Critical Review of the Assessment Methodology for James River Chlorophyll

Virginia Department of Environmental Quality

Introduction

Since Virginia's adoption of water quality criteria specific to the Chesapeake Bay and its tidal tributaries in 2006, the Commonwealth has used a procedure for assessing James River chlorophyll originally developed by the Chesapeake Bay Program Partnership. This procedure is similar to what is used to assess Bay-wide dissolved oxygen criteria attainment. A series of technical guidance documents (EPA, 2007, 2008, 2010) present the theory and step-by-step instructions of this approach. The James River chlorophyll procedure is illustrated below in Figure 1.

Figure 1. General outline of the current assessment protocol for James River chlorophyll.

a. Data are collected at sampling locations in a segment (represented as a grid). In this example, the segment has five stations, and the values represent chlorophyll samples taken at those stations. James River chlorophyll assessments are based on both monthly station visits and spatially intensive "underway" sampling (Dataflow). Three monthly monitoring runs are shown.

	30			30
1-Jul			50	
	10			10

	10			10
1-Aug			30	
	20			10

	40			20
1-Sep			20	
	10			10

b. Data are spatially interpolated to create a segment-wide "snapshot" of chlorophyll for each monitoring run. Only surface measurements are interpolated, in accordance with the default settings of the Chesapeake Bay Interpolator (EPA, 2008). Each monitoring run in the assessment period is represented by a two-dimensional interpolation grid.

	30	40	50	30
1-Jul	40	40	50	20
	30	30	40	10
	20	30	30	10

	10	20	30	10
1-Aug	20	20	30	20
	10	30	20	10
	10	10	10	10

	40	30	20	20
1-Sep	30	20	20	10
	20	20	10	10
	10	10	10	10

c. A composite "seasonal" grid representing the chlorophyll expression of a spring or summer season is created by taking the average of the interpolation grids comprising that season. For James River chlorophyll, a geometric mean is calculated.

23	29	31	18
29	25	31	16
18	26	20	10
13	14	14	10

d. The attainment status of each composite grid cell is determined by comparing values against the appropriate criterion. The spatial exceedance rate is determined by dividing the total number of exceedances (i.e., values above the criterion) into the total number of grid cells. In this example, the criterion is 15 µg/l.

X	X	X	X
X	X	X	X
X	X	X	✓
✓	✓	✓	✓

e. The spatial exceedance rate for each season-year of the 3-year assessment period is determined. (So the above steps are repeated on two additional years' worth of data).

Season-Year	Spatial Exceedance Rate
Spring Year 1	31.3%
Spring Year 2	10.0%
Spring Year 3	25.0%

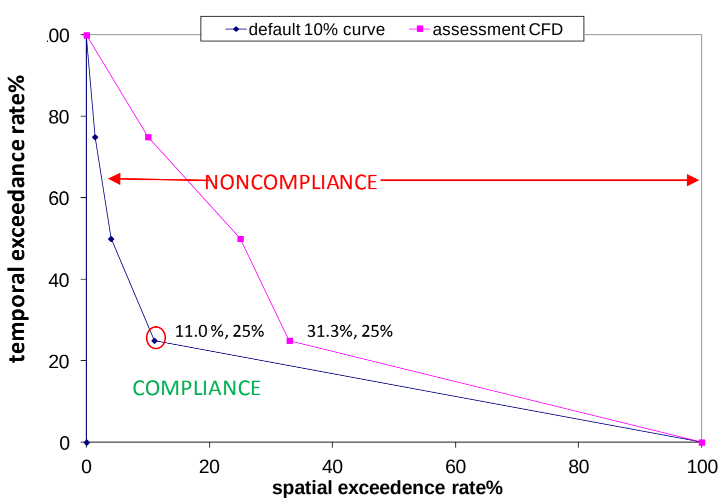
f. These spatial exceedance rates are ranked from highest to lowest.

Season-Year	Ranked Spatial Exceedance Rate
Spring Year 1	31.3%
Spring Year 3	25.0%
Spring Year 2	10.0%

g. A temporal exceedance rate is determined by using the Weibu ll equation (100*R/n+1) where R = rank and n = sample size (in this case, three).

Season-Year	Ranked Spatial Exceedance Rate	Temporal Exceedance Rate
	100%	0%
Spring Year 1	31.3%	25%
Spring Year 3	25.0%	50%
Spring Year 2	10.0%	75%
	0.0%	100%

h. Exceedance rates are plotted against a reference curve as a cumulative frequency distribution (CFD). Virginia uses a 10% hyperbolic reference curve. A segment in compliance displays a CFD which does not cross the reference CFD. Thus, the segment in this example is in noncompliance since it crosses the curve at all three points.



The statistical underpinnings of the CFD are presented in STAC (2006).

The current approach to implementing James River chlorophyll criteria is innovative. The CFD framework was developed to prevent aquatic life from losing “too much” suitable habitat to impairment while simultaneously protecting aquatic life from the impact of “too many” time intervals (days, months, seasons) spent under impairment. The CFD provides a means of defining “too much” and “too many” in a scientifically defensible, systematic way. More conventional assessment procedures usually focus solely on temporal exceedance frequency, since spatial uniformity of attainment in an assessment unit is typically assumed (EPA, 2005). Because the CFD does not require this assumption to be met, it would seem well-suited for the characterization of the large, complex segments of the Chesapeake Bay and its tidal tributaries.

While certainly qualifying as “state-of-the-art”, it is also true that the CFD approach is very experimental. In 2005, the Chesapeake Bay Program’s Scientific Technical Advisory Committee (STAC) tasked a workgroup to identify issues related to the CFD’s utility as an assessment tool. Among other “critical research tasks”, the panel recommended that further research be directed towards developing a better understanding of how well the CFD represents spatial and temporal “covariances of attainment” (STAC, 2006). But with the exception of modified reference curves for Bay-wide dissolved oxygen criteria assessments (EPA, 2010a) and the incorporation of spatially-intensive datasets (Dataflow) for use in James River chlorophyll assessments, the protocol has not changed fundamentally since the STAC report. The biases and uncertainties of the CFD, at least in the context of chlorophyll criteria attainment, remain largely unknown.

VADEQ believes that re-evaluation of the James River chlorophyll criteria necessitates scrutiny of the CFD approach as well. Both of these endeavors have been greatly enabled by the spatially- and temporally-intensive datasets that have been collected and maintained by VCU, VIMS and HRSD in recent years. But as valuable as these datasets are to chlorophyll criteria evaluation and implementation work, VADEQ anticipates that assessments will be based on a more austere monitoring program for the indeterminate future due to ever-increasing budget constraints. The enhanced monitoring conducted in the James estuary from 2011 – 2013 was only temporary in nature; going forward VADEQ will revert to the monthly (or semi-monthly) main-channel, fixed station sampling protocol that has been in traditional use. Understanding the limitations of the CFD is thus more important than ever. The CFD approach has questionable utility as an assessment tool if it requires more data than what an affordable monitoring program can provide.

VADEQ has identified two major questions concerning CFD-based assessments: 1) Are the data gathered at fixed stations sufficient for producing an accurate characterization of the spatial and spatial-temporal extent of exceedance? and 2) Does the 10% curve closely describe the exceedance distribution expected under reference conditions?

Accuracy of Spatial and Spatial-Temporal Exceedance Estimation

The question of how well the current procedure handles the widely variable nature of chlorophyll was addressed in two ways by VADEQ:

1. Evaluation of the accuracy of estimated spatial exceedance rates
2. Evaluation of the accuracy of estimated spatial-temporal exceedance rates

Spatial Exceedance Estimation

The largest source of error in the CFD approach is likely the interpolation step—wherein data collected at a few sampling locations are used to generate predications over the entire extent of a segment, allowing the proportion of "non-attaining" habitat to be estimated. One would expect to find the accuracy of spatial non-attainment estimation to be quite poor given only a small number of samples, especially for a patchy indicator like chlorophyll. But VADEQ believes it is important to confirm this assumption.

To accomplish this, spatial exceedance rates estimated from Dataflow cruise-tracks were compared to exceedance rates estimated from simulated fixed station data. Note that in this analysis, "exceedance" is defined as an instantaneous chlorophyll value greater than the appropriate season-segment criterion. These values would be seasonal means in an actual assessment. Dataflow observations were interpolated in accordance with the current assessment procedure, with predictions generated for the centroids of the Chesapeake Bay Interpolator grid (Bahner, 2006). Fixed station samples were simulated by extracting the Dataflow observations closest to actively monitored Chesapeake Bay Program (CBP) fixed stations. An interpolation grid based on those subsamples (n = 2 or 3, depending on the segment) was then created, and from this the "fixed station" spatial exceedance rate was calculated. The spatial exceedance rate derived from the interpolated Dataflow cruise was treated as the "actual" extent of exceedance, since the high resolution of these data provides a very close approximation of the true distribution of chlorophyll at any given time. The fixed station-derived spatial exceedance rate was treated as the "calculated" or "estimated" value. The percent error of each assessment was determined by the following:

$$\text{Percent error} = 100 \times \frac{|\text{Dataflow-derived spatial exceedance rate} - \text{Fixed station-derived spatial exceedance rate}|}{\text{Dataflow-derived spatial exceedance rate}}$$

Only Dataflow cruises with a low to moderate proportion of exceedances were selected for this exercise, since the reference curve that is currently used to define attainment does not "allow" any single exceedance rate greater than 11.0% (see red circle in Figure 1h). It is not so concerning if a 70% exceedance rate is calculated when the actual rate was 50%, since the segment would be judged as impaired in either case. However, an erroneous assessment decision will likely be made when, for instance, a true value of 3% is represented in the CFD with an estimate of 15%. An erroneous decision can be made even if the latter is 5%, given the current strictures of the procedure.

Table 2. Spatial exceedance rates (as percentages) estimated from interpolated Dataflow cruises and simulated CBP fixed stations samples. A low to moderate number of exceedances were observed in the selected Dataflow datasets. Only a few tidal fresh (TF) datasets could be analyzed since most of the available cruise data were collected under bloom conditions. (Data provided by VIMS and HRSD).

JMSTFU cruise date	Dataflow Spatial Exceedence Rate	CBP Stations Spatial Exceedence Rate	Percent Error
4/26/2006	2	0	100
4/26/2007	21	0	100
9/11/2008	9	100	1011
		median percent error →	**100%**

JMSTFL cruise date	Dataflow Spatial Exceedence Rate	CBP Stations Spatial Exceedence Rate	Percent Error
4/26/2006	11	0	100
9/19/2007	18	22	22
7/1/2008	26	41	58
		median percent error →	**58%**

JMSOH cruise date	Dataflow Spatial Exceedence Rate	CBP Stations Spatial Exceedence Rate	Percent Error
3/28/2006	57	28	52
8/20/2007	6	0	100
8/11/2008	2	0	100
5/21/2012	5	0	100
8/20/2012	1	0	100
3/5/2013	4	0	100
7/11/2013	17	0	100
		median percent error →	**100%**

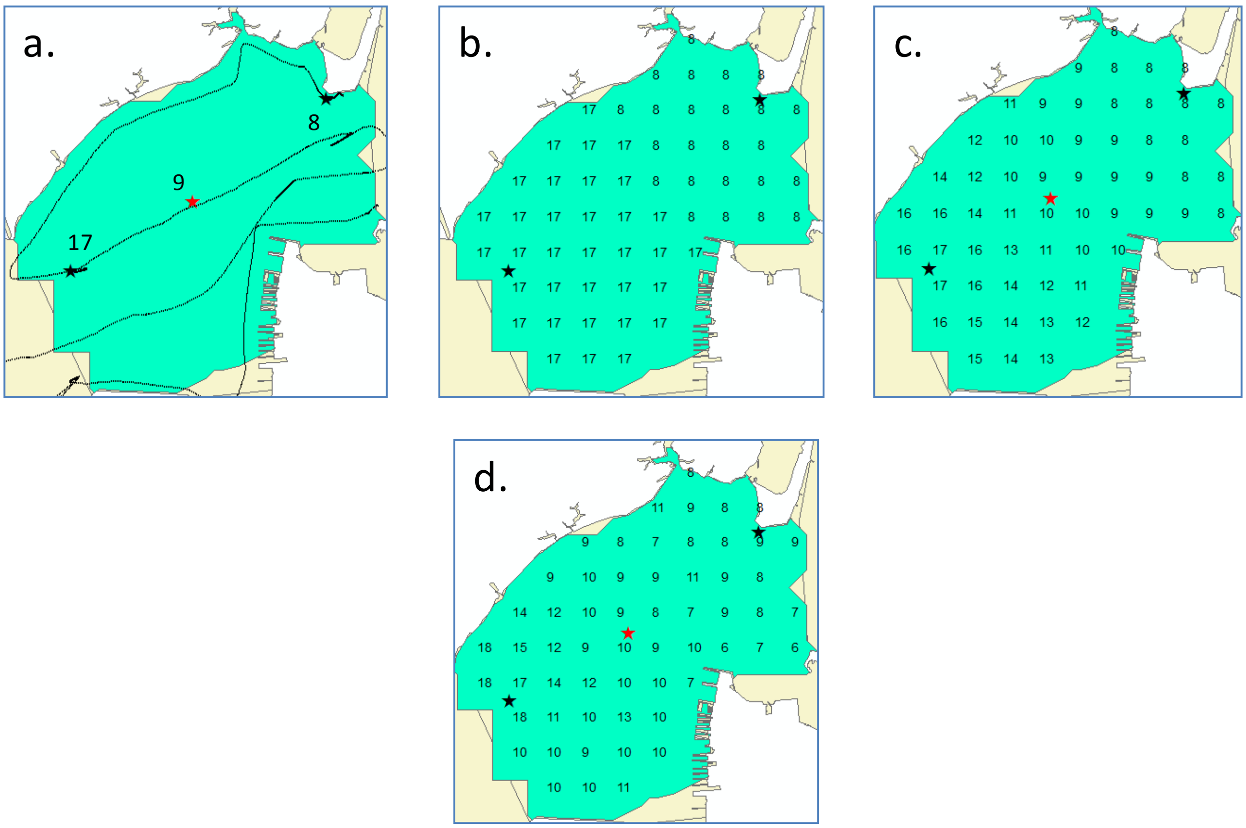
Table 2. (continued) Spatial exceedance rates (as percentages) estimated from interpolated Dataflow cruises and simulated CBP fixed stations samples. A low to moderate number of exceedances were observed in the selected Dataflow datasets. (Data provided by VIMS and HRSD)

JMSMH cruise date	Dataflow Spatial Exceedence Rate	CBP Stations Spatial Exceedence Rate	Percent Error
5/27/2010	5	0	100
4/9/2012	5	0	100
3/12/2007	20	50	150
3/17/2010	12	53	342
3/8/2006	36	0	100
9/1/2011	14	0	100
8/26/2010	13	0	100
		median percent error →	**100%**

JMSPH cruise date	Dataflow Spatial Exceedence Rate	CBP Stations Spatial Exceedence Rate	Percent Error	CBP Stations + "extra" Spatial Exceedence Rate	Percent Error
4/18/2006	5	0	100	0	100
5/24/2006	14	54	286	31	121
9/12/2006	2	0	100	0	100
3/8/2007	17	0	100	55	224
5/22/2007	17	0	100	0	100
9/15/2009	9	45	400	34	278
8/31/2011	2	0	100	0	100
3/4/2013	5	0	100	0	100
		median percent error →	**100%**	median percent error →	**100%**

As shown in Table 2, the fixed station-based interpolations produced very poor estimates of spatial exceedance rates. The samples taken at fixed stations are simply too sparse to pick up on the localized "patches" of exceedance that the Dataflow sensor is capable of detecting. Moreover, on the occasion that an exceedance is observed at a fixed station, it is exaggerated by the interpolation process, as illustrated by the differences in Figures 2b and 2d. Governed solely by the rules of arithmetic (as opposed to a geostatistical model), the Bay Interpolator applies a broad brush when the input is only two or three samples. The "strokes" of this brush are so predictable that spatial exceedance rates can be surmised just by knowing if exceedances were observed at one, two, or three stations. If a segment has two stations that are on opposite ends and an exceedance value is observed at one, the Interpolator

Figure 2. Interpolated grids for JMSPH based on the 5/24/2006 Dataflow cruise, as produced by the Bay Interpolator. A) The Dataflow cruise-track (1,928 observations) and the fixed stations where "grab samples" were taken. The black stars are the Chesapeake Bay Program stations LE5.4 and LE5.5W. The red star is an "extra" sampling location (referenced in Table 2), where an additional sample from the cruise-track was taken to illustrate how three data points affect the interpolation. The values are the chlorophyll concentrations sampled at those locations (an exceedance is any value greater than 12 µg/l). B) The interpolation grid resulting from the two CBP fixed station samples. The spatial exceedance rate is 54% C) The interpolation grid resulting from the two CBP fixed station samples plus the extra station sample. The spatial exceedance rate is 31%. D) The interpolation grid resulting from the Dataflow cruise-track. The spatial exceedance rate is 14%. Data provided by Will Hunley (HRSD).



will produce a grid with an approximate 50% exceedance rate. If an "extra" station is inserted in the middle of the segment and a non-exceeding value is sampled there, then the ~50% exceedance rate is pared down to approximately 30%. If exceedances are observed at two out of three stations, the exceedance rate will be roughly 75%. The exact values will depend on the relative position of the stations with respect to each other, the boundaries of the segment, and the magnitude of exceedances.

The results of this analysis show that there is generally a bias towards compliance when spatial exceedance rates are relatively low and fixed station-derived interpolations are assessed individually. But interpolations are almost always assessed as seasonal composites. If two interpolation grids that are biased "low" are averaged with an interpolation grid that is biased "high", the end result is likely a composite grid that closely approximates the true seasonal spatial exceedance rate. However, that "close approximation" would have to be more like "identical match" more often than not, since a difference in spatial exceedance rates as small as 0.1% can lead to an erroneous attainment decision.

Spatial-Temporal Exceedance Estimation

The CFD defines attainment in terms of space-time; the two dimensions are evaluated simultaneously. While the "proportion of space in non-attainment" axis is relatively straightforward, the temporal piece is a trickier concept to grasp. It does not represent the percent of time that a segment is in noncompliance, which is the focus of traditional criteria attainment methodologies. Rather, each y-coordinate of the CFD is the probability that the associated spatial exceedance rate (x-coordinate) was exceeded. For example, the 33% spatial exceedance and 25% temporal exceedance rates paired together in Fig 1h should NOT be read as "33% of a waterbody's area was in non-attainment 25% of the time." Instead, the two should be interpreted as "33% *or more* of a waterbody's area was in non-attainment 25% of the time." The y-axis has an interpretation similar to the cumulative distribution function (STAC, 2006).

Thus, testing the "temporal accuracy" of the CFD derived from low-frequency datasets is not possible, since the accuracy of the temporal estimate is entirely dependent on the accuracy of the spatial estimate. But testing the "space-time" accuracy (which equates to testing the "shape" of the CFD itself) is certainly feasible. To accomplish this, "estimate" CFDs derived from fixed station samples were compared to "true" CFDs derived from spatially-intensive (Dataflow) and temporally-intensive (continuous monitoring) datasets. This analysis was performed by the independent statistician Elgin Perry under contract with VADEQ. His work is presented in more detail in Appendix I (pg. 16) and summarized below.

The "true" chlorophyll expression in JMSPH during the spring and summer months of 2005-2007 was simulated by creating two models—one describing the spatial variance indicated by Dataflow cruise-tracks and one describing the temporal variance indicated by a shallow-water continuous monitor (maintained by the Moore lab at VIMS). A kriging variogram model was used for the former; an autoregression procedure was used for the latter. The information from these models was then combined to "represent a feasible realization of chlorophyll in the polyhaline James for the three year period" (Perry, 2015).

Six locations were selected to represent fixed stations (Figure 3). The simulated chlorophyll dataset was subsampled at these locations once every spring and summer month, mimicking the monthly monitoring performed at CBP fixed stations. These subsamples were then processed in accordance with the current assessment procedure (illustrated in Figure 1). To take advantage of the variability of the simulated dataset, subsampling was performed a thousand times, with each sample "collected" for a unique day-hour combination. Thus, a thousand different CFDs were created. These "sample" CFDs were then compared to the "true" CFDs derived from the full simulated dataset, as shown in Figure 4.

The “true” CFDs indicate that JMSPH was in compliance with chlorophyll standards for the 2005-2007 assessment period, as demonstrated by the positions of the “true” spring and summer curves relative to the 10% reference curve. However, the vast majority of the sample CFDs fall within the “noncompliance” zone relative to the reference—indicating that there is a heavy bias towards

Figure 3. The kriged interpolation (white-gray-black field) of a JMSPH Dataflow cruise-track (blue line). The solid blue circles are the six “fixed stations”, where the simulated chlorophyll dataset was subsampled on a monthly basis to mimic monthly site visits. Open circles are the Interpolator cell centroids. (Excerpt from Perry, 2015).

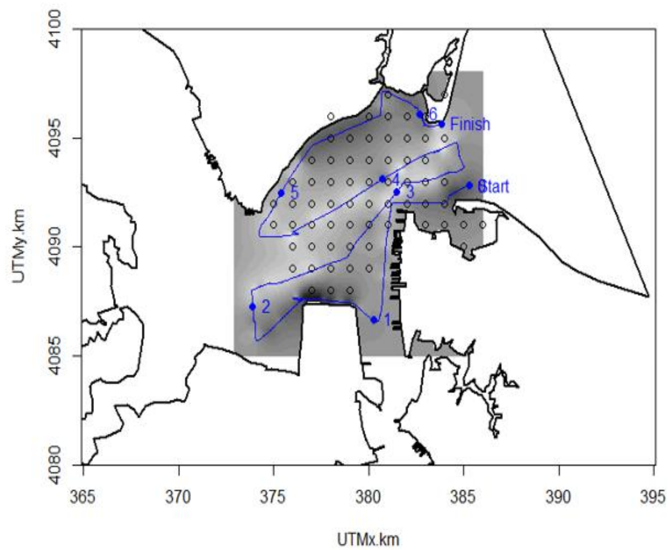
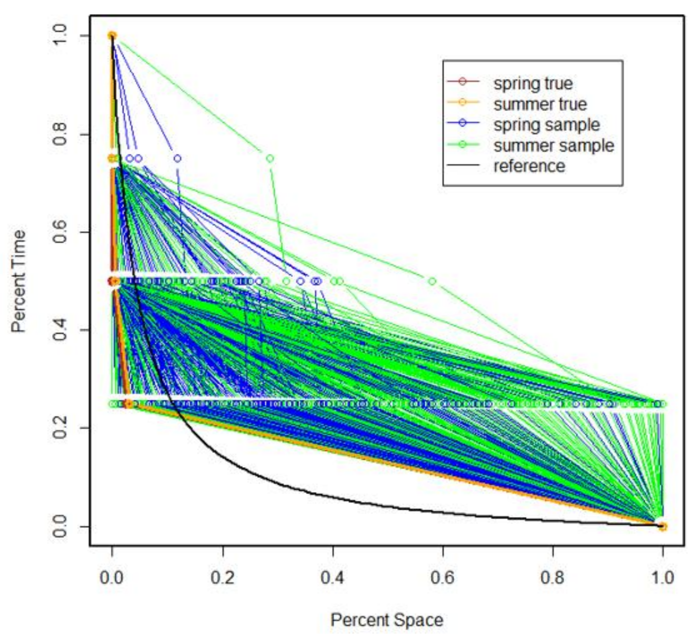


Figure 4. The 10% reference curve (black), the “true” spring and summer CFDs derived from the simulated data (red and orange), and 1,000 sample CFDs created by monthly sampling of the simulated data at the six stations. (blue and green). (Excerpt from Perry, 2015).



non-attainment when fixed station samples are used.

Perry (2015) soberly states that as long as VADEQ continues to perform fixed station-based assessments using the CFD approach, "the true state of the estuary must far exceed the requirements of its designated use in order to have a high probability of correctly identifying the passing condition (i.e. in order to remove it from the [Impaired Waters] list). At a minimum, this will require remediation beyond what is necessary. In some cases it may require a degree of improvement that is technically not achievable..."

Comparison of the 10% Curve to "Reference Condition" Exceedance Distributions

The advantage of the CFD approach is that it enables "allowable" exceedance rates to be established upon the distribution of spatial-temporal exceedances observed in reference communities. The use of a biological reference is supported by the appreciation that healthy, balanced aquatic life communities can withstand some small fraction of space and time under suboptimal water quality conditions. Thus, biologically-based reference curves provide an understanding of what level of criteria exceedances are allowable without losing support of the designated use (Tango and Batiuk, 2011). A biological reference curve is currently used to assess for Bay dissolved oxygen criteria for the deep water designated use (EPA, 2010a).

The development of biological reference curves for James River chlorophyll criteria has been thwarted by the dearth of reference phytoplankton communities. When a biological reference curve cannot be developed, EPA (2007) recommends the use of a 10% hyperbolic curve—a more sophisticated interpretation of the simple "10% rule" that EPA has long endorsed for aquatic life use support determinations (EPA, 2005). While the 10% rule governs temporal frequency of exceedance (i.e., no more than 10% of the total number of periodic samples taken during the assessment period can exceed the criteria), a 10% curve carves out an exceedance rate of 10% for space-time (spatial exceedance rate multiplied by temporal exceedance rate). The crucial assumption is that the target load reductions mandated by the Bay TMDL will foster a phytoplankton community in the James River that will exhibit spatial and temporal chlorophyll exceedance rates conforming to this "10%" space-time distribution. But it is difficult to test this assumption empirically without a reference community.

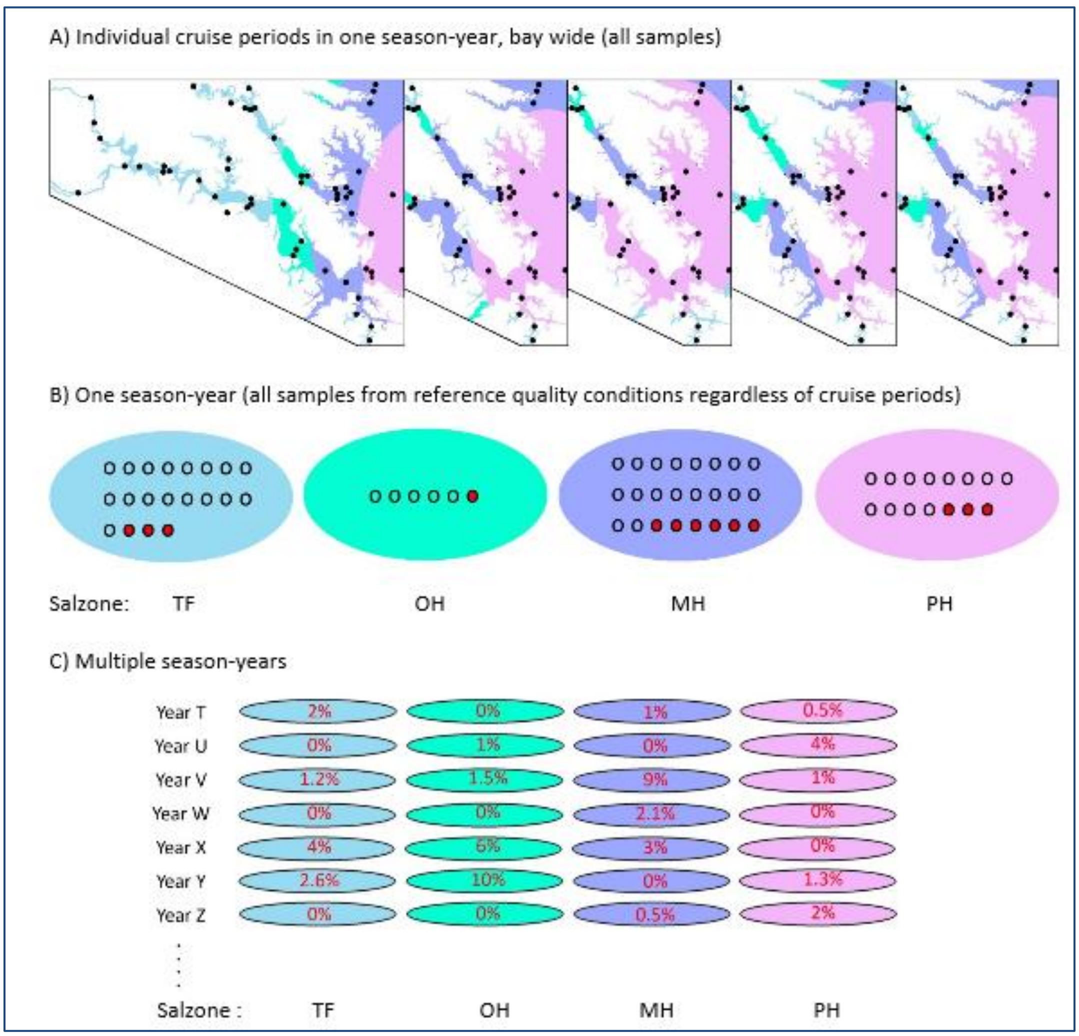
Claire Buchanan (Interstate Commission of the Potomac River Basin) was contracted by VADEQ to test this assumption through the use of reference surrogates. Her report is presented in Appendix II (pg. 51) and summarized below.

Rather than locating a reference phytoplankton community, Buchanan (2014), delving into the long-term Chesapeake Bay database, located areas throughout the Bay mainstem and tidal tributaries where water quality conditions were—however ephemerally—favorable for reference phytoplankton communities. Three parameters were chosen to define reference conditions, based on their known linkages to healthy phytoplankton communities (Buchanan, 2005)—secchi depth, dissolved inorganic nitrogen, and ortho-phosphate. If parameter thresholds (Table 3) were met during a particular

Table 3. Parameter thresholds used to define "reference condition". (Excerpt from Buchanan, 2014.)

	Spring	June	Summer	Autumn	Winter
Secchi depth (m)					
TF	>0.9	>0.8	>0.8	>0.9	>0.6
OH	>0.7	>0.6	>0.6	>0.5	>0.6
MH	>1.8	>1.45	>1.45	>2.0	>1.8
PH	>2.15	>1.85	>1.85	>2.5	>2.3
DIN (mg/liter)	≤0.07 (all seasons and salinity zones)				
PO_4 (mg/liter)	≤0.007 (all seasons and salinity zones)				

Figure 5. Illustration of the method used to develop biological reference curves. (Excerpt from Buchanan, 2014)

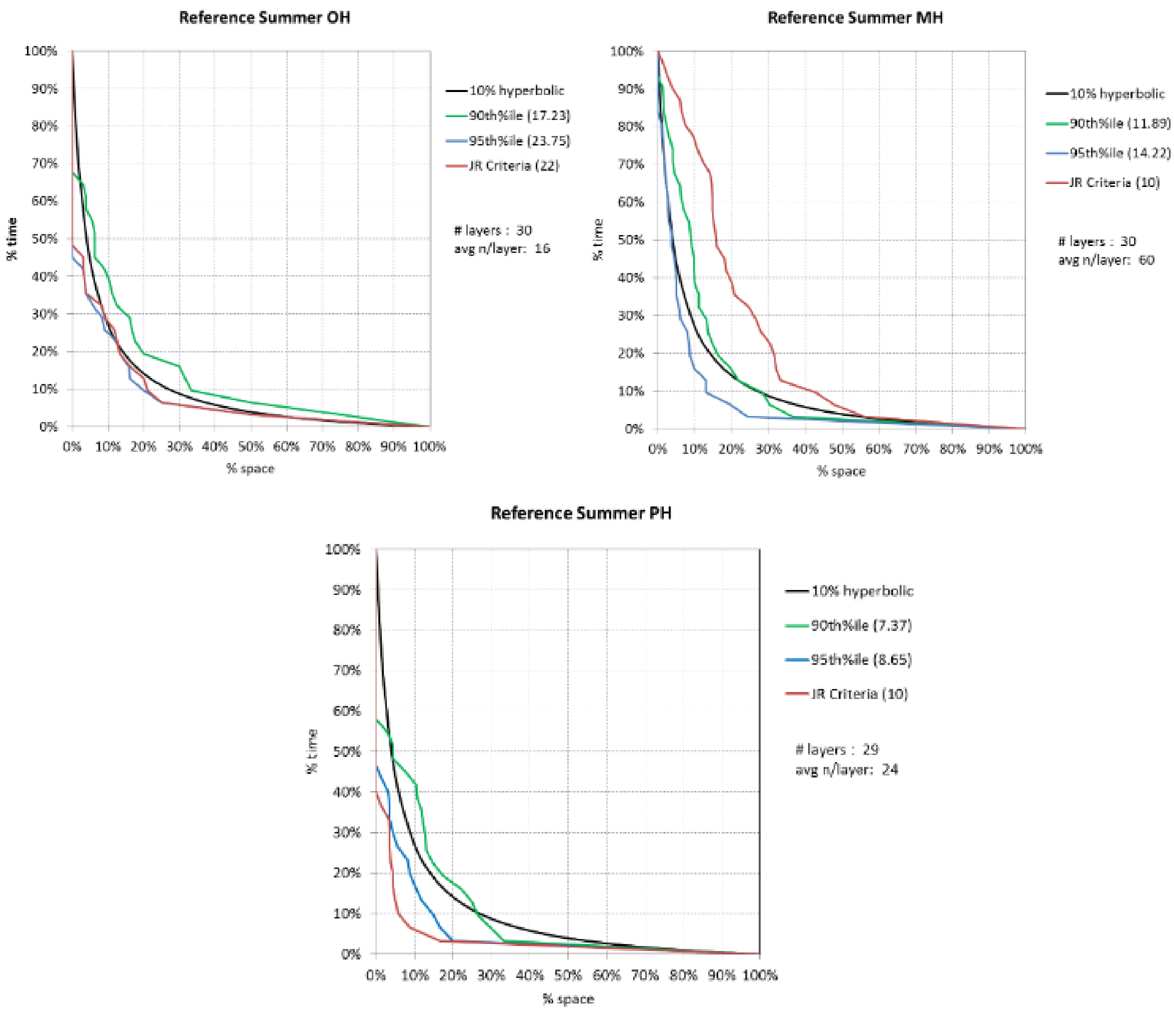


monitoring event, then the phytoplankton community existing at that station was assumed to be reference or "reference like". The observed chlorophyll concentration was compared to the chlorophyll criterion applicable to the James River for the same salinity region and season. Sampling events were aggregated by salinity regime and then spatial exceedance rates were estimated by counting the "reference" sites where an exceedance was observed and dividing by the total of "reference" sites in that salinity regime, over a calendar year (as illustrated in Figure 5.) These exceedance rates for each salinity -season combination were then ranked and matched to temporal exceedance rates (using the Weibull equation), and then plotted as CFDs. The untested assumptions underlying this analysis are duly noted by Buchanan (2015).

