PROPOSED ASSESSMENT PROCEDURE FOR BAY INSTANTANEOUS MINIMUM CRITERIA IN OPEN WATER

Abstract

The Bay Program currently does not have an approved method for assessing instantaneous minimum (IM) dissolved oxygen criteria for the Open Water and Deep Water uses. But IM criteria are used to assess the majority of Virginia's waters, using a simple procedure called the "10% rule," whereby a water is listed as impaired if monitoring data exceed the criteria more than 10% of the time. This paper presents a way of analyzing Bay monitoring data so this or a similar rule could be applied to Bay waters. Recommendations for continuous monitoring data are also discussed.

Introduction

Instantaneous minimum (IM) dissolved oxygen criteria are designed to protect against acute effects of hypoxia and are thus more stringent than criteria expressed as an average concentration over a certain duration (e.g., 30-Day Mean). "Instantaneous" refers to the "instantaneous" measurements or observations recorded at a waterbody, as opposed to averaged data. In most States, dissolved oxygen (DO) is primarily assessed against an IM since more expensive continuous monitoring is necessary to assess duration criteria. The criteria can be interpreted as an absolute limit, with no allowed exceedences. But more commonly, IM criteria are assessed with an allowable number or frequency of excursions, given the natural dynamics of water quality, the resiliency of aquatic life, and measurement uncertainty.

In Virginia, dissolved oxygen concentrations in free-flowing freshwater streams and lakes are assessed against the IM using the standard "Percentage Rule". This percentage is typically 10%. All discrete or instantaneous measurements collected in the six-year assessment window are evaluated against the minimum DO criteria, ranging from 4.0 to 6.0 mg/l depending on the official designation of the waterbody (e.g., natural trout waters). If more than 10.5% of the sampled dissolved oxygen concentrations are below the IM, an assessor concludes the waterbody experienced severe hypoxia more than 10% of the time (a condition not conducive for a viable, well-balanced aquatic community) and the water is judged to be impaired. The soundness of this approach rests on the assumption that each instantaneous measurement is independent, adequately representative, and collected with minimal bias.

Representative datasets are the result of conscientious monitoring and assessment procedures. Both are needed to ensure targeted waters are accurately characterized in time and space. Ambient monitoring addresses the latter by generating a relatively unbiased dataset[1] via pseudo-random site visits, without respect to weather events or other "biasing" phenomena. Waters are monitored on a bimonthly basis in Virginia's ambient watershed monitoring program, during both critical (e.g., warm weather months) and non-critical periods. Additionally, because the most recent six years of data are evaluated, the overall assessment is not skewed substantially by sporadic occurrences such as drought

[1] Since discrete monitoring is restricted to daytime hours during non-inclement weather, some bias is unavoidable.

or hurricanes. Spatial representativeness depends on how far monitoring station data are extrapolated. Streams are frequently broken up into “assessment units” that are homogenous enough to be represented by a single monitoring station. The extent of these assessment units varies depending on physical habitat features, surrounding land uses, existing designated uses (e.g., public water supply), and confluence with other waterbodies (see Figure 1). Rarely, a lake may be subdivided if it has a complex shoreline (e.g., inlets and coves).

