Criteria Assessment Protocol Workgroup (CAP) Meeting

Thursday, July 6, 2023
10:00 AM – 12:00 PM
Meeting Materials
This meeting was recorded for internal use to assure the accuracy of meeting notes.

Actions and Decisions

- ✓ All: Check your email for a schedule poll to schedule an early Sept. CAP meeting.
- ✓ Peter Tango (USGS) will share with the group any references/publications that show saturation of DO is more important for fish than concentration.
- ✓ Tom Parham (MD DNR) will share with the group in the future his work looking at DO and temperature impacts for striped bass resident fish and thresholds for habitat suitability.
- ✓ The CAP WG can consider presence/absence studies of species plotted with temperature, salinity and DO data to inform saturation vs concentration.
- ✓ Over the next year, small working groups will: 1) interrogate past monitoring data to understand the impact of temperature driven effects, 2) test how criteria changes would affect attainment of different segments.
    - August will coordinate. If interested in joining one or both of these discussions, please contact agoldfischer@chesapeakebay.net
- ✓ All: Investigate research options for development of new bioreference curves (research need – potential support from STAC?)
- ✓ Investigate usage of Integrated Report Category 3. Leah Ettema (EPA) and Peter will talk about this next week.
- ✓ Decision: In the interim of new bioreference curves being developed, use the 10% curve.

Participants

Amanda Shaver (VA DEQ), August Goldfischer (CRC), Becky Monahan (MDE), Breck Sullivan (USGS), Cindy Johnson (VA DEQ), Claire Buchanan (ICPRB), Clifton Bell (Brown & Caldwell), Dave Parrish (VIMS), Gary Shenk (USGS), Jacob Greene (MDE), Juan Vicenty-Gonzalez (EPA), Leah Ettema (EPA), Mark Trice (MD DNR), Matt Stover (MDE), Peter Tango (USGS), Rebecca Murphy (USGS), Sophia Grossweiler (MDE), Tish Robertson (VA DEQ), Tom Parham (MD DNR), Tony Timpano (VA DEQ), Richard Tian (UMCES)

Minutes

10:00 AM **Welcome, Introductions & Announcements – Peter Tango** (USGS), Chair

Upcoming Conferences, Meetings, Workshops and Webinars

- Chesapeake Studies Conference – September 15-16, 2023, Salisbury University, Salisbury, MD.

- Potomac Conference – September 21, 2023, Lorton, VA.
- Virginia Water Monitoring Conference – September 26, 2023, Henrico, VA.
- Chesapeake Watershed Forum – November 3-5, 2023, Shepherdstown, VA. Session proposals were due June 11. Poster proposals are due July 28.
- CERF 2023 Conference: Resilience & Recovery – November 12-16, 2023, Portland, Oregon.
- National Conference on Ecosystem Restoration – April 14-19, 2024, Albuquerque, New Mexico. Abstracts are due September 1, 2023.

Jacob Greene, new employee at Maryland Department of Environment (MDE), was introduced.

10:10 AM Considerations on creating temperature-adjusted dissolved oxygen criteria – *Peter Tango (USGS)*

Discussion:

Tish Robertson (VA DEQ): I support having temperature dependent criteria. However, I don't think it will be easy to come up with temperature dependent criteria, and we need something in the interim. We also need something to show to the public and the EPA why we need temperature dependent criteria. The analysis you've shown and Rebecca's maps are compelling, but we need a way to show that impairments we're finding are temperature driven while avoiding the perception that we are trying to dumb down our criteria or standards. I think if we can show we're seeing exceedances we wouldn't see without climate change or high temperatures that could go a long way. Coming up with an assessment process where we use our current criteria but show some proportion of our exceedances can be attributed to high temperatures would go a long way to providing the motivation or impetus for temperature dependent criteria.

Richard Tian (UMCES): Regarding saturation vs concentration, can you share any references/ publications saying saturation is more important for fish than concentration? *Action item.*

Peter: Yes. Fisheries biology tells us that it is partial pressure, or the difference between their internal pressure and external pressure, that drives the oxygen across that gradient if conditions are good. With high saturation you have a strong gradient. Let's do that comparison between pressure, saturation, and concentration.

Matt Stover (MDE): How much would it actually change if we move to saturation? Referencing our conversation from yesterday, is this the highest priority when it comes to assessment? If it's only changing the criteria by a little, for example, changing the criteria from 5 mg/l to 5.1 mg/l, changing any criteria in state standards is an onerous process that takes a lot of work. If we go down that road, we want to make sure there are substantial changes to be gained from that level of effort and it's worth it.

