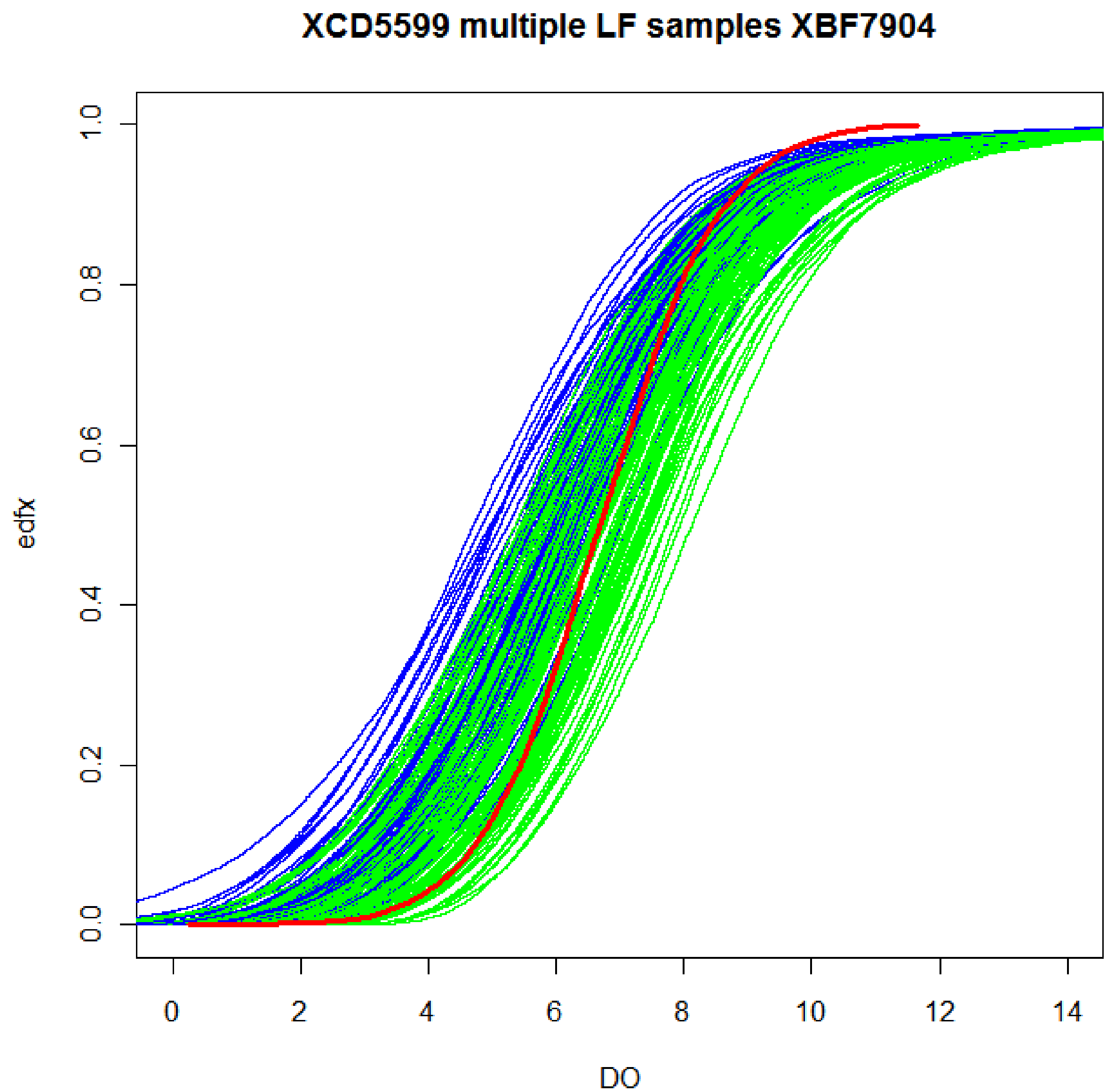


Figure 42. This figure shows the variation due to different low-frequency samples with Cubic Spline interpolation. The sending data set is held constant at one two-week interval.