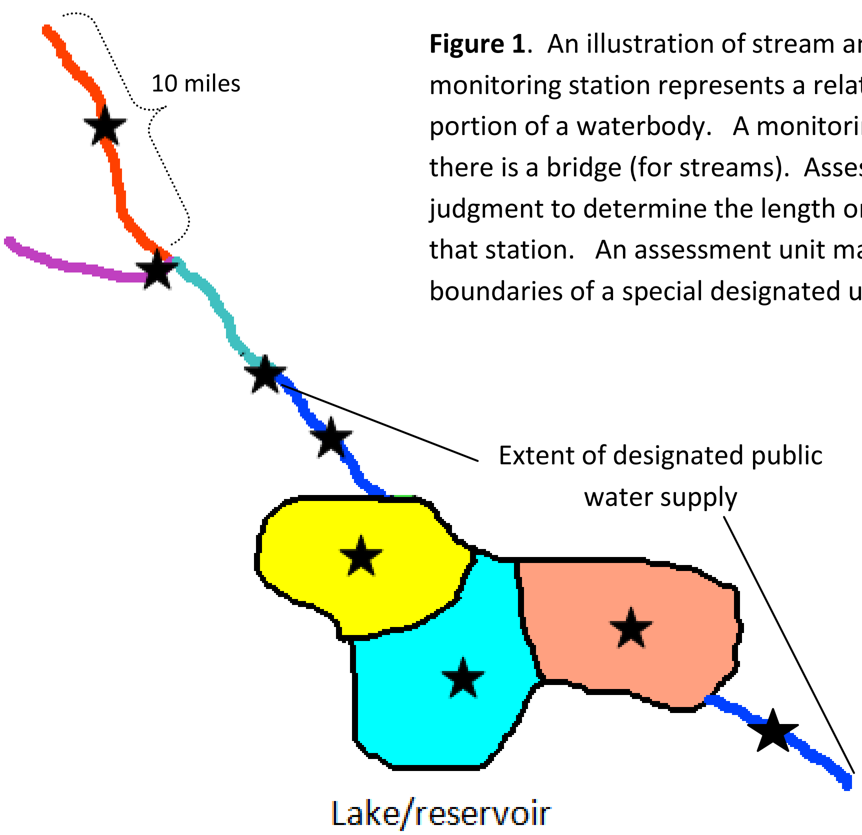


Figure 1. An illustration of stream and lake assessment units. Each monitoring station represents a relatively homogenous stretch or portion of a waterbody. A monitoring station is usually located where there is a bridge (for streams). Assessment staff use best professional judgment to determine the length or area that can be represented by that station. An assessment unit may start or end depending on the boundaries of a special designated use, such as public water supply.

In the case of lakes and reservoirs, aggregating data[2] from multiple stations often makes more sense than analyzing stations separately. The unidirectional flow characteristic of streams is missing in these systems, where pollutants are more diffuse and evenly distributed. Impacts on biological communities also tend to be spread out, especially among more mobile species. Organisms could presumably avoid any hot spots that occur as long as they are infrequent and truly spotty. Moreover, 303(d) listing of a “piece” of a lake[3] does not make sense if all the pollutant sources discharging into the lake contribute to

[2] The exceedence rate is calculated out of the total number of samples collected at all stations, with each instantaneous measurement weighted equally.

[3] In the Commonwealth, assessment units are the “waterbodies” listed on the 303(d) Impaired Waters List and thus prioritized for TMDL development.

the problem.   So from both ecological and management perspectives, pooling data from multiple locations and treating the entire lake as an assessment unit can be a reasonable approach to assessing water quality.

The Chesapeake Bay IM for Open Water and Deep Water remains unassessed due to the lack of approved assessment methodology. The most logical solution is to employ the same method that it is used for non-Bay waters—the "percentage rule"—using a hybridization of the approaches discussed above; that is, a method that incorporates spatial subdivision and data aggregation.   Like the stretches of streams that are dissimilar enough to warrant separation, disparate habitats within an estuary should also be analyzed separately so that degraded waters are precisely captured. But similar to lakes and reservoirs impacted by diffuse sources, estuaries can also be characterized "holistically", with the presumption that a few ambient monitoring stations should be sufficient to characterize its overall quality.   As long as temporal representativeness is upheld by the monitoring regime, an assessment protocol that appropriately separates and pools data would allow the IM to be assessed using the same simplistic procedure used for all other waters of the Commonwealth.

The following issues would need to be considered before implementing the percentage rule.