Peter: One way to think about it is there are a fair number of segments within a fraction of attainment. Would it make a significant difference with that point one? If it does, we can gain a rapid understanding of how many segments are coming into attainment. Frankel's paper had a statistic of uncertainty – double digit average percentage (6-20%?) of the impact of temperature on hypoxia and how much less we might have had we not had the temperature change. Impact over 30 years seems to be the important factor. Whether that translates to change in attainment calculations, I'm not sure.

Matt: Is the greater concern regarding the solubility of DO, or from the fact that fish metabolism would ramp up and they would have greater DO requirements? Knowing how the DO criteria are made, the DO requirements for a species wouldn't change, it's the DO in the water, but maybe there is research to show that striped bass's metabolism is increasing due to increased water temperatures and they have a higher DO requirement. Do you know if the solubility or metabolism impacts are a greater concern?

Peter: The sensitivity of striped bass may have been one of the drivers setting that particular criterion. In the EPA methods for criteria setting, in their analysis to gather species sensitivity and rank them, the foundation of that study comes down to four most sensitive species which weighs into guiding criteria threshold that's set. Striped bass has tolerance for temp down to FL. I don't recall the other 3 species. I'm not hearing a lot of species extinction. Thinking of applying the same method, if not reporting change in species composition, do we need to change our assessment? Is something else wrong in the assessment? I'm looking at the practical side and saying I'm not seeing rapid extinctions. If the foundation shifted, we'd have a strong case for whether the criteria are working or not. That's a devil's advocate for doing temperature without looking at the biology. Without a doubt a change to solubility will change fish ability to obtain oxygen in the worst case scenarios. There's that link but so far the biology hasn't told me we've lost species.

Richard: I calculated that DO solubility decreases from 12.4 mg/l at one degree C to 6.2 mg/l at 35 degrees C. I calculated the variability of solubility at 3.2 mg/l per degree.

- ✓ Richard's paper is linked here.

Clifton Bell (Brown & Caldwell): A New England state water body was meeting the concentration criteria but not the solubility criteria. People went back to what EPA wrote about freshwater criteria in 1986 and why they chose concentration over saturation for the DO criteria. They concluded that concentration was preferable compared to percent saturation or oxygen partial pressure and is scientifically sound because the rate of oxygen transfer across fish gills is directly dependent on the mean difference in oxygen partial pressure across the gill. The total amount of oxygen delivered to the gills is a more specific limiting factor than oxygen partial pressure per se. There are a few other references from the EPA where they say you don't need saturation. I bring this up because if you pursue saturation, we'd need to deal with some of this preceding guidance.

Tom Parham (MD DNR): For the past couple years we have been looking at DO and temperature impacts for striped bass resident fish and thresholds for habitat suitability. We scoured tons of papers to pull together what different habitat conditions for the fish. The majority of information is in mg/l. Saturation is a great way to go, but the data is not there. I'm all for using it but when you're trying to set these according to living resources it's going to be hard to build these thresholds because at least for striped bass that data is not there.

Peter: If there's a reference list, please share it. I've seen San Francisco Bay trawl survey data with temperature, salinity, DO data plotting presence/absence of species. That would be another body of research to look at presence/absence and whether there are any red flags or changes through time. *Action items.*

Tom: We're trying to wrap up the manuscript by end of the month. We've worked with modeling group looking at these thresholds. We looked at bioenergetic studies and a whole suite of research and it's just not there for the saturation.

Peter: Sounds like there's interest in putting together a summary of changes in temperature driven impairments and what influence changing the standards and criteria may have on this.

Mark Trice (MD DNR): We don't record percent saturation. That could be back calculated, though.

Tish: As a workgroup, we should think about coming up with a way to interrogate our monitoring data to see temperature driven effects. That would answer the question on how much of an impact this is causing. Before we commit to something as ambitious as temperature dependent criteria, we need to establish a practical reason. I think there's a good reason, but we don't really know how many of our exceedances can be traced to temperature.

Richard: I cannot expect that they are significant, because when we talk about water temperature increases the maximum increase would be 1 degree. 1 degree changes solubility about 0.1-0.2. That's the level of change we're talking about. I think Matt is right in that it's not a big change in solubility from temperature over the time we've collected data.