Comparing Cubic Spline interpolation (Figure 11.) with Fourier Series interpolation (Figure 8.) for the low-frequency component of the synthetic data appears to make little difference.

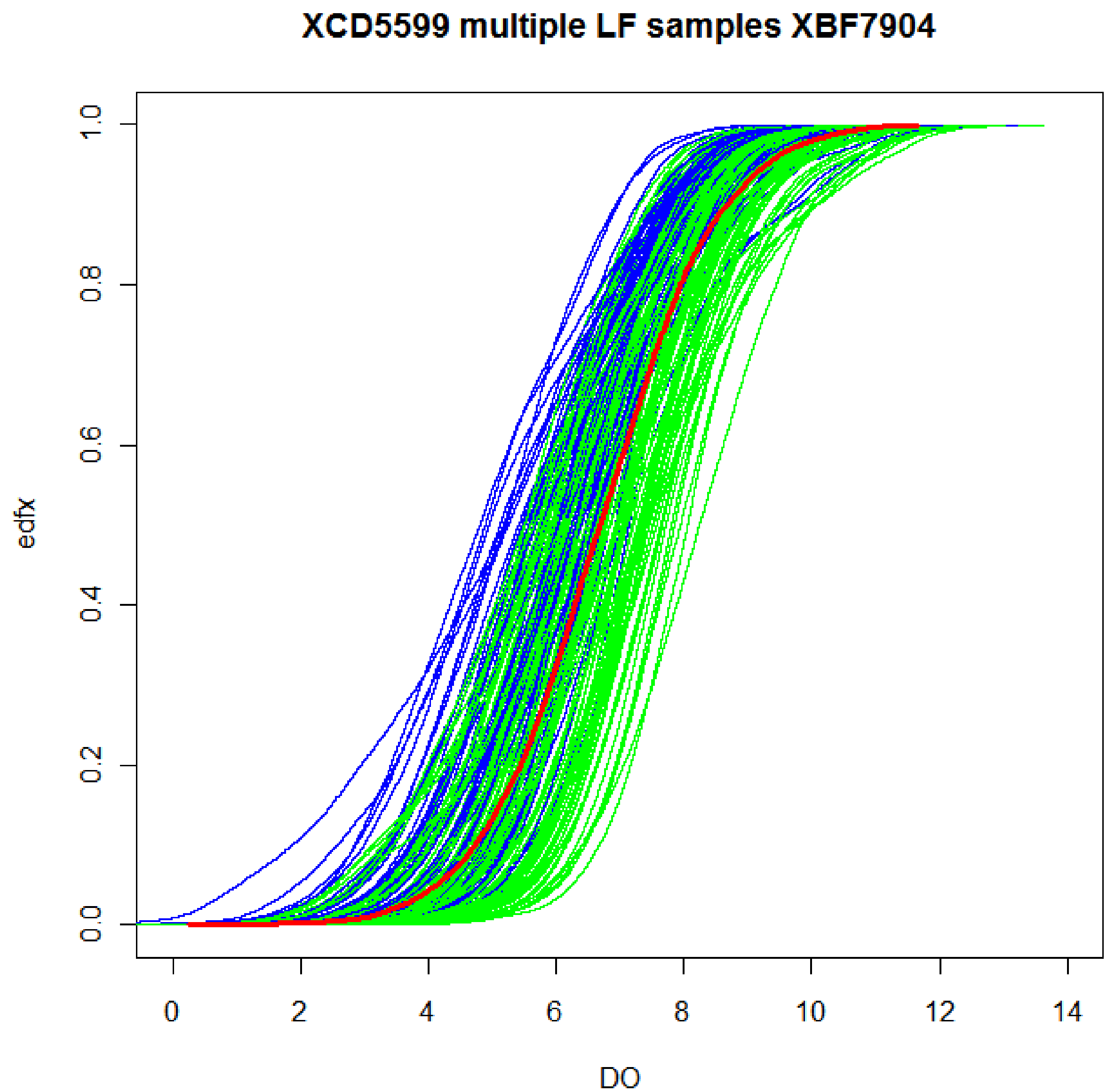


Figure 43. This figure shows the variation due to different low-frequency samples with Simple Linear interpolation. The sending data set is held constant at one two-week interval.