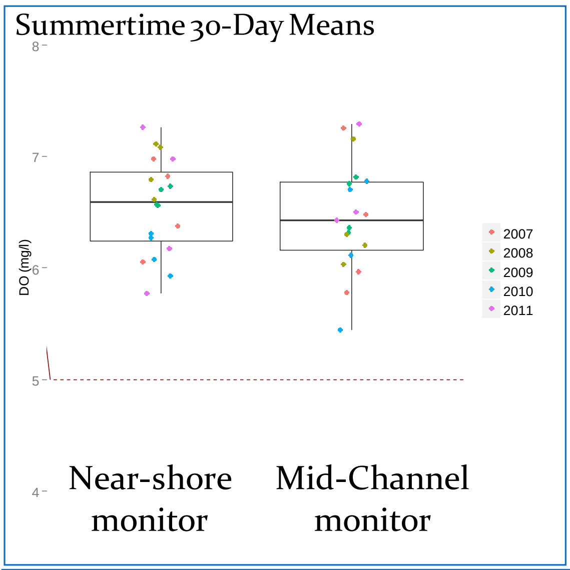
- Subdivision of segments
- Separate evaluation of discrete (fixed station) and continuous monitoring data

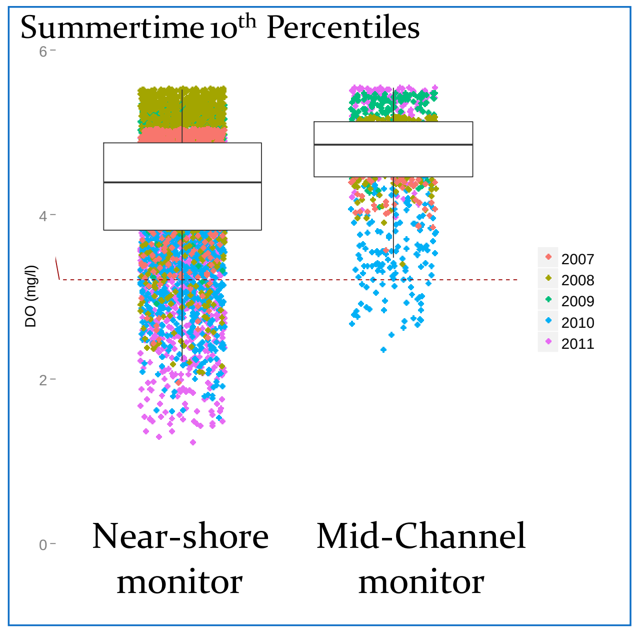
Subdivision of Segments

Most Bay segments are large and heterogeneous, encompassing dissimilar habitats and areas impacted by different pollutant sources.   A robust characterization of these complex waters requires the use of multiple monitoring stations, distributed in such a way that a full "snapshot" of that segment is captured in assessment.   But ideally, as described earlier, "different" waters would be separated from one another so that snapshots aren't so generalized as to be meaningless.   An assessment shouldn't overlook a troubled water for the sake of convenience.

Shallow, near-shore waters are potentially "troubled", due to their closer proximity to pollutant sources and since respiration rates tend to be naturally quite high in these areas. Temperature extremes are also more likely here.   Since these waters also provide critical nursery and refuge habitat, aquatic life are especially sensitive to hypoxia when it occurs in these areas.   As Figure 2 shows, when continuous monitors were placed in near-shore and mid-channel habitats, excursions of the IM were much more frequent and of greater magnitude in the former versus the latter, even though 7-day and 30-day mean concentrations were similar in the two habitats.   Pooling the two datasets together would "dilute" the near-shore exceedence rate, masking the severity of the hypoxia affecting aquatic life residing there. Thus, a sound argument can be for separating these habitats in assessment.

Minor tributaries, as opposed to the major tributaries for which the Bay segments are named after, present another challenge. While it is true that water quality in minor tributaries is impacted by mainstem waters via tidal incursion, these waters are also impacted by their own local pollutant sources.   Indeed, because they are smaller and relatively isolated, reduction in pollutant loads at the local level (subwatershed) will be more evident in minor tributaries before they become manifest in the overall segment. On the other hand, “hot spots” are also more likely to be found in minor tributaries. These hypoxic events may not be evident in the mainstem, but a degraded minor tributary still constitutes a partial loss of the beneficial use. Thus, just as with near-shore and mid-channel habitats, there is justification for separating minor and major tributaries for the assessment of the IM.





