

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 it possible to obtain near continuous measurements of chlorophyll in a temporal domain (ConMon) or spatial domain (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 which 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 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 chlorophyll. 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.).

DataFlow JMSPH_05042006.csv

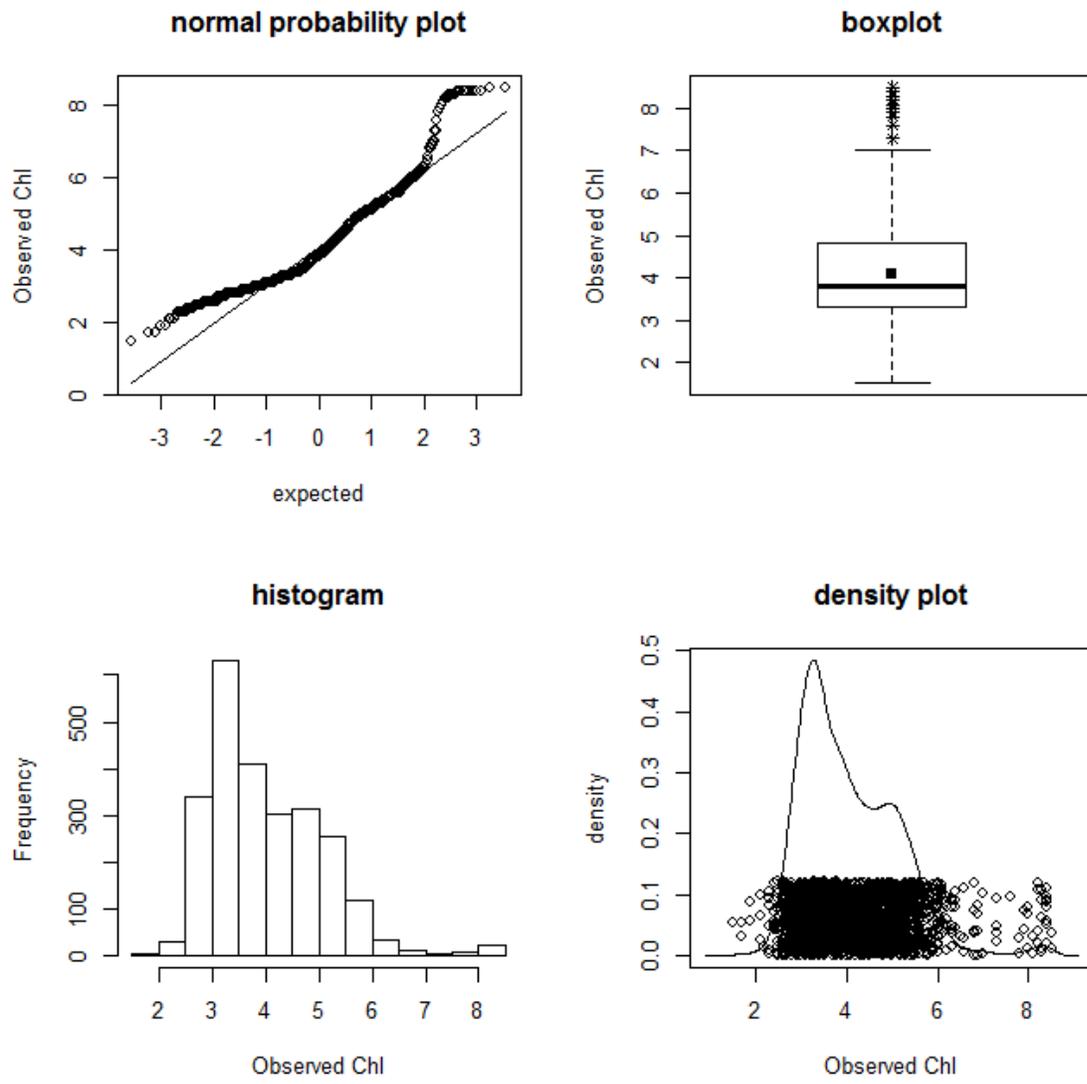


Figure 1. Distribution plots for observed chlorophyll from Data Flow cruise James River Polyhaline 05/04/2005.

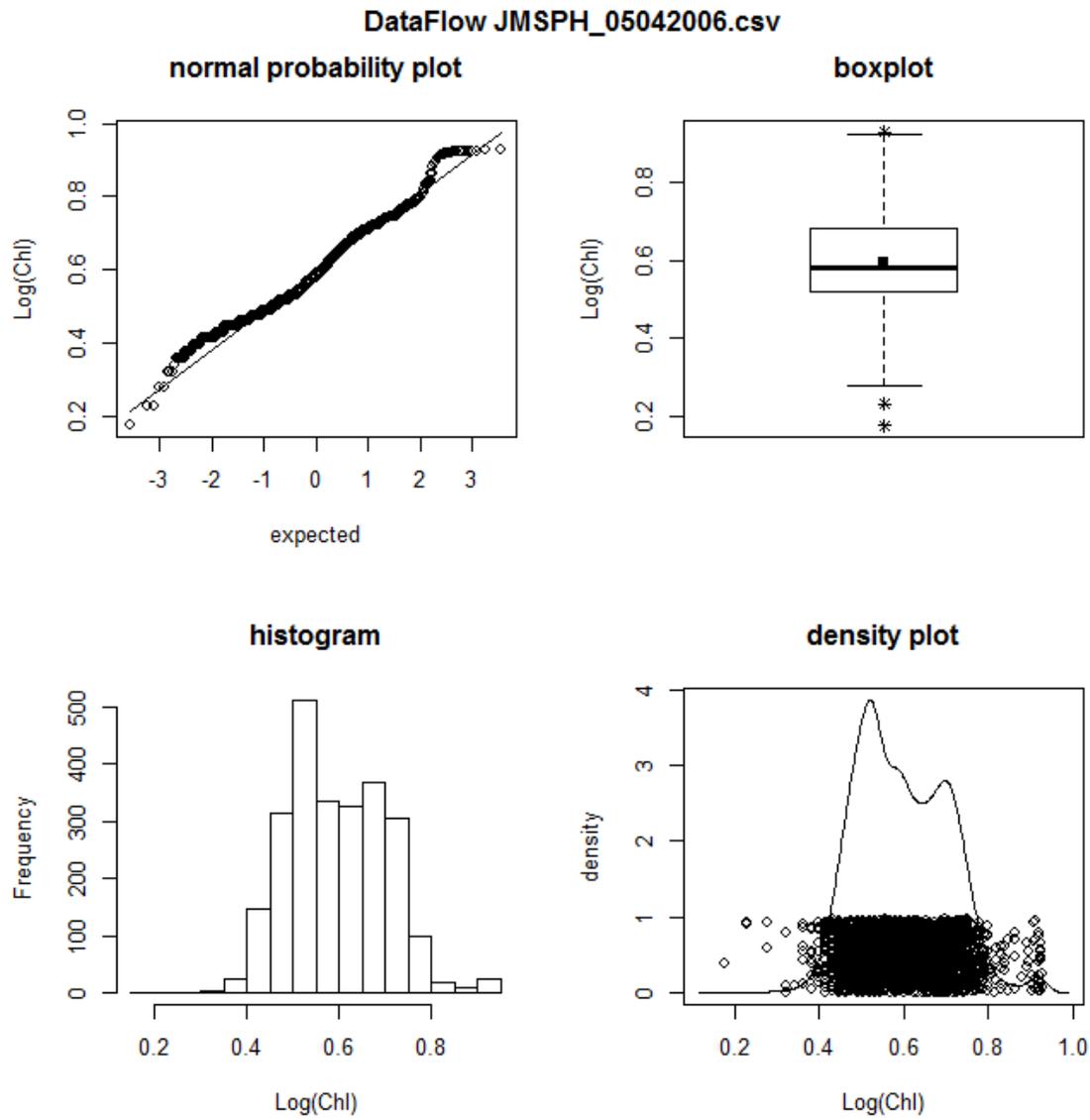


Figure 2. Distribution plots for logarithm transformed chlorophyll from Data Flow cruise James River Polyhaline 05/04/2005

For each cruise, the Lat/Long data are converted to UTM_y/UMT_x 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 UTM_y/UMT_x 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.