Comparing linear interpolation (Figure 12.) to Cubic Spline interpolation (Figure 11.) or Fourier Series interpolation (Figure 8.), it appears that linear interpolation yields a better fit.

APPENDIX 12

A parametric simulation approach to assessing the umbrella concept for the instantaneous minimum criteria.

Elgin Perry
3/10/2011

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. Here we develop and show results from a simulation approach to addressing this question.

Methods

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. Each fitting results in a vector of 7 parameters as follows:

b_Int - the intercept which reflects the mean because other covariates are centered,
b_cday - linear trend term for the week fitted as a coefficient of centered day,
b_sin, b_cos - coefficients for diel trend fitted to trig-transformed time with a 24 hour period,
b_ar1,b_ar2 - autoregressive terms at lags 1 and 2, and
mse - residual mean square error

These parameter estimates were obtained for each short time series to yield 251 sets of parameters. These 251 vector observations were analyzed by Multivariate Analysis of Variance (MANOVA) using Proc GLM in SAS. The MANOVA model included 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. For the CB4 data, the MANOVA model was simplified by dropping Latitude and Longitude which leaves terms for Month, Total Water Depth, and Sensor Depth. Coefficients from this MANOVA model were used to estimate a mean predicted value for the time series parameter vector which seeded the parametric simulation. A multivariate normal random number generator (R-package) was used to generate 1000 realizations of the time series parameter vector using the mean vector and the Variance-Covariance matrix estimate from the MANOVA. Each of these 1000 realization of the time series parameter vector were passed to a function which simulated a 1-week time series based on the simulated parameter vector values. The percent of violations of the instantaneous minimum criterion (3.2 ppt) were tabulated for each 1-week time series yielding 1000 estimates of this percentage. The range and frequency of these percentages are compared for various mean vectors associated with different conditions specified by different values of the independent variables in the MANOVA model.

Results:

Results are presented for all buoy locations to examine general trends and for the subset of locations in segment CB4 that supports the simulation experiment.

MANOVA results all locations

When examining data from all buoy locations, in a multivariate sense, all of these terms are statistically significant (Table 1).

Table 1. Manova test results for dependent vector (b_int,b_cday,b_sin,b_cos,b_AR1,b_AR2,mse) for all Buoy sites.

Source	Pillai's Trace	Pr > F
month	0.2895	0.0191
TotDep	0.1018	0.0007
SampDep	0.2063	<.0001
lat	0.0592	0.0451
long	0.2102	<.0001

Table 2 shows which independent variables appeared to have an effect on which dependent variables for all Buoy sites and Table 3 gives the coefficient estimates for the covariates to show the direction of association.

Table 2. P-values for each manova term and for each dependent variable.

Source	b_int	b_cday	b_sin	b_cos	b_AR1	b_AR2	mse
month	0.0861	0.9041	0.3811	0.4845	0.0130	0.0909	0.1277
TotDep	<.0001	0.4168	0.9888	0.7560	0.1728	0.2066	0.1374
SampDep	<.0001	0.4214	0.0381	0.5415	0.1808	0.2711	0.0331
lat	0.2065	0.3651	0.2688	0.0563	0.9958	0.2387	0.1713
long	0.7956	0.0432	0.9265	0.9906	<.0001	0.2204	0.0290

Table 3. Coefficient estimates for covariates for all Buoy sites.