Some of the "reference condition" CFDs were found to be similar to the 10% hyperbolic curve, such as the curves based on exceedances of summer JMSTFL and JMSOH criteria (the latter shown in Figure 6). However, the majority of the CFDs did not conform to the 10% distribution. The reference CFDs based on criteria for JMSTFU (both seasons), spring JMSOH, and JMSMH (both seasons) were more lenient in their "allowable" space-time exceedance rates relative to the 10% curve, while the other CFDs (spring JMSTFL criterion, spring and summer JMSPH criteria) were more stringent. Interestingly, there was more conformity when 90th and 95th percentile chlorophyll thresholds for chlorophyll were used in lieu of the James River criteria—suggesting that there is some empirical basis for a 10% reference curve, but that the James River criteria themselves may not be reflective of "reference", at least in a general Bay-wide sense.

One important caveat to Buchanan's analysis is that her definition of "exceedance" is not in complete accordance with the James River criteria. The criteria are expressed as seasonal means, not the instantaneous values that her analysis is based on (though it must be noted that the 10% hyperbolic curve was not developed with seasonal mean exceedances in mind, but rather exceedances of upper/lower thresholds(STAC, 2006)). Thus, it is not possible to infer from her results what a reference curve truly looks like with regards to James River criteria in their current form. All that can be confidently surmised is that the 10% hyperbolic curve is likely an overly simplistic representation of "allowable" exceedance rates for James River chlorophyll—too lenient in some cases, too stringent in others. Among other suggestions, Buchanan (2014) recommends the development of season and segment-specific hyperbolic curves that better characterize the distribution of space-time exceedance under reference conditions.

Figure 6. Some of the "reference condition" CFDs generated by Buchanan (2014). Red curves are the CFDs derived from reference exceedances of the salinity-specific James River criterion. Green and blue curves are derived from exceedances of the 90th and 95th percentiles of the "reference condition" chlorophyll samples specific to that salinity and season. (Excerpt from Buchanan, 2014).



Conclusion

VADEQ finds that there are severe limitations of the CFD approach in the context of James River chlorophyll criteria. In summary:

1. The approach does not lead to accurate attainment determinations when sampling is conducted through a conventional monitoring program. Moreover, the approach appears to be biased towards nonattainment, which may explain (at least in part) why the criteria are not fully attained at the target nutrient loading scenario of the Bay TMDL (Appendix O., Bay TMDL, 2010b). It appears that the only way to minimize this bias is to employ expensive spatially-intensive monitoring. Budget constraints make this an untenable position for the Commonwealth.

2. The current reference curve is probably not a good target for the current James River chlorophyll criteria. It is likely the 10% curve is overly stringent for some James River criteria. Since there is evidence the approach is already biased towards nonattainment when only conventional datasets are available, the use of an overly stringent reference curve cannot be justified. Furthermore, the use of a metric that is laden with untested assumptions threatens the defensibility of assessments.

It is not VADEQ's position that the CFD approach is unsalvageable or that its short-comings with respect to chlorophyll criteria necessarily apply to dissolved oxygen criteria. But VADEQ does believe it has sufficient cause to propose an alternative assessment method—one that offers several advantages over the CFD.

Literature Cited

Bahner, L. 2006 User guide for the Chesapeake Bay and tidal tributary interpolator version 4.6. NOAA Chesapeake Bay Office.

Buchanan, C., R. V. Lacouture, H. G. Marshall, M. Olson, and J. M. Johnson. 2005. Phytoplankton Reference Communities for Chesapeake Bay and its Tidal Tributaries. Estuaries 28(1):138-159.

Buchanan, C. 2014. Biological Reference Curves for Assessing the James River Chlorophyll *a* Criteria . Interstate Commission for the Potomac River Basin. Rockville, MD.

Perry, E. 2015. Notes on James River Chlorophyll Simulator and CFD validation. Final report submitted to Virginia Department of Environmental Quality November, 2015.

Scientific Technical Advisory Committee. 2006. The Cumulative Frequency Diagram Method for Determining Water Quality Attainment: Report of the Chesapeake Bay Program STAC Panel to Review of Chesapeake Bay Program Analytical Tools. STAC Publication 06-003. 74 pp.

Tango, P. and R. Batiuk. 2013. Deriving Chesapeake Bay Water Quality Standards. Journal of the American Water Resources Association 1-18.

U.S. Environmental Protection Agency. 2005. Guidance for 2006 Assessment, Listing and Reporting Requirements Pursuant to Sections 303(d), 305(b) and 314 of the Clean Water Act. U.S. Environmental Protection Agency, Washington, D.C.

U.S. Environmental Protection Agency. 2007. Ambient Water Quality Criteria for Dissolved Oxygen, Water Clarity, and Chlorophyll *a* for the Chesapeake Bay and its Tidal Tributarities: 2007 Technical Support for Criteria Assessment Protocols Addendum. 903-R-07-003. EPA-CBP Annapolis, MD.

U.S. Environmental Protection Agency. 2008. Ambient Water Quality Criteria for Dissolved Oxygen, Water Clarity, and Chlorophyll *a* for the Chesapeake Bay and its Tidal Tributarities: 2008 Technical Support for Criteria Assessment Protocols Addendum. 903-R-001. EPA-CBP Annapolis, MD.

U.S. Environmental Protection Agency. 2010a. Ambient Water Quality Criteria for Dissolved Oxygen, Water Clarity, and Chlorophyll *a* for the Chesapeake Bay and its Tidal Tributarities: 2010 Technical Support for Criteria Assessment Protocols Addendum. 903-R-10-002. EPA-CBP Annapolis, MD.

U.S. Environmental Protection Agency. 2010b. Chesapeake Bay Total Maximum Daily Load for Nitrogen, Phosphorus and Sediment. U.S. Environmental Protection Agency Chesapeake Bay Program Office, Annapolis, MD.

Notes on James River Chlorophyll Simulator and CFD validation

Elgin Perry
eperry@chesapeake.net
377 Resolutions Rd.
Colonial Beach Va. 22443
410-610-1473
Revision 30 Oct. 2015

Overview

This project is to develop tools that will enable the estimation of false positive and false negative error rates for the procedure in place for assessing the numerical chlorophyll criteria in the tidal James River. The key tool is a computer simulation that creates chlorophyll data with spatial and temporal patterns that are typical of observed chlorophyll in the James River. This simulation makes it possible to create data on a dense spatial and temporal grid that covers an entire river segment for a three year assessment period. The state of this simulated data is known to be compliant or non-compliant with the numerical criteria. To test for an assessment error, a subset of data are sampled from the simulation with the same spatial and temporal frequency as ship-board sampling of the James River. The full assessment procedure is applied to this sub-sample of simulated data and the result of the assessment is compared to the known state of the simulation. If the assessment finds a result of non-compliance when it is known that the simulated data are compliant, this is a false positive error. If the assessment finds compliance when the known state is non-compliant, this is a false negative error. The rates for these errors are estimated by repeatedly sampling the simulated data and tabulating the results of the assessments. The proportion of errors observed in these repeated assessments estimates the error rate.

Background

Along with many Chesapeake Bay tributaries, the tidal James River was listed as impaired under the Clean Water Act by the U.S. Environmental Protection Agency (EPA) in 1999 for violation of Virginia's Water Quality Standards. The primary driver for this listing was the assessment that tidal waters of James River do not contain diverse, healthy and balanced populations of many expected aquatic life forms including phytoplankton at the base of the food chain. Like all primary producers, phytoplankton contain chlorophyll which is essential for converting the sun's energy into food.). Thus measures of chlorophyll are frequently used as surrogate measures of phytoplankton biomass. Scientists have developed a technology for easily measuring the quantity of chlorophyll in the water column by using light to excite the phytoplankton and then measuring fluorescence as the phytoplankton return to non-excited state. This meter approach makes is possible to obtain near continuous measurements of chlorophyll in a temporal domain (ConMon) or spatial domation (DataFlow). This ease of measuring chlorophyll coupled with the connection of chlorophyll to phytoplankton, and in turn phytoplankton to nutrient enrichment, make chlorophyll an important parameter for assessing the impaired state of the tidal James River and determining when sufficient improvement has been attained for delisting this tributary.

Nutrient enrichment is a pervasive problem in the Chesapeake Bay watershed. Virginia's existing Water Quality Standards require that "substances which nourish undesirable or nuisance aquatic plant life will be controlled" (9 VAC 25-260-20). To meet that requirement, Virginia adopted the Nutrient Enriched Waters (9 VAC 25-260-330-350) and Policy for Nutrient Enriched Waters (9 VAC 25-40) in 1988. These existing regulations also recognized that nutrients contribute to undesirable growths of aquatic plant life, classified waters as nutrient enriched and imposed phosphorus limits on discharges to waters classified as nutrient enriched. It would seem that if the primary goal is to control nutrient enrichment, the criteria for assessing the goal should be defined in terms of nutrient concentrations. However, with current technology, this is not a practical approach.

With our current monitoring program, chlorophyll is a more reliable indicator of nutrient enrichment than nutrient concentrations. The difficulty with using nutrient concentration as a criterion of excess nutrients is that the presence of excess nutrients in the water column is very ephemeral. Within a matter of hours of nutrient delivery, phytoplankton can increase their assimilation rate and begin to consume these nutrients. Within a matter of days, phytoplankton can double their population and further consume available nutrients. Because excess nutrients are often consumed in less than a week, it is unlikely that the fixed station monitoring program with collects samples once or twice a month will intercept the presence of excess nutrients. However, once the nutrients are assimilated into phytoplankton, they will persist for weeks. For this reason, phytoplankton biomass and the surrogate measure chlorophyll are more reliable indicators of nutrient enrichment. For this reason, chlorophyll a was also recognized in the Nutrient Enriched Waters sections of the regulation as an indicator of nutrient enrichment.

Water quality criteria guidance prepared by the EPA Chesapeake Bay Program (CBP) makes clear that States are expected to adopt narrative chlorophyll a criteria (USEPA,2003). Furthermore, the EPA strongly encourages states to develop and adopt site-specific numerical chlorophyll a criteria. In response to this guidance, in 2005 Virginia became the first Bay jurisdiction to adopt numerical chlorophyll a standards promulgated for the Tidal James River. This same EPA guidance that encouraged numerical chlorophyll a criteria also put forth a nascent numerical procedure for assessing water quality based on a Cumulative Frequency Diagram (CFD).

The CFD assessment methodology evolved from a need to allow for variability in water quality parameters due to unusual events. For chlorophyll a, a threshold criterion is established for which it is determined that chlorophyll a that exceeds this threshold is in a degraded state. Because chlorophyll a is highly variable in space and time, it is unlikely that a healthy waterbody will remain below the threshold in all places at all times. In the spatial dimension, there will be small regions that persistently exceed the threshold due to poor flushing or other natural conditions. It is recognized that these small regions of degraded condition should not lead to a degraded assessment for the segment surrounding this small region. Similar logic applies in the temporal dimension. For a short period of time, water quality in a large proportion of a segment may exceed the threshold, but if this condition is short lived and the segment quickly returns to a healthy state, this does not represent an impairment of the designated use of the segment. Recognition that ephemeral exceedances of the threshold in both time and space do not represent persistent impairment of the segment leads to an assessment methodology that will allow these conditions to be classed as acceptable while conditions of persistent and wide spread impaired condition will be flagged as unacceptable. The assessment methodology was developed by first quantifying how much of the segment is not in compliance with the criteria (percent of space) for every point in time. In a second step the process quantifies how often (percent of time) a segment out of compliance by more than a fixed percent of space. The results from these calculations can be presented in graphical form where percent of time is plotted against percent of space (Figure 17). It is arbitrary to treat space first and time second. A similar diagram could be obtained by first computing percent noncompliance in time and then considering the cumulative distribution of percent time over space. This new assessment procedure was named the Cumulative Frequency Diagram or CFD.

It is against this background of newly defined numerical chlorophyll criteria and a novel assessment process that the current study was conducted.

Organization of Reporting

The process of estimating false positive and false negative error rate involve three separate data analyses. The first is a spatial analysis to assess spatial dependence of DataFlow observations of chlorophyll and develop a tool for recreating the random spatial variability of chlorophyl. The second analysis is a time-series analysis to assess the serial time-dependence of ConMon observations of chlorophyll and develop an auto-regression procedure to model this auto-correlation. The third step brings together these two elements into a simulator that reflects the spatio-temporal character of chlorophyll data and conducts a sampling experiment to estimate the assessment error rates. In what follows, the first two analyses are treated as research projects and methods, results, and conclusions are presented for each. Finally the results of the first two analyses are combined to and presented in methods, results, and conclusions for the simulation and sampling analysis which estimates the assessment error rates.

At this point in time, this work is being presented as a proof of concept. It is hope that after stakeholders have an opportunity to review and comment on this tool, it can be used to address the “Issues to be considered” that have been raised concerning the current implementation of the James River Chlorophyll Criteria. In what follows, each step is treated as a research project with methods and results and then discussion is presented for all steps together. In this initial implementation, this simulator has a spatial resolution defined by the cells of the CBP interpolator (roughly 1 km), and a temporal resolution of 1 hour.

Spatial Analysis Methods

The spatial structure of the simulated chlorophyll data is based on using Kriging estimation of the DataFlow data. Kriging is a statistical interpolation procedure originally developed by the South African mining engineer Danie G. Krige and later given a more complete mathematical development by French mathematician, Georges Matheron (http://en.wikipedia.org/wiki/Kriging). In this application, Kriging has the advantage that once statistical estimates of the spatial dependence are obtained, it is possible to simulate new data with similar random properties. This capability is not available through nearest neighbor averaging which is employed the Chesapeake Bay Program interpolator.

The kriging algorithms employed for this work are found in the GeoR package for the analysis of geostatistical data (Ribeiro and Diggle, 2012) of the R statistical programming language (R Core Team, 2013). The specific functions in the geoR package are: variog() for estimating the variogram, variofit() for obtaining a mathematical model of the variogram, krige.conv() to obtain an interpolation of the data, grf() to produce simulated data, and image() to produce graphical output.

The data for this are the polyhaline DataFlow data from 2005-2007 as downloaded from VECOS. Individual observations of chlorophyll tend to follow a log-normal distribution (Figure 1.) and thus the data analysis conducted here will all be done with base 10 log-transformed data to improve the normality of the observations (Figure 2.).

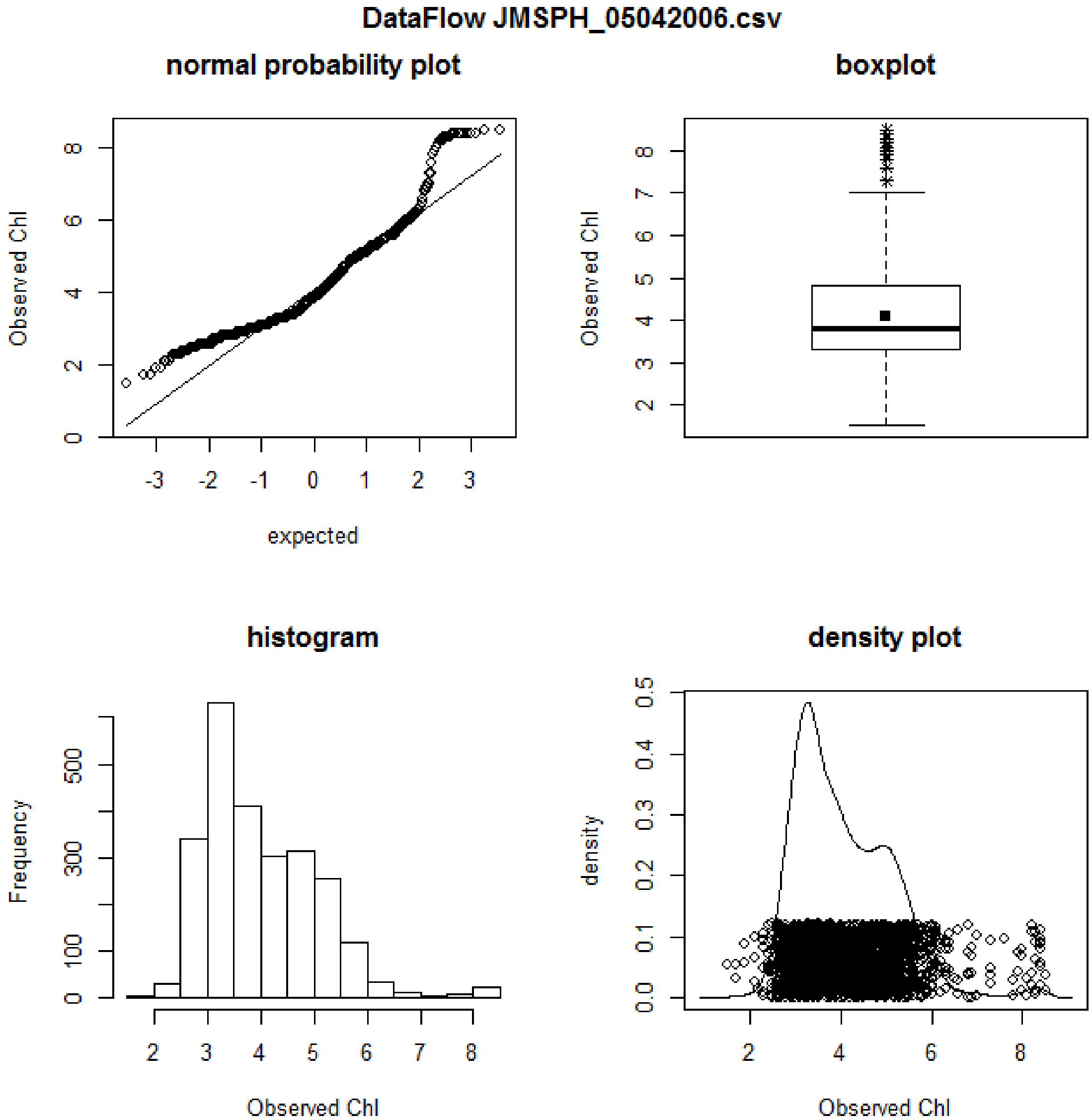


Figure 1. Distribution plots for observed chlorophyll from Data Flow cruise James River Polyhaline 05/04/2005.

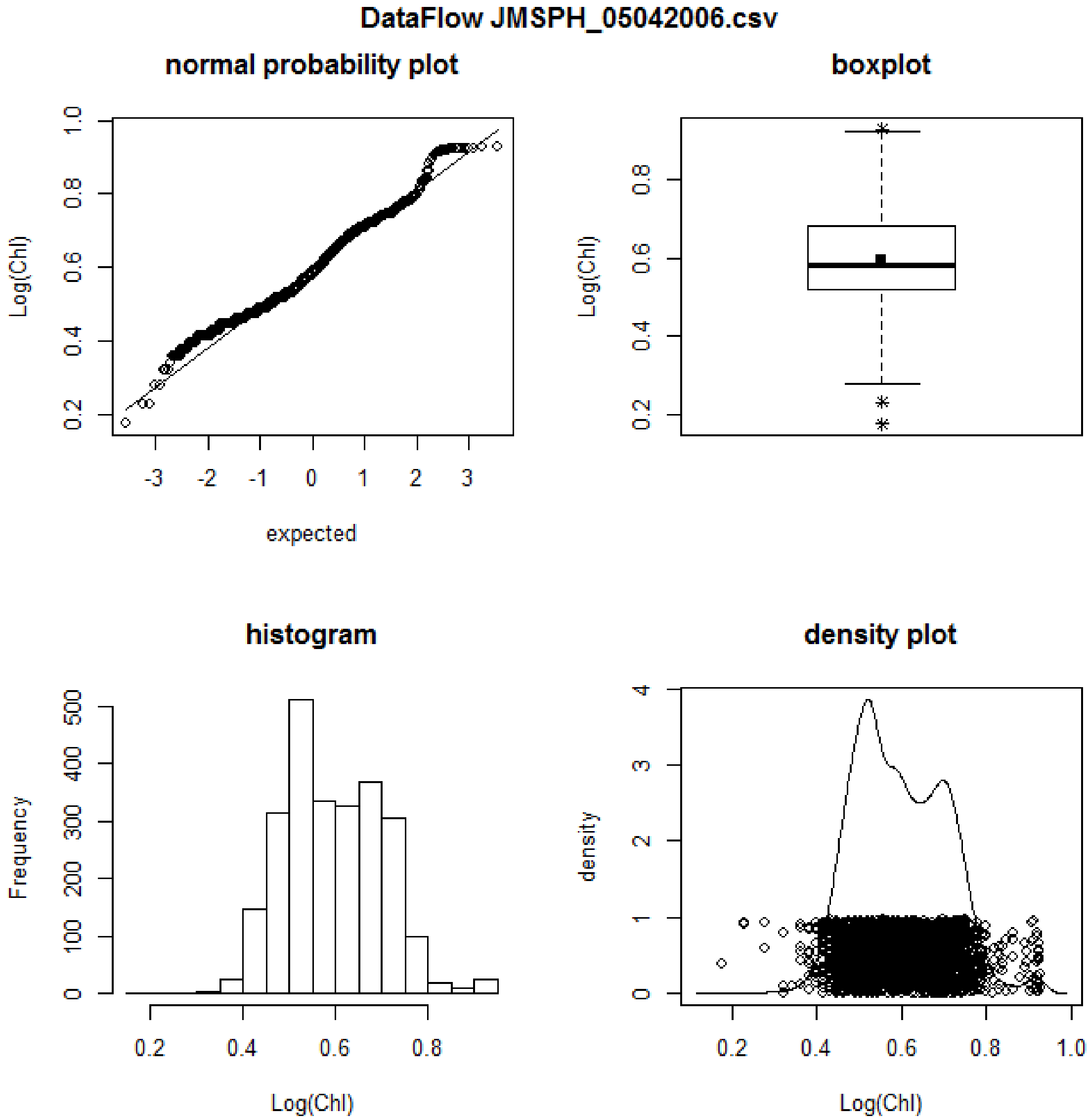


Figure 2. Distribution plots for logarithm transformed chlorophyll from Data Flow cruise James River Polyhaline 05/04/2005

For each cruise, the Lat/Long data are converted to UTMy/UMTx scaled to kilometers because the large numbers recorded in meters cause numerical scaling issues with the variogram estimation software. In order to match data with interpolator cells, the UTMy/UMTx values are rounded to the nearest kilometer where needed.

Spatial Analysis Results:

The variogram estimator quantifies the degree of dependence among near chlorophyll measurements as a function of distance (Figure 3.) The nugget level quantifies the level of variability for measurements in the same location and is essentially measurement error. Each circle shows variability for the differences of pairs of observations observed at specified distances. The solid red curve is a smooth mathematical function used to approximate spatial dependence as a function of distance. The sill shows the level at which the variability is approaching a maximum. The range is the distance associated with the sill. For these data, observations collected more the 2.0 kilometers apart are essentially independent.

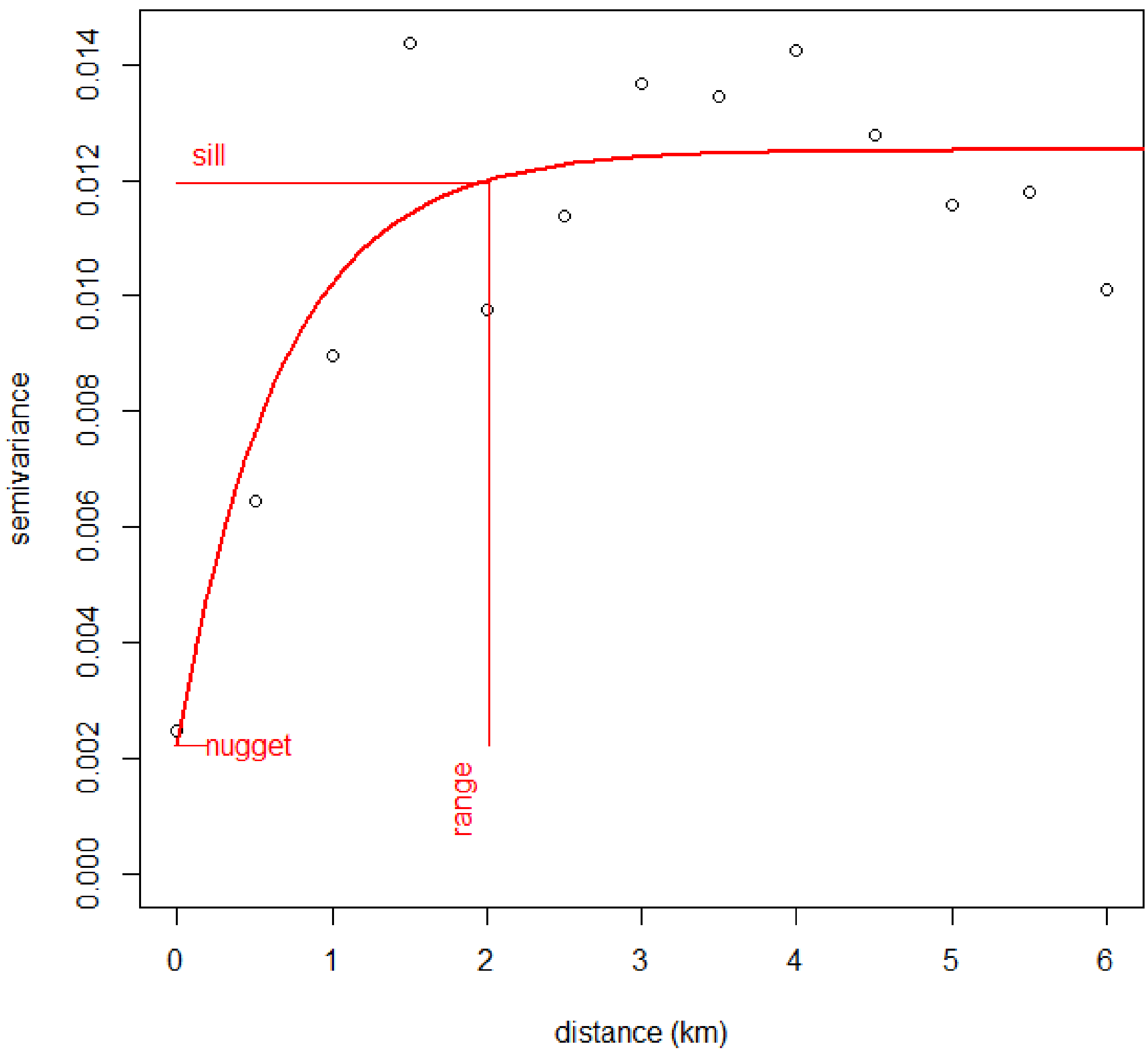
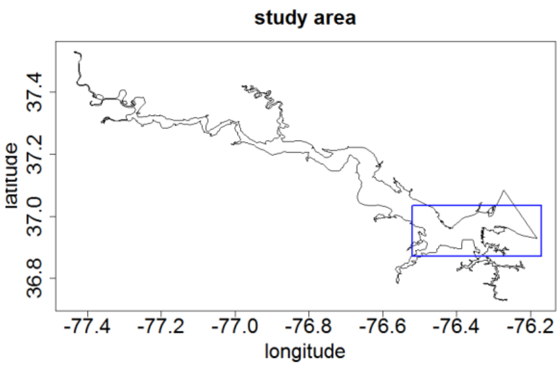


Figure 3. Variogram estimates (circles), estimated variogram function (solid line), and spatial data attributes of partial sill, nugget, and range for James River Polyhaline DataFlow 05/04/2005.

Using the smooth variogram estimator fitted by the R-function variofit(), kriging is used to obtain interpolated chlorophyll for the James Polyhaline (Figure 4.). Super-imposed on the gray-scale contour map of log(chlorophyll) is the cruise track (blue line) of the dataflow data that was interpolated to create the contoured image. The open black circles show the CBP interpolator cell centers that will become the loci for the chlorophyll simulator. Along the cruise track are locations marked at about 10 km intervals (blue dots) than show the association with the plot of observed log(chlorophyll) along the cruise track (Figure 4.)



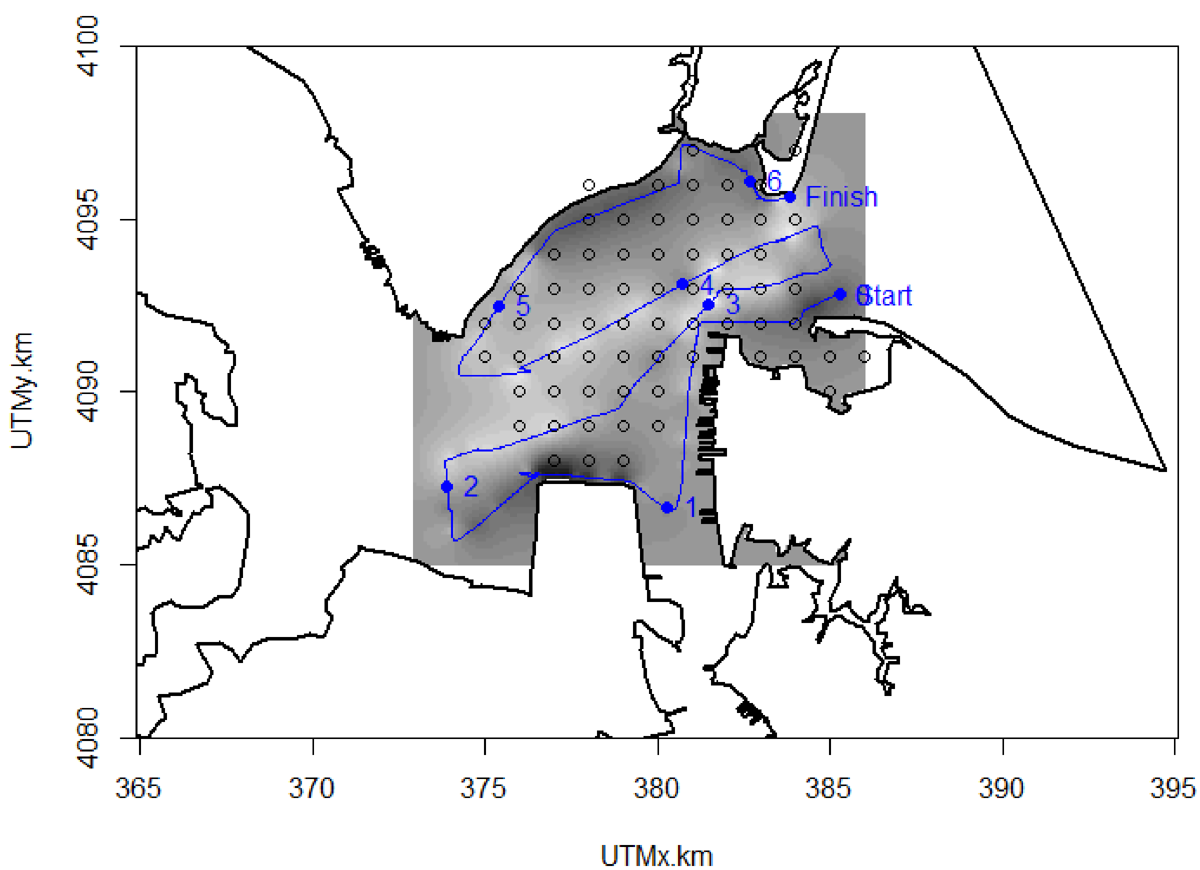


Figure 4. Kriged interpolation at a scale of 0.1 km for DataFlow log(chlorophyll) collected on 5/04/2005. Blue line shows the cruise track. Open black circles show the CBP interpolator cell centers.