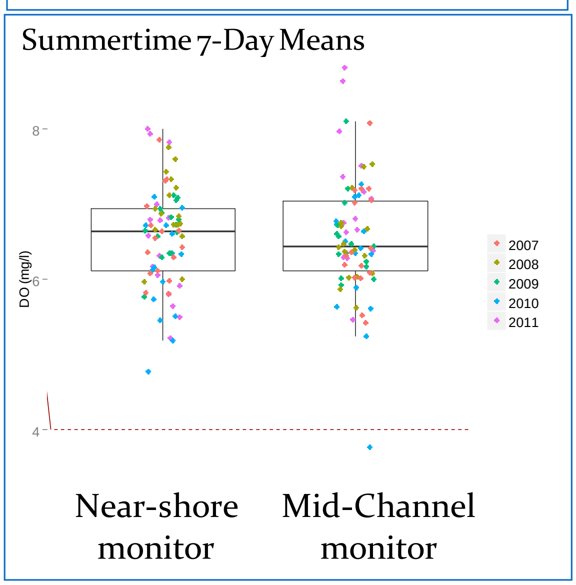


Figure 2. A comparison of summertime 30-day mean, 7-day mean, and instantaneous DO concentrations in near-shore and mid-channel monitors simultaneously deployed in YRKPH. The dashed horizontal lines represent the respective criteria (note that the IM actually has two temperature-dependent criteria but only the lowermost value of 3.2 mg/l is shown in the graph above).

Separate Evaluation of Discrete and Continuous Datasets

A wealth of continuous monitoring data has been collected in Bay waters, providing scientists and managers with insight into the dynamics of DO that would otherwise be elusive. However, despite their high value, continuous monitoring (ConMon) data have not yet been incorporated into the assessments of DO in the Bay. Currently, assessment is based solely on discrete "grab" sampling conducted at a mixture of fixed and probabilistic stations. The assessment of the IM presents an opportunity to take full advantage of ConMon, particularly since most ConMon stations are located in the sensitive near-shore habitat.

However, ConMon data should not be evaluated the same as discrete data. Assumptions that are appropriate to make with a low-frequency dataset break down with a high-frequency dataset. One such assumption is independence of observation. The 10% rule presumes that the aquatic life use can be supported as long as criteria are not exceeded more than 10% of time. Thus, every sample that is evaluated by this rule is assumed to represent a distinct, somewhat randomly-drawn, slice of time. A low-frequency discrete monitoring regime has very little chance of violating this assumption. High-frequency ConMon, however, easily can. To illustrate, the typical ConMon instrument collects data every 15 minutes, corresponding to a dataset of 105,120 observations over three years. Monthly monitoring at a discrete station generates 36 data points. Under the 10% rule, four exceedences at the discrete station would indicate a problem. But if the same rule were applied to a ConMon dataset, a water could experience 100 days of sustained hypoxia and still be assessed as "meeting." This would be unacceptable. So if a percentage rule were to be applied to ConMon, it would have to be much more conservative than the threshold applied to discrete samples.

The "Straw Man"

I combined the above recommendations with the percentage rule to assess the Open Water IM in VA's tidal waters. The separation of the segments was done by applying a 0.5-km buffer to the shoreline of all segments (see Figure 3). This resulting "zone" is quite similar to the 2-meter depth contour used to delineate the Shallow Water-SAV use, and thus captures all the shallow water monitoring program stations. This zone also encapsulates almost all of the minor tributary stations served by DEQ and citizen groups. Thus, a 0.5-km buffer is an easy, though admittedly arbitrary, way to subdivide segments. (For lack of better terminology, I will refer to the stations within the buffer as "near-shore" and stations outside the buffer as "mid-channel.")