Source	b_int	b_cday	b_sin	b_cos	b_AR1	b_AR2	mse
TotDep	0.2224	0.0060	0.0001	-0.0031	-0.0106	0.0080	0.0148
SampDep	-0.4079	-0.0072	0.0309	-0.0074	0.0125	-0.0083	-0.0255

lat	-0.2449	0.0244	-0.0496	0.0703	0.0001	0.0271	0.0493
long	0.1058	-0.1157	0.0087	-0.0009	-0.3149	0.0595	0.1666

Some general trends to notes from tables 2 and 3 are the following:

- There are seasonal trends to mean DO and the AR parameters.
- DO seems to improve as water depth increases other things held constant.
- DO degrades as sensor depth increases.
- AR1 terms are stronger in the western bay
- mse decreases with sensor depth

Table 4. Partial Correlation Coefficients from the Error SSCP Matrix / Prob > |r| DF = 239 for all Buoy sites.

	b_Int	b_cday	b_sin	b_cos	b_AR1	b_AR2	MSE
b_Int	1.000000	-.052225 0.4206	-.116969 0.0705	0.113032 0.0805	0.252967 <.0001	-.225183 0.0004	-.078779 0.2240
b_cday	-.052225 0.4206	1.000000	0.128183 0.0473	-.019640 0.7621	0.083167 0.1992	-.026105 0.6874	-.132840 0.0398
b_sin	-.116969 0.0705	0.128183 0.0473	1.000000	-.074374 0.2511	-.296165 <.0001	0.205687 0.0014	0.020856 0.7479
b_cos	0.113032 0.0805	-.019640 0.7621	-.074374 0.2511	1.000000	0.095132 0.1417	-.089933 0.1649	-.185441 0.0039
b_AR1	0.252967 <.0001	0.083167 0.1992	-.296165 <.0001	0.095132 0.1417	1.000000	-.816881 <.0001	-.297462 <.0001
b_AR2	-.225183 0.0004	-.026105 0.6874	0.205687 0.0014	-.089933 0.1649	-.816881 <.0001	1.000000	0.264092 <.0001
MSE	-.078779 0.2240	-.132840 0.0398	0.020856 0.7479	-.185441 0.0039	-.297462 <.0001	0.264092 <.0001	1.000000

Notes on Table 4.

Strongest correlation is among parameters that model the error process. The autoregressive terms b_AR1 and b_AR2 have an inverse dependence. The mse term is - correlated with AR1, + correlated with AR2 and shows some association with the diel cycle terms.

There is little correlation among terms that model the mean (i.e. b_int, b_cday, b_sin, b_cos)

MANOVA results CB4

MANOVA results for CB4 are generally similar to those for the whole Buoy set.

Table 5. Manova test results for dependent vector (b_int,b_cday,b_sin,b_cos,b_AR1,b_AR2,mse) for CB4.

Source	Pillai's Trace	Pr > F
month	1.9229	<.0001
TotDep	0.6465	<.0001
SampDep	0.5022	<.0001

Table 6. P-values for each manova term and for each dependent variable.

Source	b_int	b_cday	b_sin	b_cos	b_AR1	b_AR2	mse
month	<.0001	0.9053	<.0001	0.0221	<.0001	<.0001	0.4314
TotDep	<.0001	0.7663	0.1743	0.3736	<.0001	<.0001	0.0008
SampDep	<.0001	0.9410	0.0633	0.1758	0.9755	0.8605	0.0461

Table 7. Coefficient estimates for covariates for CB4.

Source	b_int	b_cday	b_sin	b_cos	b_AR1	b_AR2	mse
TotDep	0.4657	-.0045	0.0512	-0.0275	-0.0596	0.0533	0.1121
SampDep	-0.6674	-.0016	0.1042	-0.0620	0.0004	-.0025	-.0948

Table 8. Partial Correlation Coefficients from the Error SSCP Matrix / Prob > \|r\| for CB4.