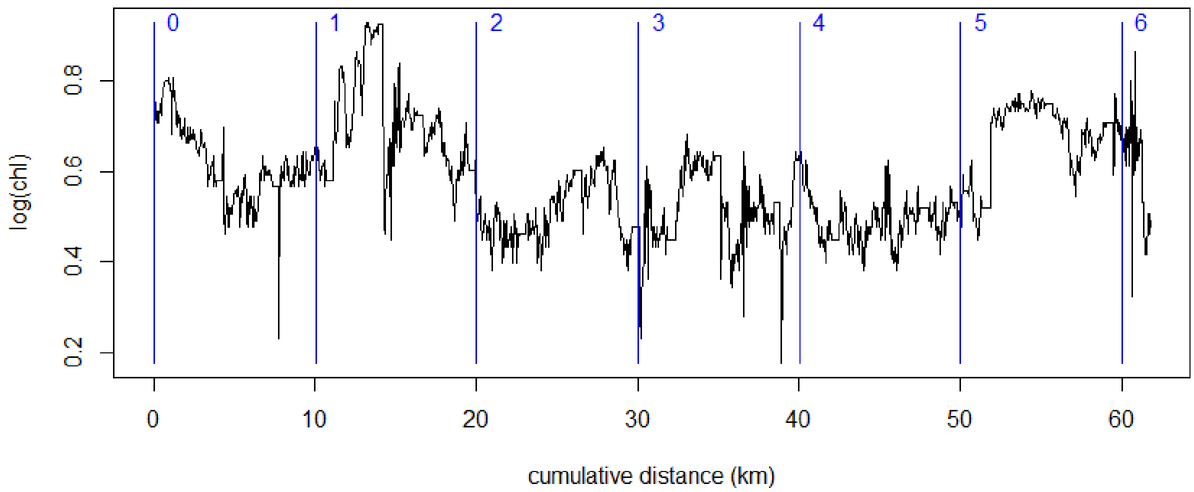


Figure 5. Observed log(chlorophyll) as a function of cruise track distance. The verticle blue lines correspond by number to the blue dot locations of Figure 4.

For this project, the kriging interpolation is done at a 1km scale to match the CBP interpolator cells (Figure 6.)

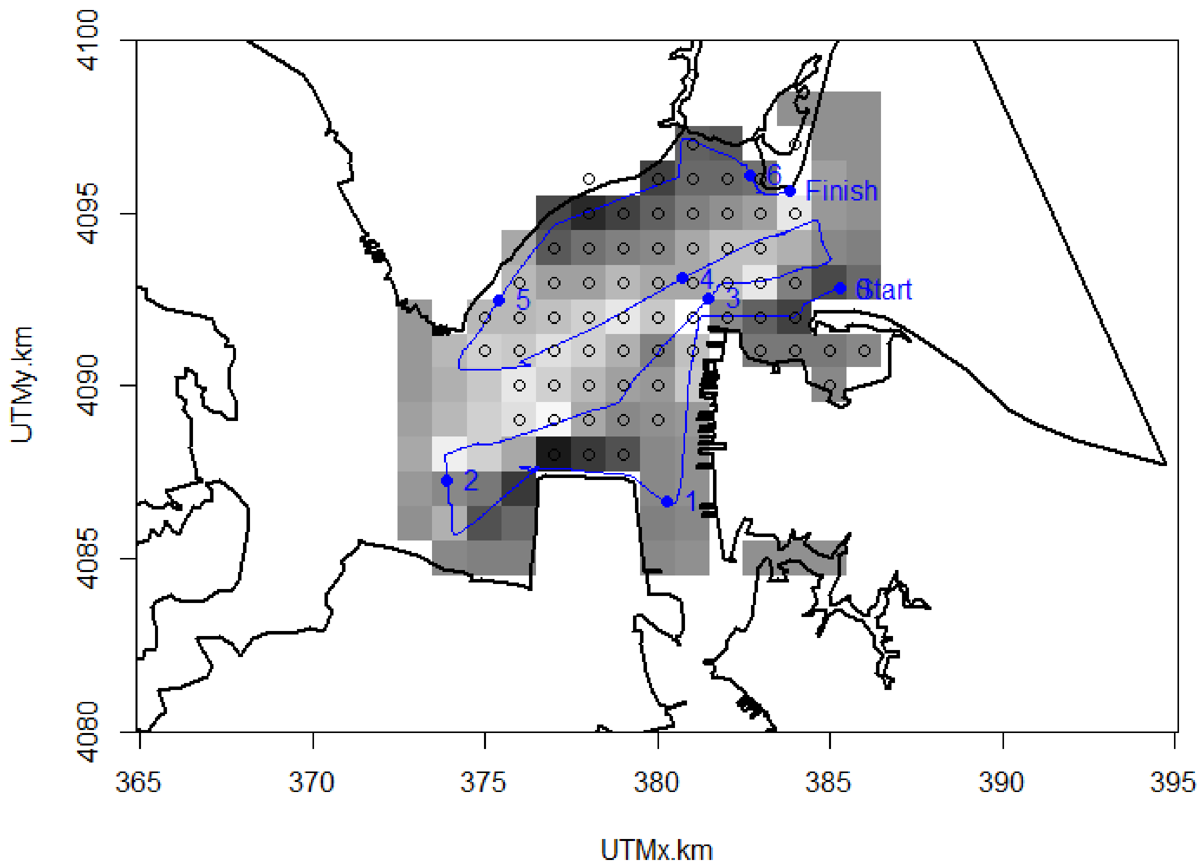


Figure 6. Krig/ed interpolation at a scale of 1 km for DataFlow log(chlorophyll) collected on 5/04/2005.

A full set of kriging analyses of polyhaline chlorophyll and figures for the year 2006 can be found in the file PH06KrigMaps.rtf.

It is important to understand that the interpolation of the data is a smoothed estimator of the observed chlorophyll (Figure 7., black). This estimator is trying to tract the mean of chlorophyll as it varies over space. Thus the interpolated surface does not represent the natural variability of chlorophyll. The simulated data obtained using the grf() function (Figure 7., red) do re-create the small scale variability. The grf() function recreates the simulated random field by starting with a vector of independent normal random variates with mean zero and variance 1 and multiplying this vector by what is essentially a square-root of the variance-covariance matrix of the spatial observations. The variance-covariance matrix is estimated from the variogram and its square-root is determined by a Cholesky decomposition. When the resulting multivariate normal vector is summed with the spatial means estimated by Kriging interpolation, the resulting data have the same spatial mean structure and the same spatial variability and dependence as the observed data.

The variability shown in the simulated data is similar to the variability that is evident in the observed data along the cruise-track (Figure 7.) The observed data are from the red portion of the cruise track shown in Figure 8. The simulated data are from the green transect in Figure 8. The observed data show some more variability just because it is observed at a finer scale than the simulated data. The simulation behind Figure 7 produces data at intervals of 100 meters and the observed DataFlow data in Figure 7 are collected at intervals of 30-40 meters.

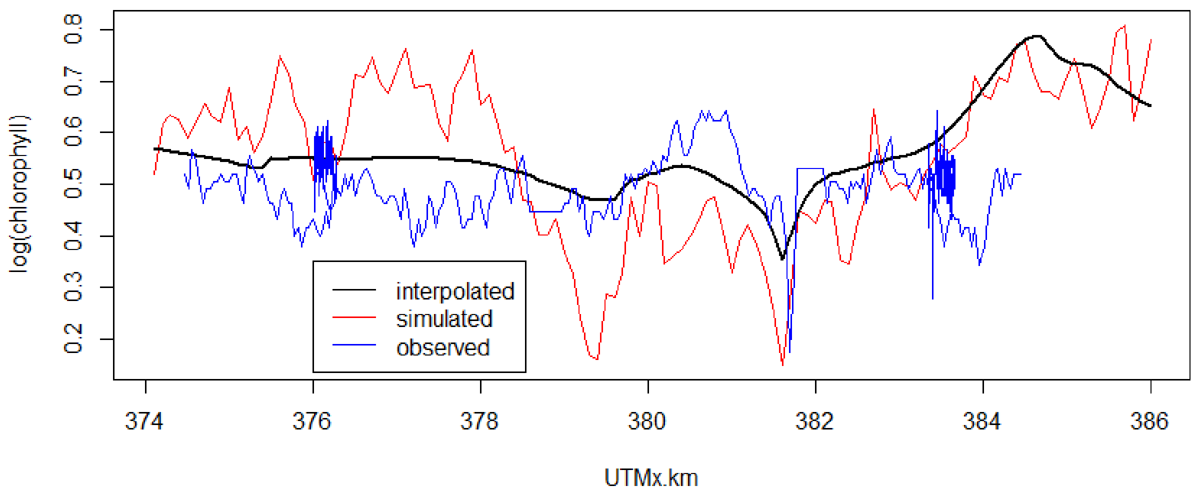


Figure 7. Interpolated, simulated, and observed log(chlorophyll) data for a transect where UTMy.km = 4092.6 which is about at the level of the cruise start (Figure 8- green transect). The observed data are

from the cruise track distance interval (35.3, 47.5) (Figure 8, red).

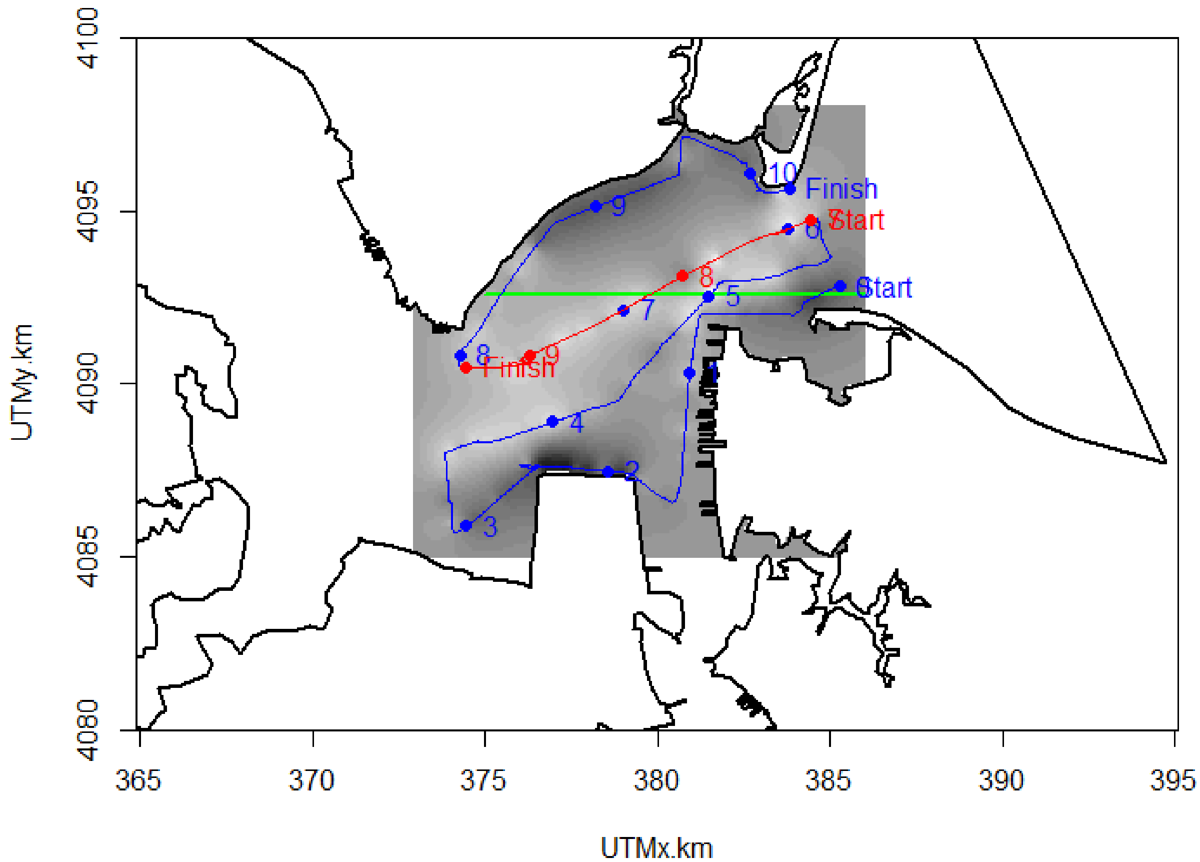


Figure 8. Illustration of the spatial paths of cruise track data (red) and simulated data (green) being compared in Figure 7.

Temporal Analysis Methods

While the Kriging simulator helps to produce simulated data with realistic spatial dependence, it is also important to reconstruct realistic temporal dependence in the simulated data. The strategy employed in this study is to assess the serial dependence of ConMon data that are concurrently collected in the James Polyhaline near Wythe Point(Figure 9). The Wythe Point ConMon data are collected at 15 minute intervals and these data are transformed to base 10 logarithms and averaged to 1 hour time periods for this analysis (Figure 10.). A generalized additive model (GAM) is used to estimate the seasonal trend and diel trend for these data (Figure 11., Table1.0).

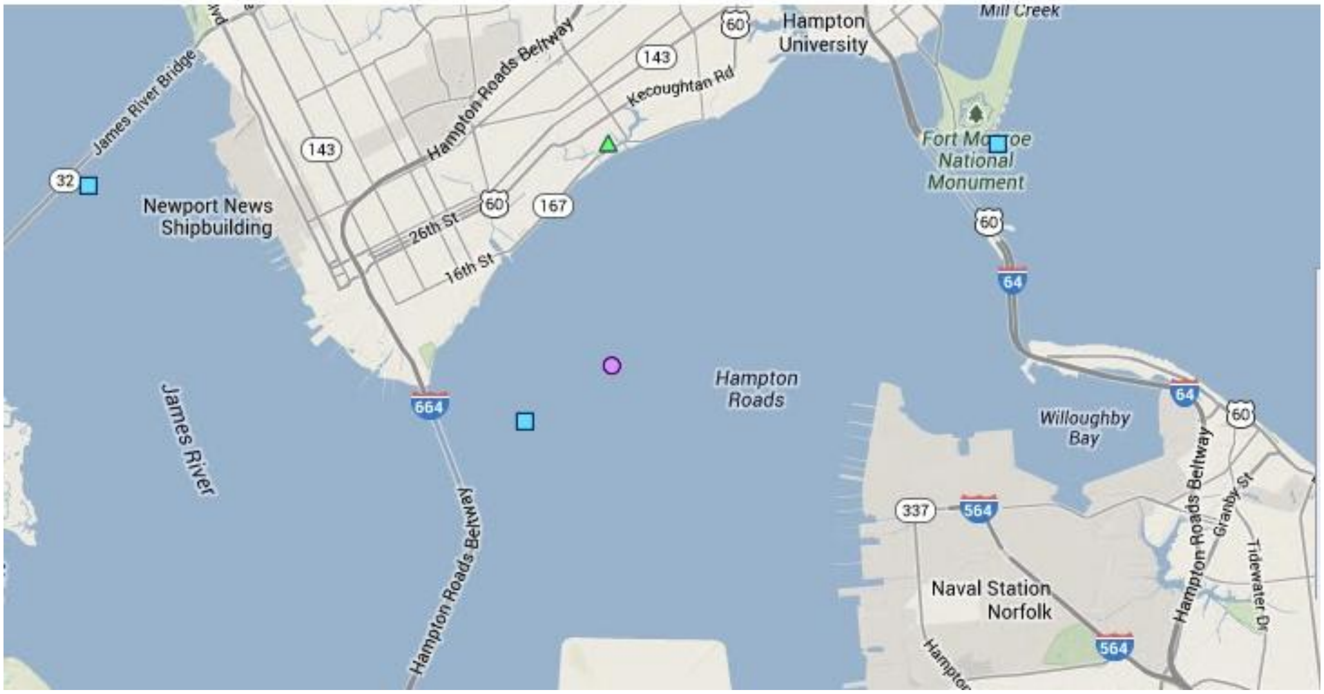


Figure 9. The Wythe Point ConMon is located at the green triangle on the north shore below the Hampton Roads Beltway and above the I64 bridge tunnel.

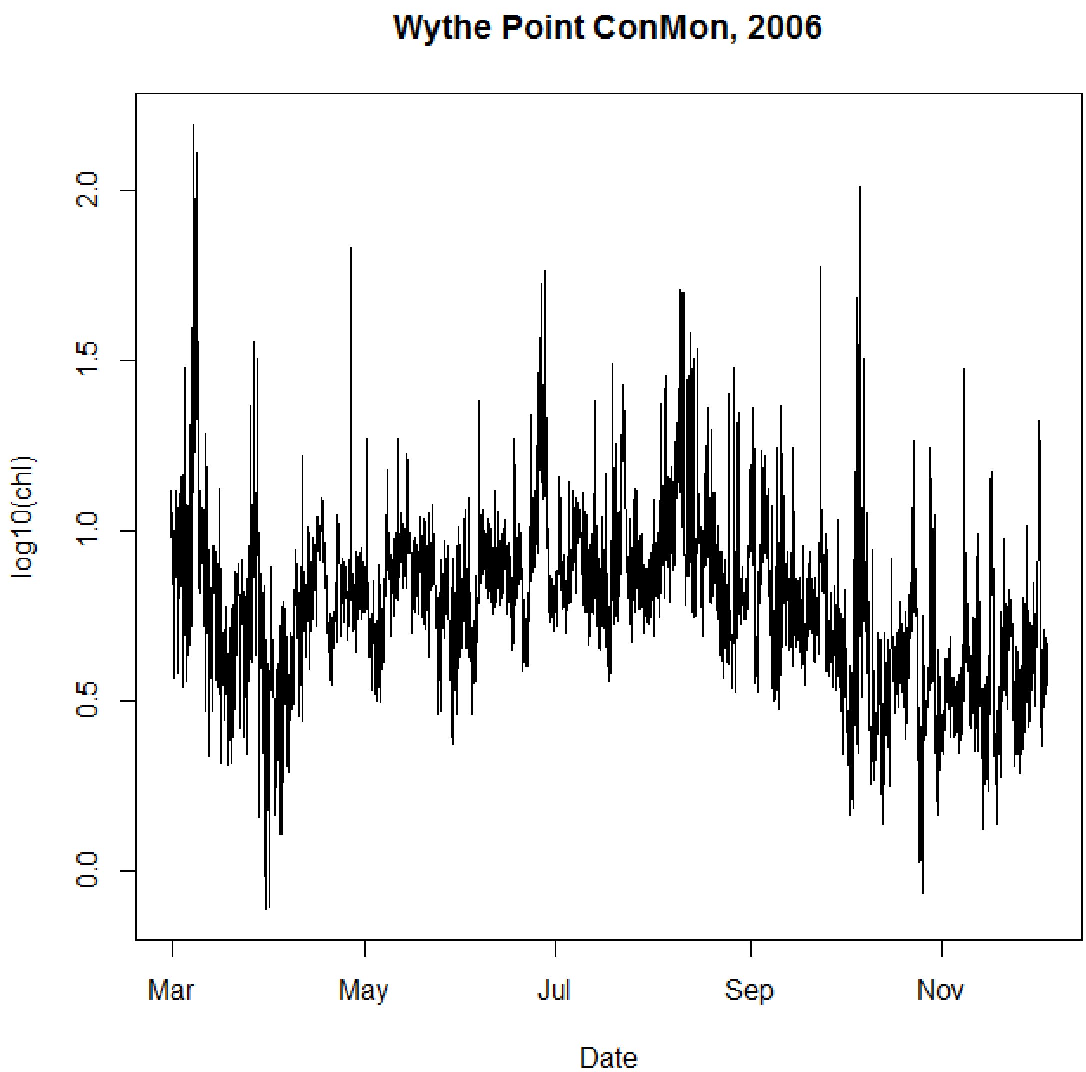


Figure 10. Time series of the base 10 logarithm of chlorophyll collected at the James Polyhaline ConMon Station near Wythe Point.

Temporal Analysis Results

Analysis of the Wythe Point data shows that a large majority of the serial dependence can be explained by an auto-regressive model with one degree of lag (AR1 model). The apparent auto-correlation of these raw data is 0.9. In the gam, both the seasonal term and the diel term are statistically significant (Table 1.0). The seasonal term (top panel Figure 11.) clearly explains a lot of variability in these data. The chlorophyll is high in March and then decreases toward early June. The low point in June is followed by a slow rise to high chlorophyll in late August and then a decrease in the fall. The model was implemented to capture the average diel cycle over the 2006 season. The diel variability is small compared to the seasonal variability and the diel pattern is characterized by a slight increase in chlorophyll in the late afternoon. It is possible that this pattern of chlorophyll results from diel migration of phytoplankton in the water column. However, there are studies (e.g., YSI 2015) which show that bright sunlight can suppress the fluorescence signal of chlorophyll, and thus this trend might possibly be a measurement artifact due to depression of fluorescence In mid-day. Thus, while the diel trend is statistically significant, the magnitude of this trend seems sufficiently small that it can be ignored for this exercise without invalidating results.

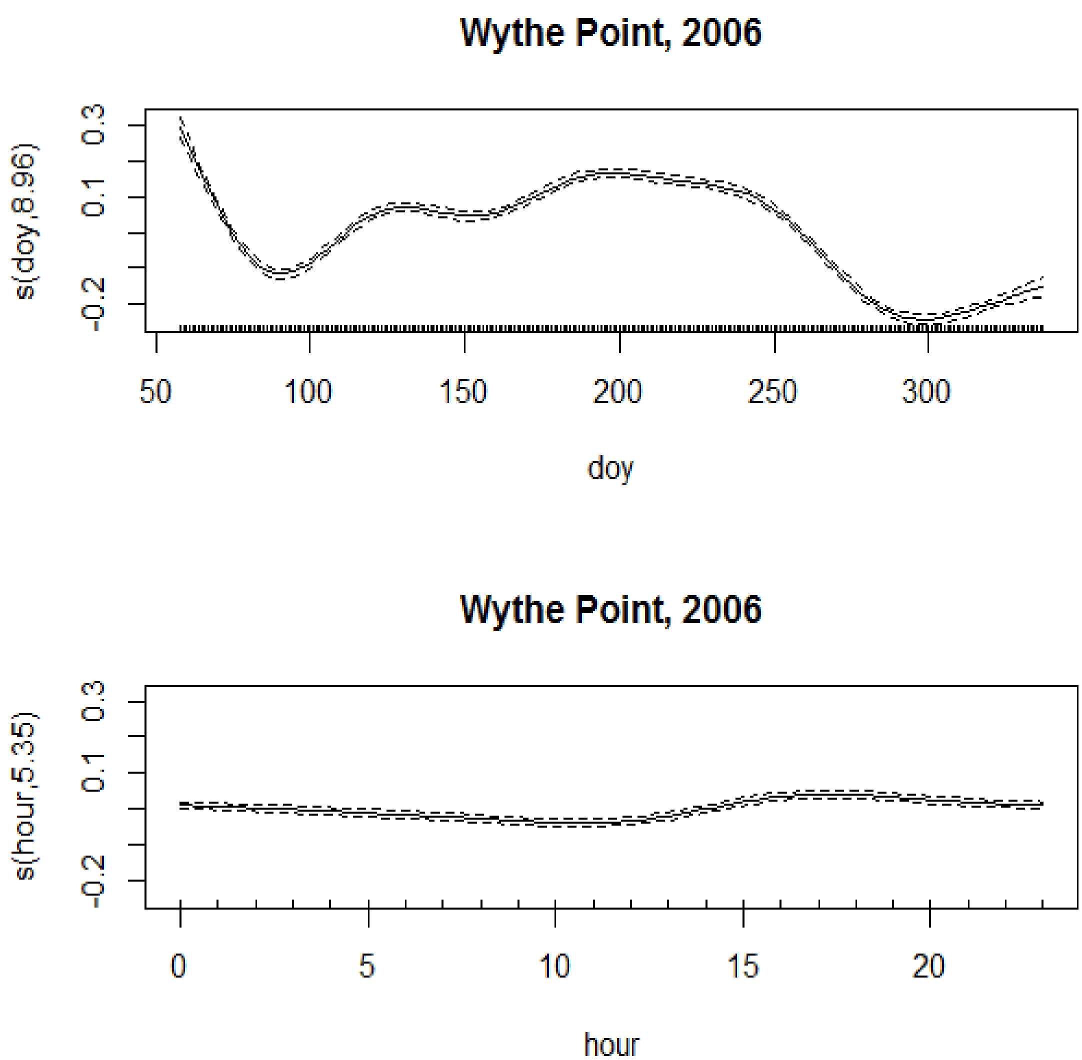


Figure 11. Graphical illustration of GAM smooth components. The top panel is seasonal pattern expressed as a smooth trend is day of year. The bottom panel is diel trend expressed as a function hour within the day.

Table 1. Analysis of Variance for the Smooth Terms of the GAM for 2006 Wythe Point ConMon data.

Type	Source	edf	F-stat	p-value
smoothed terms	s(doy)	8.96	393.148 7	<0.0001

	s(hour)	5.35	15.0806	<0.0001

The error structure of the GAM was specified as an AR1. The resulting estimate of the auto-regressive parameter is 0.86 which indicates a strong dependence of deviations from one hour to the next (Figure 12).

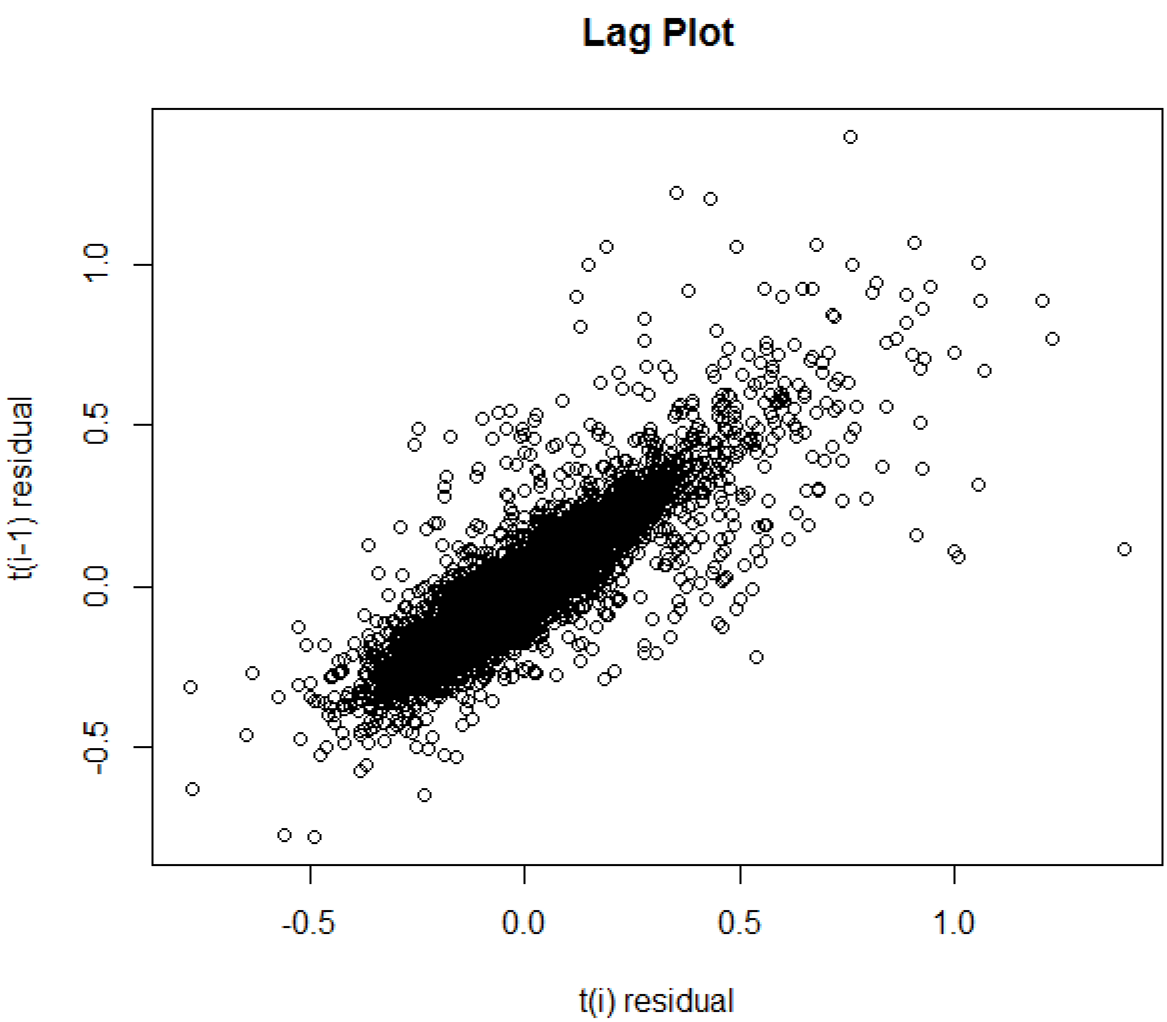


Figure 12. Lag plot of residuals from the Gam with smooth terms for seasonal cycle and diel cycle.

To assess the consistency (i.e. assumption of stationarity) of the AR1 model, the GAM model was fitted to each month of the 2006 data and the monthly AR parameters examined for pattern (Figure 13.). The highest autocorrelation was observed in June and the lowest autocorrelation was observed in August. For the remaining months the autocorrelation was about 0.8. At this point there is not sufficient information of establish whether the departures from 0.8 in June and August are a seasonal trend or just stochastic variation for this data set. Thus we proceed with the assumption that a stationary AR1 model with parameter = 0.8 is a reasonable process for generating a simulated chlorophyll time series.

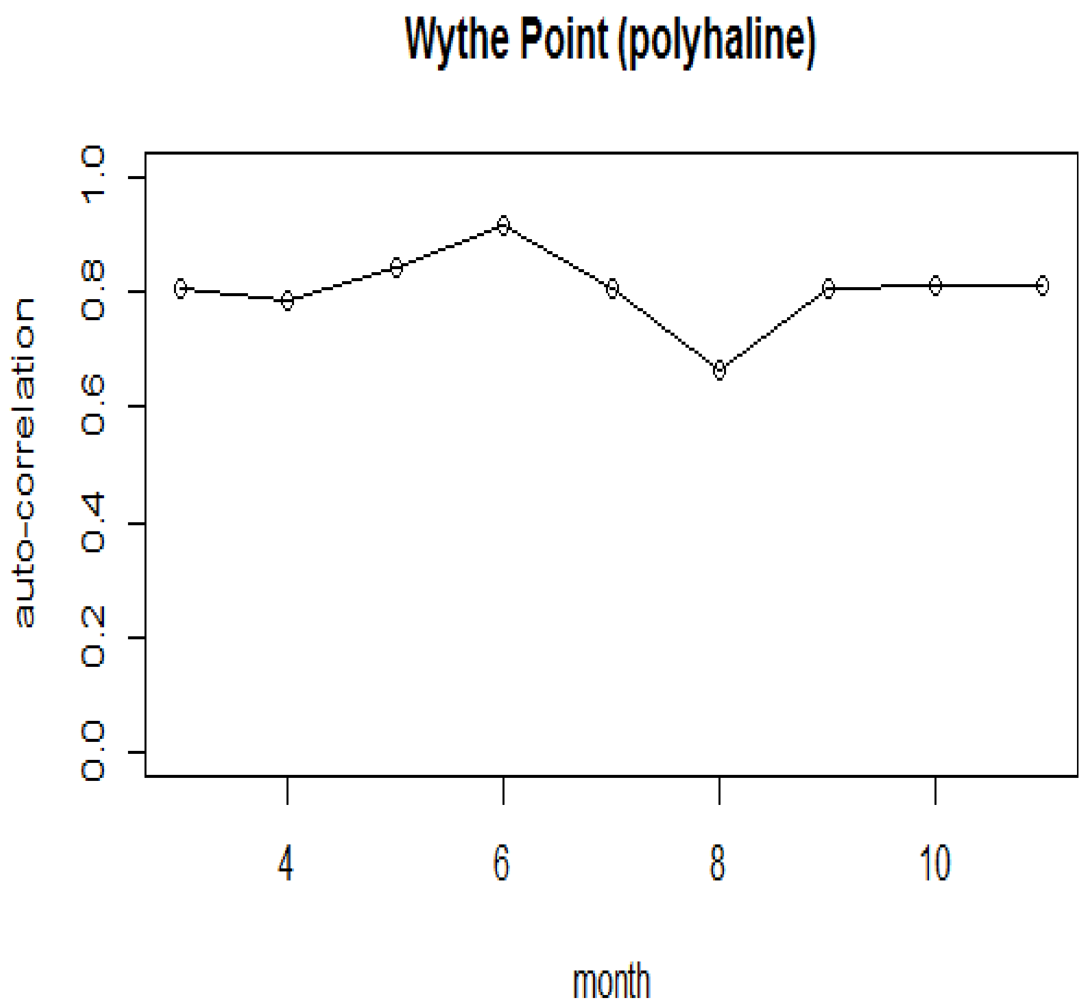


Figure 13. Autocorrelation parameters estimated for individual months from the Wythe Point 2006 ConMon data.