Furthermore, I included both discrete and ConMon datasets into the assessment. Samples of the same data type were aggregated, within their respective segment and habitat. For instance, for the YRKMH segment (see Table 1), three exceedence rates were calculated: 1) one for near-shore discrete samples (pooled datasets from fixed near-shore stations), 2) one for the mid-channel discrete samples (pooled datasets from fixed mid-channel stations), and 3) one for the near-shore ConMon station. For YRKMH to be assessed as meeting the IM, all three exceedence rates must "pass" the appropriate percentage rule. Data collected within the most recent assessment window (2010-2012) were evaluated for this exercise.

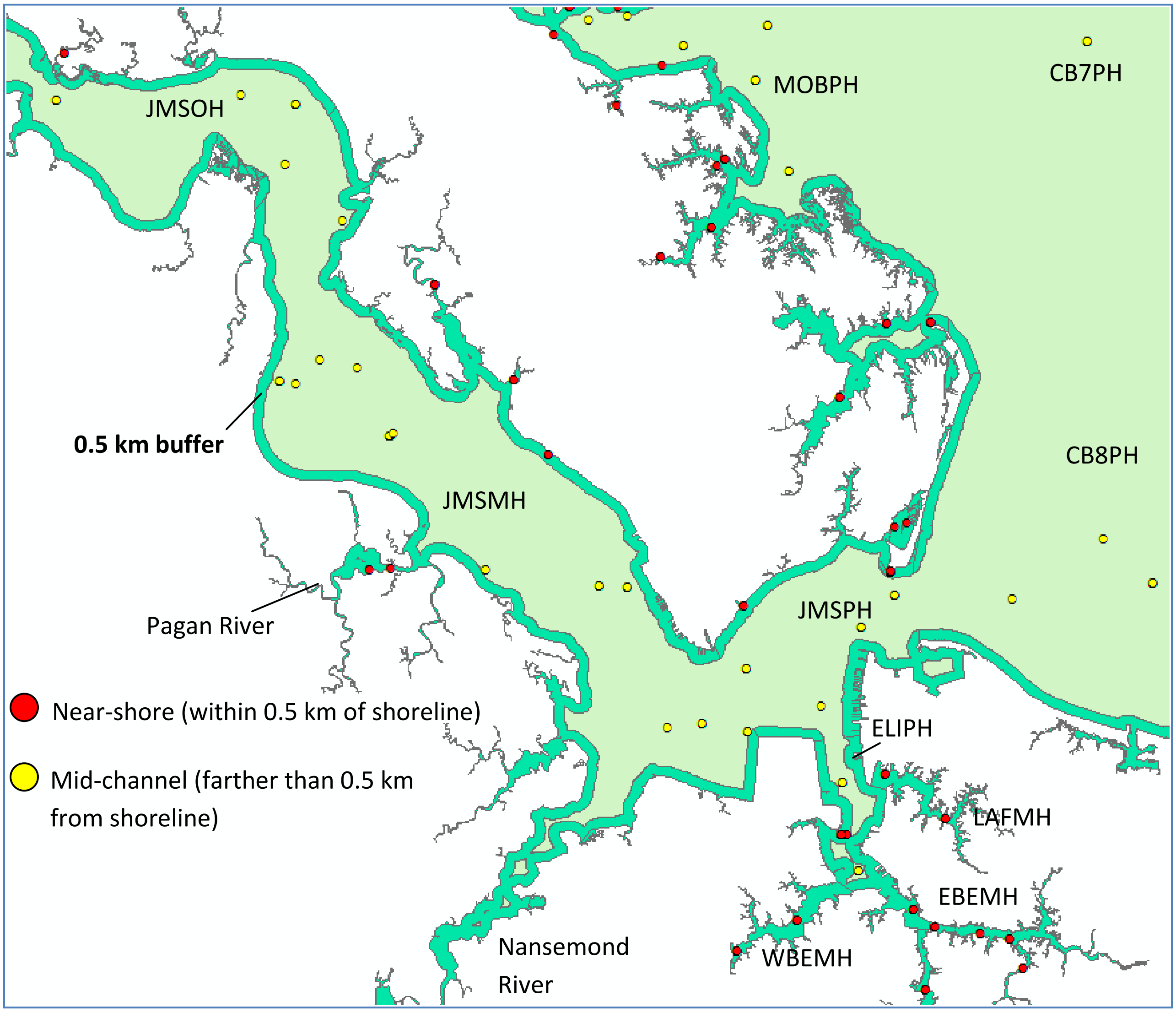


Figure 3. Segment subdivision in a portion of the lower Bay. The 0.5-km buffer effectively captures shallow water monitoring program stations, as well as many non-CBP DEQ stations (particularly in the minor tributaries) and citizen monitoring stations (which also tend to be conducted in minor tributaries). Within some "skinny" segments, such as EBEMH and LAFMH, all stations are considered "near-shore."

(Since this is just a "straw man", only summertime data collected in the first 6 meters of the water column were assessed. But in practice, Open Water should be distinguished from Deep Water whenever appropriate. A "rest-of-the-year" assessment would also need to be done.)

Table 1 summarizes the results.

	EXCEEDENCE RATE%				
SEGMENT	MIDCHANNEL DISCRETE	NEARSHORE DISCRETE	ALL DISCRETE	SHALLOW CONMON	NO. OF MIDCHANNEL/ NEARSHORE DISCRETE DATA POINTS
APPTF	nodata	0	0		0/62
CB5MH	1.6	3.3	1.8		364/30
CB6PH	0.5	0	0.4	2.2	585/87
CB7PH	0.5	0	0.5		1005/24
CB8PH	0	nodata	0		336/0