DF = 49	b_Int	b_cday	b_sin	b_cos	b_AR1	b_AR2	MSE
b_Int	1.000000	-0.023319 0.8723	0.021175 0.8840	0.259961 0.0683	-0.166289 0.2484	0.267643 0.0602	-0.066908 0.6443
b_cday	-0.023319 0.8723	1.000000	0.168352 0.2425	0.447174 0.0011	0.113026 0.4345	-0.087273 0.5467	0.034049 0.8144
b_sin	0.021175 0.8840	0.168352 0.2425	1.000000	0.061012 0.6738	-0.228950 0.1098	0.120024 0.4064	-0.137566 0.3408
b_cos	0.259961 0.0683	0.447174 0.0011	0.061012 0.6738	1.000000	-0.100788 0.4862	0.171738 0.2330	-0.129673 0.3694
b_AR1	-	0.113026	-	-	1.000000	-	-

	0.166289 0.2484	0.4345	0.228950 0.1098	0.100788 0.4862		0.900034 <.0001	0.073316 0.6129
b_AR2	0.267643 0.0602	- 0.087273 0.5467	0.120024 0.4064	0.171738 0.2330	- 0.900034 <.0001	1.000000	- 0.074231 0.6084
MSE	- 0.066908 0.6443	0.034049 0.8144	- 0.137566 0.3408	- 0.129673 0.3694	- 0.073316 0.6129	- 0.074231 0.6084	1.000000

CB4 - violation rate results-

Using the MANOVA model for CB4 we can obtain a predicted value of the time series parameter vector as a function of month, water depth, and sensor depth. In this simuluation I have started with a choice of month, water depth, and sensor depth for which the mean DO is just greater than the 30 day mean criterion of 5.0.

The independent variable vector that yields this prediction is

May	Jun	Jul	Aug	Sep	Oct	WaterDepth	SensorDepth
0	0	1	0	0	0	10	6

for which the predicted vector of time series parameters is

b_Int	b_cday	b_sin	b_cos	b_AR1	b_AR2	mse
5.0058	-0.0493	-0.4072	-0.0527	0.9333	-0.0319	0.3164

This predicted vector and the estimated Variance-Covariance matrix is used to seed a multivariate normal random number generator that creates 1000 realizations of the time series parameter vector. A one week time series of 15 minute observations is generated for each realization. The b_Int term of this predicted vector is the weekly mean of the one week time series. Based on the 15 minute observations, the percent of observations below the instantaneous minimum criterion is computed. The umbrella concept is assessed by comparing the true monthly mean (5.0058), the simulated weekly means (b_Int) in the 1000 realizations, and the violation rates of the instantaneous minimum in the 15 minute observations.

By changing the SensorDepth of the independent variable vector, the longterm mean can be adjusted to assess the effect of this parameter on the relationship among the three criteria assessments. Thus by raising the sensor depth from 6m to 3m the mean DO is increased from 5.0058 to 7.0082 (Table 9). The time series parameters for diel signal and the mse term increase as well. The linear trend term and the AR terms remain fairly constant.

Table 9. Predicted values of the time series parameter vector as a function of decreasing Sensor Depth.

Sensor Depth	b_Int	b_cday	b_sin	b_cos	b_AR1	b_AR2	mse
6	5.0058	-0.0493	-0.4072	-0.0527	0.9333	-0.0319	0.3164
5	5.6733	-0.0476	-0.5114	0.0094	0.9328	-0.0294	0.4112
4	6.3408	-0.0460	-0.6156	0.0714	0.9324	-0.0268	0.5060
3	7.0082	-0.0443	-0.7198	0.1335	0.9320	-0.0243	0.6008

To compare violation rates of the 7-day criterion and instantaneous criterion I crosstabulate cases where the 7 day mean < 4.0 against cases where the violation rate of the instantaneous minimum exceeds 10% in each 1 week time series.