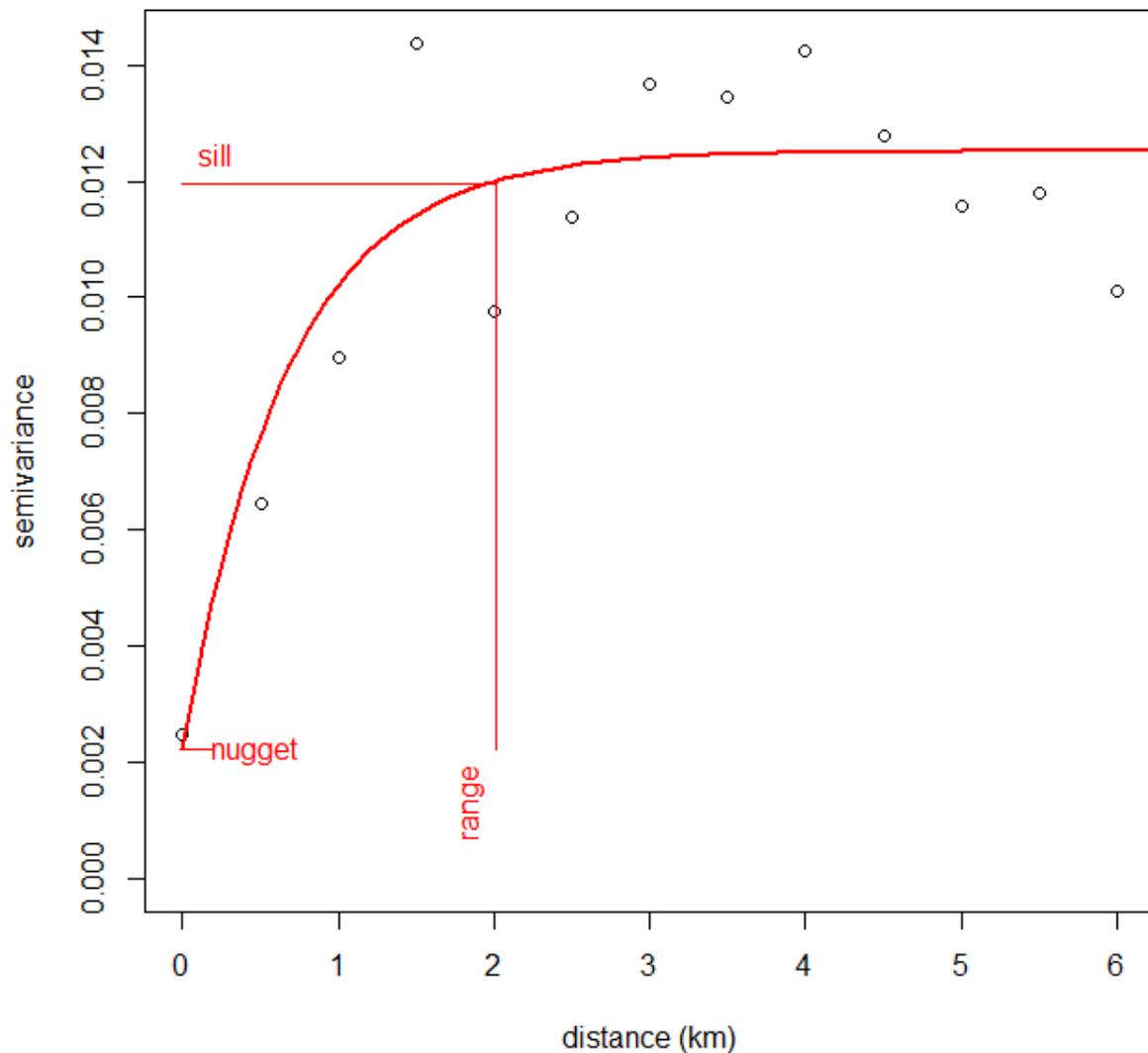


Figure 3. Variogram estimates (circles), estimated variogram function (solid line), and spatial data attributes of partial sill, nugget, and range for James River Polyhaline DataFlow 05/04/2005.

Using the smooth variogram estimator fitted by the R-function `variostat()`, kriging is used to obtain interpolated chlorophyll for the James Polyhaline (Figure 4.). Super-imposed on the gray-scale contour map of $\log(\text{chlorophyll})$ 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 $\log(\text{chlorophyll})$ along the cruise track (Figure 4.)

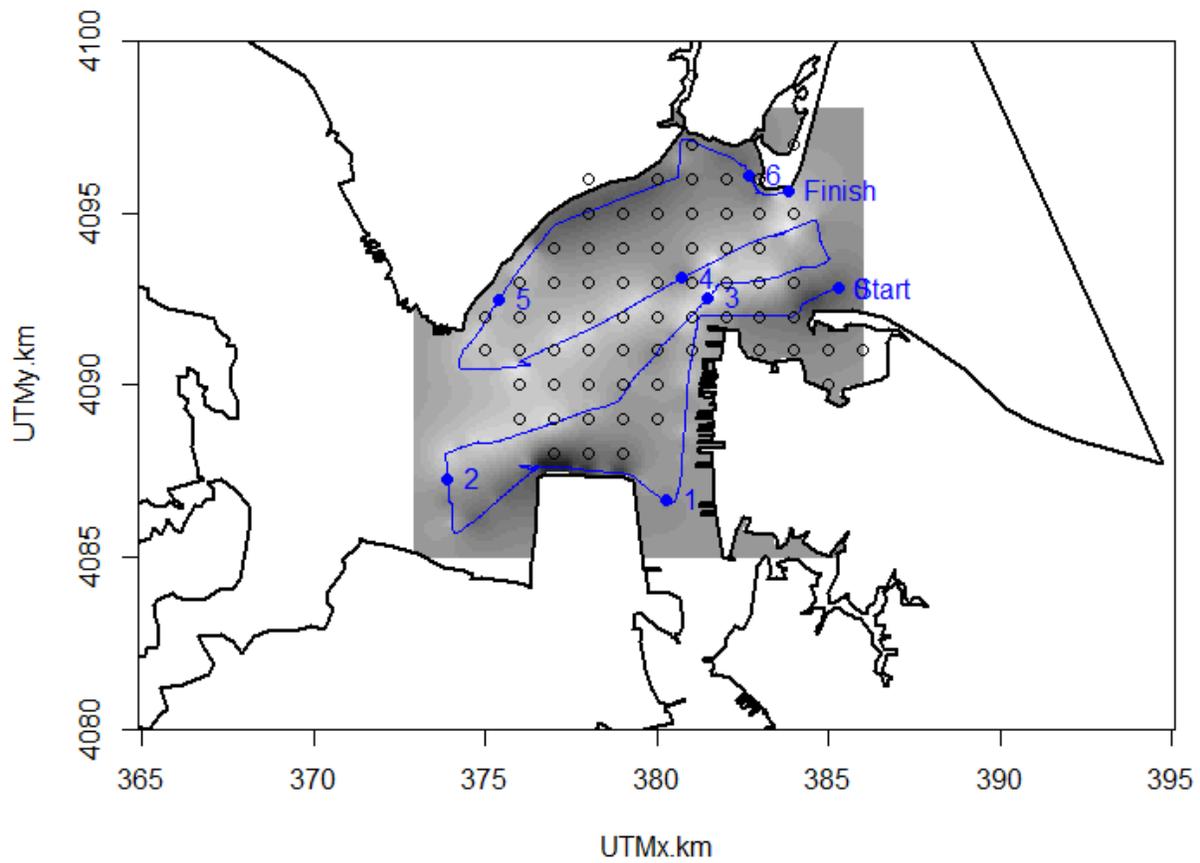
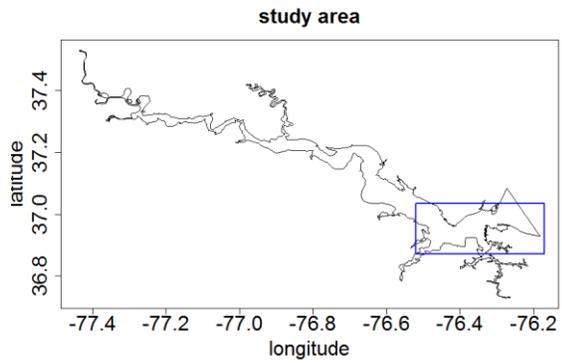


Figure 4. Kriged interpolation at a scale of 0.1 km for DataFlow log(chlorophyll) collected on 5/04/2005. Blue line shows the cruise track. Open black circles show the CBP interpolator cell centers.

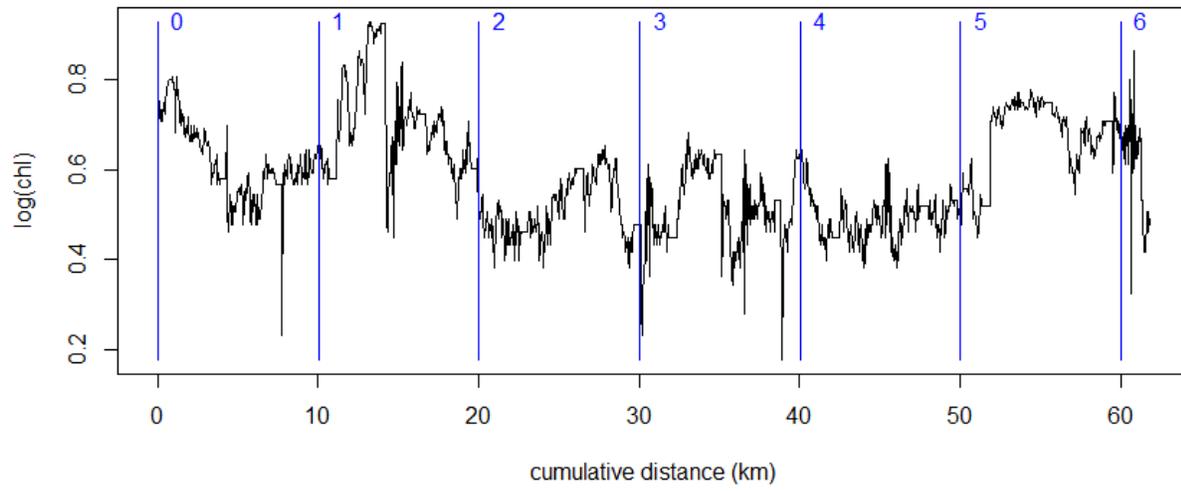


Figure 5. Observed log(chlorophyll) as a function of cruise track distance. The verticle blue lines correspond by number to the blue dot locations of Figure 4.