Simulation Analysis

Simulation, Sampling and Assessment Methods

The next step of this development process is to create a procedure to combine the spatial and temporal components into a simulator to generate data in a spatial-temporal domain. The procedure brings together standard methods of kriging in the spatial domain and auto-regression in the temporal domain. The simulated chlorophyll observations have three components: seasonal mean, spatial error, and temporal error. The seasonal mean is determined by the interpolated surface of each DataFlow cruise (e.g. smooth black line As shown in Figure 7). The spatial error is obtained from the grf() function of the geoR package. The time series component is created by repeatedly generating matrices of spatial random error and combining them using an autoregressive formula (equation 1).

$$\omega_t = \varphi \times \omega_{t-1} + \epsilon_t \qquad 1$$

where, ω_t are the correlated errors, ϵ_t are independent errors, φ is the autoregressive coefficient, and t is the ordinal number for time. For this simulator the spatial domain is determined by the cell centers of the CBP interpolator. The time steps are set at 1 hour. Data are simulated for the 2005-2007 time period to provide a dataset for executing a 3 year assessment using the methods published in ().

The data created by this simulation are treated as if they represent a feasible realization of chlorophyll in the polyhaline James for the three year period. In this exercise, the chlorophyll for each interpolator cell center and every hour of time are known. Using these data, we compute seasonal averages for every interpolator cell and then compute to proportion of space violating the relevant season criteria for the three year period. These proportions are then ranked and processed into a Cumulative Frequency Diagram (cfd) to represent the true state of nature. The cfd for the true state is compared to the reference cfd curve to define whether the true state is in compliance with the criteria. This true state is the basis for comparison for different levels of sampling.

The first sampling experiment explored here is to assess the precision of cfd's estimated from monthly sampling at the CBP fixed stations. The experiment is based on repeatedly drawing samples from each month of the 3 year simulation. These samples are treated as fixed station sample data and are processed into a family of estimated CFDs. Each set of samples from a month are processed by IDW interpolation, the interpolated data within a season are averaged by interpolator cell id, and seasonal proportion of space exceeding the seasonal criteria is estimated, ranked and scored to obtain repeated estimates of the CFD.

In order to simulate sampling at the fixed stations from the simulated data, we first identify the interpolator cell centers that are closest to the published latitude/longitude of the fixed stations (Figure 13.) This figure shows each fixed station location as a black dot with a line connecting to the nearest interpolator cell which is shown with a red center. The remaining interpolator cells for the segment are shown as open circles in aquamarine. Interpolator cells outside of the segment are shown in gray. The locations shown in red are designated as simulated fixed stations.

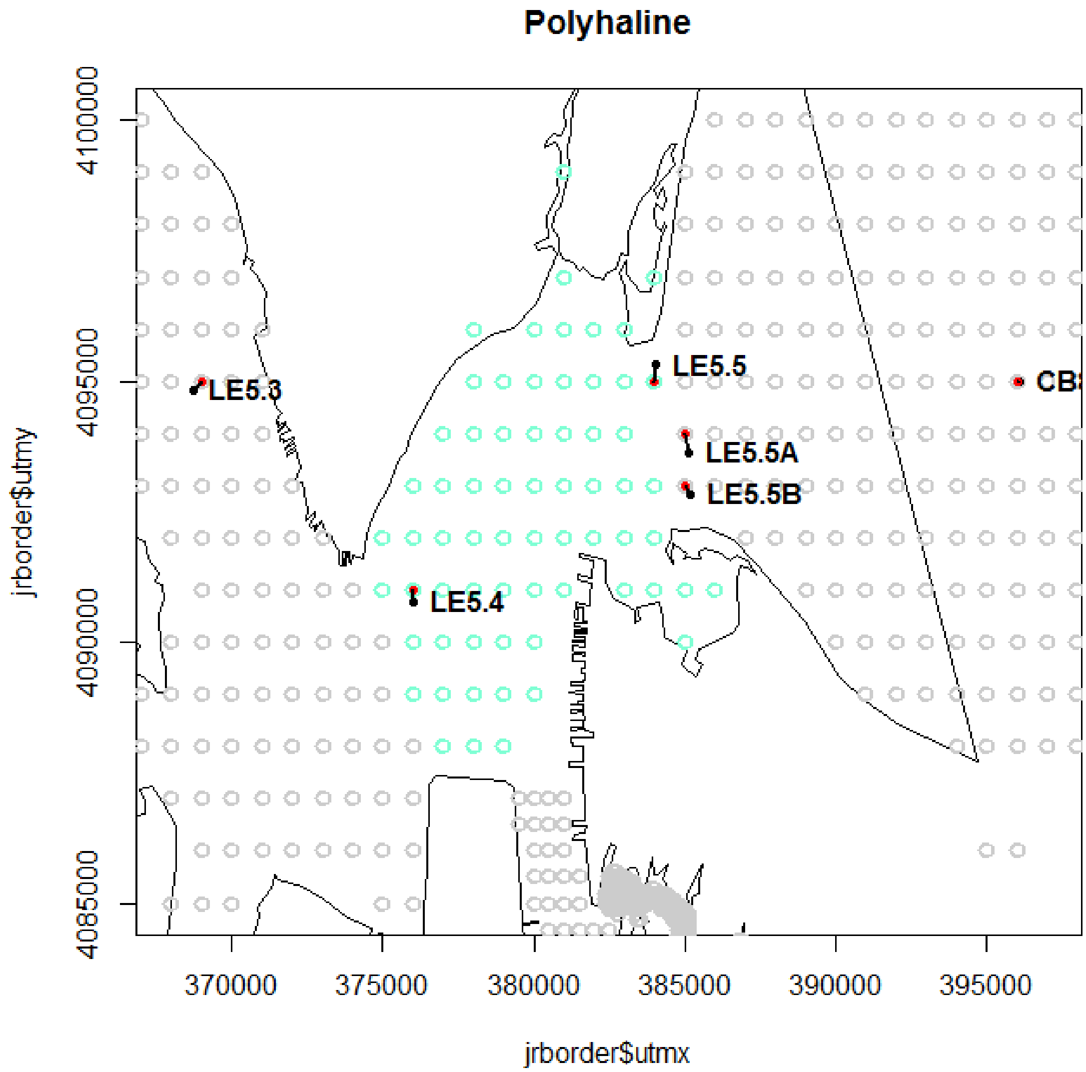


Figure 14. Each fixed station location is shown as a black dot with a line connecting to the nearest interpolator cell which is shown with a red center. The remaining interpolator cells for the segment are shown as open circles in aquamarine. Interpolator cells outside of the segment are shown as open circles in gray.

The simulated fixed stations in the simulated data are sampled once a month. The day of the month is chosen randomly where all days have equal probability and the hour of the day is chosen at random from the hours between 9:00 a.m. and 4:00 p.m. All of these simulated fixed stations in the segment are sampled simultaneously for the day and hour selected. These simulated fixed station observations are then interpolated to obtain the estimate chlorophyll surface for the segment. The interpolation algorithm for this step employs an Inverse Distance Squared method like that used by the CBP interpolator.

The monthly interpolations are executed for each month between March and September for each year between 2005 and 2007. The resulting interpolation estimates are averaged by season where spring is March –May and summer is June-September. These cell by cell seasonal averages are compared to seasonal criteria in the logarithm base 10 scale: spring criteria = log(12)= 1.079181 and summer criteria = log(10) = 1.0. The seasonal proportion failing is estimated by the number of cells for which the seasonal average exceeds the seasonal mean criterion divided by the total number of polyhaline interpolator cells. For each season, the three estimates for the three years are ranked and proportion of time is estimated by the score (rank/(N+1)) where N is the total number of years equal to three in this case.

Results of Simulation and CFD assessment

The procedure described above was used to generate hourly data from March through September for each interpolator cell in the polyhaline segment for the years 2005 through 2007. ConMon data from Wythe Point is available for 2006 and 2007 and can be compared to simulated data from the closest interpolator cell (Figure 15.) which is about 0.6 km distance from the Wythe Point monitor. The ConMon data appear to have somewhat greater variability than the simulated data, but the seasonality and serial dependence of the simulated data appears to reflect the patterns in the ConMon data quite well. It is important to remember that the variability and the seasonality of the simulated data are based on the DataFlow cruises. The ConMon data were used only to obtain and estimate of serial dependence. Thus the ConMon data offer a relatively independent confirmation of the simulation process. At this point it is undetermined whether the greater variability of the ConMon data is because the chlorophyll of the shallow water location is inherently more variable than channel locations or because the simulation process is underestimating variability.

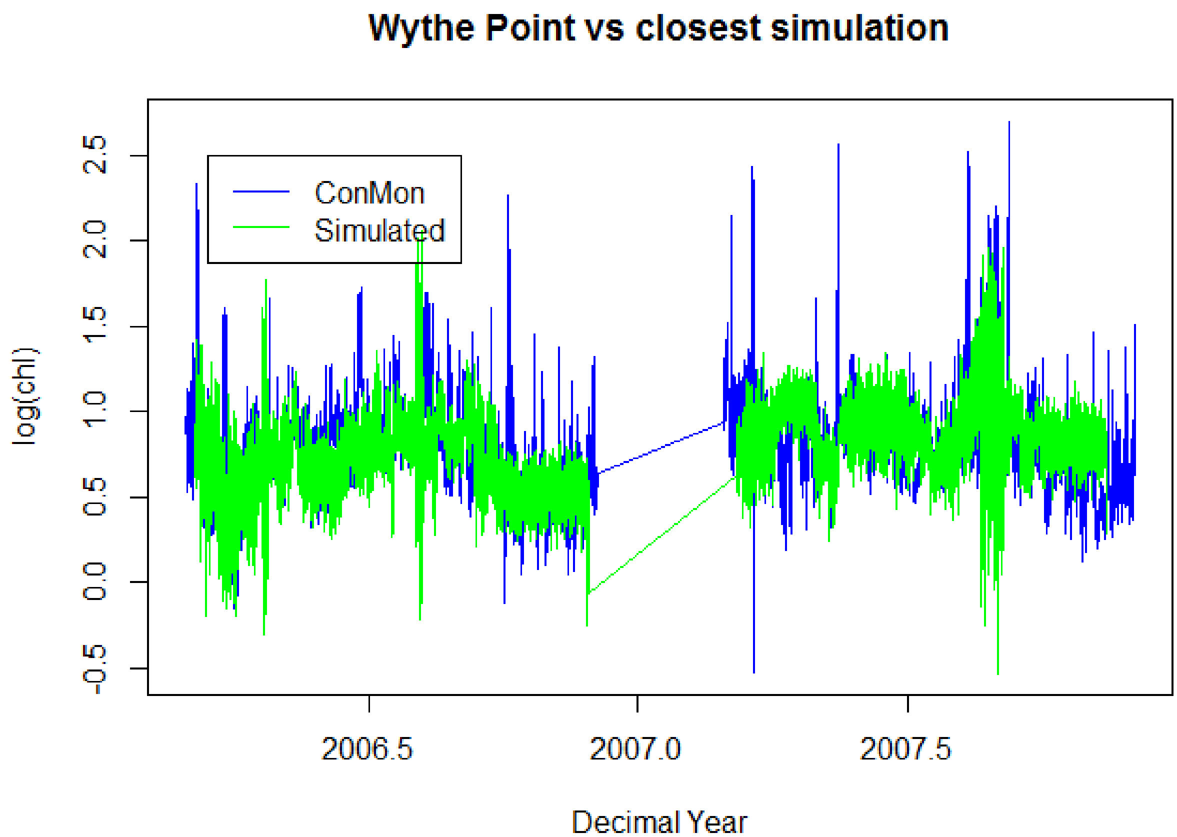


Figure 15. Two years of simulated log-chlorophyll data from this simulation process compared to the same two years of observed time series from the Wythe Point ConMon station.

Zooming in on the time axis of the simulated time series as compared to the observed ConMon time series (Figure 17.), we can examine in more detail the temporal dependence apparent in these data. To my eye, the ConMon data appear to have somewhat smoother oscillations than the simulated data indicating stronger serial dependence. On the whole, the simulated data are a reasonable reflection of the observed data.

As described above, the simulated data represent log(chlorophyll) for each interpolator cell location and for each hour of each day. These data are averaged over the spring and summer seasons for each interpolator cell (e.g. Figure 16) and then compared to seasonal criteria to obtain percent of space out of compliance. Using standard methods, these compliance rates are processed to CFD curves (Figure 18). Because of the high spatial and temporal density of the simulated data, these curves represent the true state of the simulated data.

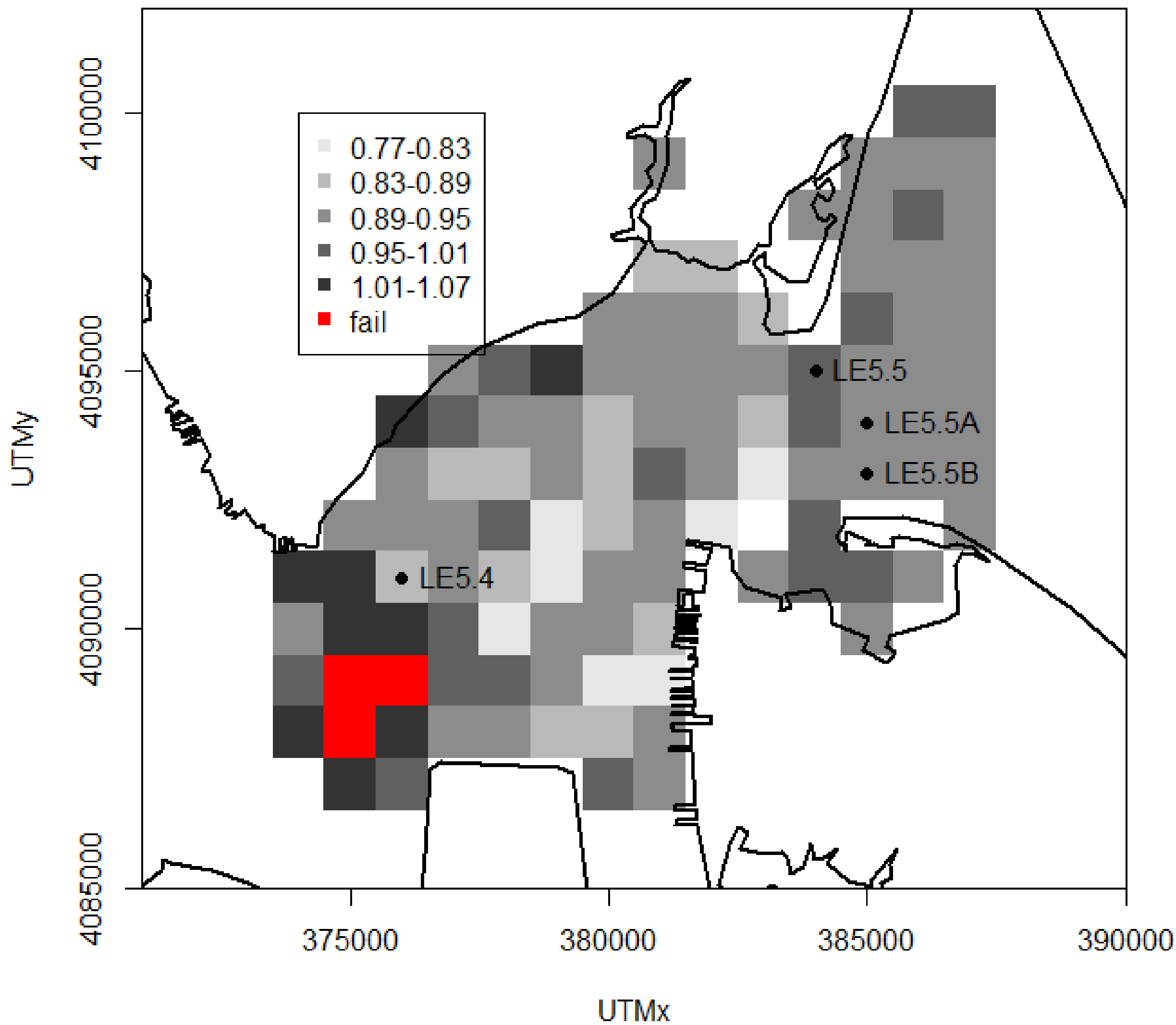


Figure 16. Spring time means of the simulated data for the year 2005.

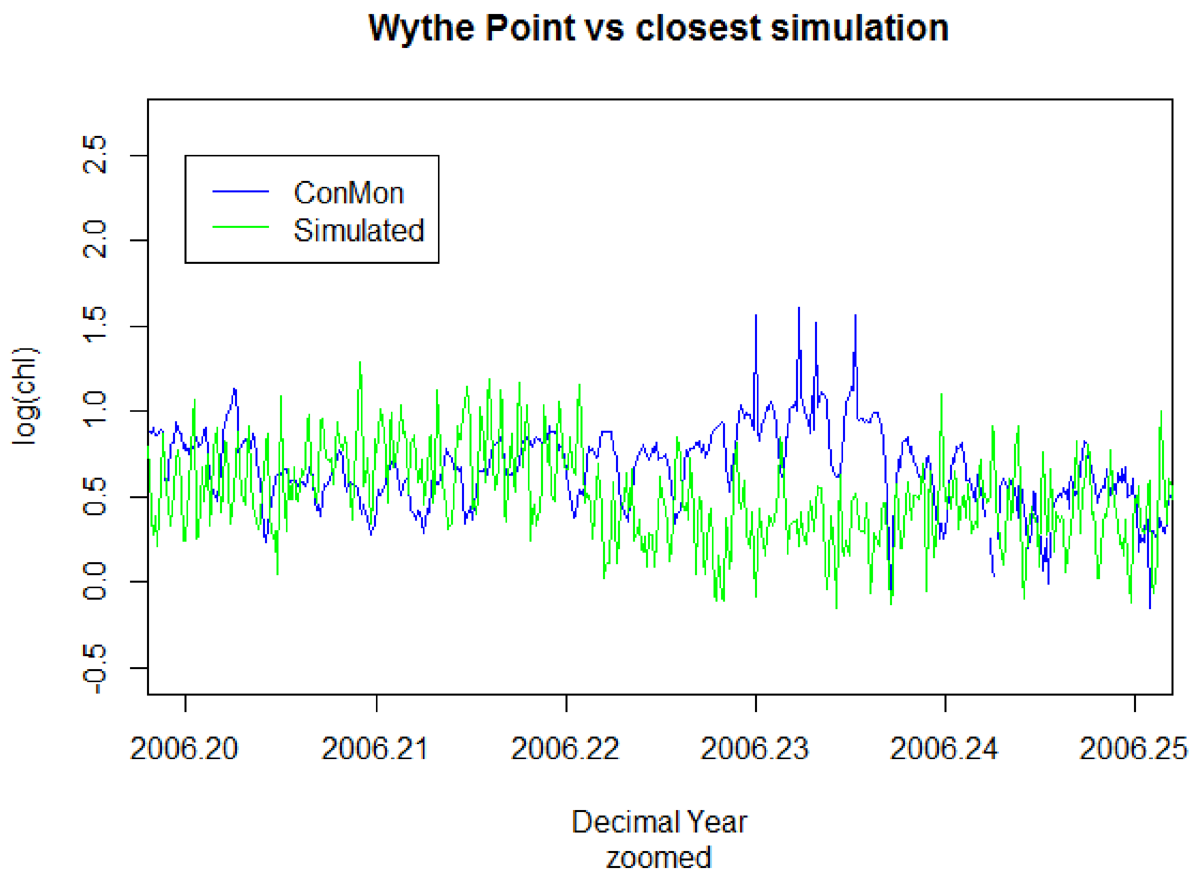


Figure 17. Zooming in to examine 0.05 years (about 18 days) of the Wythe Point ConMon log(chlorophyll) time series as compared to the simulated time series. The blue curve shows a partial time series from the Wythe Point ConMon data. The green curve shows simulated data from the interpolator cell that is closest to the Wythe Point ConMon.

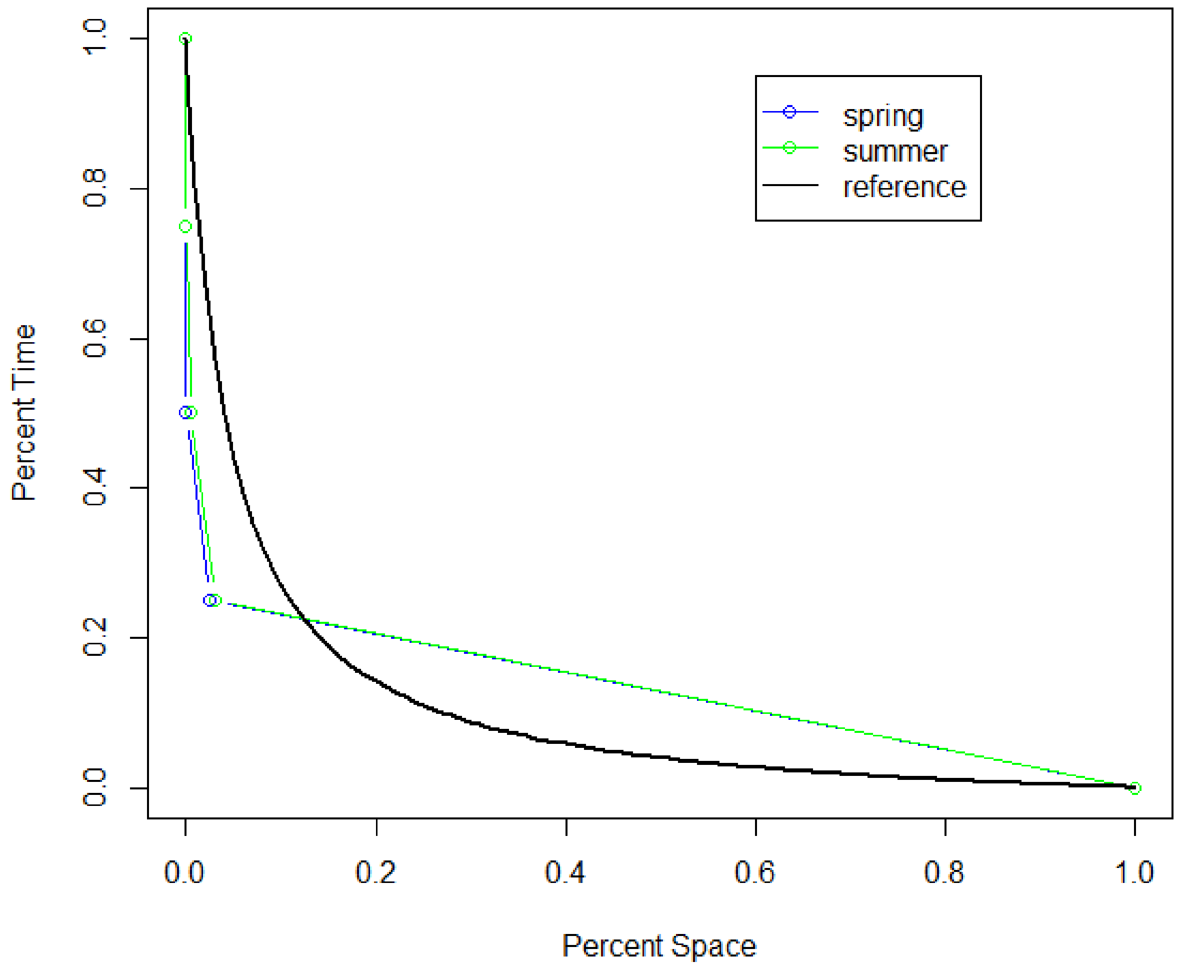


Figure 18. True seasonal CFD curves for the simulated data without manipulation as compared to the standard reference curve.

The true state of the simulated data without manipulation is passing by a comfortable margin with respect to the standard reference curve. Note that the mean of the simulated data is set by the Kriging estimates obtained from the sequence of dataflow cruises, and thus suggests that the true state of the polyhaline segment is in compliance for the 2005-2007 time period.

In a first sampling experiment, we study the uncertainty of the CFD assessment when once a month fixed station sampling is employed to estimate the CFD. The methods for this experiment are those described above where one day and hour per month is randomly selected from the simulated data and processed through IDW interpolation and seasonal averaging to obtain estimated CDF curves. This process is repeated 1000 times and the resulting CFD curves are compared to the reference curve (Figure 19.)

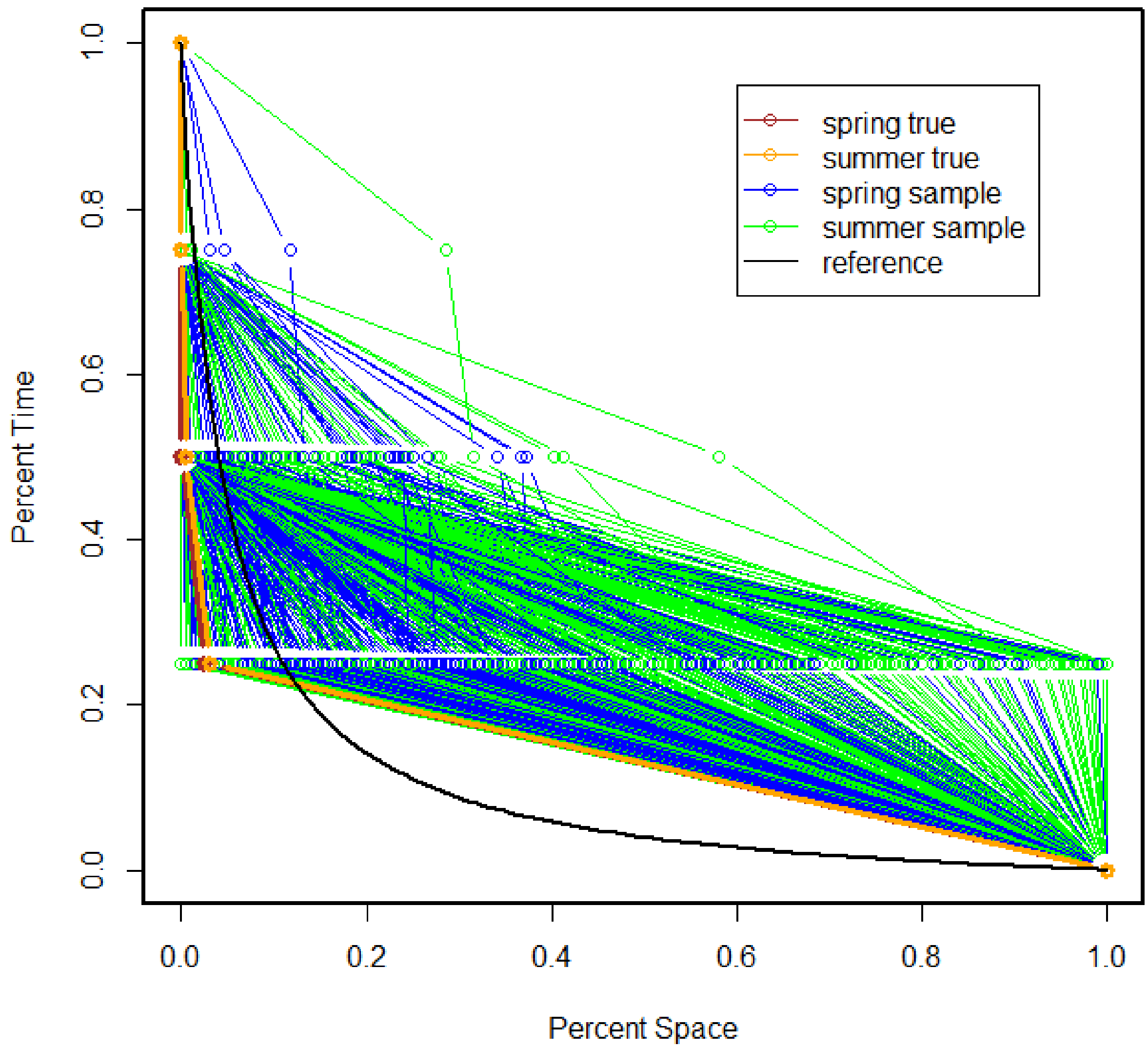


Figure 19. The standard reference curve (black), the true spring and summer CFD for the simulated data (brown and orange), and 1000 estimated CFDs created by monthly sampling of the simulated data (blue and green).

In Figure 19, the black curve is the reference CFD. The brown and orange curves are the true CFDs for spring and summer (respectively) for the 2005-2007 assessment period based on the simulated data.

The true CFDs are in compliance with the reference curve. The family of blue and green curves represents CFDs for spring and summer (respectively) that were computed by repeatedly drawing once a month samples from the fixed stations and going through the CFD process. The uncertainty of once a month sampling creates a lot of failing CFD curves (false positives) even when the true curves are passing.

The false positive rate for one-a-month fixed station sampling is in the neighborhood of 40% (Table 2.0). Thus there is a high probability of requiring remediation when none is needed. In this assessment using seasonal compliance for a three year period, there are only three observations of proportion of space out of compliance that are ranked to create the CFD. It is instructive to review which of the 1st, 2nd, or 3rd rank is responsible for CFD violation. It is clear for both spring and summer, that the large majority of the time, it is the highest ranked percent of non-compliance that crosses the CFD reference curve. In fact, the occurrence of the rank 2 percent of space being out of compliance when the rank 3 percent of space is in compliance is very rare, less than 1% of cases. Thus the CFD process is almost equivalent to just assessing compliance based on the maximum of the percent of space assessments during the three year period. It is not clear that this conclusion would hold when there are more than three compliance rates being ranked because the reference curve begins to swing strongly to the right for percent of time less than 25.

Table 2.0. Proportion of CFD violations (false positives) by Rank and total for the sampling experiment where the true CFD represents a passing condition.

Spring				Summer			
rank 1	rank 2	rank 3	Total	rank 1	rank 2	rank 3	Total
0.005	0.076	0.375	0.385	0.001	0.070	0.410	0.413

In second sampling experiment, we manipulate the simulated data to create a data set where the true CFD is failing. This is accomplished by simply increasing all chlorophyll observations by 20 percent. With this manipulation, the spring data just barely exceed the reference curve while the summer data exceed the reference curve by substantial margin (Figure 20.)