Table 10. Crosstabulation of violations of 7-day criterion with an indicator variable for violations of the instantaneous minimum exceeding 10% for a given level of sensor depth and mean DO.

		7-day mean >= 4.0	7-day mean < 4.0	marginal failure instantaneous minimum
Table 10a.				
Sensor Depth = 6	failure Instantaneous	542	9	551
mean DO = 5.0058	minimum < 10%	65.86 %	5.08 %	55.1 %
	failure Instantaneous	281	168	449
	minimum > 10%	34.14 %	94.92 %	44.9 %
	marginal for failure	823	177	1000
	of 7-day mean	100%	100%	100%
Table 10b.				
Sensor Depth = 5	failure Instantaneous	656	1	657
mean DO = 5.6733	minimum < 10%	69.42 %	1.82 %	65.7 %
	failure Instantaneous	289	54	343
	minimum > 10%	30.58 %	98.18 %	34.3 %
	marginal for failure	945	55	1000
	of 7-day mean	100%	100%	100%
Table 10c.				
Sensor Depth = 4	failure Instantaneous	746	0	746
mean DO = 6.3408	minimum < 10%	76.2 %	0 %	74.6 %
	failure Instantaneous	233	21	254
	minimum > 10%	23.8 %	100 %	25.4 %
	marginal for failure	979	21	1000
	of 7-day mean	100%	100%	100%
Table 10d.				
Sensor Depth = 3	failure Instantaneous	834	0	834
mean DO = 7.0082	minimum < 10%	83.57 %	0 %	83.4 %
	failure Instantaneous	164	2	166
	minimum > 10%	16.43 %	100 %	16.6 %
	marginal for failure	998	2	1000
	of 7-day mean	100%	100%	100%

Table 11. Summary of violation rates over levels of sensor depth for ease of assessing trend.

sensor depth	6	5	4	3
Monthly Mean DO	5.0058	5.6732	6.3407	7.0082
7 day criterion failure rate	17.7%	5.5%	2.1%	2%
rate of instantaneous criterion > 10%	44.9%	34.3 %	25.4 %	16.6%

When the long term mean DO is at a 'just passing' level, the simulation predicts that the 7-day mean criterion will be violated only about 17.7% of weeks (Tables 10a, 11) . If the long term mean DO increases to 5.7 then we expect about 5%.5 weeks with failure of the 7-day criterion (Tables 10b, 11). Thus if the 30-day mean criterion is satisfied, it is quite likely that violations of the 7-day mean criterion will be satisfied unless the 30 day mean hovers in the 'just passing' zone for an extended period.

Looking at the violations of the instantaneous minimum is not so encouraging. When the long term mean is 'just satisfied', the simulation predicts that the instantaneous minimum criterion exceedance rate will exceed 10% in about 45% of weeks (Tables 10a., 11, Figure 1). Even when the long term mean DO is 7, the simulation predicts 16.6% of weeks will have an instantaneous minimum criterion exceedance rate in excess of 10% (Tables 10d, 11, Figure 1). This result suggests that the 30-day mean will not serve as an umbrella for the instantaneous minimum even when we add considerable cushion to the 30-day mean.

To assess whether the 7-day mean is and umbrella for the instantaneous minimum, consider the failure rate of the instantaneous minimum given that the 7-day mean criterion is satisfied (column 2 of tables 10a-10d, Figure 2). These results show that failure of the instantaneous minimum criterion can remain high (more than 10% failure in greater than 34% of weeks, Table 10a) even when the 7-day mean criterion is satisified. Furthermore, as the degree by which the 7-day criterion is satisfied increases to more that 95%, the failure of the instantaneous minimum criterion can remain high (more than 10% failure in greater than 16% of weeks, Table 10d.)