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

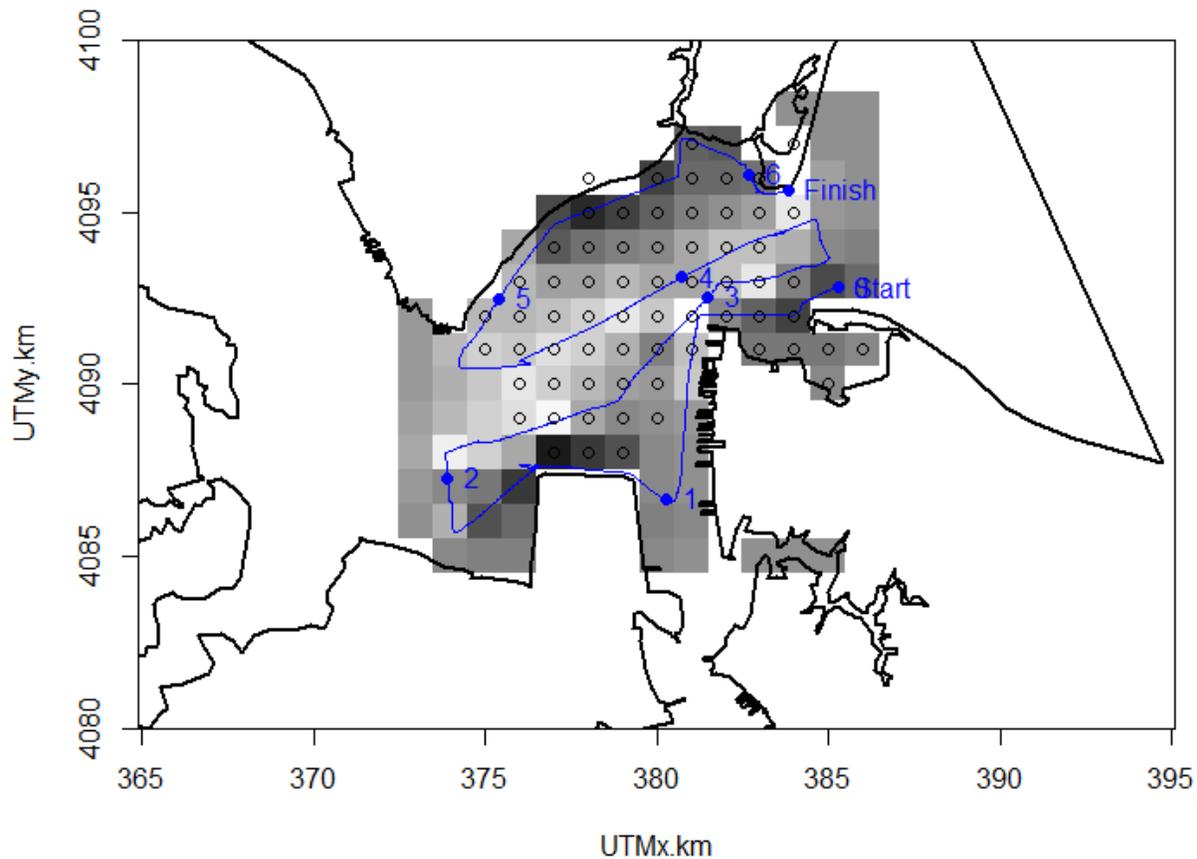


Figure 6. Krig/ed interpolation at a scale of 1 km for DataFlow log(chlorophyll) collected on 5/04/2005.

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.

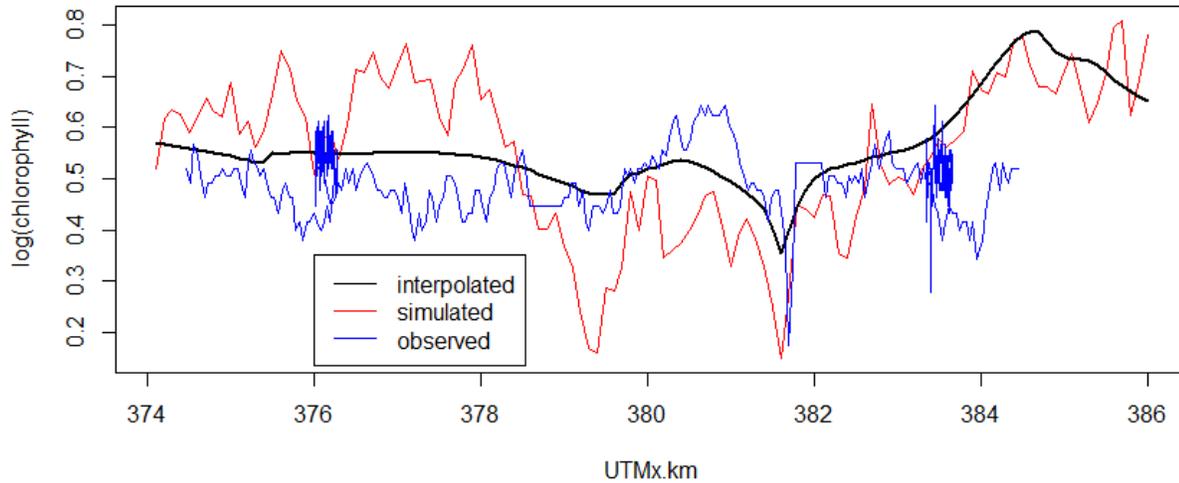


Figure 7. Interpolated, simulated, and observed log(chlorophyll) data for a transect where UTM_y.km = 4092.6 which is about at the level of the cruise start (Figure 8- green transect). The observed data are

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

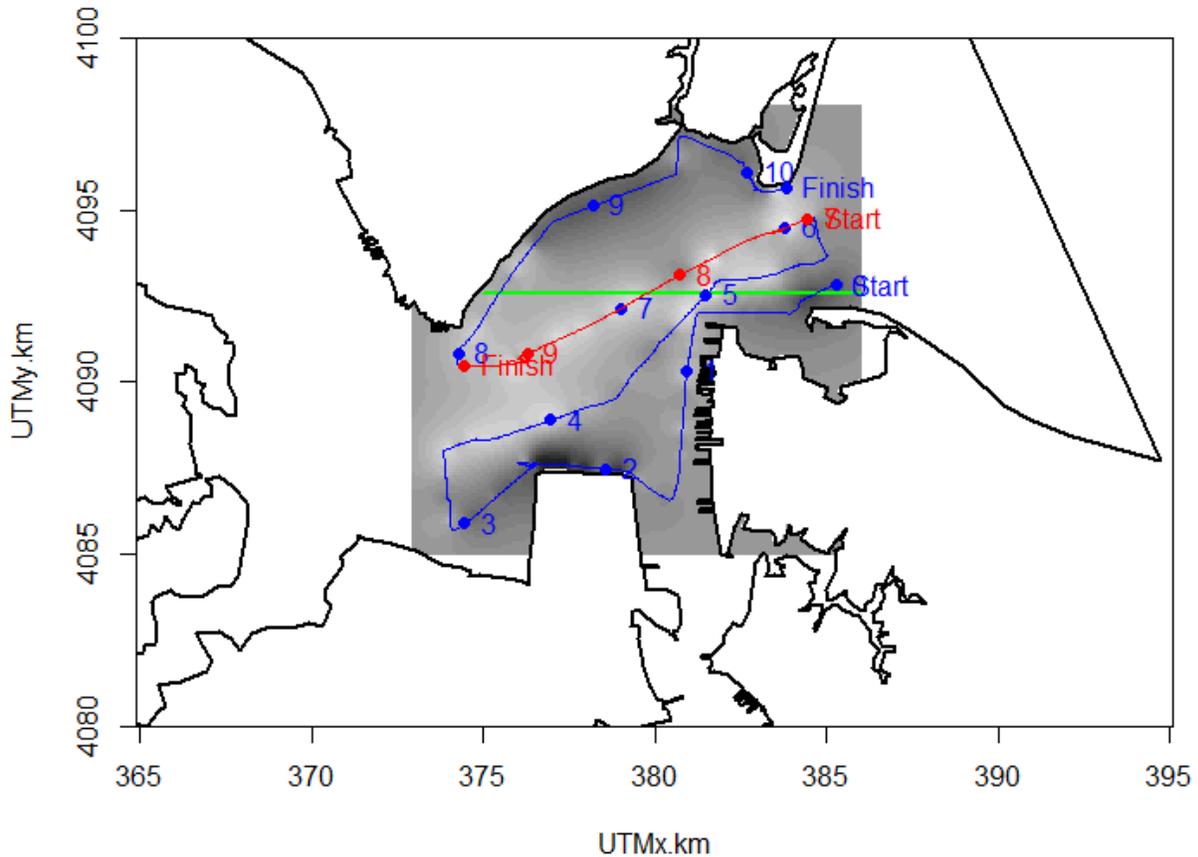


Figure 8. Illustration of the spatial paths of cruise track data (red) and simulated data (green) being compared in Figure 7.

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., Table 1.0).

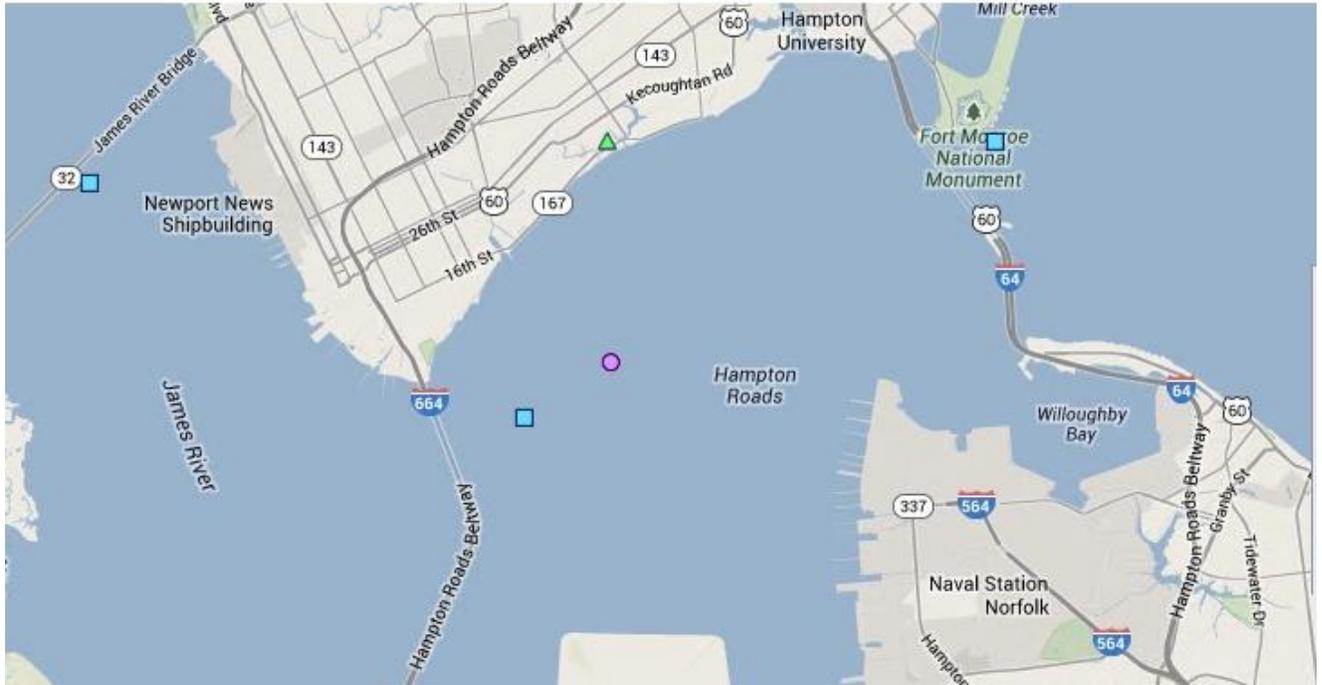


Figure 9. The Wythe Point ConMon is located at the green triangle on the north shore below the Hampton Roads Beltway and above the I64 bridge tunnel.