CHKOH	nodata	0	0		0/44
CRRMH	0	13	11.4		9/52
EBEMH	nodata	0	0		0/51
ELIPH	2	1.4	1.9		88/71
JMSMH	0	1.9	0.7		164/108
JMSOH	0	0	0		70/5
JMSPH	0	0	0		92/159
JMSTF	0	0.8	0.6		263/62
LAFMH	nodata	2.9	2.9		0/68
MOBPH	0.2	0.4	0.3	4.1	2163
MPNOH	nodata	17.4	17.4		0/69
MPNTF	nodata	0	0		0/49
PIAMH	0	0	0		80/15
PMKOH	nodata	32	32		0/19
PMKTF	0	0	0		1/308
POCMH	0	0	0		55/5
POTMH	nodata	3.3	3.3		0/30
POTOH	0	0	0		6/44
POTTF	nodata	0	0		0/75
RPPMH	13	0	11.3		324/39
RPPOH	0	0	0		1/57
RPPTF	nodata	0	0		0/155
SBEMH	nodata	23	23		0/39
WBEMH	nodata	0	0		0/25
YRKMH	4.7	1.3	1.5	4.9	21/302
YRKPH	7.5	1.9	2.3	3.4	40/474

Table 1. Exceedence rates of samples aggregated based on their location and data type (discrete versus continuous). The far-right column shows the total number of discrete observations collected in either habitat. Segments in “red” are those that fail the Open Water 30-Day Mean.

Even though Virginia uses a "10% rule" to evaluate exceedence rates of dissolved oxygen criteria, Bay waters don't have to be held to the same threshold. But a 10% rule is a good place to start. If we apply a 10% cut-off to the exceedences rates in Table 1, the following segments fail the IM: CRRMH, MPNOH, PMKOH, RPPMH, and SBEMH. There are no "surprises" in this list, for all of these segments consistently fail the 30-Day Mean. However, if failure of the 30-Day Mean is a good indicator of a "problem" water, a lot of problem waters also meet the IM. Furthermore, as discussed above, a 10% rule is very lenient when applied to ConMon data. MOBPH could have experienced many days of sustained hypoxia (or even anoxia) between 2010 and 2012, as recorded by at the multiple monitors deployed in its shallow waters, and yet still be judged as "meeting."