Peter: That's the variability and uncertainty we're dealing with. At the moment, with so few data points with the station resolution we have, it's tough to detect that with a high degree of certainty. I agree with Tish that interrogating some of the monitoring data would be a good idea. At the UMCES monitoring pier, there's a significant temperance difference from the 1930s to now. With the level of uncertainty in our analysis, maybe significant solubility changes are not detectable for another 50 years. It sounds like there is an opportunity to put together some sort of test to take the attainment information we have, seeing how many are within a fraction of attainment, and get a sense of those that may be most sensitive to a small adjustment of criteria.

Action item for workgroup over next year. If there are volunteers to work on that, let Peter know.

Matt: Any consideration of establishing water quality criteria, it's a heavy lift. Before we go down that road, there's many considerations and different factors to evaluate. It's a good idea to do some tests with the data we have first.

10:40 AM Beyond 2025 – Considerations for retaining or revising the Water Quality Monitoring and Assessment Outcome language of the 2014 Watershed Agreement – *Peter Tango (USGS)*

Discussion:

Tish: Something that speaks to our ability to assess all of the criteria would be good. That would probably go under the reporting component in the Outcome. I get why we don't have standards attainment as a goal, but at least being able to assess all of the standards could be a goal.

Peter: We could at least put a 0-100 scale of how are we doing in our Outcome language. The Outcome itself doesn't specify when we say monitoring what dimensions are we targeting, and there's lots of monitoring going on.

Breck (in the chat): I think the reporting part should be kept because I think that is something STAR, ITAT, and NTN does very well. And our partners release great materials!

Matt: I would second what Tish mentioned. If you wanted to put something more measurable. Being able to comprehensively monitor a segment, and show our stakeholders that we can comprehensively assess, is what we want. It's tough to message around this Outcome if we can't do that.

Peter: If it's not measurable, it's tough to manage. The history of work suggests we had fewer data points over time and level funding, so there was a lot of tweaking to still be able to get valuable information. We have the 2003 EPA document. And we have new tools and insights. I think we can do a better job of explaining our capacity to monitor.

Matt: There's positive stuff happening – more nontraditional partner monitoring is coming in through the Chesapeake Data Explorer and I hope more of that can be used in assessments. MDE has ramped up monitoring. We're monitoring more in the Bay than we have ever have before in a collective sense despite level funded budgets. Putting a numeric target on something doesn't look great if we don't meet it, but if we show we're not meeting a numeric target, it provides an impetus to get more funding.

Peter: For water clarity and Submerged Aquatic Vegetation (SAV) there's a lot of work on use of satellite imagery. We are lagging a bit on the water clarity side. But chlorophyl and Harmful Algal Blooms (HABs) quantification has continued to evolve because of coordination between satellite imagery and on the ground info. We could even set up a StoryMap of where we are and where we're headed in our monitoring capacity. That should offer an opportunity to make it scalable.

11:20 AM Developing decision rules on dissolved oxygen criteria attainment using 4-D interpolator output – *Peter Tango (USGS)*

Discussion:

Tish: The deep water 30-day mean criterion is assessed with the bioreference curve, and the open water 30-day mean criterion is assessed with the 10% reference curve. And the deep channel instantaneous minimum criterion is assessed with a 10% reference curve. And that's the only instantaneous minimum we currently assess.

Peter: Yes. We don't have a bioreference curve in the open water.

Tish: All 3 uses have instantaneous minimum criterion, but we're only currently able to assess the deep channel.

Action for Peter: Update slides/tables accordingly with clarification on these details.

Peter: How are you thinking about recommending what we do for curves for 1-day, 7-day and instantaneous minimum criterion? Is this something we need research on? Or are we ready to adopt 10% as interim criteria, if possible?

Tish: The open water are difficult to have bioreference curves for. Jeni Keisman documented in a 2010 addendum – she tested 10% reference curves we were using and found for the deep water 30-day mean, the bioreference curves she developed work for capturing the distribution of the good BIBI scores that we had in the database at that time. Whereas we couldn't do something like that for open water, which makes sense because the open water use is not to protect benthic communities, it's more to protect pelagic life, and we don't have a reference fish community. I don't know how we would do open water with bioreference curves unless we become more specific with fish communities. We'd have to have an IBI for fish which we don't have. The bioreference concept works for deep water and deep channel. Deep channel is only one criterion, instantaneous minimum, and I think we're already using a 10% curve. Jeni concluded that was good enough. It is the deep water we have to figure out what we'll do with. We're definitely using a bioreference curve for 30-day mean for that. And we have the migratory fish spawning use we're supposed to be assessing and that has the 7-day mean and the instantaneous criterion we're supposed to be assessing as well.