Wythe Point ConMon, 2006

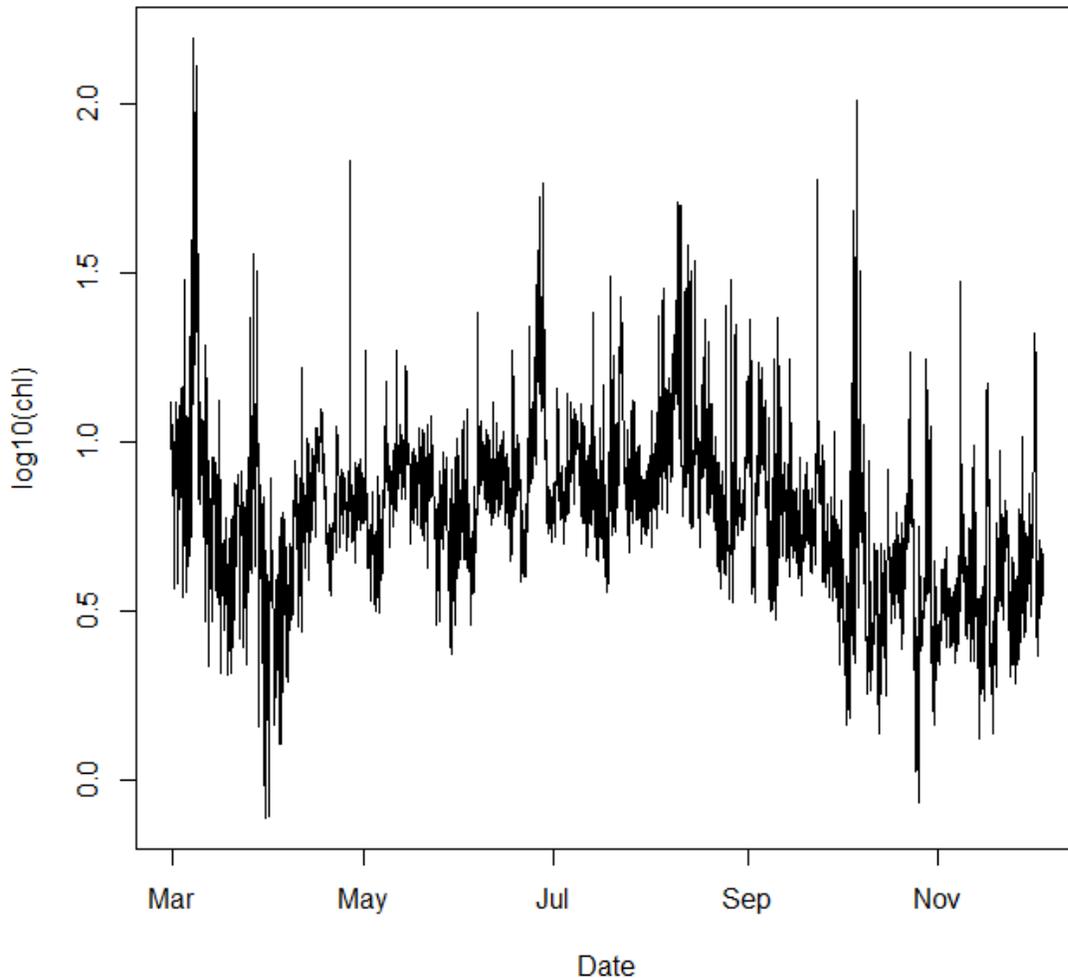


Figure 10. Time series of the base 10 logarithm of chlorophyll collected at the James Polyhaline ConMon Station near Wythe Point.

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 (citation ??) 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.

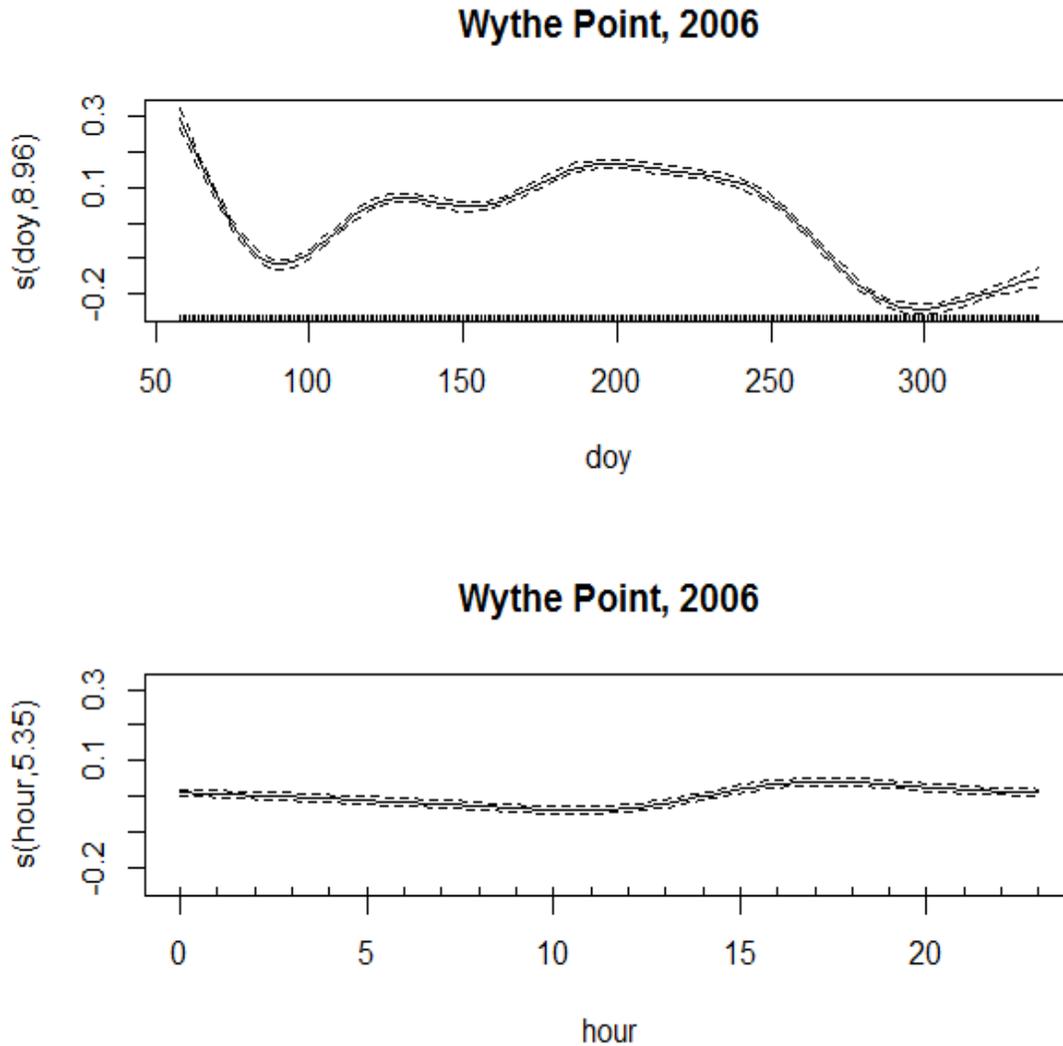


Figure 11. Graphical illustration of GAM smooth components. The top panel is seasonal pattern expressed as a smooth trend is day of year. The bottom panel is diel trend expressed as a function hour within the day.

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).