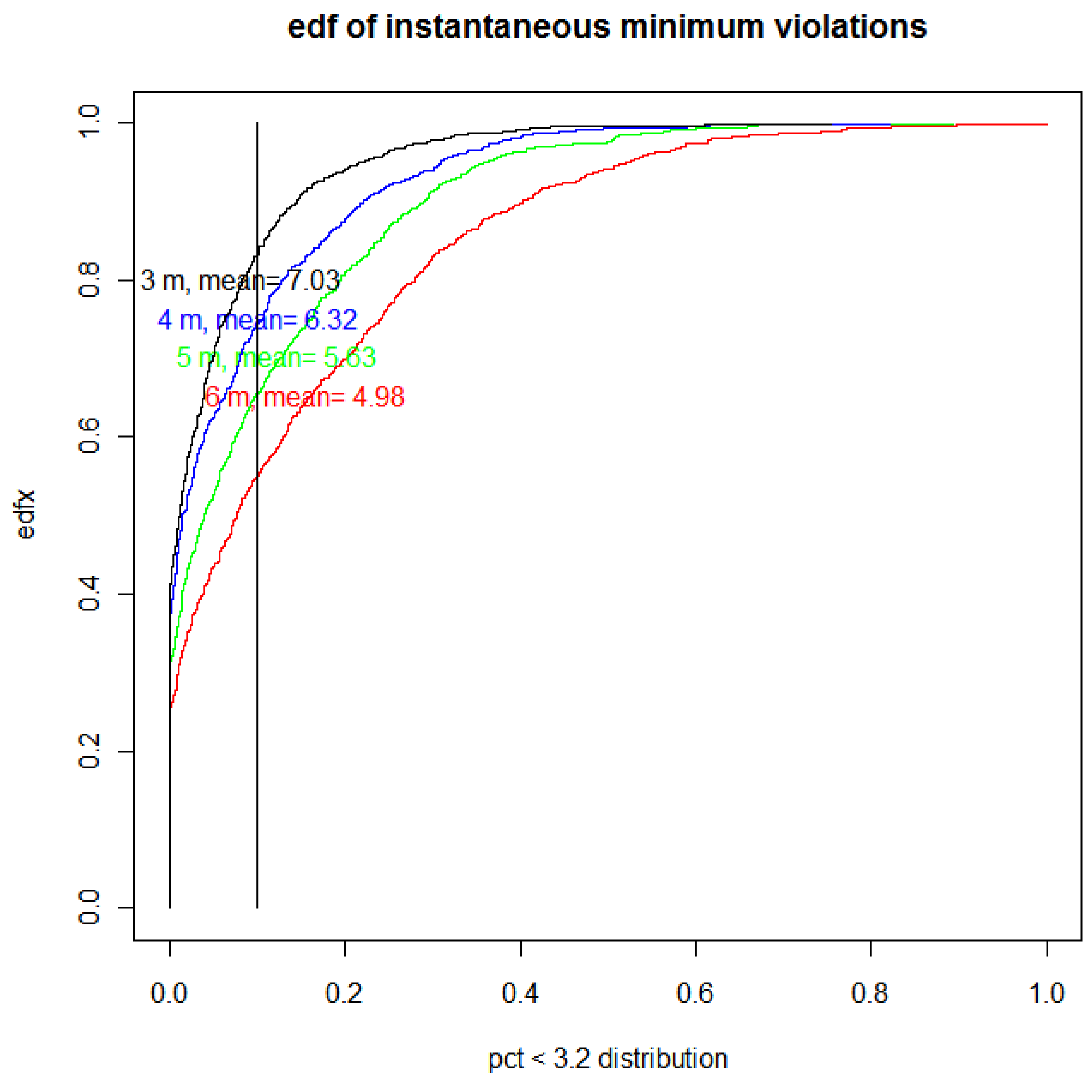


Figure 44. CEDF of percent of violation of the instantaneous mimimum as a function of sample depth and mean DO.