Richard: The deep water bioreference curve is approximately 10% (correction to what Tish said). When tracking, develop some kind of research program, and develop a bioreference curve for each criterion, if possible. And use 10% as interim to proceed with criteria assessment. But in the long term if we can get a specific bioreference curve we should, I think.

Peter: Maybe STAC can put a team together for revisiting this.

Tish: To me the interim decision rule would be using a 10% curve, if we're sticking with a CFD approach. That's what they 10% default is developed for. Then as the science develops, we can insert a bioreference curve.

Peter: Sounds like a good place to start. The full 4-Dimensional Interpolator's assessment won't be applied until 2027/28. So, what do we do for the next 5 years. This will be helpful for testing the 4-D interpolator. Right now, we're limited to select criteria; can we use a weight of evidence from continuous monitoring data to speak to those short duration criteria evaluation that have been on hold, and how would EPA view that. Maybe we need to get a small team getting together with key EPA people to approach that in the meantime. As well as planning for 4-D application and setting up the decision criteria.

Science need and action item: Research to support development of additional bioreference curves. Research to support reference for most sensitive fish groups for setting each of the criteria.

Tish: We've been waiting 2 decades. It seems like now we're just kicking the can…time to do something.

Peter: I'm fine putting that (10%) as a solid starting point.

Tish: If we do use 10% CFD…I'm totally agnostic on whether we keep using the CFD. I'm open to other approaches and I don't think the 2003 document bars us form using other options. If we want to move swiftly, then using the existing framework is fine. In the past you mentioned about rounding. I think we should. I am interested in seeing what other jurisdictions think. I think there is a strong basis for us to include in our framework a nod at the uncertainty there.

Peter: Richard, is there inherent rounding in your programming? The only thing we're not rounding is that actual final result.

Richard: The rounding is not in the program, but we practice that in the modeling world. If we are within one point of attainment, we are considered attained.

Gary: 1.49 was a policy decision by the PSC, based on their judgement that we weren't looking at an assessment from data, we're looking at model results and projecting to the future. If you get within 1% that's as far as we're going to go with the TMDL. They're not saying that 1% nonattainment equals attainment; they said we won't set our nutrient allocations any stricter than 1%. The idea behind that was we got closer, we could zero in on what we need to do for nutrient reductions. Then in the second round, they said 1% rounded down; 1.49 was the exact number. But that was completely a policy decision.

Gary: I completely understand that we can't tell if we're in attainment or not if we're in violation by .01% of the space time. We can't tell if we're not in violation if we're just missing it by .01% of the space time. So why do we round and give extra violation where we don't say we were almost in violation maybe we should do something.

Tish: We don't have the ability to measure going to the gazillionth place so how we can defend a decision going out to 6 zeros.

Gary: I agree it's hard to defend. There's no way to absolutely say we are or are not in violation. If we are making an estimation of uncertainty, do we have the uncertainty of making sure the ecosystem is protected or not. To me that's a policy decision, not a technical decision.

Peter: We did factor in the 10% curve as a buffer but then seems to be uncertainty on top of that whether or not we're making the 10%.

Tish: That CFD approach is really unforgiving because we're not calculating the full area under curve and saying it's 10% over. We're looking at each point along the CFD, and whether it is exceeding at each of those points. If it's over at all, even 0.001 over, we're saying that means something. My point is that it doesn't mean something. If we go to court over our assessments, we'd need to say we can measure to that degree, but we can't measure to that degree. Maybe in the future we will be able to. But not right now. I want an assessment rubric that is technically sound and that we can defend, not making stuff up. I like defending how we're doing on a technical basis. That's why I think of it as a technical matter, not as a policy matter, because we can measure uncertainty and we don't have to come up with an arbitrary number. We can make it so we're titrating to the bare minimum. I'm not saying we would round to a whole number. We just shouldn't be treating all our digits as significant because we don't do that in the regulatory setting.