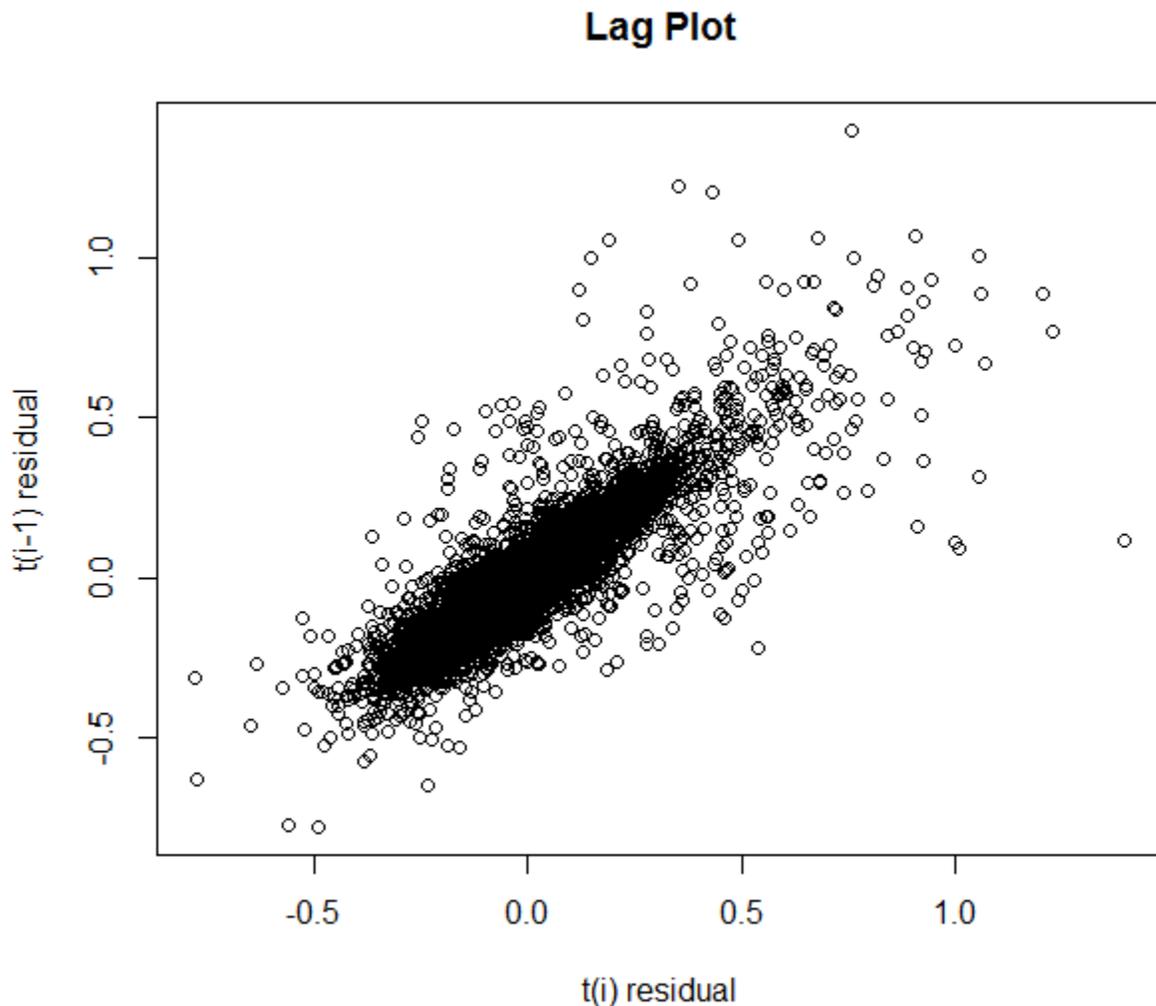


Figure 12. Lag plot of residuals from the Gam with smooth terms for seasonal cycle and diel cycle.

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.

Wythe Point (polyhaline)

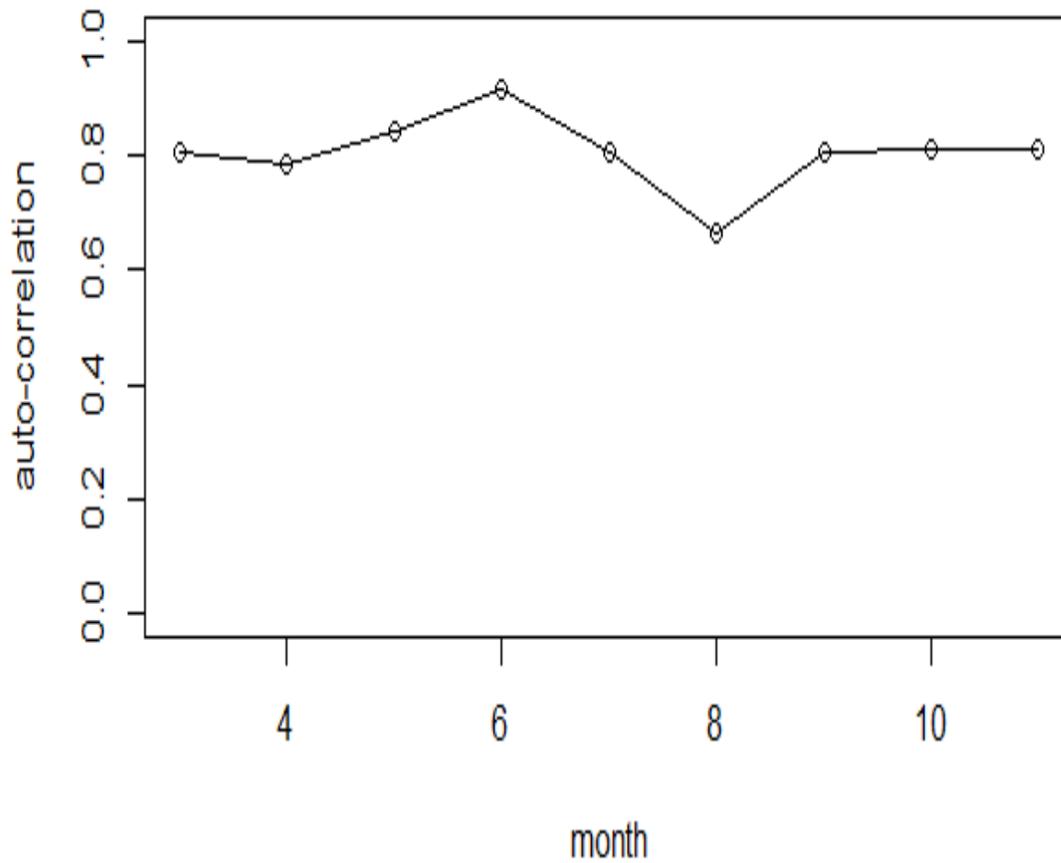


Figure 13. Autocorrelation parameters estimated for individual months from the Wythe Point 2006 ConMon data.

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).

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

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.

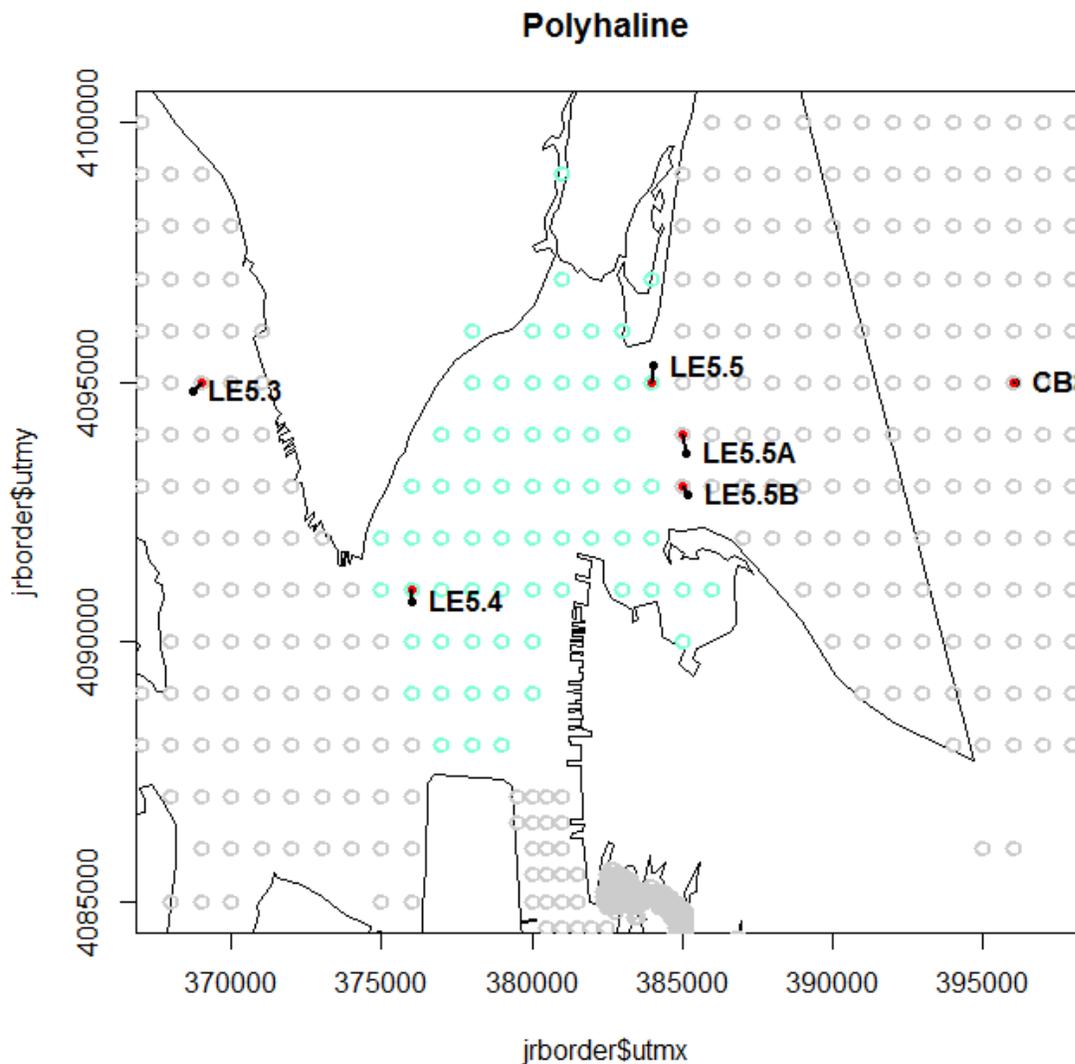


Figure 14. Each fixed station location is shown as a black dot with a line connecting to the nearest interpolator cell which is shown with a red center. The remaining interpolator cells for the segment are shown as open circles in aquamarine. Interpolator cells outside of the segment are shown as open 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 $(\text{rank}/(N+1))$ 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.