Instead of a 10% rule, we could use a 1% rule, citing the complexity of Bay waters as justification to be extra cautious. Applying this threshold results in 14 segments failing the IM: CB5MH, CB6PH, CRRMH, ELIPH, JMSMH, LAFMH, MOBPH, MPNOH, PMKOH, POTMH, RPPMH, SBEMH, YRKMH, and YRKPH. This rule not only captures more "problem" waters, but also two segments which meet the 30-Day Mean criterion (i.e., CB5MH and JMSMH). It also results in the failure of a segment based solely on "near-shore" exceedences (JMSMH). Most of these exceedences were observed in the Pagan River, which is known to have poor water quality. CB6MH and MOBPH fail based solely on ConMon data. A 1% violation rate in ConMon is still quite lenient. But it can be argued that because the error rate is high in deployed instruments and because observations are usually collected just from a single location (which may not be very representative), a lenient "fudge" factor is justified for ConMon.

An alternative to a blanket threshold is to apply a hybrid of the two rules, one that evaluates the data types differently. If we apply a 10% rule to the discrete data and a 1% rule to the ConMon data, nine segments fail: CB6PH, CRRMH, MOBPH, MPNOH, PMKOH, RPPMH, SBEMH, YRKMH, and YRKPH. Three of these fail based solely on ConMon; these also happen to be the only segments where ConMon was deployed during the assessment window. However, one shouldn't conclude that assessing ConMon will always result in a segment failing the IM. For instance, no IM exceedences were detected when ConMon was deployed in APPTF from 2006 to 2008.

Considerations

- *Do we need to have a minimum sample size? For instance, only one DO value was recorded in the mid-channel of RPPOH. If it had exceeded the IM, the segment would have a 100% exceedence rate in its near-shore habitat.*

  In Virginia, we do not make listing decisions based on a single data point. Assessors have the discretion to list a water based on two exceedences, even if only two data points were collected. But usually we try to assess ten or more samples. We could require that a "failing" exceedence rate be associated with at least ten samples (or whatever number is deemed appropriate for that segment and/or habitat). Or we could assess the water "impaired" based on two exceedences but require ten samples for delisting (I like this idea because it provides motivation to collect more data!)

- *Do we need to require a certain "coverage" of station? For instance, none of the near-shore habitat was sampled in CB8PH. Most of the near-shore data in JMSMH was collected in the Pagan River, rather than in the mainstem or the Nansemond River (see Figure 3).*

Ideally, we would want to have stations distributed in a balanced fashion, so that the amount of data collected in each habitat is proportionate to the size of that habitat. This is more of a monitoring issue than an assessment one, however. States are mandated by the Clean Water Act to evaluate all existing and readily available data and information. We have some leeway in how these terms are defined, but we don't have the luxury of assessing until we get a perfectly representative dataset. What we can strive for is a monitoring regime that varies somewhat from cycle to cycle, so that we can approximate full representativeness over the course of multiple assessments. Perhaps DEQ will switch resources from the Pagan River to the Nansemond River in the 2012-2014 assessment. Eventually some benthic probabilistitic stations will land in the near-shore area of CB8PH, allowing that habitat to be assessed. Citizen groups could also be recruited to fill in gaps within their subwatershed of interest.

- *Don't we need to tackle the question of uncertainty? How can you really know that PMKOH has poor water quality when it was only observed 19 times over three years? And if we don't have any information the near-shore of CB8PH, how can we really know that it didn't experience hypoxia?*

A pragmatic way of appreciating uncertainty is by viewing assessment as the "best" professional judgment drawn from the best information available at any given time—a judgment that is restricted to only *that* time. If PMKOH had an excessive number of IM violations between 2010 and 2012, it does not automatically follow that it will still have them in the next assessment interval (2012-2014) or the one after that. However, if there really is a serious issue of hypoxia there, it will likely be detected in more than one assessment. If the hypoxia is not evident in the next assessment and the one after that, we can probably conclude that water quality has either improved or it wasn't that seriously degraded to begin with. It is widely understood that assessments are time-specific and easily subject to change. This is why EPA allows States to remove waters from the 303(d) List as soon as new data become available. This policy mechanism serves to mitigate uncertainty in assessment. Expanding the assessment window from three years would also be a good way to reduce uncertainty.