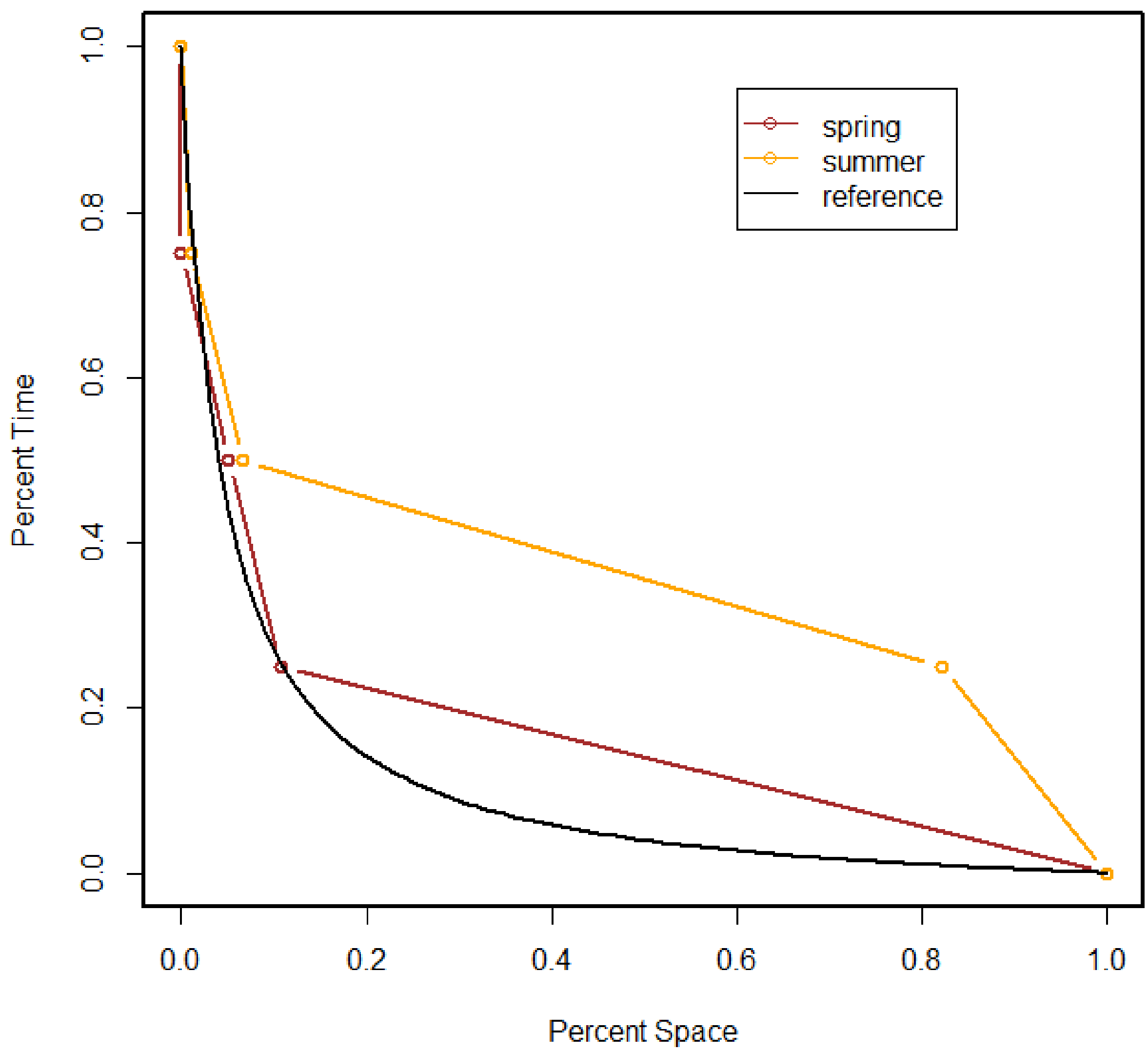


Figure 20. True seasonal CFD curves for the simulated data that has been manipulation to create failure when compared to the standard reference curve.

Using the manipulated data, the once-a-month fixed station sampling experiment is repeated (Figure 20, Table 3.0). Again, we see remarkable variability in the estimated CFDs relative to the true CFDs. The false negative rate for both spring and summer is in the neighborhood of 20%. That suggests, that even when the segment is in a failing condition, there is about a 20% chance that a CFD created from once-a-month fixed station sampling will indicate passing. This is a high probability of leaving the environment unprotected.

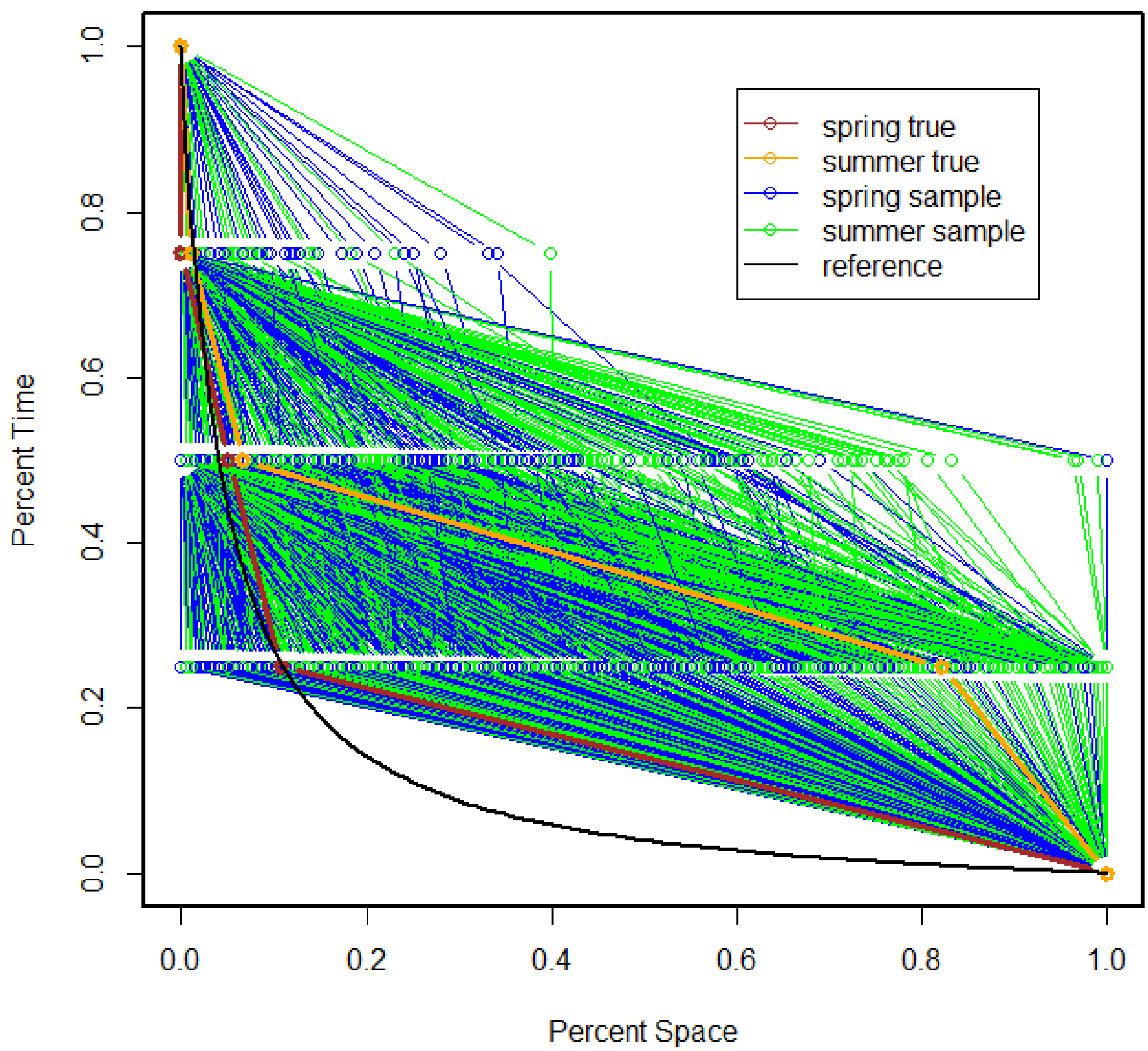


Figure 21. The standard reference curve (black), the true spring and summer CFD for the simulated data that has been manipulated to fail (brown and orange), and 1000 estimated CFDs created by monthly sampling of the manipulated data (blue and green).

Table 3.0. Proportion of CFD passes (false negatives) by Rank and total for the sampling experiment where the true CFD represents a failing condition.

Spring				Summer			
rank 1	rank 2	rank 3	Total	rank 1	rank 2	rank 3	Total
0.964	0.633	0.231	0.219	0.943	0.625	0.208	0.192

In a third sampling experiment, we manipulate the simulated data to create a data set to where the true CFD is passing by sufficient margin that there is a reasonably high probability that a CFD constructed from once-a-month fixed station sampling will result in pass. This is accomplished decreasing all chlorophyll observations in the simulation to 75% of the simulated value. With this manipulation, both the spring and summer show zero percent of space in violation for all three years (Figure 21.), and yet the false positive rate is still between five and ten percent (Table 4.0). Thus while there is a high probability that a CFD based on once-a-month fixed station sampling will result in passing, there are greater than 1 in 20 odds that it won't. Furthermore, the simulated data were already appreciably below the criteria before the 25% reduction. The 25% reduction put the chlorophyll level at about 65% (35% reduction) of what would be a marginal level. This shows the true cost of inadequate monitoring. If assessments rely on once-a-month fixed station monitoring, then chlorophyll will have to be reduced 35% below an acceptable level in order have a reasonably high probability that the assessment will show that it is acceptable.

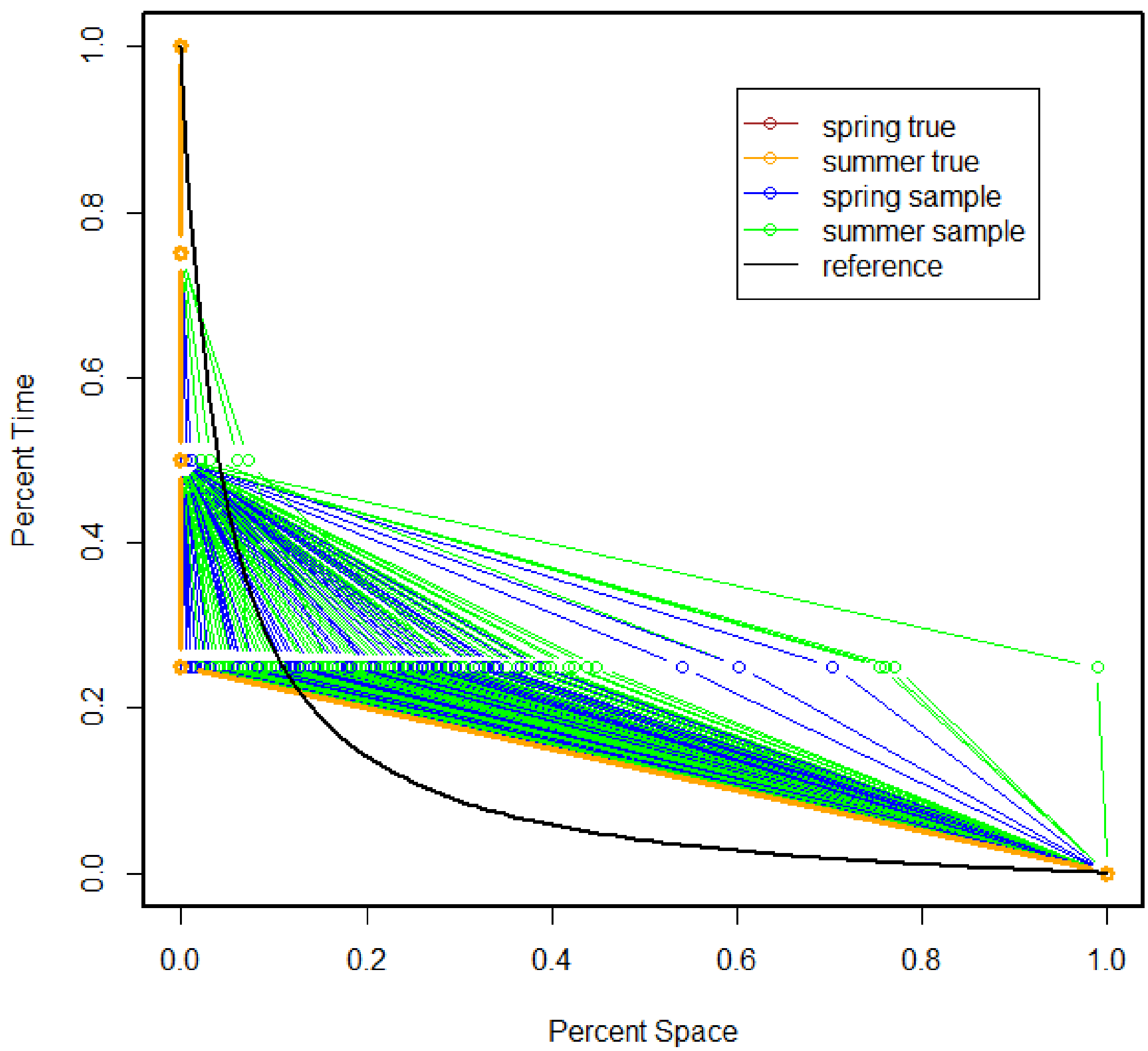


Figure 22. The standard reference curve (black), the true spring and summer CFD for the simulated data that has been manipulated to have a high probability that the estimated CFDs will pass (brown and orange), and 1000 estimated CFDs created by monthly sampling of the simulated data (blue and green).

Table 4.0. Proportion of CFD violations (false positives) by Rank and total for the sampling experiment where the true CFD represents a passing condition for simulated data reduced by a factor of 0.75.

Spring				Summer			
rank 1	rank 2	rank 3	Total	rank 1	rank 2	rank 3	Total
0.000	0.000	0.076	0.076	0.000	0.003	0.088	0.089

Discussion

The CFD is a nascent assessment tool that has been implemented by CBP. It is intuitively appealing because it quantifies the spatial and temporal aspects of criteria violations separately. Because it is novel, the study of its statistical properties such as uncertainty and bias is immature. The simulation study presented here provides a tool for empirical evaluation of the variability and bias of the CFD under different sampling plans.

While the CFD is intuitively appealing, there has been little study of its reliability and in particular how much data is required to make it reliable. In this limited application, only the once-a-month fixed station sampling has been evaluated for the polyhaline James River segment. This analysis shows that the variability of the CFD based on this limited sampling plan is very high. When the true condition of the estuary is either passing or failing, the sample CFD has a high probability of reaching the wrong conclusion. The odds of making the right decision are very little better than if the decision were reached by flipping a coin. The costs of this uncertainty can be very high. When the estuary is not meeting the designated use, there is a high risk of failing to identify this failing condition so that the resource remains unprotected as illustrated by sampling experiment 2. On the other hand, when the estuary is meeting the designated use, there is a high risk of concluding that remediation is needed which places an unnecessary burden on the stakeholders responsible for that remediation. This risk is illustrated by sampling experiment 1. Experiment 3 demonstrates that when uncertainty is high, the true state of the estuary must far exceed the requirements of its designated used in order to have a high probability of correctly identifying the passing condition (i.e. in order to remove it from the TMDL list). At a minimum, this will require remediation beyond what is necessary. In some cases it may require a degree of improvement that is technically not achievable such as Dissolved Oxygen concentrations above the saturation level. In other cases trends in an improving direction may have to go so far that system performance is sub-optimal. For example chlorophyll may need to be reduced to a point where it is symptomatic of an oligotrophic estuary. In summary, when an assessment procedure is implemented, it is always important to ask 'How much improvement is required to have a high probability of satisfying this assessment?'. The tools developed in this study provide a means of answering this question for the CFD.

In addition to assessing the variability of the CFD, bias in this assessment procedure should also be considered. In experiment 2 the spring case produced a true CFD that is close to the reference curve. Given that the true state is very close to the reference condition, one would expect the pass:fail odds to be close to 50:50. Yet the observed odds are 22:78. This suggests that the CFD is biased in favor of failing. This could be because the fixed stations are a biased representation of the simulated chlorophyll field or it could be because of some mathematical bias of the CFD process with small samples. This warrants investigation.

This proof of concept exercise combines the spatial properties of dataflow with the temporal properties of ConMon to create an analysis combining the assets of both. To my knowledge, this is the first attempt to combine these two aspects of the shallow water monitoring program into a coherent spatial-temporal analysis. The cursory validation examples presented indicate the simulated data does mimic the spatial and temporal dependence that is present in the observed DataFlow and ConMon data. Some deficiencies are noted, and some improvement might be achieved with additional research on more appropriate time-series and spatial statistical models.

The possibilities of additional applications based on these methods are many. In addition there are areas where additional research might improve this methodology. A partial list of potential research is given here in bullet format:

- Test the generality of these results by applying these methods to other James River segments.
- Examine the degree of improvement attained when interpolations are based on DataFlow sampling rather than fixed station sampling.
- The tools developed here focus on variability. Assessments of bias should be developed as well.
- Apply the methods developed here to other important assessment parameters such as dissolved oxygen and water clarity.
- Improve the simulation through additional comparisons of the simulated data to ConMon and DataFlow.
- Explore using trends in ConMon to inform simulation temporal interpolation between DataFlow cruises.
- Develop an analytical framework for the CFD and confirm its validity using this simulation process.
- The methods developed here are somewhat complex. Some form of quality assurance is needed to confirm that the results are not invalid because of programming or conceptual errors.

In the near future I hope to meet with the sponsors of this research to assess the utility of this line of research and prioritize these research options.

Citations:

Cressie, Noel A. C. (1993)
Statistics for Spatial Data. John Wiley & Sons, Inc., New York. pp900.

Ribeiro, Paulo J. Jr and Peter J. Diggle, 2012
Package 'geoR', http://cran.r-project.org/web/packages/geoR/geoR.pdf.

R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Environmental Protection Agency. 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 CBP/TRS. Region III Chesapeake Bay Program Office.

YSI Incorporated. YSI 6_Series Environmental Monitoring Systems Operations Manual. In website: https://www.ysi.com/. Downloaded February 13, 2015. 122 pp.

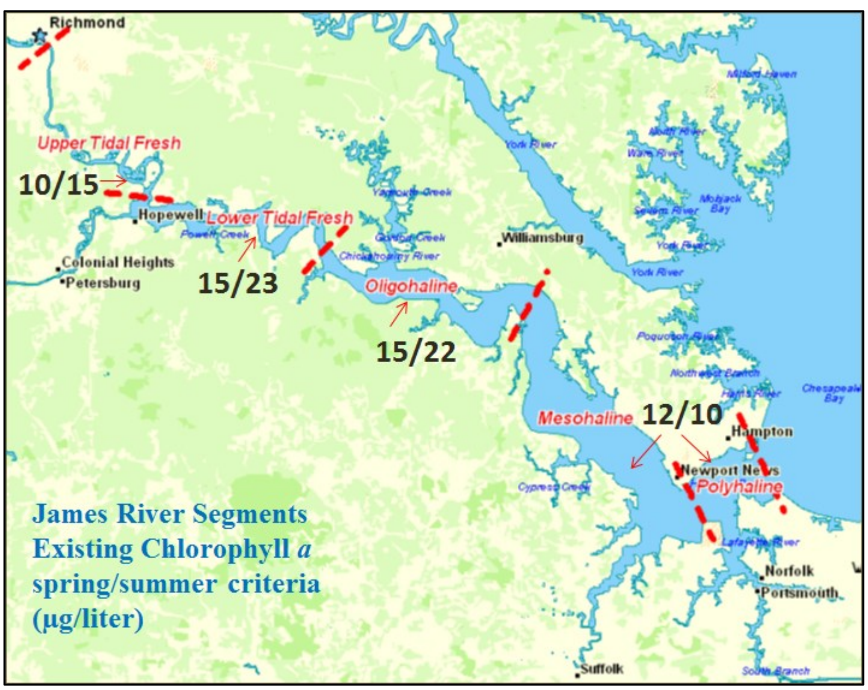
Biological Reference Curves for Assessing the James River Chlorophyll *a* Criteria

Final Report
September 8, 2014

Prepared by
Claire Buchanan, PhD
Interstate Commission on the Potomac River Basin

for

Chesapeake Environmental Communications
and
Virginia Department of Environmental Quality



ICPRB Report 14-3

Interstate Commission on the Potomac River Basin
30 West Gude Dr., Suite 450
Rockville, MD 20850
301-984-1908

Disclaimer

The opinions expressed in this report are those of the author and should not be construed as representing the opinions or policies of the United States government, or the signatories or Commissioners to the Interstate Commission on the Potomac River Basin.

Acknowledgements

This project was supported by funds from the Virginia Department of Environmental Quality (VADEQ), by way of Chesapeake Environmental Communications (CEC), and from CWA§106 grant 98339412-0. The Interstate Commission on the Potomac River Basin, an interstate compact river basin commission of the United States government and the compact's signatories: Maryland, Virginia, Pennsylvania, West Virginia, and the District of Columbia.

Biological Reference Curves for Assessing the James River Chlorophyll *a* Criteria

Executive Summary

Crucial to the application of any criteria is a reference for measuring criteria attainment. A simple, symmetrical 10% hyperbolic curve is currently the default reference for Virginia's James River chlorophyll *a* criteria. The curve is intended to represent the spatial and temporal exceedances of the criteria in healthy phytoplankton populations. This study compares the 10% hyperbolic curve to "biological reference curves" derived from actual healthy populations living in high quality waters that support Chesapeake Bay open water designated uses. The study then explores some of the statistical properties of the biological curves.

Criteria approximating the upper percentiles (90^{th} – 95^{th}) of chlorophyll *a* concentrations in high quality waters produce biological reference curves that track the default 10% hyperbolic curve. Four of the ten season- and segment-specific chlorophyll *a* criteria for the James River approximate the 90^{th} – 95^{th} percentiles. In these cases, the default 10% hyperbolic curve is a good reference for measuring criteria attainment. The other James River criteria are slightly higher than the 95^{th}%ile or lower than the 90^{th}%ile, and their biological reference curves depart somewhat from the default curve. Overall, James River criteria are generally protective of high quality environmental conditions in Chesapeake Bay open water designated uses.

When Virginia's current 3-year assessment procedures are applied to the Chesapeake data set of high quality tidal waters, small statistical biases and artifacts inherent to the method can be seen. For example, instances of forced non-compliance may occur due to a statistical artifact ("bottleneck") inherent in the reference curves of assessment units having relatively few interpolator cells. The existing method could be slightly modified or clarified to avoid these issues.

The report's findings suggest the following changes be made to the existing Virginia Department of Environmental Quality procedures for assessing chlorophyll *a* criteria:

1) Continue to use simple, symmetrical hyperbolic curves as default reference curves in chlorophyll *a* criteria assessments, but adjust these curves so that the percent of allowable exceedances in each assessment unit reflects the biological reference curve derived with the James River criteria.

2) Increase spatial and temporal coverage with the use of data from other sampling technologies (e.g., DATAFLOW, satellite imaging, and continuous monitoring buoys) and improve the accuracy of interpolated chlorophyll assessment layers.

3) Given that the existing WQS require seasonal means in Virginia chlorophyll *a* assessments, use a longer assessment period *and* develop reference curves derived from seasonal means. This minimizes the biases introduced into the assessments by too few data layers and seasonal averaging.

4) Develop confidence intervals for the hyperbolic reference curves that account for statistical biases and artifacts inherent to the assessment method. Establish rules for assessing attainment rates that fall outside the reference curves but inside these confidence intervals.

Introduction

The Commonwealth of Virginia has developed numeric chlorophyll *a* criteria for the tidal James River to use in CWA§303 assessments (VADEQ 2004). The state also adopted a Cumulative Frequency Distribution (CFD) methodology to measure attainment of its chlorophyll *a* criteria (USEPA 2007). Essential to this CFD method are reference curves which delineate allowable exceedances of the criteria in space and time for a given assessment period. CFD curves constructed from actual monitoring data (attainment curves) are then compared to these reference curves to determine whether or not the criteria were met.

A simple, symmetrical 10% hyperbolic curve can be used as a default reference curve for chlorophyll *a*. However, reference curves based on phytoplankton populations inhabiting high quality conditions in the upper, sunlit layer of the water column (reference conditions) would be preferred. "While there is mathematical and statistical logic underpinning the [10% hyperbolic] chlorophyll reference curve, it is important to remember that it is based on parametric models and simplifying assumptions. It is recommended that validation exercises be performed to insure that the general shape of CFD curves generated from data collected in near reference conditions is approximated by the proposed curve (USEPA 2007)." At a minimum, the reference conditions should have water column transparency deep enough to support unstressed photosynthesis in phytoplankton and nutrient concentrations low enough to limit the formation of nuisance algal blooms. Individual water samples exhibiting these conditions are found fairly often in some open water habitats of Chesapeake Bay. The samples represent discrete parcels of water flowing through the estuary. Since phytoplankton cells are short-lived and their populations adapt quickly to their surrounding environment, populations found in reference conditions are believed to represent healthy, desirable communities.

With some care, reference water samples found in the Chesapeake Bay Program (CBP) tidal water quality monitoring database can be used to create biological reference curves for Virginia's chlorophyll *a* criteria. The objective of this analysis is to investigate the properties of these reference curves, and to compare them to the default 10% hyperbolic curve.

Methods

Data preparation

In April 2014, D. Jasinski (Chesapeake Environmental Communications) downloaded from the CBP website (www.chesapeakebay.net, Data Library) data for Chesapeake tidal water samples collected above the pycnocline or in the surface layer between 1984 and 2013. Parameters selected for downloading included: salinity, chlorophyll *a*, dissolved inorganic nitrogen (DIN), ortho-phosphate (PO_4), Secchi depth, total nitrogen (TN), total phosphorus (TP), dissolved organic carbon (DOC), total organic carbon (TOC), total suspended solids (TSS), dissolved oxygen (DO), water temperature, and pheophytin. Station information included station name, water body name, CBP segment (2003), latitude, and longitude. For each station-date event, multiple measurements of a parameter in the above-pycnocline layer were averaged.

Station total depths obtained from the CBP 2012 Water Quality Users Guide or from M. Mallonee, the Water Quality Data Manager at CBPO, were incorporated. Stations with total depths less than 2 meters were then removed (they are not considered open water environments). Sampling events were assigned to one of five seasons based on date: spring (March – May), June, summer (July – September), autumn (October – November), and winter (December – February). Sampling events were assigned to one of four salinity zones based on the average above-pycnocline salinity measured on the sampling date: tidal fresh (<0.5 ‰), oligohaline (>0.5 - 5.0 ‰), mesohaline (>5.0 - 18.0 ‰), and polyhaline (>18.0 ‰). Sampling events with no Secchi depth or chlorophyll *a* measurement were removed. Events with Secchi depth equal to zero were also removed. Five sampling events with suspicious, inconsistent values for chlorophyll *a* and pheophytin or DIN and PO_4 were removed. If three or more sampling events occurred in a 7-day window, all but one of the records was removed to avoid over-weighting measurements from a particular location and sampling time. The QA/QC'ed data set at this point contained 64,200 records. These sample events are considered representative of Chesapeake tidal, open water environments between 1984 and 2013.

Each sampling event was grouped by season and salinity zone and classified into one of four water quality categories using the classification thresholds for Secchi depth, DIN and PO_4 developed by Buchanan *et al.* (2005). The classification thresholds and the four water quality categories are described below. Gaps in the nutrient data prevented definite classification of 11,672 records and they were removed, leaving 52,528 records.

Reference conditions and populations

Chesapeake Bay Program (CBP) partners have qualitatively described in various inter-agency agreements and reports those Bay environments and designated uses they are striving to recover. The environments have nitrogen and phosphorus concentrations low enough to limit the formation of nuisance algal blooms, water column transparency clear enough to promote healthy growth of vascular plant (underwater grasses), and dissolved oxygen levels adequate for fish and bottom-dwelling communities (e.g., CBP 1987, 2000; USEPA 2003). These restoration goals are considered attainable under the present circumstance of a Bay watershed heavily influenced by humans. No longer attainable are the pre-Colonial water quality conditions, when water transparency was much deeper and the dominant primary producers were not planktonic algae (phytoplankton) but rather benthic algae (e.g., Cooper and Brush 1993) and more expansive beds of underwater grasses (e.g., Miller 1986).

Building on earlier research and data analysis results, Buchanan *et al.* (2005) developed quantitative thresholds to classify existing water quality conditions in Chesapeake Bay open waters (**Table 1**) and create distinct water quality categories relevant to phytoplankton (**Table 2**). The data are grouped into habitats defined by season and salinity zone (see above) to minimize the recognized influences of season and salinity on phytoplankton. The nutrient thresholds in **Table 1** are based on nutrient bioassays performed by Fisher and Gustafson (2003). They separate bloom-limiting and excess nutrient concentrations. The Secchi depth thresholds are from an application of the Relative Status Method to data from the 1985 – 1990 (spring and summer) and 1985 – 1999 (autumn and winter) monitoring periods as described in Buchanan *et al.* (2005) and Olson (2009). They generally separate adequate and inadequate water clarity conditions.

Phytoplankton communities in waters meeting all three thresholds (Better/Best water quality category in **Table 2**) are presumed to be the healthiest in the Bay at this time. They have consistently low and less variable total biomass, chlorophyll *a* and pheophytin (another photopigment). Their ratios of chlorophyll to biomass (Chla:C) are also low and less variable, indicating underwater light levels are high enough to avoid stressing cellular photosynthesis pathways. Their populations have relatively stable proportions of taxonomic groups, larger average cell sizes, and low biomasses of key bloom-forming taxa. Finally, median values for total biomass of the phytoplankton size fractions important to grazers (2

– 200 µm) are the same or higher than those in the degraded categories in 12 of the 16 season-salinity habitats, suggesting that the ongoing nutrient reductions will not "starve" grazers in the future.

Phytoplankton populations in the Mixed Better Light (MBL) water quality category (**Table 2**) prove to be good surrogates for those in the Better/Best category. Secchi depths in the MBL category meet their classification criteria but one or both nutrients fail their classification criteria. Phytoplankton photochemical, biomass, and taxonomic metrics in the Better/Best and MBL categories are indistinguishable in most cases (Buchanan *et al.* 2005, Lacouture *et al.* 2006, Johnson and Buchanan 2013). This is true even when samples in the MBL category have excess nitrogen *and* excess phosphorus concentrations. **Figures 1a - e** illustrates the chlorophyll *a* properties of phytoplankton in the Better/Best and MBL categories as compared to those in the degraded categories. Degraded categories have Secchi depths that fail their classification criteria and nutrient concentrations that fail one or both of their classification criteria.

Samples representing the Better/Best category in the 1984 – 2013 timeframe were rare in the tidal fresh and oligohaline and seasonally rare in the mesohaline and polyhaline (**Table 3**). For this reason, MBL populations are used in combination with the Better/Best populations in this analysis to represent reference conditions and develop the biological reference curves. Including the MBL category as a reference water quality condition increases sample numbers in each season and salinity zone and avoids giving unfair latitude to the reference classifications in tidal fresh and oligohaline salinity zones. However, when numbers of Better/Best category samples increase in response to ongoing nutrient and sediment load reductions to tidal waters, chlorophyll *a* concentrations in this category will best represent stable, desirable phytoplankton populations in a recovered Chesapeake Bay.

All open water designated uses appear to be supported in the conditions meeting reference classification thresholds for phytoplankton. Dissolved oxygen concentrations associated with Better/Best and MBL samples meet the Chesapeake Bay 30-day mean criteria for open waters often, i.e., $\geq$ 5.5 mg/liter for 0 – 0.5 ‰ salinities and $\geq$ 5.0 mg/liter for > 5.0 ‰ salinities. In spring, the success rate is more than 99.4% in samples from all salinity zones. In summer, success rates are lowest in mesohaline salinities (82.6%) and highest in oligohaline salinities (96.1%). Dissolved oxygen is $\geq$ 3.0 mg/liter in 99.2% of all samples, the exceptions being 43 mesohaline and 3 polyhaline summer samples. Water clarity criteria do not exist for open water environments of Chesapeake Bay. However, the classification thresholds used to delineate adequate Secchi depth for reference phytoplankton populations (**Table 1**) are approximately the same or higher than the original water clarity requirements for submersed aquatic vegetation, which were 0.8 m in tidal fresh and oligohaline salinities and 1.0 m in mesohaline and polyhaline salinities (Batiuk *et al.* 1992).

Virginia's chlorophyll criteria pertain only to the spring and summer seasons. Therefore, only the spring and summer reference data (9,415 records) were used in the biological reference curve analyses. The occurrence of MBL samples in tidal fresh and oligohaline waters and Better/Best and MBL samples in mesohaline and polyhaline waters is fairly evenly distributed and single stations or small groups of stations do not dominate in reference conditions.

Criteria attainment curves

Chlorophyll *a* criteria assessment procedures are described in Chapter 5 of USEPA (2008). CFD curves for Virginia James River assessments are currently generated from multiple data sources (routine shipboard sampling, DATAFLOW, calibration data) as follows:

1. Compile and QA/QC data set of chlorophyll *a* values for the 3-year assessment period.

2. Group data by date and segment.
3. Apply the CBP interpolation program and populate an assessment layer for each segment and sampling date with estimated chlorophyll *a* values (an assessment layer for chlorophyll *a* is the grid of surface water quality model cells in a segment).
4. For each interpolation cell, calculate a season-year arithmetic mean, or simple average, across all dates (**Figure 2**).
5. For each cell, determine if the season-year average violates the criteria.
6. Calculate the percent of all cells violating the criteria in the segment.
7. Determine the cumulative probability of the space violation rate (Weibull formula).
8. Construct a CFD.
9. If any point of this CFD crosses the reference curve, the segment is deemed "impaired."

Reference curves

The simple, symmetric 10% hyperbolic curve used as a default reference curve for chlorophyll *a* is calculated as follows:

$$(x+b) * (y+b) = a$$

where x is %space in violation, y is %time in violation, b = 0.0429945, and $a = b^2 + b$ (CBP 2007). One data analysis objective is to compare this default curve to various biological reference curves.