Wythe Point vs closest simulation

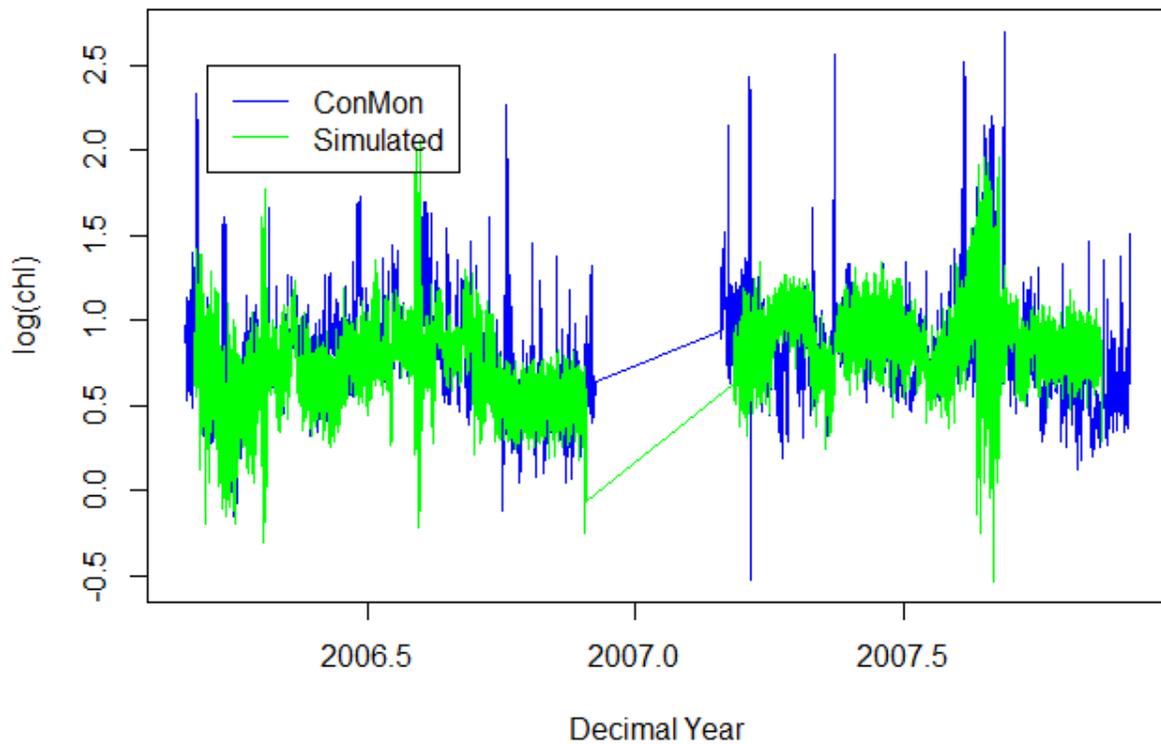


Figure 15. Two years of simulated log-chlorophyll data from this simulation process compared to the same two years of observed time series from the Wythe Point ConMon station.

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.

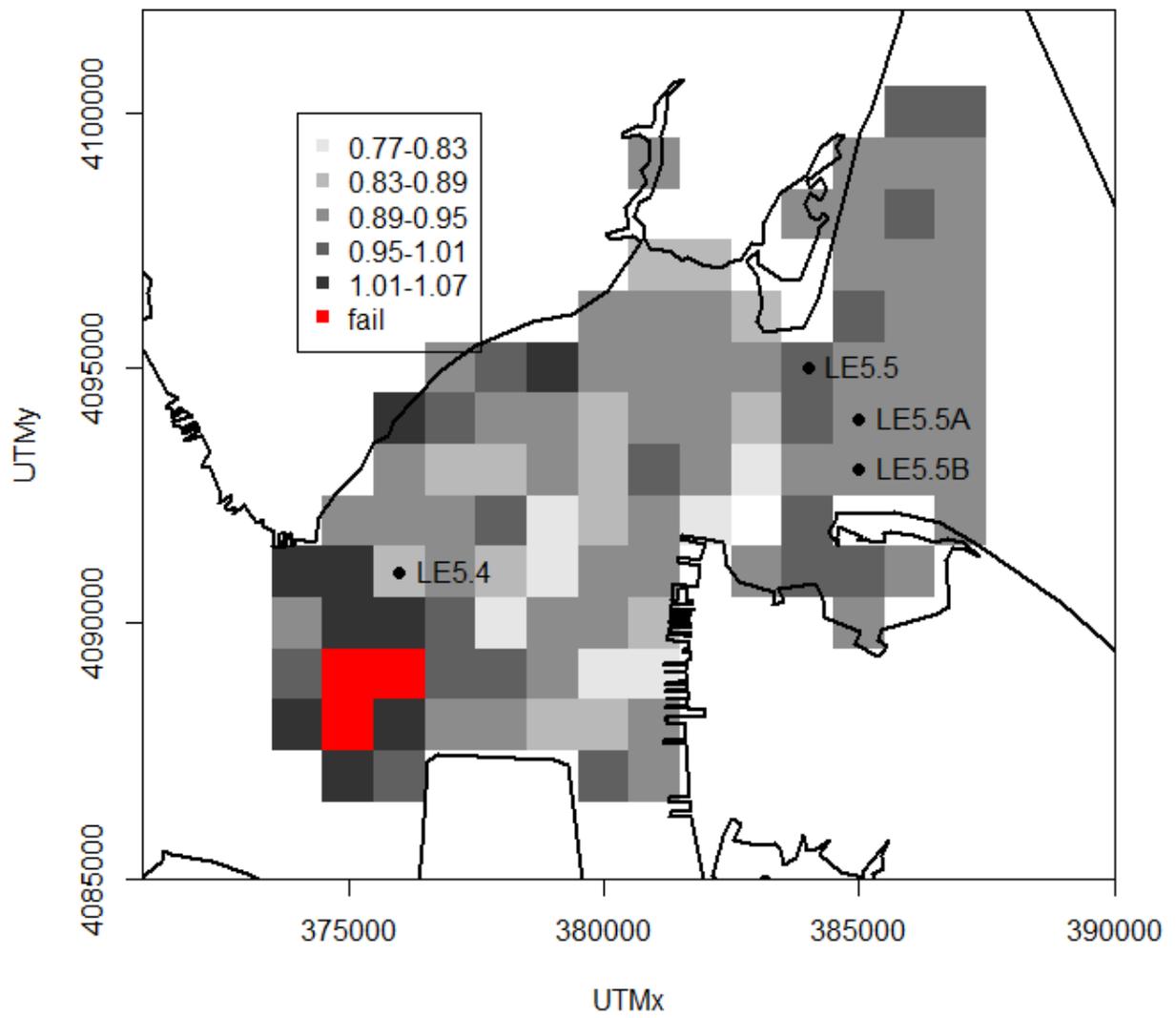


Figure 16. Spring time means of the simulated data for the year 2005.

Wythe Point vs closest simulation

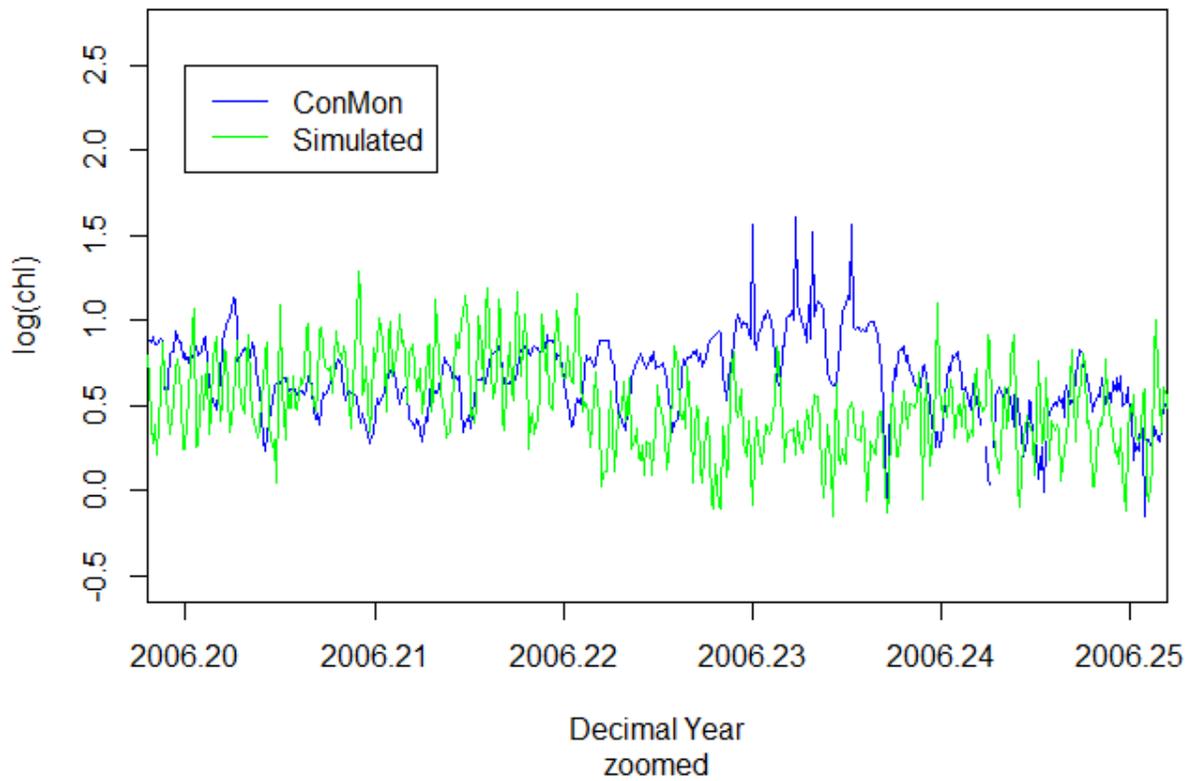


Figure 17. Zooming in to examine 0.05 years (about 18 days) of the Wythe Point ConMon log(chlorophyll) time series as compared to the simulated time series. The blue curve shows a partial time series from the Wythe Point ConMon data. The green curve shows simulated data from the interpolator cell that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