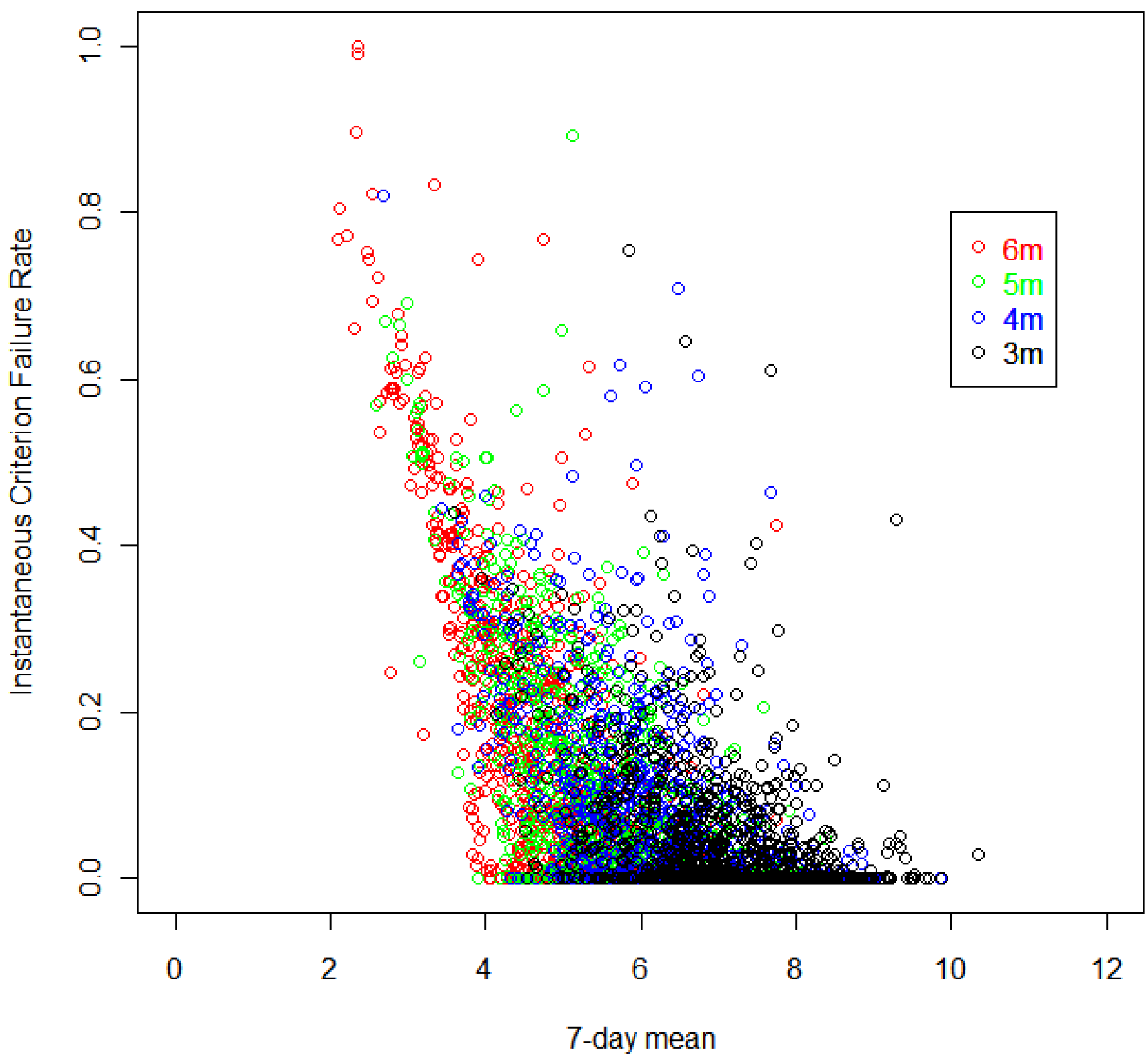


Figure 45. Percent of violation of the instantaneous mimimum as a function of 7-day mean and sample depth.