Biological reference curves should technically be developed in the manner described above in "criteria attainment curves" (CBP 2007). However, reference-quality conditions would have to occur throughout the spring and summer, across entire assessment units (segments) or salinity zones, over multiple years in order to create the necessary assessment layers. Actual reference conditions are sporadic and not widespread in the Bay at this time and sufficient samples to create these layers cannot be found. Two assumptions are made that overcome the lack of coverage and sample density, maintain some degree of year-to-year natural variability found in reference-quality samples, and allow development of biological reference curves:

Assumption 1 Reference samples collected in a particular habitat *in a given year* are assumed to represent the spatial distribution of chlorophyll *a* concentrations *in a single assessment layer* for that habitat type, regardless of which year the data were collected. For example, chlorophyll *a* concentrations in reference samples collected in mesohaline waters during summer 1992 are assumed to represent the spatial distribution of reference chlorophyll *a* concentrations across a single assessment layer for the summer mesohaline habitat. Likewise, chlorophyll *a* concentrations in reference samples collected in the segment JMSMH in summer 1992 are assumed to represent the spatial distribution of reference chlorophyll *a* concentrations across a single assessment layer for the JMSMH segment.

Assumption 2 The proportion of chlorophyll *a* concentrations failing a criterion in a particular habitat *in a given year* represents the percent of failures in *a single assessment layer* for that habitat. By extension, the failure rates of a criterion for multiple years are assumed to represent the temporal distribution of failure rates for the corresponding number of assessment layers. For example, 12 annual failure rates calculated from summer mesohaline data correspond to the failure rates of 12 assessment layers for that habitat.

CFDs representing biological reference curves can be constructed when these assumptions are applied to the pool of reference samples and the season-salinity-year results can be used as individual assessment layers. Up to 30 assessment layers can be assembled from the 1984 – 2013 monitoring data for the 8 season- salinity habitats (**Table 4**)

Three approaches were used to develop and investigate the properties of different biological reference curves. The steps involved in creating the different curves are outlined here.

Method 1 Treat all reference quality sampling events for a given salinity zone, season and year as representing an individual assessment layer and comparable to a layer created in step 3 of "criteria attainment curves" above. The percent of samples failing the season- and salinity-specific criteria of that salinity zone/season/year combination represents the violation rate of a single assessment layer. All years between 1984 and 2013 collected from Chesapeake Bay open water environments and containing more than two reference quality samples are utilized. For each season-salinity habitat:

1. Determine the percent of samples failing the chlorophyll *a* criteria in each assessment layer.
2. Sort the assessment layers from largest to smallest % failure rate (% space violation rate).
3. Determine the cumulative probability of failure (Weibull formula).
4. Construct a CFD curve.

A schematic of this method is shown in **Figure 3**.

Method 2 Data are treated as in Method 1 but only assessment layers with ten or more reference quality sampling events are considered. Analysis using Method 2 was only done for the summer mesohaline habitat which has the largest numbers of reference quality sampling events per year for all 30 years of the CBP monitoring program. Method 2 was designed to investigate the influence of number of assessment layers on the shape and position of a biological reference curve. For this season-salinity habitat:

1. Assessment layers are randomly selected, with replacement, from the actual pool of 30 assessment layers.
2. Criteria violation rates of these randomly selected layers are used to constructed 50 combinations of 24 layers, 50 combinations of 18 layers, 100 combinations of 15 layers, 100 combinations of 12 layers, 100 combinations of 9 layers, 100 combinations of 6 layers and 100 combinations of 4 layers.
3. Sort each individual combination from largest to smallest failure rate.
4. Calculate the average (mean) failure rate from the 50 or 100 values associated with each rank (% time violation).
5. Determine the cumulative probability of the average failure rate, as well as that of the 10^{th}%ile and 90^{th}%ile of the distribution around the average (Weibull formula).
6. Construct a CFD curve.

Method 3 Data are treated as in Method 1. Analysis using Method 3 was only done for summer mesohaline habitat. Method 3 was designed to investigate the influence of annual averaging (e.g., **Figure 2**) on the shape and position of the CFD curves. The average failure rate (% space violation) of a given rank (% time violation) is assumed to represent the failure rate for one summer month of one year in an assessment period. For this season-salinity habitat:

1. Return to the 100 randomly selected combinations of 15, 12, and 9 assessment layers produced in Method 2 step 2.
2. For each unique combination of 15 assessment layers, calculate an annual average from 3 randomly selected layers, calculate another annual average from 3 more randomly selected layers, and so forth; continue until 5 annual averages have been calculated for that particular combination of 15 assessment layers; repeat for all 100 combinations. The result is 100 series of

5 annual averages. These are intended to represent possible annual average in a 5-year assessment period.

3. Repeat step 2 on the 100 random selected combinations of 12 assessment layers and construct 100 series of 4 annual averages. These are intended to represent possible annual averages in a 4-year assessment period.
4. Repeat step 2 on the 100 random selected combinations of 9 assessment layers and construct 100 series of 3 annual averages. These are intended to represent possible annual averages in a 3-year assessment period.
5. Sort the annual averages in the 3-, 4-, and 5-year series from largest to smallest failure rate.
6. Calculate the average (mean) failure rate from the 100 values associated with each rank in an assessment period.
7. Determine the cumulative probability of the average failure rate, as well as that of the 10^{th}%ile and 90^{th}%ile of the distribution around the average (Weibull formula).
8. Construct a CFD curve for each of the assessment periods.

Chlorophyll *a* criteria tested

Three sets of chlorophyll *a* criteria were tested (**Table 5)**. Two are based on the 90^{th} and 95^{th} percentiles of all reference quality samples (i.e., Better/Best and MBL water quality categories combined) in each season-salinity habitat. The third set is the existing James River chlorophyll *a* criteria. The 90^{th}ile and 95^{th}ile were chosen as test criteria because the US Environmental Protection Agency recommends an approximately 10% allowable criteria exceedance if a default CFD reference curve is used (USEPA 2003).

The above-pycnocline salinity measured at the time of sampling decides which 90^{th}ile and 95^{th}ile criterion applies to a given sampling event, regardless of where the station is located. This approach differs from CBP procedures which use segment-specific rather than salinity-specific criteria. The James River segment JMSMH, for example, is designated mesohaline and chlorophyll *a* data collected in that segment are assessed only with the mesohaline criteria of 12 (spring) and 10 (summer). However, the segment's routine monitoring stations (LE5.2, LE5.3) experience salinities between 0.2‰ to 26.4‰. For the purpose of developing biological reference curves, the James River segment-specific criteria were used as if they were salinity-specific criteria. The James River polyhaline segment criteria of 12 (spring) and 10 (summer) were applied to all samples in the reference data set associated with >18‰ salinity and not simply to CBP segments designated as polyhaline (PH). Similarly, James River segment-based criteria for mesohaline, oligohaline, and tidal fresh were applied according to the salinity measured at time of sampling and not simply to CBP segments designated as mesohaline (MH), oligohaline (OH) and tidal fresh (TF), respectively.

Results

1. What nutrient concentrations occur in reference water quality conditions?

Reference-quality waters are defined for this analysis as the Better/Best and MBL water quality categories. The Better/Best category has water clarity adequate for unstressed phytoplankton photosynthesis and nutrient concentrations known to be low enough to limit bloom formation in open water environments (i.e., $\leq$ 0.07 mg DIN/liter, $\leq$ 0.007 mg PO_4/liter). The MBL category, which is used as a surrogate for reference quality conditions to increase sample numbers for the analysis, has adequate water clarity but one or both nutrients are above bloom-limiting concentrations. MBL is considered reference-quality because its phytoplankton communities are essentially indistinguishable from those in the Better/Best category.

Analysis of nutrient concentrations in the MBL category shows the following:

- when one of the two nutrients is limiting in the MBL category, it is most often PO_4 (**Table 6**);
- the exceptions are summer mesohaline and polyhaline which are mostly limited by DIN;
- when both nutrients are present in excess concentrations in the MBL category, PO_4 is rarely greater than 10x (4.3% of all samples) and usually less than 5x (80.6% of all samples) the bloom-limiting concentration of 0.007 mg PO_4/liter; and
- when both nutrients are present in excess concentrations in the MBL category, DIN is less than 5x the bloom-limiting concentration of 0.07 mg/liter in about 50% of all samples and greater than 10x of the threshold in about 25% of all samples.

The fact that neither DIN nor PO_4 is hugely greater than its bloom-limitation threshold when both are present in excess amounts suggests additional phytoplankton growth in these particular MBL samples would have been limited soon by one of the nutrients.

2. Can biological reference curves be developed from chlorophyll a concentrations observed in reference water quality conditions? Yes.

Figures 4a - h show the biological reference curves produced with Method 1 when the season- and salinity-specific 90^{th}%ile and the 95^{th}%ile criteria (**Table 5**) are applied to all available reference-quality samples from Chesapeake open water environments. Assessment layers with as few as 3 samples per layer were used in this analysis, so some bias in the CFD curves due to small sample sizes is expected. The CFDs produced with the 90^{th}%ile and 95^{th}%ile criteria generally follow the default 10% hyperbolic curve. Percentiles *between* the 90^{th}%ile and 95^{th}%ile produce CFDs more closely overlay the 10% hyperbolic curve (not shown). Limiting the curves to assessment layers with 10 or more samples did not greatly change the general shape or position of most curves, although the bias created by the layers with small sample sizes became more evident.

3. Do James River chlorophyll criteria produce reference curves that follow the default 10% hyperbolic curve? Sometimes.

Figures 4a – h also show the biological reference curves produced when James River criteria are applied using Method 1. When James River criteria range between the 90^{th}%ile and 95^{th}%ile criteria, they produced CFDs that closely follow the default 10% hyperbolic curve, i.e., spring tidal fresh (upper, lower), summer tidal fresh (lower), and summer oligohaline. James River criteria for the spring and summer polyhaline are higher than their corresponding 95^{th}%ile criteria and their CFD curves fall noticeably below the 10% hyperbolic curve. This suggests the James River polyhaline criteria may be somewhat under-protective of reference conditions. James River criteria for summer tidal fresh (upper), spring oligohaline, and spring and summer mesohaline are to varying degrees lower than their corresponding 90^{th}%ile criteria and their CFD curves are noticeably above the 10% hyperbolic curve. These criteria may be somewhat over-protective of reference conditions.

The James River criteria correspond to the following percentiles of the reference chlorophyll data in the corresponding season-salinity zone (asterisk * indicates the CFD curve approximates the 10% hyperbolic curve):

Spring TF upper 89.1^{st}%ile*

	TF lower	95.5^{th}%ile*
	OH	73.8^{th}%ile
	MH	82.8^{th}%ile
	PH	97.3^{rd}%ile
Summer	TF upper	86.6^{th}%ile
	TF lower	94.0^{th}%ile*
	OH	94.4^{th}%ile*
	MH	81.0^{nd}%ile
	PH	97.8^{th}%ile

Hyperbolic curves can be formulated to specifically overlay biological reference curves derived with the existing James River chlorophyll criteria. These hyperbolic curves have percents of allowable exceedances ranging from 26.2% to 2.2% instead of the default allowance of 10%.

4. Does the number of individual assessment layers affect the shape and position of the biological reference curve? Yes, when there are fewer than 9 assessment layers.

Figures 5a – c show biological reference curves developed from 24, 15, 12, and 9 individual assessment layers and **Figures 6a – c** show biological reference curves developed from 6 and 4 individual layers. Reference data for the summer mesohaline habitat were used in the analysis because this habitat had the most data. (We assume the other season-salinity habitats will behave the same.) All curves were developed using Method 2. The James River (10 µg/liter), 90^{th}%ile (11.89 µg/liter) and 95^{th}%ile (14.22 µg/liter) criteria were each tested. When 9 or more individual assessment layers are used to create the CFDs, the number of assessment layers does not appear to influence the overall shape and position of the curves relative to the observed 30-layer curve. When fewer than 9 assessment layers are used (i.e., 6 or 4), the points on the CFD curve in the middle and lower right corner begin to pull up and away from the observed 30-layer curve into the non-compliance zone. This is likely an artifact reflecting the statistical limitations of using few assessment layers to construct the CFD curves.

5. Do annual averages (each calculated from 3 assessment layers) produce biological reference curves comparable to those constructed from individual assessment layers? To varying degrees, no.

Figure 7 uses a hypothetical example to illustrate the effect of annual averaging on a CFD curve. In the example, nine assessment layers – one for each month-year – are generated for a 3-year assessment window (e.g., March-Yr1, April-Yr1, May-Yr1, March-Yr2, … May-Yr3) and are used to calculate three annual means. Both CFD curves are plotted in the figure's graph. Distributions of chlorophyll *a* values are typically log-normal meaning the mean value of several measurements is higher than most of the measurements. This is the case in the hypothetical example. Year 3 has the lowest annual mean (14.2%) so it is positioned at %time = 75% on the 3-point, annual mean CFD curve. Two of the three monthly values making up Year 3's average are lower than 14.2%. They are in fact the two lowest monthly values and as such are plotted at %time = 90% and 80% on the monthly, 9-point CFD curve. Their low values coupled with their high positions on the y-axis locate them closer to the reference curve (default 10% hyperbolic curve) than their corresponding annual mean on the 3-point, annual mean CFD curve. Similarly, Year 2 has the highest annual mean (37%) and is positioned at %time = 25% on the 3-point, annual mean CFD curve. Only one of the three monthly values making up Year 2's average is higher than 37% while the other two values are lower. So, Year 2's mean value ends up closer to the reference curve than the monthly, 9-point CFD curve.

This pattern of bias is found in the summer mesohaline biological reference curves developed with 3-, 4- and 5-year annual means (Method 3). When compared to the CFD curves made from the original 9, 12 and 15 assessment layers, the annual mean curves are biased outward toward non-compliance in the upper left "bottleneck" area and biased inward toward compliance in the lower right side of the curve (**Figure 8**). Annual averaging essentially compresses the natural variability seen over time in the data. The bias is an artifact of averaging the individual assessment layers. Its affect diminishes for criteria values closer to the 90th%ile – 95th%ile, whose biological reference curves approach the 10% hyperbolic curve (e.g., compare the James River criteria with the 95th%ile criteria results in summer mesohaline habitats).

6. Can criteria exceedances rate be affected by the number of interpolator cells used to generate an assessment layer? Yes.

This can be demonstrated with an example. For a 3-year assessment period, a reference curve "bottleneck" occurs at %space (x-axis) = 1.355% and %time (y-axis) = 75% (**Figure 7**). Assume a segment has 100 interpolator cells and, in the year having the lowest annual average, one cell in one month of the year fails and no cells fail in the other two months.

March of Low Year - 1% (1 of 100 cells fail)
April of Low Year - 0% (0 of 100 cells fail)
May of Low Year – 0% (0 of 100 cells fail)

This produces an annual average failure rate of 0.333% from a total of 300 evaluated cells. While the lowest possible failure rate is 0%, the second lowest possible rate for a 100-cell segment and three monthly interpolations would be 1/300 or 0.333%. This failure rate will be in compliance (not exceed the reference curve) at %time = 75%, which is the y-axis position of the year having the lowest annual average in a 3-year assessment window.

Now, let's say that the segment had 24 interpolator cells and once again one cell in one month of the year fails and no cells fail in the other two months.

March of Low Year – 4.167% (1 of 24 cells fail)
April of Low Year - 0% (0 of 24 cells fail)
May of Low Year – 0% (0 of 24 cells fail)

This produces an annual average failure rate of 1.389% from a total of 72 evaluated cells. This failure rate is positioned to the right of the 3-year assessment reference curve point at %time = 75% and thus is slightly out of compliance. For a 3-year assessment period, if a segment has fewer than 72 interpolator cells evaluated in a season (fewer than 24 cells per month) then the resulting CFD curve is destined to fail at the bottleneck. Likewise, if two cells fail in one month and no cells fail in the other two months, then the season-year having the lowest annual average will fail at the bottleneck if the segment has fewer than 144 cells per season (48 cells per month).

The minimum number of interpolator cells per segment required for the year with the lowest annual average to simply pass through the bottleneck changes as the assessment period lengthens. For a 4-year assessment period, the bottleneck occurs at %time (y-axis) = 80% and corresponds to %space in violation (x-axis) = 1.020%. A minimum of 98 interpolator cells per segment per season (or 33 cells per segment per month) are required to pass the reference curve bottleneck if one cell in any month fails.

For a 5-year assessment period, the bottleneck occurs at %time values = 83.3% and corresponds to %space = 0.818%. A minimum of 123 interpolator cells per segment per season (or 41 cells per segment per month) are required to pass the reference curve bottleneck if one cell in any month fails.

7. Is there a relationship between mean chlorophyll a and the frequency of criteria exceedance regardless of how the data are grouped? Yes.

The 1984 – 2013 CBP monitoring data for chlorophyll *a* from all water quality conditions in open water environments were used to address this question. Data were grouped three ways: season – station (a), season – CBP segment – year (b), and season – salinity zone – year (c). Each sample was scored if it exceeded 10, 20, or 30 µg/liter. The means and frequencies of exceedance of the 10, 20, and 30 µg/liter thresholds were then determined for each group. Relationships between the mean and the frequency of exceeding the three thresholds were tight and often nearly identical regardless of how the data were grouped. **Figure 9** shows the relationships across all Chesapeake Bay tidal waters. Coefficient of determination (r^2) values ranged between 0.85 and 0.97. **Figure 10** shows the (b) and (c) relationships for just the James River.

Discussion

Reference water quality conditions as defined for this study have water clarity adequate for unstressed photosynthesis, as indicated by low, stable Chla:C ratios. Reference condition concentrations of two key nutrients, DIN and PO_4, can exceed the bioassay thresholds known to limit algal bloom formation (Fisher and Gustafson 2003). However, when this occurs one or the other of these two nutrients is usually limiting, or the quantities of excess PO_4 are comparatively low. It is important to recall that these reference conditions represent the best available at this time. Ongoing nutrient and sediment load reductions to tidal waters should begin to increase numbers of samples meeting all three classification criteria. Chlorophyll *a* concentrations in that category (Better/Best) will most accurately represent stable, desirable phytoplankton populations in a recovered Chesapeake Bay.

Biological reference curves for chlorophyll *a* can clearly be developed from phytoplankton populations inhabiting reference quality conditions in open water environments. Criteria values between the 90th and 95th percentiles of chlorophyll *a* concentrations in present-day reference conditions generate CFD curves that closely approximate the default 10% hyperbolic reference curve for the two seasons and four salinity zones assessed by Virginia. This indicates the 10% hyperbolic curve is a reasonable representation of the natural spatial and temporal extent of algal blooms in Chesapeake Bay under reference conditions. Another CFD analysis based on 1960s Chesapeake chlorophyll *a* data shows a similar closeness to the 10% hyperbolic curve (Curve 3 in Figure 4.1, Appendix A, USEPA 2007). The analysis used the observed means and 90th%iles of chlorophyll *a* in the different habitats and made certain assumptions about spatial and temporal variances.

In general, all of the existing James River chlorophyll *a* criteria are protective of high quality habitat conditions. James River criteria values approximating the 90th to 95th percentile values in reference populations will produce biological reference curves that closely follow the default 10% hyperbolic curve. James River criteria values lower than the reference 90th%ile may be slightly over-protective of reference conditions; criteria values larger than the 95th%ile may be slightly under-protective of reference conditions. This is not to say that the James River criteria which differ from the 90th – 95th percentile range are *not* protective against impairment. It only means they are more or less protective of the reference water quality conditions that support the healthiest phytoplankton populations presently found in Bay tidal waters.

Several avoidable biases and artifacts were found in the biological reference curves.

- When just six or four assessment layers are used to construct biological reference curves, points on the middle and lower right portions of the CFD curve tend to shift toward the non-compliance zone. This is in comparison to curves constructed with nine or more layers. It may be advisable for Virginia to use nine or more assessment layers to build attainment CFD curves if a 10% hyperbolic curve serves as the default reference.
- The shape of the biological reference curve changes when it is constructed from 3, 4, or 5 annual averages instead of the 9, 12, or 15 corresponding monthly assessment layers. Points in the upper left region of the 3, 4, and 5 point CFD curves shift toward the non-compliance zone whereas points in the lower right region shift towards the compliance zone.
- The reference curve bottleneck (upper left region of the CFD curve) can force out of compliance segments having few interpolator grids, even if their criteria failure rates are very low. This artifact of the interpolator approach may be avoided if individual (monthly) assessment layers are used in Virginia assessments rather than annual averages.

It is apparent in the results that monitoring data with more intensive spatial and temporal coverage will improve the CFD approach. CFDs constructed from higher spatial and temporal density data will be less subject to bias and artifact. Regardless, chlorophyll *a* criteria somewhere between the overall 90^{th} and 95^{th} percentiles of each season and salinity zone habitat will most closely parallel the default 10% hyperbolic curve. More stringent selection of the reference quality samples from which the reference curves are developed could further align the biological reference curves and the 10% hyperbolic curve.

Strong relationships occur between the mean chlorophyll *a* concentration and the frequency of exceeding chlorophyll *a* threshold concentration, regardless of how the data are analyzed. The same mean concentration of chlorophyll *a* at a station, in a CBP segment, or in an entire salinity zone moving longitudinally with wind and flow over time appears to have about the same probability of exceeding a specific threshold. Documenting the relationships may appear tangential to the development of biological reference curves. However, a concern was that salinities higher or lower than a segment's designated salinity would affect the segment's exceedance frequencies. Tidal waters are in no way bound to the salinity designations of CBP segments. Severe droughts shift oligohaline waters well into segments designated as tidal fresh and large storms push low salinity waters into segments designated as polyhaline. Another concern was the James River's shallow bathymetry and its possible effect on exceedance frequencies.

Finding the relationships has several implications. First, the strong similarities between relationships in different data groupings indicate it was acceptable to use the James River criteria as salinity-specific in developing and characterizing biological reference curves. When used as salinity-specific criteria, James River criteria are applied according to the salinity observed at the time of sampling. This is how the 90^{th}%ile, 95%ile, and James River criteria were used to develop the biological reference curves. When used as segment-specific criteria, the criterion applied to a sample is decided by the salinity designation of the sampling station's segment. Second, the relationships might assist in future chlorophyll *a* assessments. Assessment units with questionable exceedance frequencies can be check against the expected frequencies calculated from established relationships between the mean and the frequency of exceeding specific chlorophyll *a* criteria. Third, and perhaps most importantly, the relationships demonstrate that exceedance frequencies are not strongly controlled by salinity or locational features.

This further support the idea that chlorophyll *a* concentrations and bloom frequencies are most strongly controlled by the water quality conditions surrounding the phytoplankton population.

Recommendations

The report's findings suggest the following changes to the existing Virginia Department of Environmental Quality procedures for assessing chlorophyll *a* criteria:

1) Continue to use simple, symmetrical hyperbolic curves as default reference curves in chlorophyll *a* criteria assessments, but adjust these curves so that the percent of allowable exceedances in each assessment unit reflects the biological reference curve derived with the James River criteria.

2) Increase spatial and temporal coverage with the use of data from other sampling technologies (e.g., DATAFLOW, satellite imaging, and continuous monitoring buoys) and improve the accuracy of interpolated chlorophyll assessment layers.

3) Given that the existing WQS require seasonal means in Virginia chlorophyll a assessments, use a longer assessment period and develop reference curves derived from seasonal means. This minimizes the biases introduced into the assessments by too few data layers and seasonal averaging.

4) Develop confidence intervals for the hyperbolic reference curves that account for statistical biases and artifacts inherent to the assessment method. Establish rules for assessing attainment rates that fall outside the reference curves but inside these confidence intervals.

Literature cited

Batiuk, R. A., P. Heasley, R. Orth, K. Moore, J. C. Stevenson, W. Dennison, L. Staver, V. Carter, N. B. Rybicki, R. Hickman, S. Kollar, S. Bieber and P. Bergstrom. 1992. Chesapeake Bay Submerged Aquatic Vegetation Habitat Requirements and Restoration Targets: A Technical Synthesis. CBP/TRS 83/92. U.S. EPA Chesapeake Bay Program, Annapolis, Maryland.

Buchanan, C., R. V. Lacouture, H. G. Marshall, M. Olson, and J. M. Johnson. 2005. Phytoplankton Reference Communities for Chesapeake Bay and its Tidal Tributaries. Estuaries 28(1):138-159.

Chesapeake Bay Program (CBP). 1987. 1987 Chesapeake Bay Agreement. Chesapeake Bay Program Office, Annapolis, MD. Available online at www.chesapeakebay.net.

Chesapeake Bay Program (CBP). 2000. Chesapeake 2000 Agreement. Chesapeake Bay Program Office, Annapolis, MD. Available online at www.chesapeakebay.net.

Cooper, S. R. and Brush, G. S. 1993. A 2,500-year history of anoxia and eutrophication in Chesapeake Bay. Estuaries 16(3):617-626.

Fisher, T. R. and A. B. Gustafson. 2003. Nutrient-addition bioassays in Chesapeake Bay to assess resources limiting algal growth. Progress report: August 1990 – December 2002. Prepared for Maryland

Department of Natural Resources, Chesapeake Bay Water Quality Monitoring Program, by University of Maryland Horn Point Laboratory, Cambridge, MD.

Johnson, J. M. and C. Buchanan. 2014. Revisiting the Chesapeake Bay Phytoplankton Index of Biotic Integrity. Environmental Monitoring and Assessment 186:1431-1451.

Lacouture, R. V., J. M. Johnson, C. Buchanan, H. G. Marshall. 2006. Phytoplankton Index of Biotic Integrity for Chesapeake Bay and its tidal tributaries. Estuaries 29(4):598-616.

Miller, H. M. 1986. Transforming a "Splendid and Delightsome Land": Colonists and Ecological Change in the Chesapeake 1607 - 1820. Journal of the Washington Academy of Sciences 76(3):173-187.

Olson, M. 2009. Relative Status Indicator. Development and Evolution of a Relative Measure of Condition for Assessing the Status of Water Quality and Biological Parameters Tracked in the USEPA Chesapeake Bay Program Long Term Monitoring Programs, Final Report. Prepared for the Interstate Commission on the Potomac River Basin. ICPRB Report 09-4. Available online at www.potomacriver.org.

Robertson, T. 2014. Chlorophyll *a* Assessment: a General Overview. Slide 14 of presentation made to the James River Chlorophyll Criteria Scientific Advisory Panel on 11 April, 2014.

U.S. Environmental Protection Agency (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. Region III Chesapeake Bay Program Office, Annapolis, MD. Available online at www.chesapeakebay.net.

U.S. Environmental Protection Agency (USEPA). 2007. Ambient Water Quality Criteria for Dissolved Oxygen, Water Clarity and Chlorophyll *a* for the Chesapeake Bay and its Tidal Tributaries. 2007 Addendum. EPA 903-R-07-003. Region III Chesapeake Bay Program Office, Annapolis, MD. Available online at www.chesapeakebay.net.

U.S. Environmental Protection Agency (USEPA). 2008. Ambient Water Quality Criteria for Dissolved Oxygen, Water Clarity and Chlorophyll *a* for the Chesapeake Bay and its Tidal Tributaries. 2008 Technical Support for Criteria Assessment Protocols Addendum. EPA 903-R-08-001. Region III Chesapeake Bay Program Office, Annapolis, MD. Available online at www.chesapeakebay.net.

Virginia Department of Environmental Quality (VADEQ). 2004. Chlorophyll *a* Numerical Criteria for the Tidal James River, November 30, 2004 (revised 1/12/2005). Technical Report prepared by Virginia Department of Environmental Quality.

Table 1. Classification thresholds used to delineate adequate water clarity (Secchi depth) for phytoplankton and concentrations of dissolved inorganic nitrogen (DIN) and ortho-phosphate (PO_4) that limit the formation of nuisance algal blooms. Seasons: spring (March – May); June*; summer (July – September); autumn (October – November); winter (December – February). Salinity zones: TF, tidal fresh (≤0.5 ‰); OH, oligohaline (>0.5 - 5.0 ‰); MH, mesohaline (>5.0 - 18.0 ‰); PH, polyhaline (>18.0 ‰). From Buchanan *et al.* (2005). * June was not included in the original classification scheme. It was subsequently added and assigned the summer thresholds.

	Spring	June	Summer	Autumn	Winter
Secchi depth (m)					
TF	>0.9	>0.8	>0.8	>0.9	>0.6
OH	>0.7	>0.6	>0.6	>0.5	>0.6
MH	>1.8	>1.45	>1.45	>2.0	>1.8
PH	>2.15	>1.85	>1.85	>2.5	>2.3
DIN (mg/liter)	≤0.07 (all seasons and salinity zones)				
PO_4 (mg/liter)	≤0.007 (all seasons and salinity zones)				

Table 2. Water quality categories. See Table 1 for classification thresholds.

Category name	Description
Better/Best	meets all thresholds for Secchi, DIN, & PO_4
Mixed Better Light (MBL)	meets Secchi threshold, fails DIN and/or PO_4 threshold
Mixed Poor Light (MPL)	fails Secchi threshold, meets DIN and/or PO_4 threshold
Poor/Worst	fails all thresholds for Secchi, DIN, & PO_4

Table3. Number of water quality sampling events in each season-salinity habitat category of the Chesapeake open water environment (i.e., ≥ 2 meter depth) after data preparation. All sampling events have Secchi depth and chlorophyll *a* measurements, can be associated with a salinity zone, and can be definitely classified into one of the four water quality categories. Total number of sampling events is 52,528. See Tables 1 and 2 headings for details.

Salinity zone	Water quality category	Spring	June	Summer	Autumn	Winter
TF	Better/Best	0	0	1	1	1
	MBL	350	168	523	229	433
	MPL	723	240	619	191	121
	Poor/Worst	2630	825	1781	1018	1397
OH	Better/Best	1	1	20	14	0
	MBL	275	120	469	389	182
	MPL	385	100	475	177	118
	Poor/Worst	800	291	663	346	471
MH	Better/Best	100	96	792	99	93
	MBL	1061	396	1007	293	589
	MPL	5117	1852	3950	1767	2287
	Poor/Worst	1071	490	2485	1464	796
PH	Better/Best	132	27	79	63	194
	MBL	203	95	622	139	158
	MPL	966	504	1787	1185	1333
	Poor/Worst	247	165	1158	865	253

Table 4. Years in the reference water quality data set that can represent assessment layers for a given season-salinity habitat and their corresponding sample numbers. **Bold black text** with gray highlight indicates years with greater than 10 samples. Some analyses used only these samples; others included all season-salinity habitats with more than 2 samples. Season-salinity habitats with 1 or 2 samples are considered insufficient to calculate a meaningful %failure of a criterion (red text).

Assessment Layer (Year)	Spring TF	Spring OH	Spring MH	Spring PH	Summer TF	Summer OH	Summer MH	Summer PH
1984	6	0	0	0	3	**12**	**46**	**20**
1985	6	0	**61**	**28**	**20**	4	**50**	**59**
1986	**16**	6	**103**	**21**	**20**	6	**81**	**34**
1987	**10**	7	**64**	6	**13**	**10**	**113**	**73**
1988	**12**	8	**98**	8	**17**	**15**	**101**	**35**
1989	7	6	**84**	**27**	**16**	**18**	**53**	8
1990	7	2	**31**	**18**	**22**	**12**	**61**	**34**
1991	**26**	**10**	**98**	**30**	**18**	**19**	**68**	**54**
1992	**16**	7	**76**	**38**	**28**	**17**	**83**	**34**
1993	7	**22**	**32**	7	**29**	**29**	**75**	**23**
1994	9	**44**	**71**	6	**20**	**17**	**76**	**28**
1995	8	5	**56**	**32**	**12**	**26**	**68**	**45**
1996	**10**	**22**	**35**	5	6	**12**	**45**	4
1997	**18**	9	**33**	2	**16**	**16**	**57**	**29**
1998	8	**36**	**12**	3	**11**	**25**	**81**	**12**
1999	**28**	7	**41**	5	**34**	**16**	**62**	**17**
2000	**27**	3	**18**	**10**	**41**	**16**	**56**	9
2001	**14**	1	**38**	**22**	**21**	**20**	**95**	**31**
2002	**15**	3	**28**	**34**	**23**	**26**	**77**	**66**
2003	2	**12**	**17**	1	**11**	**19**	**41**	4
2004	1	**10**	**24**	4	**15**	**22**	**51**	5
2005	**13**	6	**17**	0	**19**	**18**	**44**	9
2006	**12**	2	**11**	1	5	6	**23**	5
2007	**19**	8	7	0	**17**	**32**	**68**	9
2008	**10**	**10**	**19**	1	**28**	**32**	**50**	6
2009	**15**	6	**42**	**14**	**28**	**20**	**81**	3
2010	8	7	**18**	8	**10**	**10**	**24**	**17**
2011	7	**11**	**12**	1	8	4	**10**	1
2012	7	4	**13**	2	**10**	5	**33**	**23**
2013	6	2	2	1	3	5	**26**	4

Table 5. Chlorophyll *a* criteria investigated. The 90[th] and 95[th] percentile criteria (rounded to 2 decimals) are calculated from all reference quality samples in each season-salinity habitat (i.e., Better/Best and MBL categories combined). The James River tidal fresh is divided into two segments with different criteria for assessments: the JMSTF1 (upper) segment between Richmond and Hopewell, and the JMSTF2 (lower) segment between Hopewell and the JMSOH boundary. See text for details of how criteria are applied to chlorophyll *a* measurements.