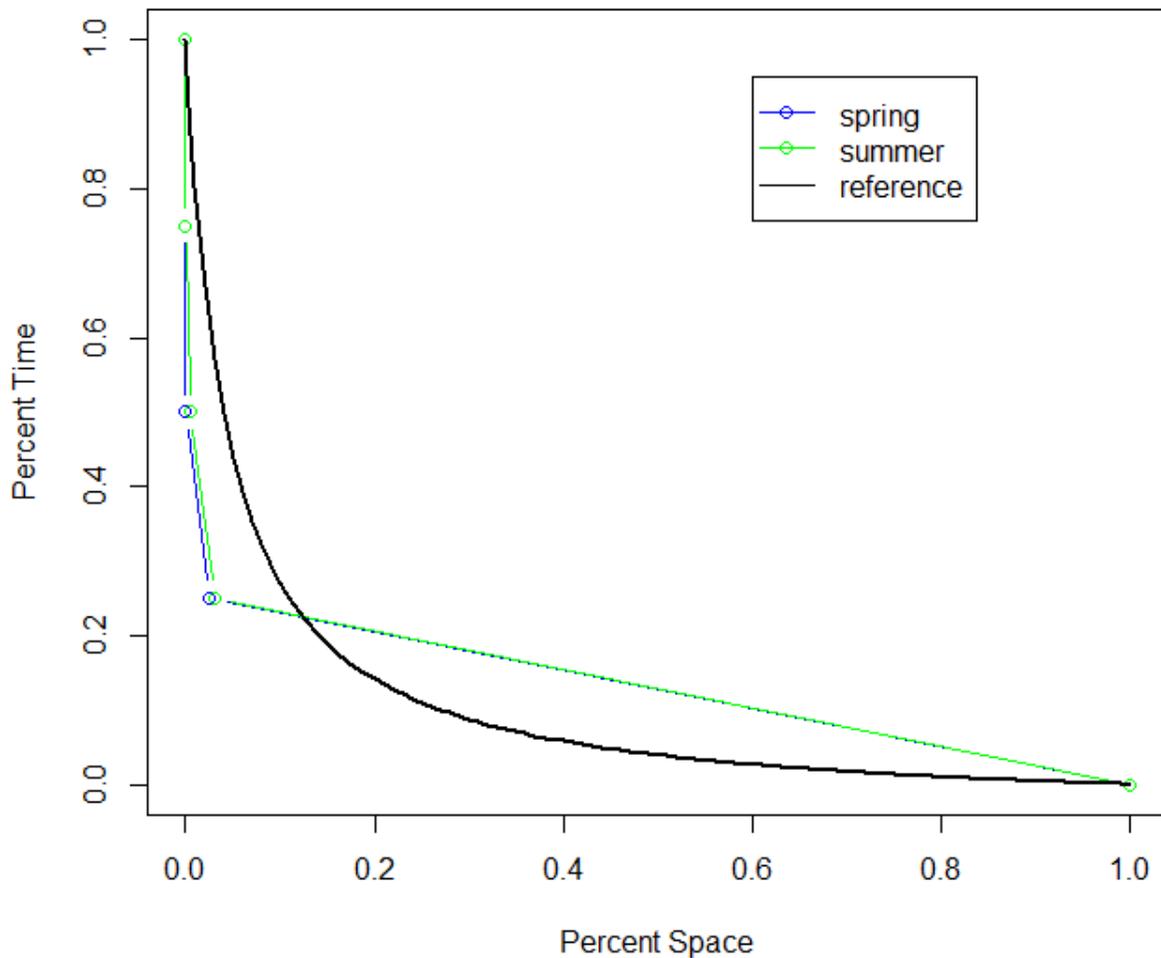


Figure 18. True seasonal CFD curves for the simulated data without manipulation as compared to the standard reference curve.

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.)

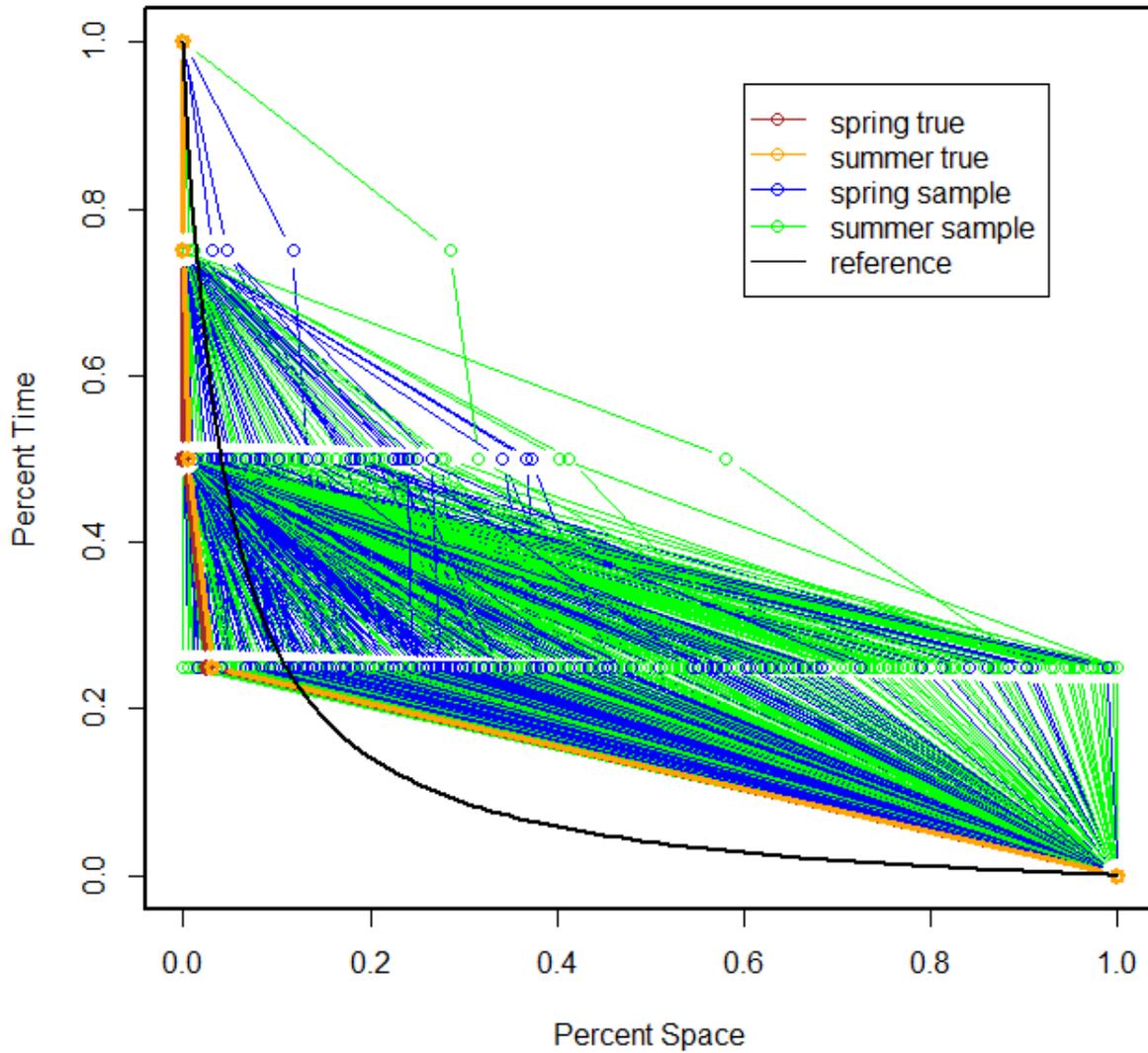


Figure 19. The standard reference curve (black), the true spring and summer CFD for the simulated data (brown and orange), and 1000 estimated CFDs created by monthly sampling of the simulated data (blue and green).

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.)