- *Should every segment be required to have ConMon data? If so, how much data are sufficient?*

A rotating deployment schedule of ConMon could be designed so that an optimal number of segments would be monitored within an assessment period. What constitutes "optimal" ultimately depends on the funding available, but it is also shaped by the amount of value added by ConMon at any given time, for any set of segments.

Rather than deploying in all segments, we could target the subset that is best positioned for "delisting." Currently, these would be segments like APPTF, PMKTF, and PIAMH, which meet the Open Water 30-Day Mean and have no recently observed exceedences of the IM. If ConMon doesn't reveal problems in these areas, we would have sufficient grounds to delist these waters (assuming the 30-Day is more protective of the 7-Day Mean). Then the ConMon could be deployed to the "second-tier" of delist candidates (segments like JMSMH). As long as ConMon is eventually re-deployed in the future and the fixed station monitoring program is maintained, the Aquatic Life use in these delisted waters would continue to be strictly protected and subject to "re-listing" at any time.

Currently, Bay Program-funded ConMons are deployed in a segment for three years before being deployed elsewhere. Three years' worth of continuous data, even if just limited to summer months, represents an embarrassment of riches for 303(d) assessments! In DEQ's assessment guidance, the minimum duration for continuous monitoring datasets is thirty 24-hour periods, with a focus on the "critical period." These conditions allow assessments to be based on the most representative datasets that a constrained budget can buy. If a similar minimum was established for Bay waters targeted by ConMon, this would enable more segments to be monitored within an assessment period.

While I think ConMon datasets provide highly useful information and that we should make a strong effort to assess them, I also think it would be quite unwise to *require* ConMon. This practically guarantees that Bay waters will remain unassessed, despite all the high quality data that are already routinely collected. It is also treats Bay waters with a lot more preciousness than we give the other waters of the Commonwealth. Almost all of Virginia's waters are monitored via low-frequency discrete sampling. Both listing and decisions are made from rather limited datasets, even for parameters (such as bacteria) which are have major public health implications. If ConMon was required for our other waters, our assessments would be rather sparse.

- *A percentage rule doesn't take advantage of all the information that ConMon provides.*

ConMon not only allow us to look at frequency and magnitude, but also duration of exceedences. This is possibly its biggest advantage over discrete monitoring. Assessment methodology can specify a minimum duration even if this is not explicitly expressed in regulation. We could allow a segment's ConMon to violate 1% (or 0.1% or 0.01%) of the time, but if a monitor records hypoxia lasting 1 hour or longer, we could deem that water as "failing" the IM. Or, we could allow for a set minimum number of "hypoxic hours", provided concentrations never drop below X and there is a minimum "recovery period" of Y hours/days/weeks. I don't recommend a rubric that is much more complicated than this, but we do have some leeway in tailoring a procedure that takes full advantage of ConMon's strengths.

- *What about the Deep Water Use?*

The percentage rule could also be applied to segments containing the Deep Water use. Because this use tends to be much more "mid-channel" than "near-shore", it is probably not even necessary to subdivide segments. The only challenge would be in how to distinguish Open Water from Deep Water in the absence of a full vertical profile.

During most visits of stations located in the Deep Water Use, a vertical profile is taken, thus allowing for the calculation of an "instantaneous" pycnocline depth. However, this won't always be the case. For instance, benthic probabilistic monitoring usually samples DO only at the bottom. For 30-Day Mean assessments, these single-depth observations are placed into either the Open Water or Deep Water use based on pycnoclines calculated from nearby vertical profiles sampled within the same two-week interval. A similar procedure could be used in the IM assessment. For instance, the presence/absence of the pycnocline at LE4.3 on August 15 can be used to estimate whether a benthic monitoring grab sample collected elsewhere in YRKPH on August 18 should be assessed as an Open Water or Deep Water observation.