	TF	OH	MH	PH
90th%ile criteria				
Spring	10.39	22.54	15.49	7.94
Summer	16.93	17.23	11.89	7.37
95th%ile criteria				
Spring	13.53	28.61	22.5	10.05
Summer	24.16	23.75	14.22	8.65
James River criteria (upper, lower)				
Spring	10, 15	15	12	12
Summer	15, 23	22	10	10

Table 6. Analysis of the Mixed Better Light (MBL) category samples where one nutrient concentration is bloom-limiting and the other is not. Bloom-limiting concentrations are 0.07 mg DIN/liter and 0.007 mg PO_4/liter (from Fisher and Gustafson 2003).

	Spring		Summer	
Salinity zone	% of samples limited by DIN	% of samples limited by PO_4	% of samples limited by DIN	% of samples limited by PO_4
TF	0.00%	100.00%	9.84%	90.16%
OH	0.00%	100.00%	16.42%	83.58%
MH	0.13%	99.87%	59.91%	40.09%
PH	21.21%	78.79%	92.95%	7.05%

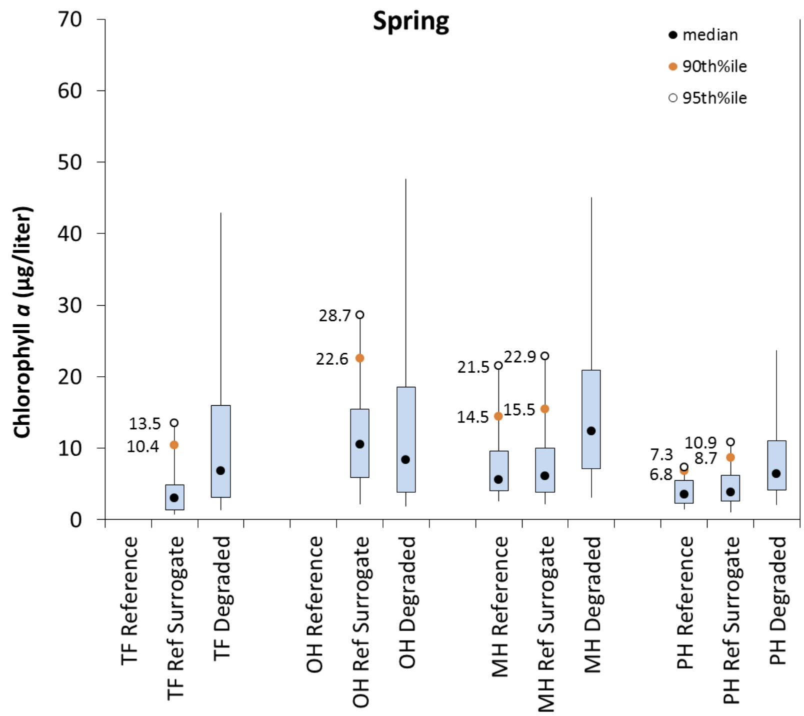


Figure 1a. Comparisons of spring (March – May) chlorophyll *a* concentrations in the salinity-specific Better/Best categories ("Reference"), Mixed Better Light categories ("Ref Surrogate"), and combined Mixed Poor Light and Poor/Worst categories ("Degraded") for all Chesapeake open water habitats, 1984-2013. Box, 25th – 75th percentile; whiskers, 5th – 95th percentile. Values of 90th and 95th percentiles are shown for Reference and Ref Surrogate categories. Habitats with < 20 samples are not shown.

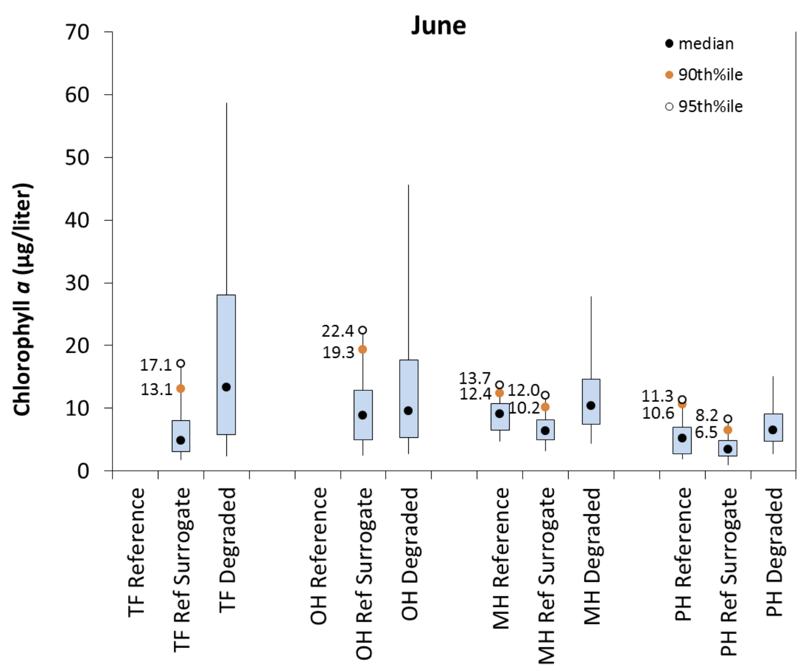


Figure 1b. June chlorophyll *a* concentrations.

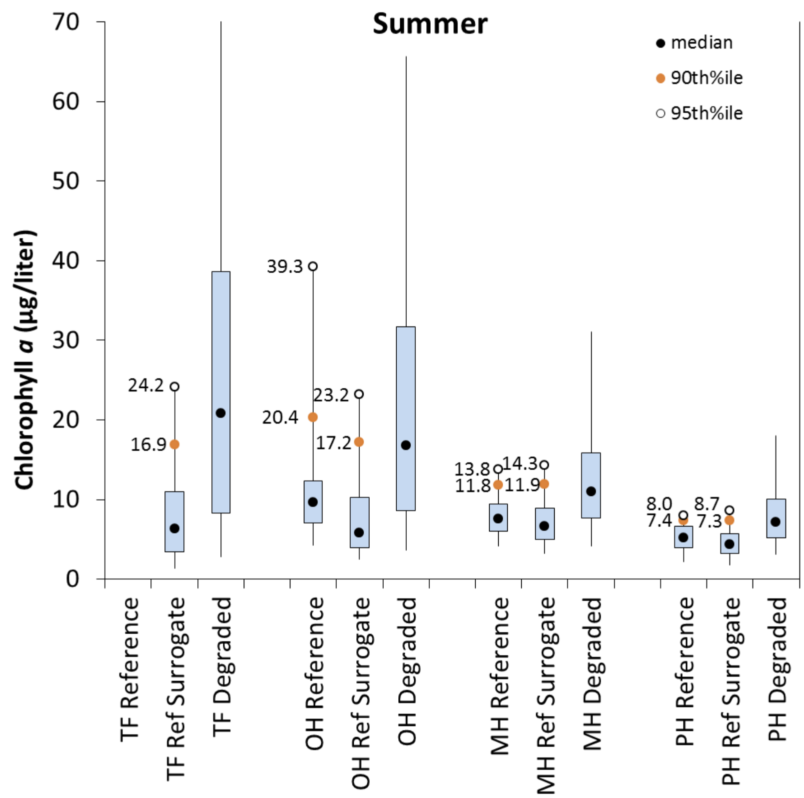


Figure 1c. Summer (July – September) chlorophyll *a* concentrations.

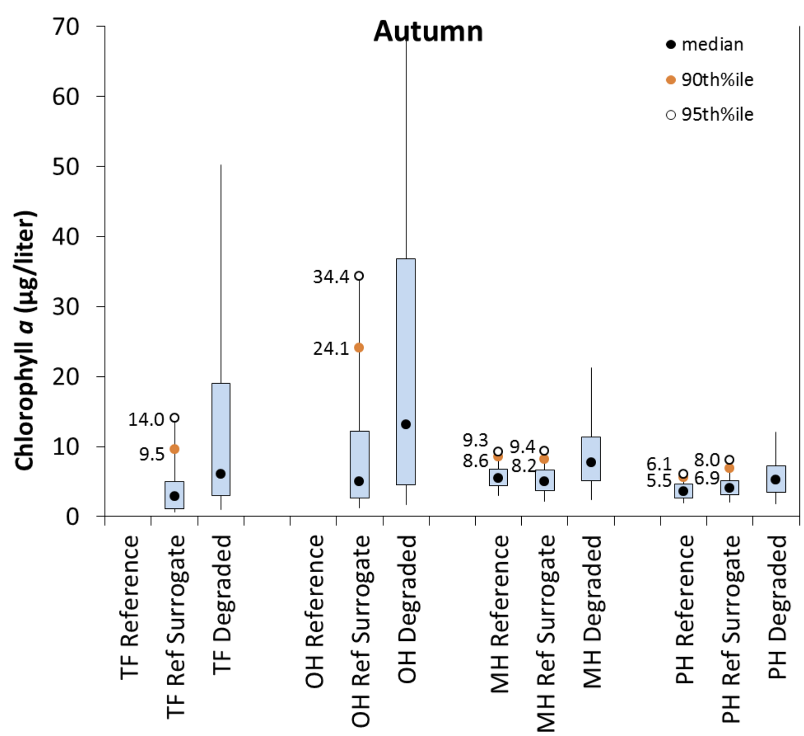


Figure 1d. Autumn (October – November) chlorophyll *a* concentrations.

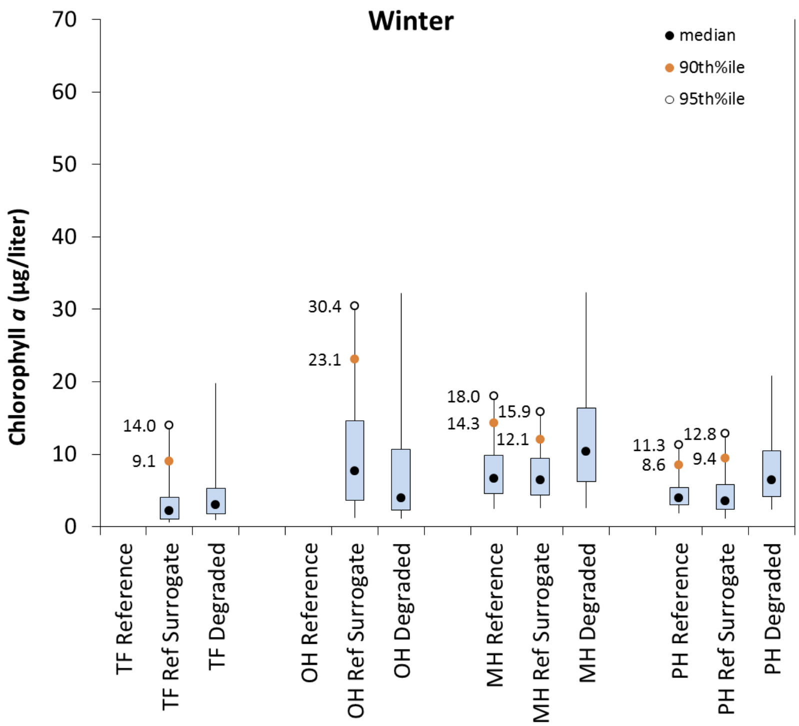


Figure 1e. Winter (December – February) chlorophyll *a* concentrations.

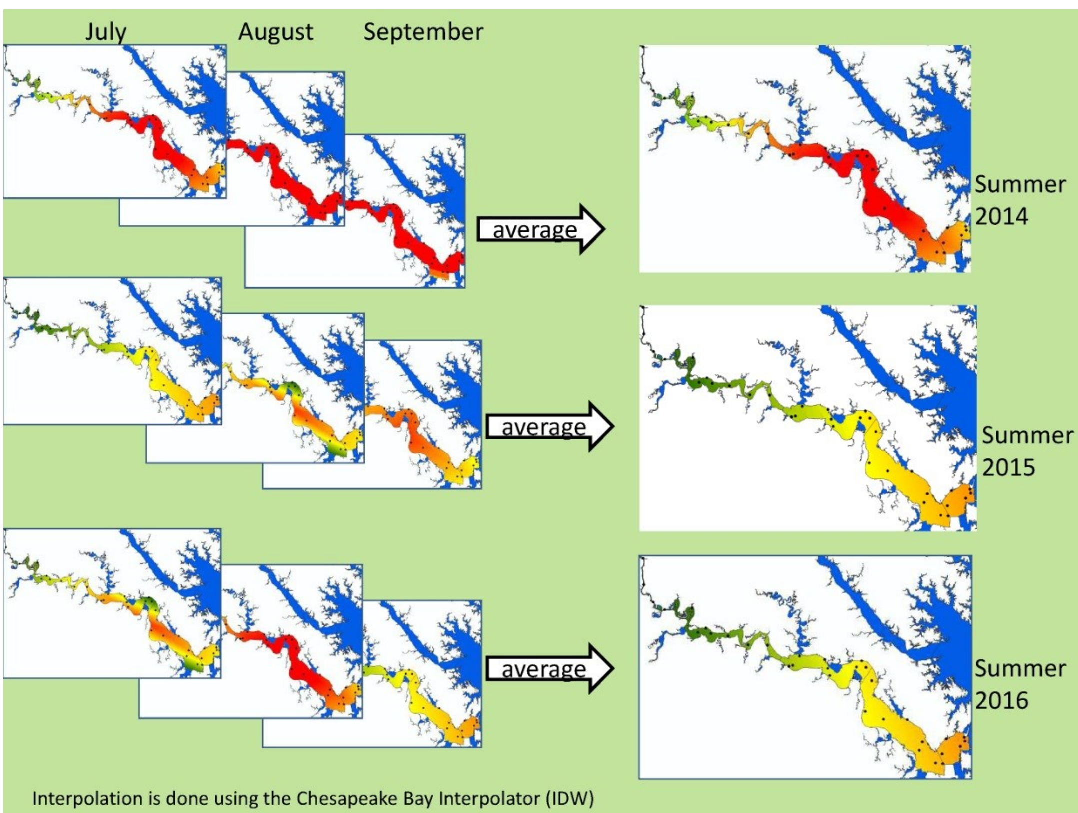
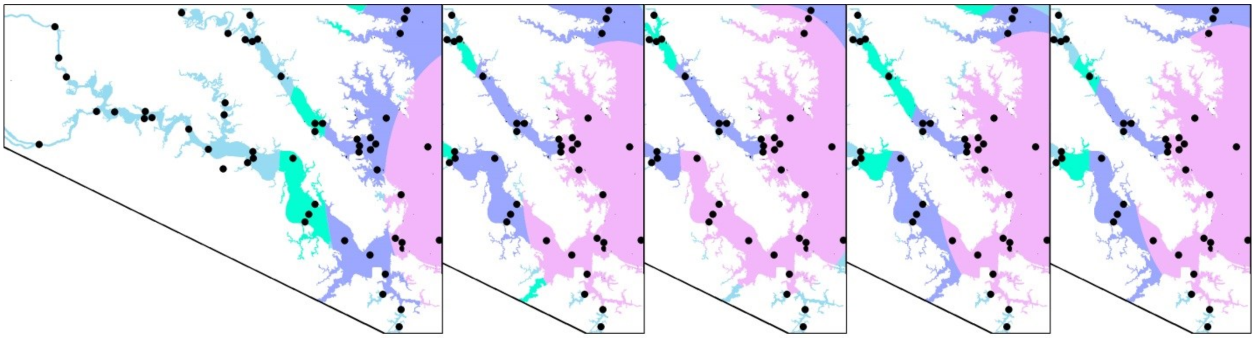
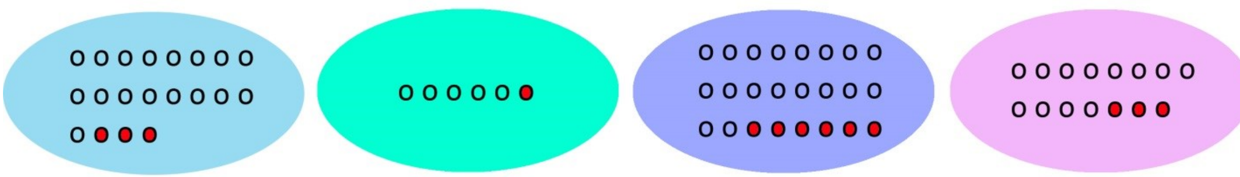


Figure 2. Illustration of Virginia method of annual averaging from monthly chlorophyll *a* interpolations (from T. Robertson 2014).

A) Individual cruise periods in one season-year, bay wide (all samples)



B) One season-year (all samples from reference quality conditions regardless of cruise periods)



C) Multiple season-years

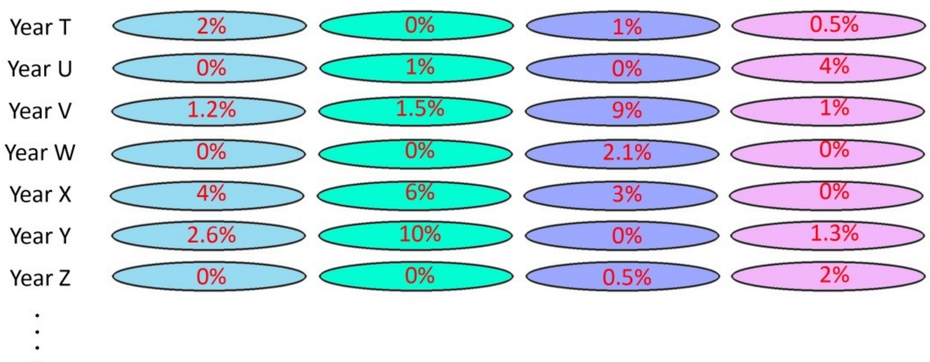


Figure 3a-c. Illustration of Method 1 used to develop biological reference curves for chlorophyll *a*. The method assumes that the proportion of observations violating a criterion in a given season-year is comparable to the exceedance rate in space of a single assessment layer developed with the prescribed VADEQ method. A) All sampling events in a given season-year are classified according to the salinity observed at the time of sampling (light blue = tidal fresh; green = oligohaline; blue = mesohaline; purple = polyhaline). B) All sampling events in the season-year whose water quality conditions meet the reference classification criteria are extracted. Samples are scored by the appropriate season-salinity criterion (dots = total number of reference sampling events; red dots = criterion failed). C) Each year is treated as if it were an individual assessment layer representative of the given season anywhere in the given salinity zone. Attainment rates for the individual layers can be used to construct biological reference curves from multiple layers.

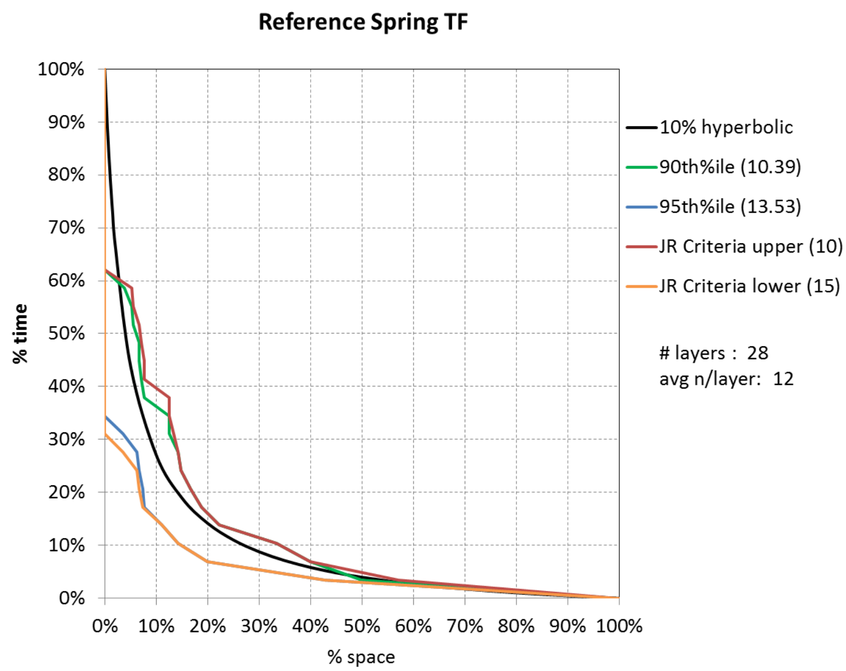


Figure 4a. Spring tidal fresh CFD curves for chlorophyll *a* from reference water quality conditions found in Chesapeake Bay open waters between 1984 and 2013. The number of assessment layers used to create the CFD curves and the average number of samples per layer are indicated on the right. The values of the criteria applied to the data are indicated in parentheses.

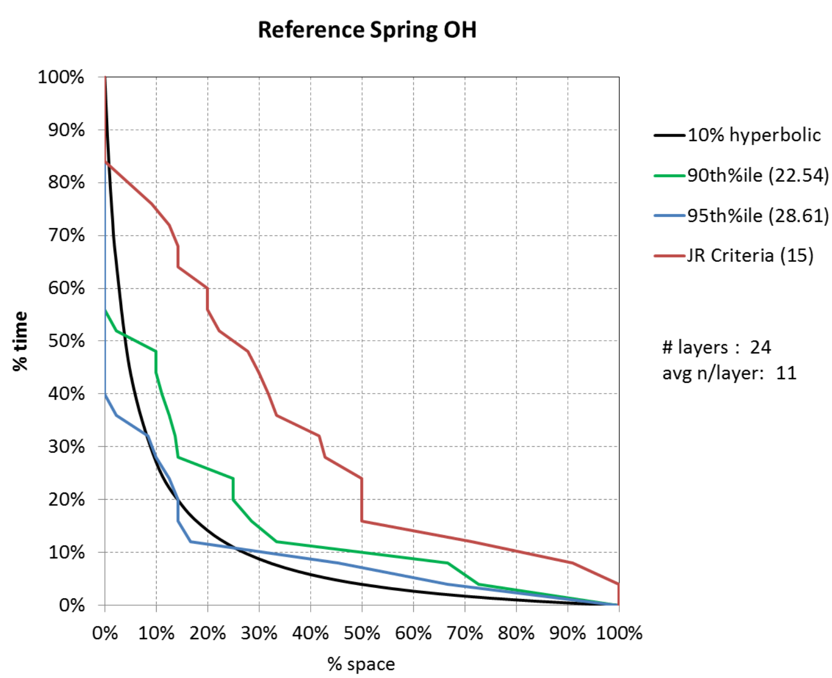


Figure 4b. Spring oligohaline CFD curves for chlorophyll *a* from reference water quality conditions.

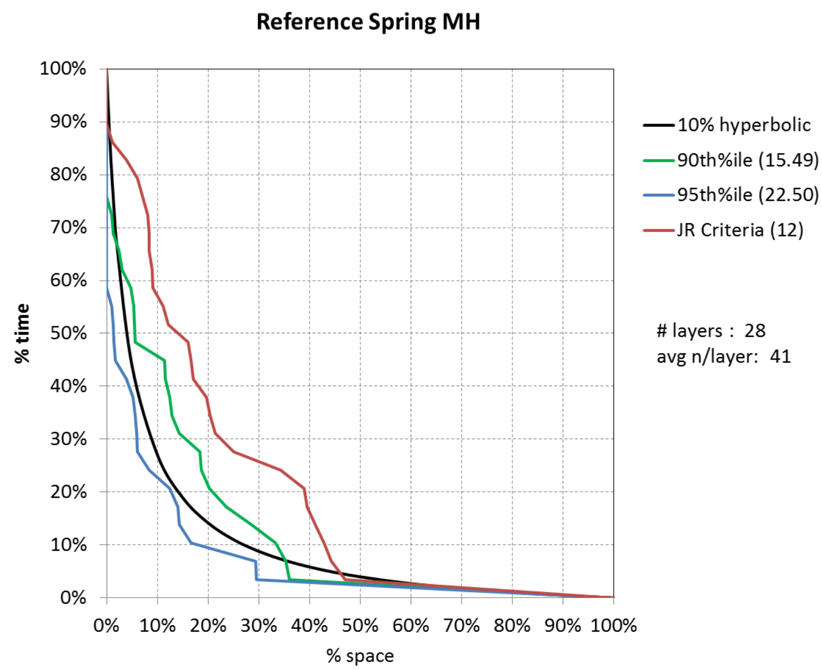


Figure 4c. Spring mesohaline CFD curves for chlorophyll *a* from reference water quality conditions.

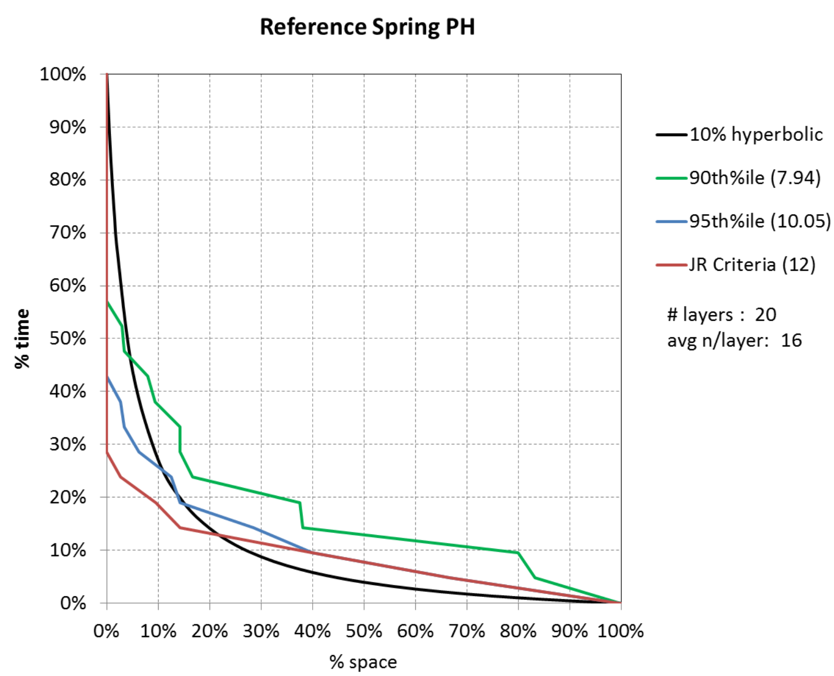


Figure 4d. Spring polyhaline CFD curves for chlorophyll *a* from reference water quality conditions.

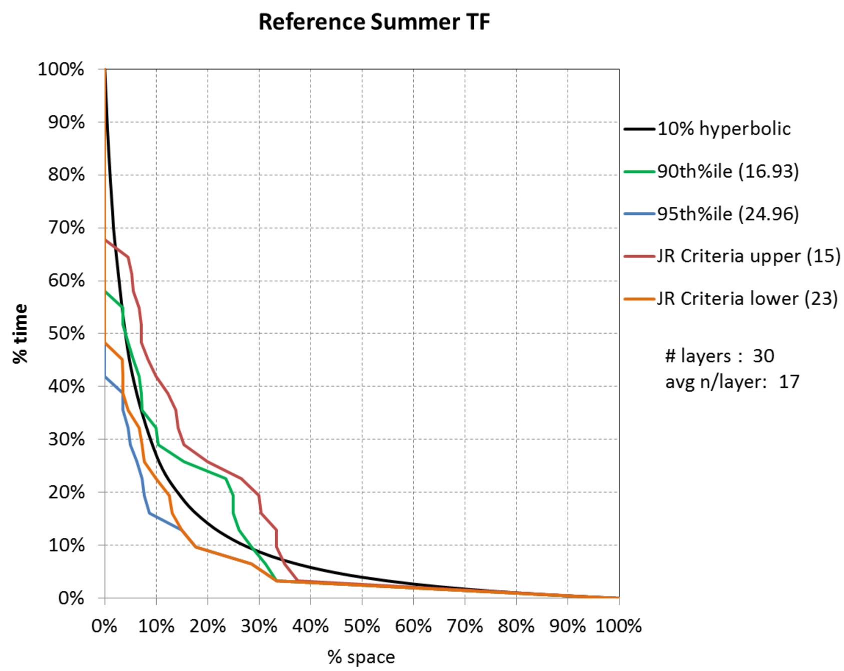


Figure 4e. Summer tidal fresh CFD curves for chlorophyll *a* from reference water quality conditions.

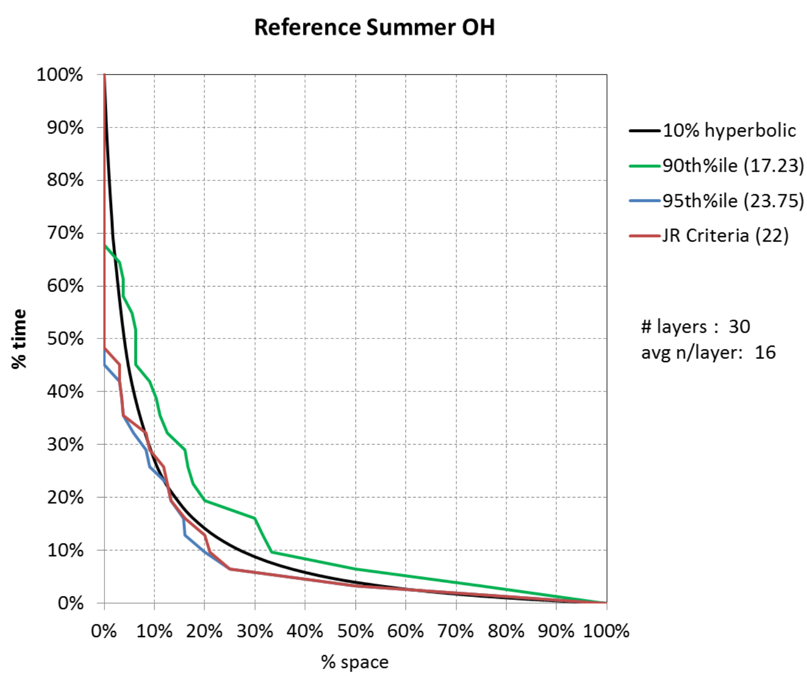


Figure 4f. Summer oligohaline CFD curves for chlorophyll *a* from reference water quality conditions.

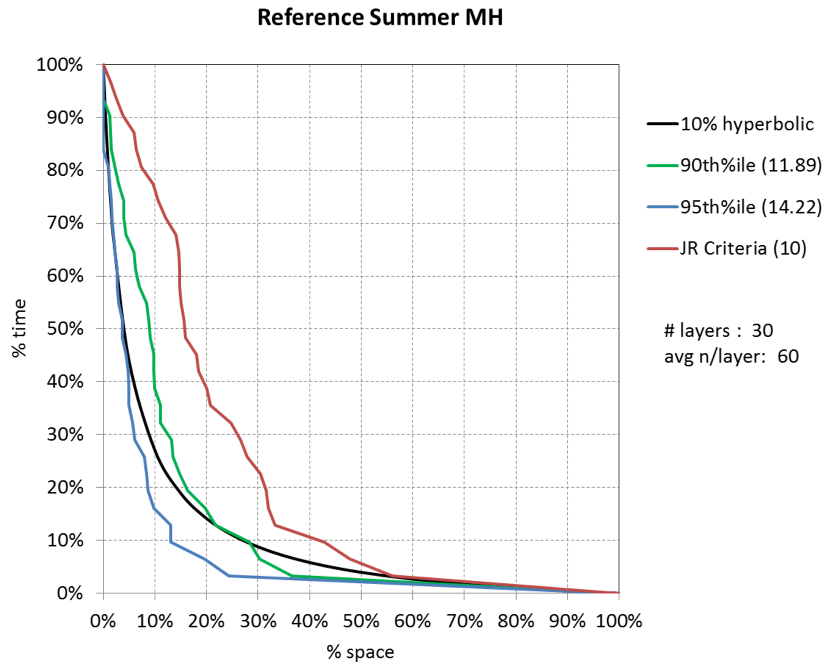


Figure 4g. Summer mesohaline CFD curves for chlorophyll *a* from reference water quality conditions.