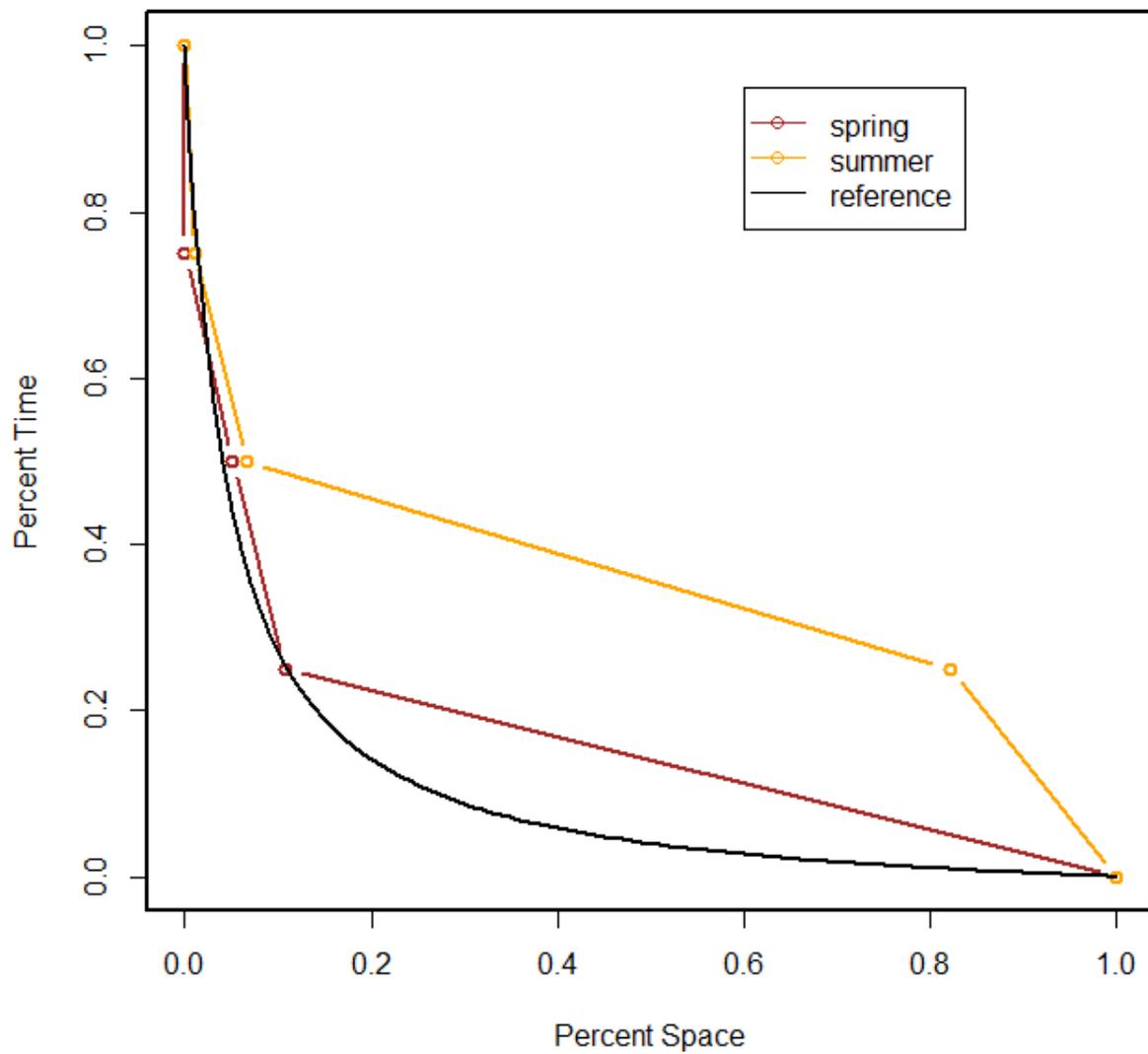


Figure 20. True seasonal CFD curves for the simulated data that has been manipulation to create failure when compared to the standard reference curve.

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.

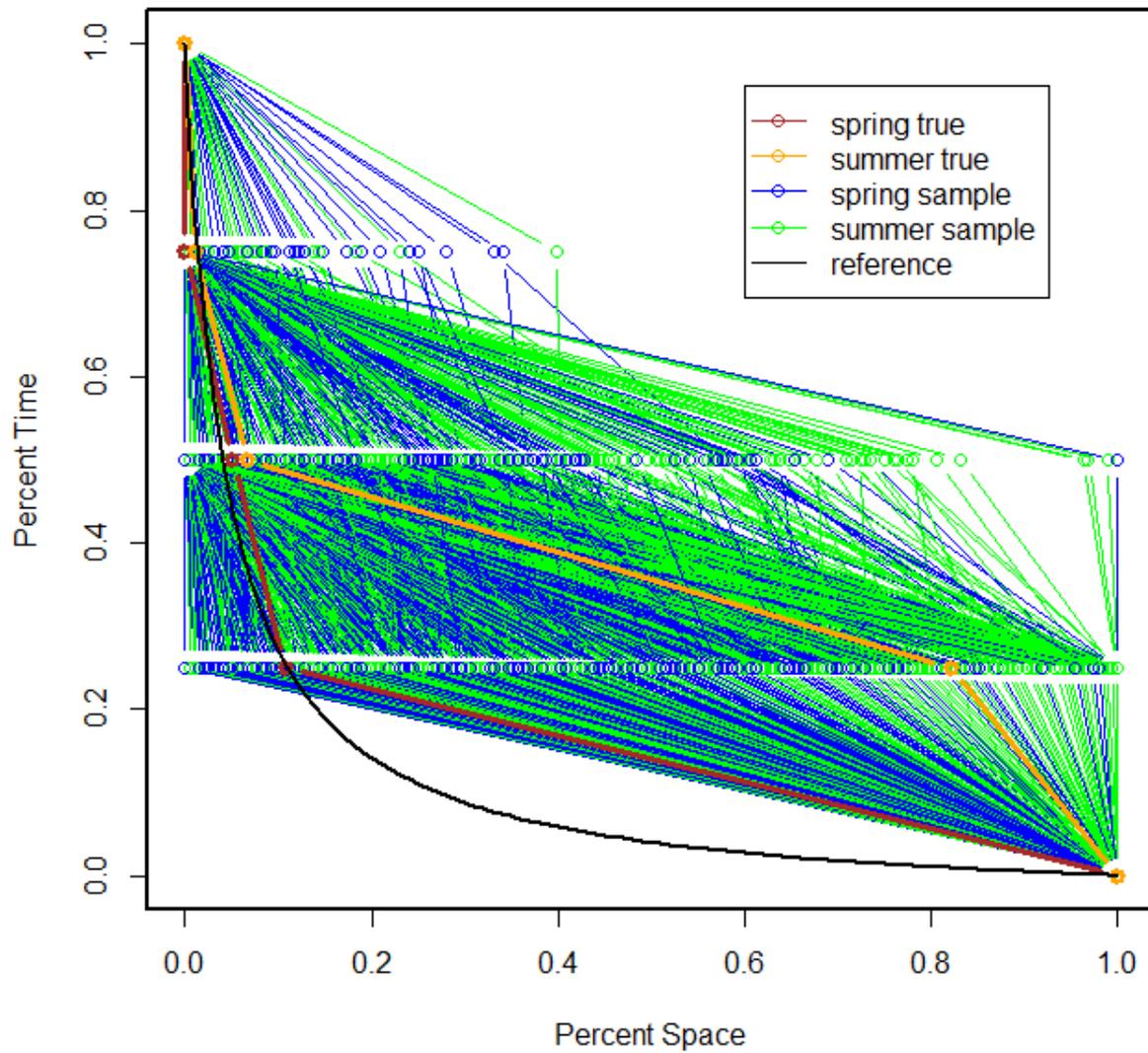


Figure 21. The standard reference curve (black), the true spring and summer CFD for the simulated data that has been manipulated to fail (brown and orange), and 1000 estimated CFDs created by monthly sampling of the manipulated data (blue and green).

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.

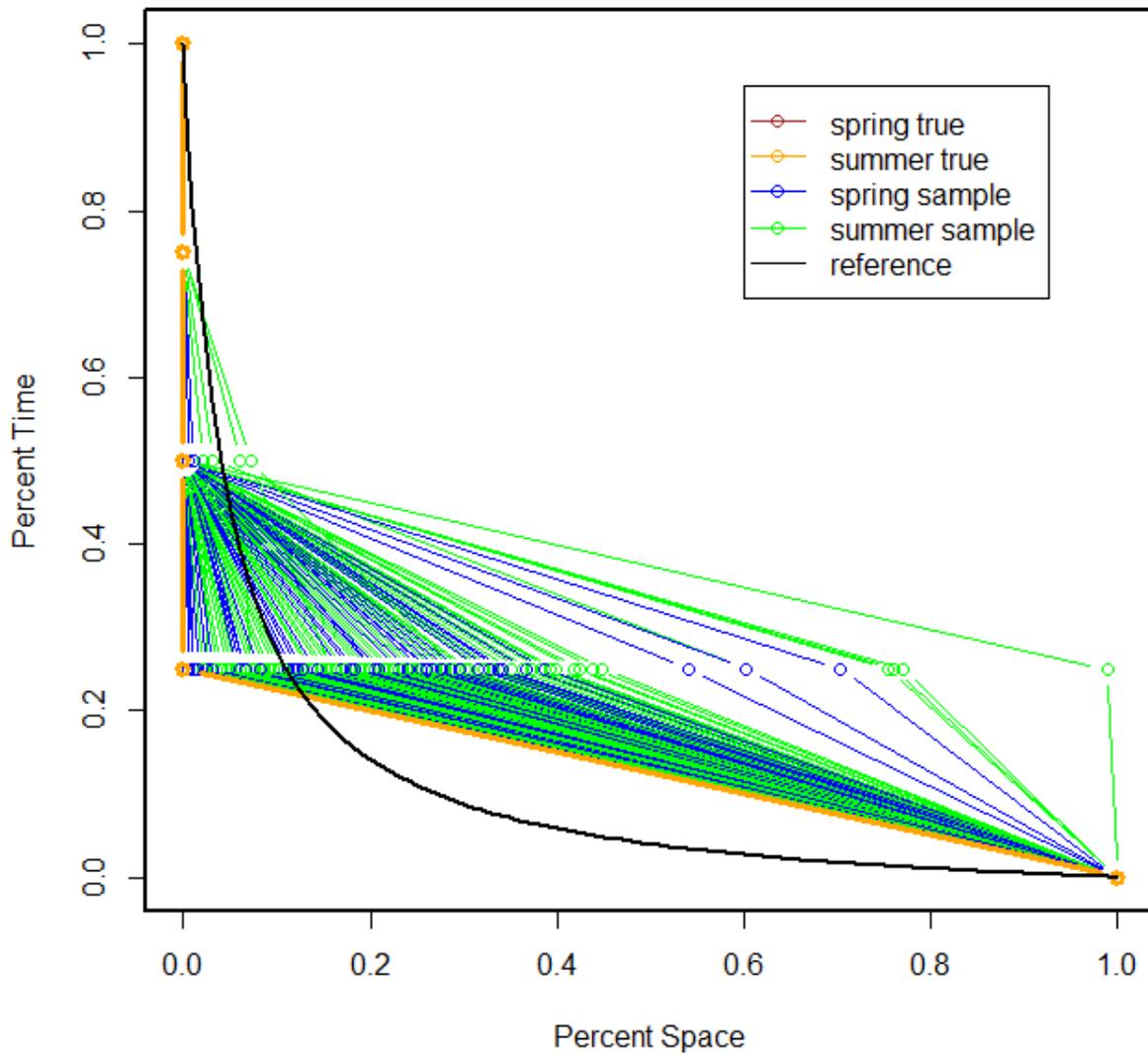


Figure 22. The standard reference curve (black), the true spring and summer CFD for the simulated data that has been manipulated to have a high probability that the estimated CFDs will pass (brown and orange), and 1000 estimated CFDs created by monthly sampling of the simulated data (blue and green).

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 use 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.