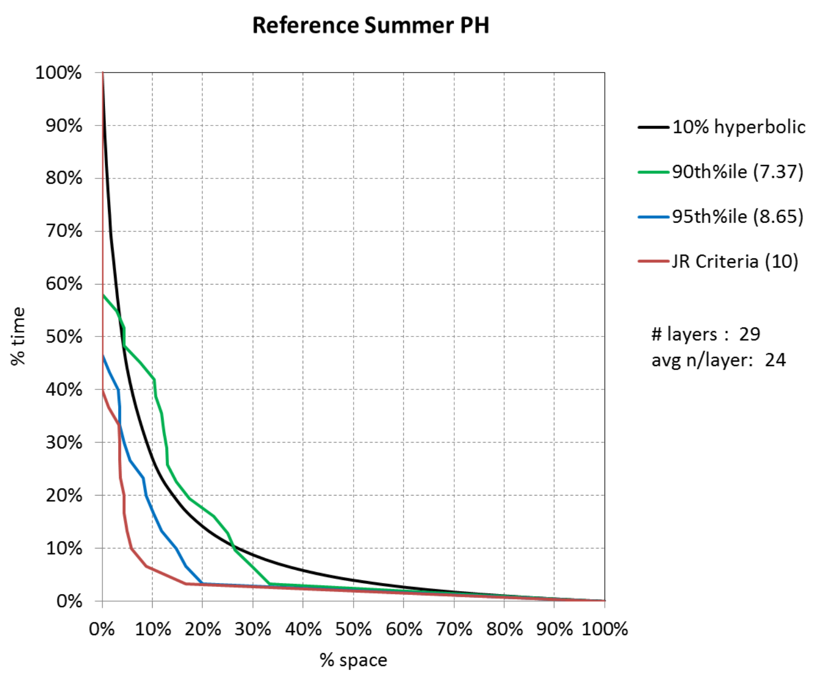


Figure 4h. Summer polyhaline CFD curves for chlorophyll *a* from reference water quality conditions.

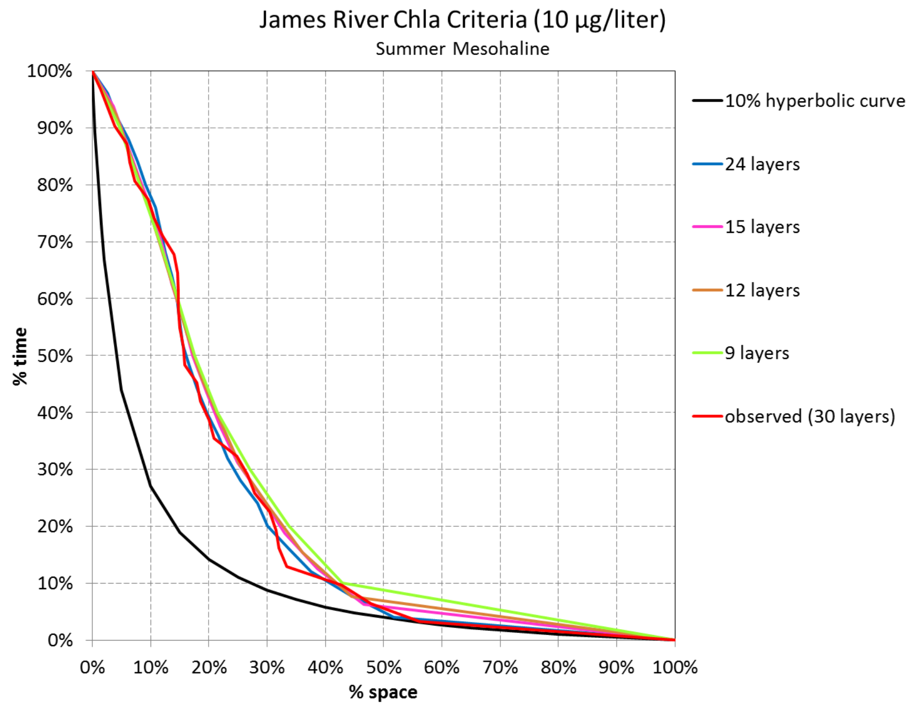


Figure 5a. Biological reference curves based on the James River criterion for summer mesohaline habitat and developed from 24, 15, 12, and 9 assessment layers. Layers were created using random sampling with replacement (Method 2). See text for details.

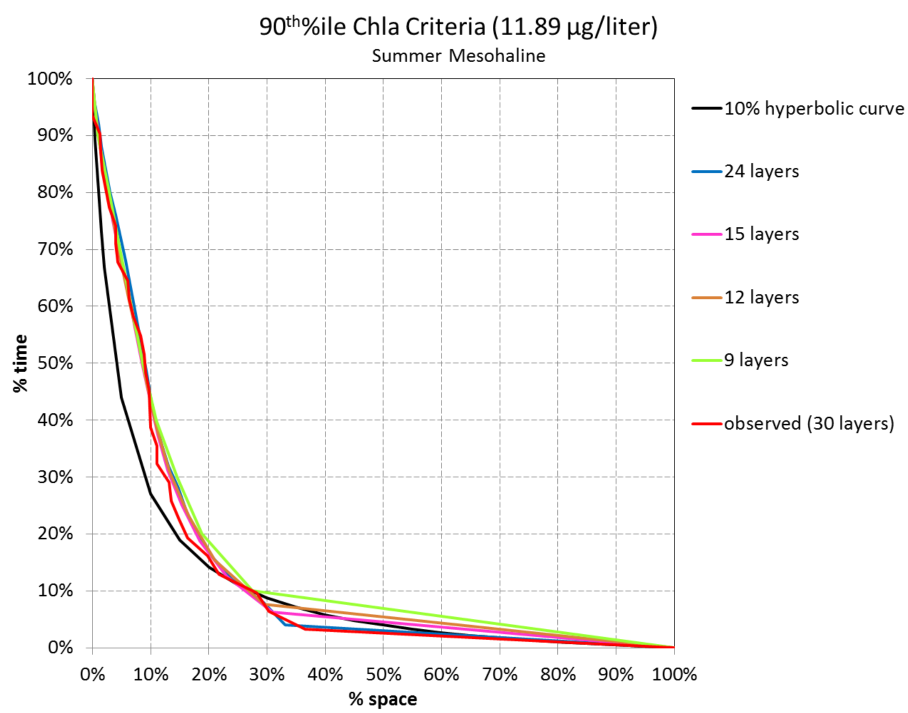


Figure 5b. Biological reference curves based on the 90th%ile criterion for summer mesohaline habitat and developed from 24, 15, 12, and 9 assessment layers.

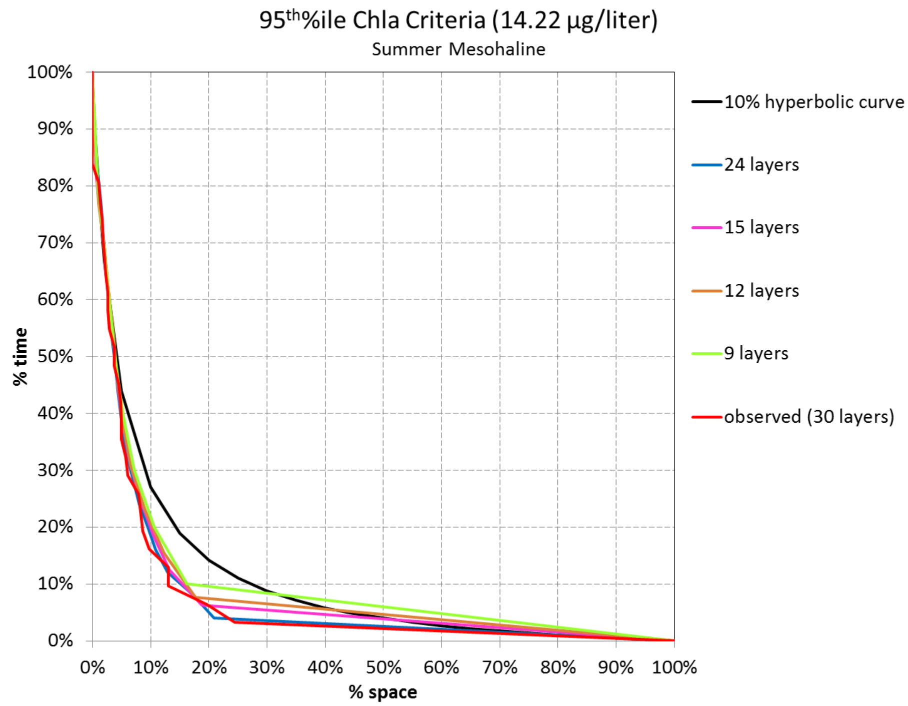


Figure 5c. Biological reference curves based on the 95th%ile criterion for summer mesohaline habitat and developed from 24, 15, 12, and 9 assessment layers.

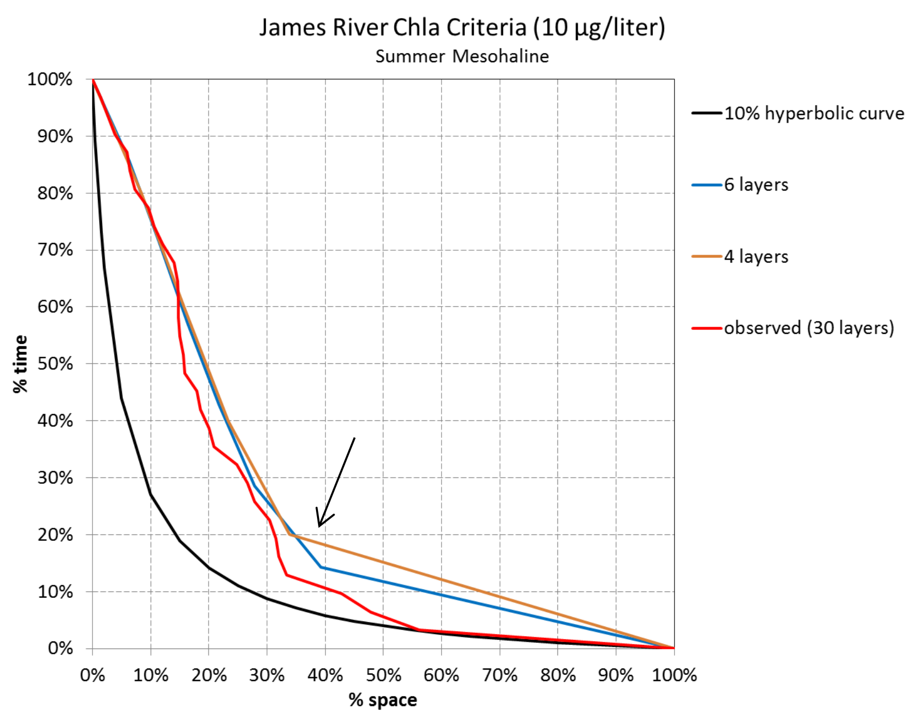


Figure 6a. Biological reference curves based on the James River criterion for summer mesohaline habitat and developed from 6 and 4 assessment layers. Layers were created using random sampling with replacement (Method 2). Arrow indicates bias caused by too few layers (see text for details).

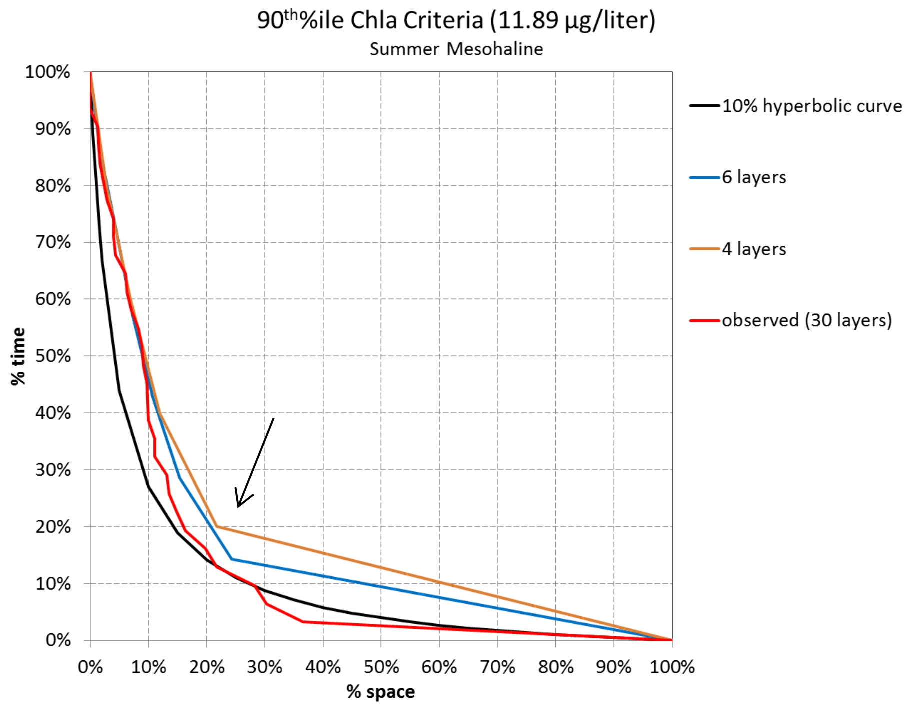


Figure 6b. Biological reference curves based on the 90th%ile criterion for summer mesohaline habitat and developed from 6 and 4 assessment layers.

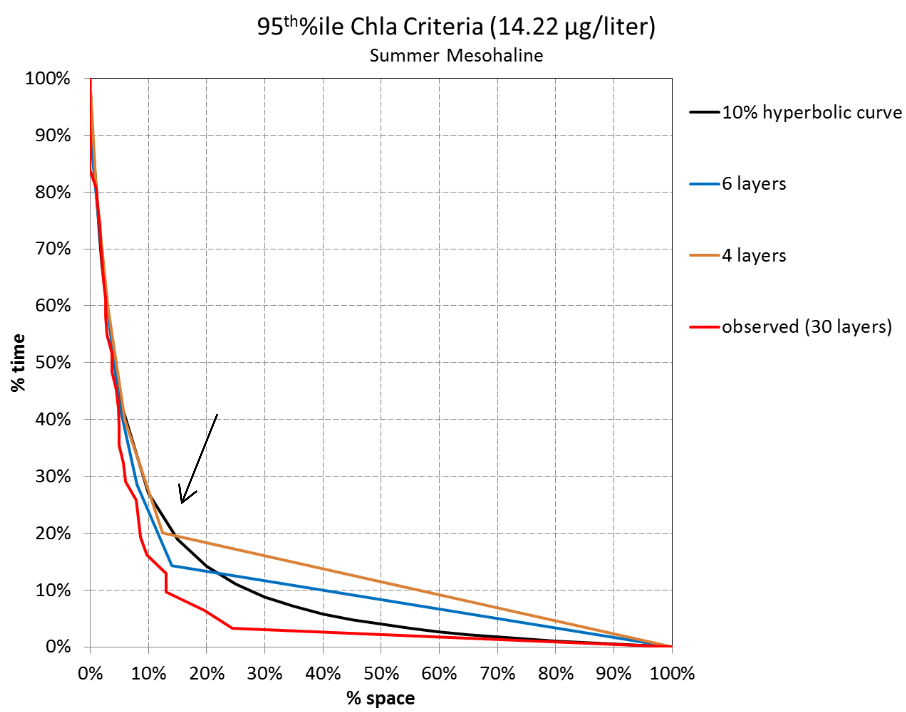


Figure 6c. Biological reference curves based on the 95th%ile criterion for summer mesohaline habitat and developed from 6 and 4 assessment layers.

Individual Layers

%time	%space	Year
0.0%	100.0%	
10.0%	85.0%	2
20.0%	65.0%	1
30.0%	36.0%	3
40.0%	20.0%	2
50.0%	18.2%	1
60.0%	14.8%	1
70.0%	6.0%	2
80.0%	5.0%	3
90.0%	1.5%	3
100.0%	0.0%	

Annual Means

%time	%space	Year
0.0%	100.0%	
25.0%	37.0%	2
50.0%	32.7%	1
75.0%	14.2%	3
100.0%	0.0%	

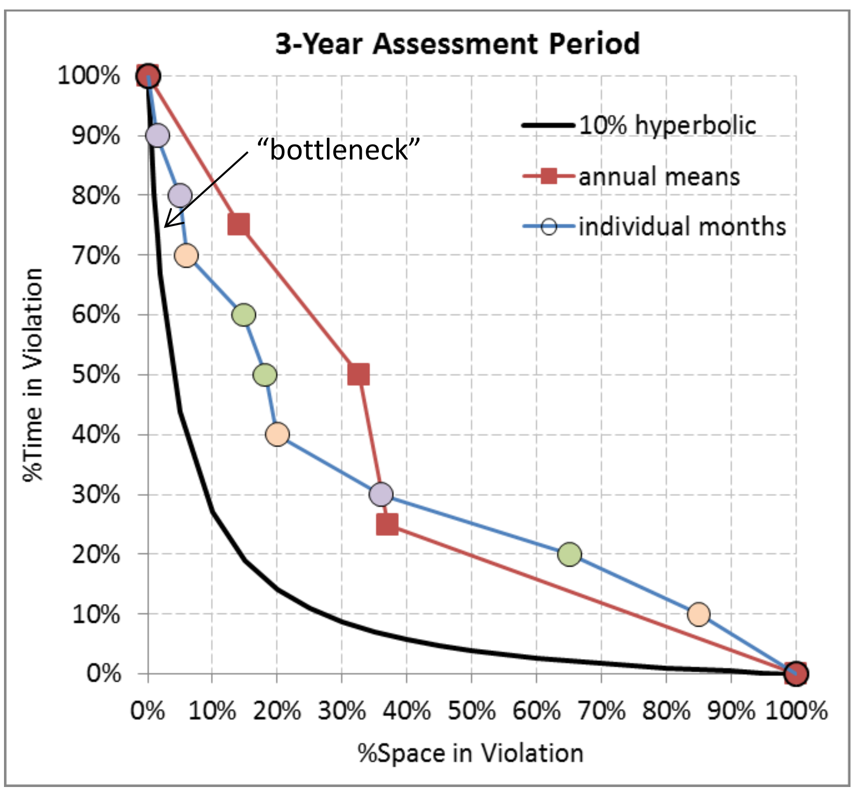


Figure 7. Hypothetical example of CFD curve biases created when annual (seasonal) means are used.

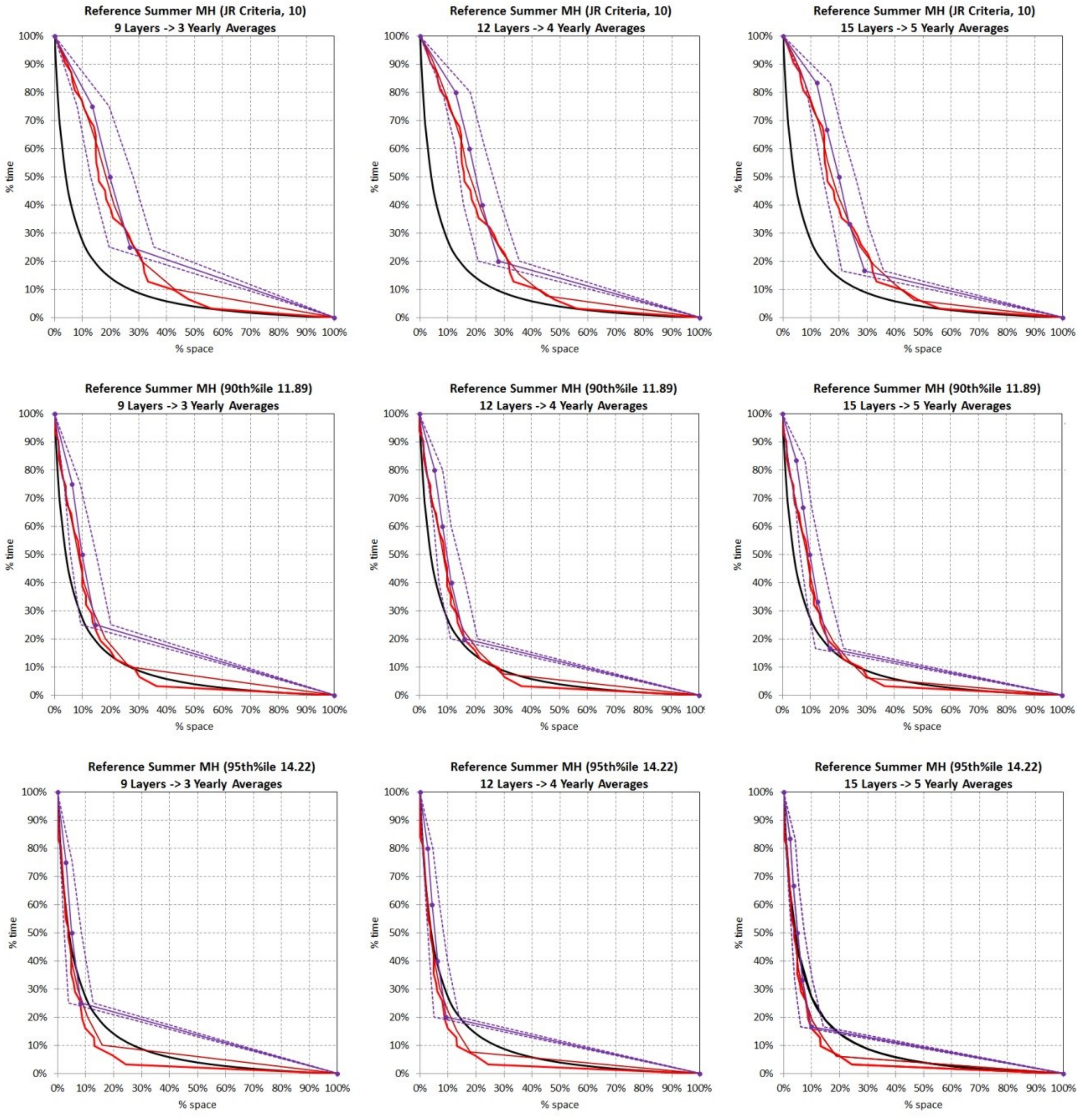
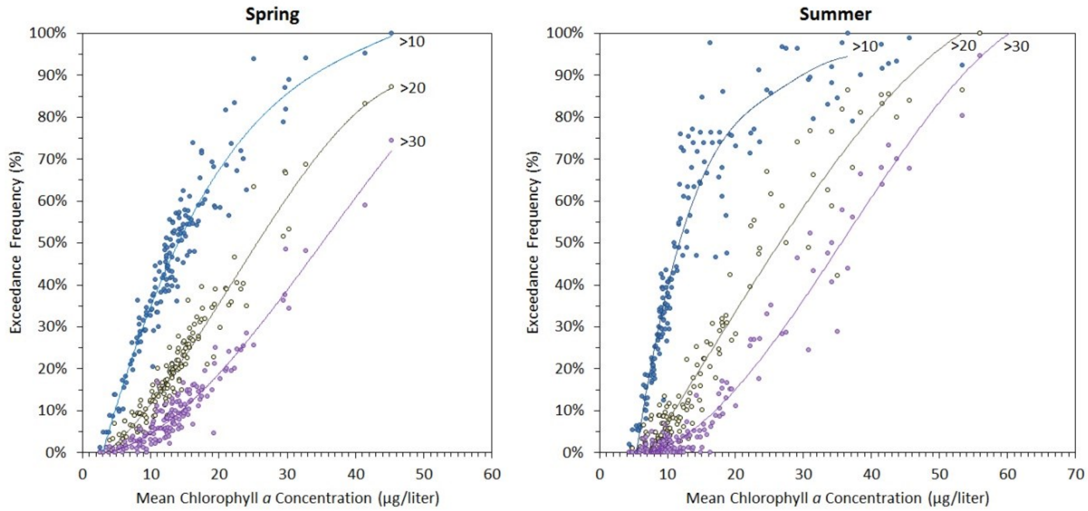
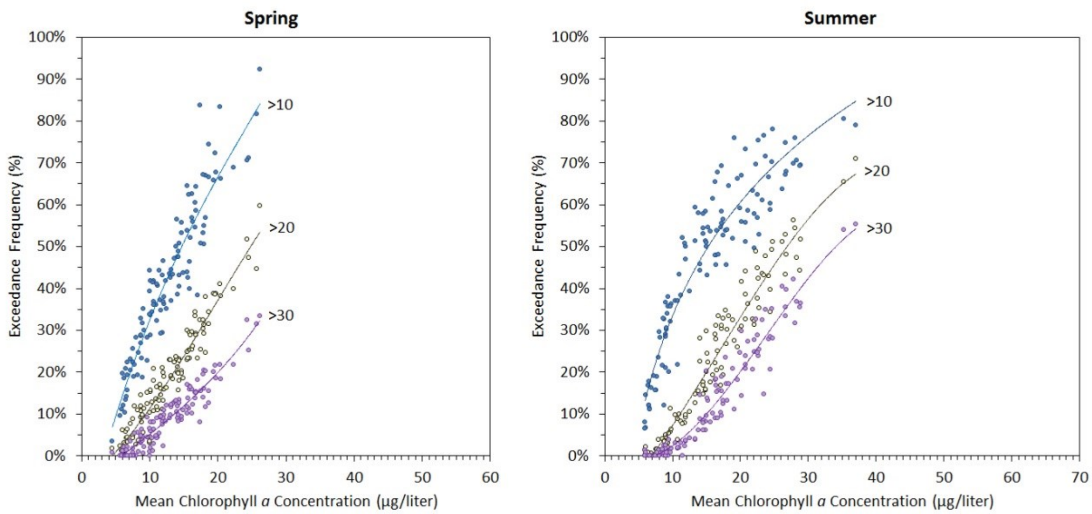


Figure 8. Biological reference curves for summer mesohaline built from 3, 4, and 5 annual averages. The 10% hyperbolic curve, the complete 30-point CFD curve, and the CFDs built from the underlying 9, 12, 15 individual assessment layers are shown for comparison. Assessment layers for annual averages were created using Method 3. See text for details. Red solid line: CFD curve of all 30 summer mesohaline assessment layers (representing different summer months in this example). Orange solid line: CFD curve base on the 9, 12, and 15 assessment layers used to calculate the 3, 4, and 5 yearly averages for a season, respectively. Purple solid line: CFD curve built from the yearly averages for a season (see Method 3 for details). Purple dashed line: 10^{th}%ile and 90^{th}%ile around the CFD curve.

A)



B)



C)

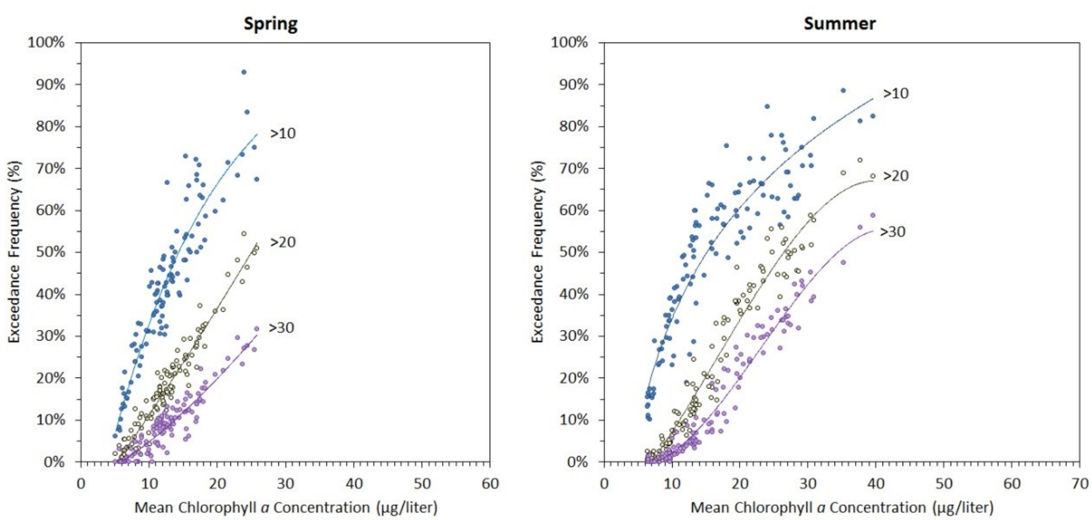
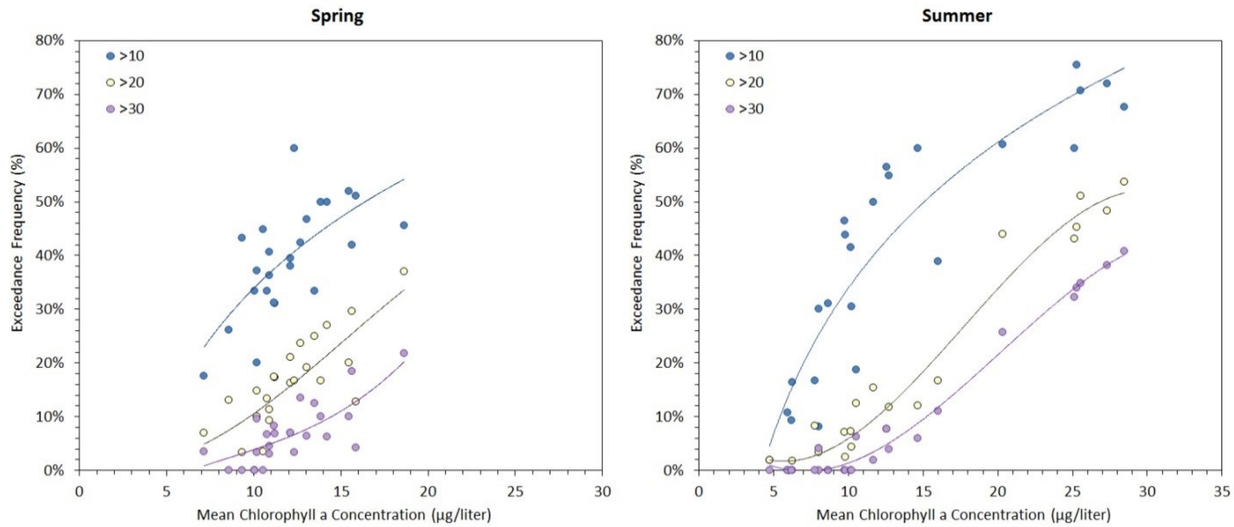


Figure 9. Relationships in entire Chesapeake Bay between mean chlorophyll *a* and frequency of exceeding 10, 20, and 30 µg/liter. Seasonal data grouped by station (A), CBP segment and year (B), and salinity zone and year (C). Points with n<30 not shown.

A)



B)

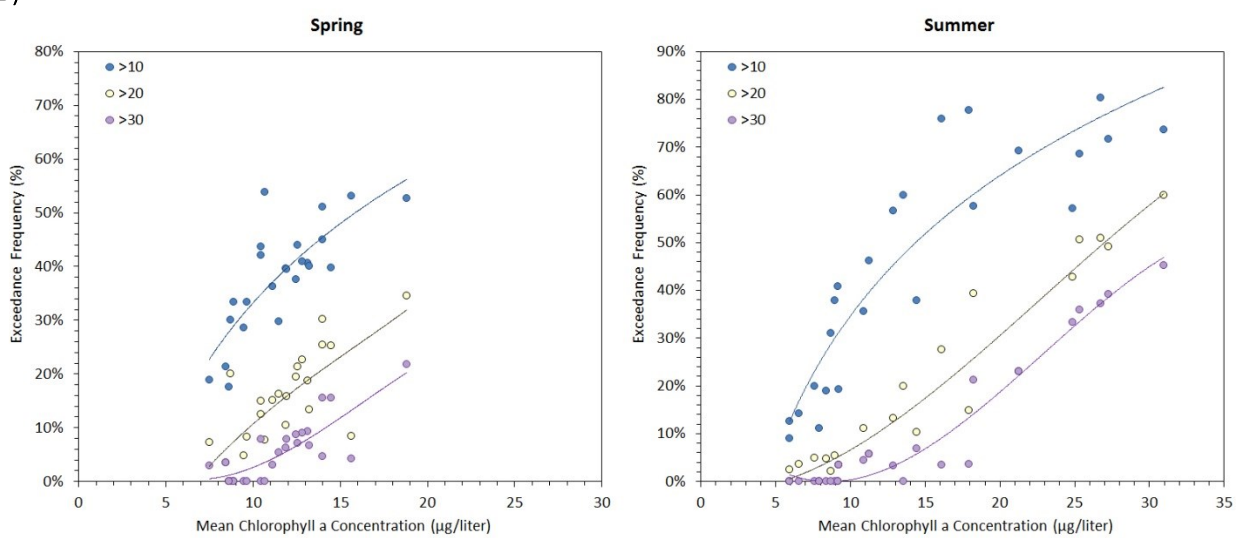


Figure 10. Relationships in the James River between mean chlorophyll *a* and frequency of exceeding 10, 20, and 30 µg/liter. Points with n<24 not shown. Data were grouped by season and the James River CBP segment JMSTF, JMSOH, JMSMH, and JMSPH (A) or by season and salinity zone (B). Data were then further divided into six time periods (1985 – 1989, 1990 – 1994, 1995 – 1999, 2000 – 2004, 2005 – 2009, and 2010 – 2013) instead of by year in order to ensure sufficient numbers of samples per point.