Gary: I agree with being honest about our uncertainty. In the load allocation part of TMDL, you have to prove that your limits will meet water quality standards, not other way around. TMDL load = point source plus non-point source plus margin of certainty. The greater the margin, the more reductions you have to make. If we say our uncertainty is +/- 3-4%, I don't know if we automatically round down but from standpoint of TMDL we round up. Anything not proven to be in violation is in violation. My concern is if we open this can of worms, we make it impossible to meet these criteria.

Tish: I'm not a TMDL expert but I know here what we do for assessment is the same as what we do for the TMDL – in other words if we round in assessment, we round in the TMDL simulation. When we have differences between the assessment and the TMDL, it's usually that the TMDL is more stringent. We usually have no exceedances allowed in the TMDL, but in the assessment we allow 10%. In the Bay TMDL, it's almost like we're less stringent in compliance tests than in assessments, which is reverse of what I'm used to.

Gary: Agreed.

Peter: It's insightful to compare insights from the research and regulatory worlds.

Tish: When I look at the spreadsheet that Richard and Qian produced, when I look at those exceedance rates that are ridiculously small, I can't wrap my head around it. Especially if it is consistently like that. I hear what you're saying about erring on the side of caution. But we have processes in monitoring and assessment if we make the wrong decisions in assessment, in the next assessment we'll capture that. We're not walking away from any of these segments when

we meet a criterion. We have infrastructure to constantly check up on these segments. I don't think it's a huge consequence to say it's meeting it, if it's not. How much do we care about being right or wrong?

Peter: I am fairly sure that rounding will not suddenly turn the entire Bay to a healthy Bay in our attainment assessment. We can look at how many of these segments are less than 0.1 or 0.1 and think about the impact of temperature adjusted criteria as well as this issue of rounding. It will help us think about the soundness and defensibleness of our decisions.

Matt: I agree with Gary in some respects. If we draw a line, whether policy or technical based, then we're beholden to that line. Our tools and criteria are imperfect but they're the best we have. With the integrated report, we do have another category we can put them in, which we haven't really considered previously: Category 3, for those straddling the boundary of attainment. Category 3 is "we don't know". Admittedly it's not great to say we're doing all this monitoring and that we don't know the actual attainment results. But if we are open and honest, maybe it's not a big deal. We can tell them we're almost near attainment. We would have to have an addendum or overwrite old documents because we default to the worst Category, Category 5. Some of ours, if we're only getting the 30-day mean criteria, which we've been assessing for most of this time, even if we're meeting it in some cases we haven't been able to delist those waters. That's where I feel like it's super critical to be able to assess the full DO suite if we're going to stick with that decision process. Otherwise, maybe we need to revisit that.

Claire Buchanan (in the chat): A fuzzy line is needed?

Amanda Shaver (in the chat): We think definitely a little thicker of a line, Claire :)

Gary: I like that idea. Not sure how it would be used in policy. As Tish was saying, being honest about that fact for some segments we don't know if they are in attainment or not for a 3-year period. If we had categories of 90% sure in attainment/90% sure it's not in attainment would be something to consider. I agree with Tish that the 10% reference curve is unforgiving. When you have those incursions above causing these small numbers, I have seen that a lot. Maybe we need to revisit that reference curve and work on more bioreference curves. The deep-water curve was very successful in that we had 100% of designated use years that met a BIBI of 3 or greater, and didn't have violation. And 100% of those with BIBI of 1 or 2 had a violation. It was a great reference curve. And that reference curve is not shaped like the 10% reference curve. So the two things that would be helpful are 1) having a category of we don't know, and 2) having more bioreference curves that are shaped differently than that unforgiving 10% curve.

Peter: In our small group conversation yesterday we talked about the need for new bioreference curves. On the other hand, when we're doing the 4-D interpolator output without new bioreference curves yet, we'd apply the 10% curve in the interim so we can make a complete assessment using the tool.

Peter's summary: Going forward, we will investigate concentration vs saturation vs partial pressure. We will summarize existing work including publications and EPA standards reporting. As a workgroup, over the next year, especially after Tom Parham's work comes out, we should aim for a summary paper that captures reasons for using one or the other, and the significance. We can utilize Richard and Qian's work on the attainment deficient, and investigate how sensitive our trends are likely to be/what the impact would be with tweaks such as temperature or rounding. The challenge we are dealing with is changing criteria is a big decision, and not to be taken lightly.

Action: August will organize a small group to discuss the impact of temperature further. Interested: representative from MDE, and Gary Shenk (USGS). Let August know if you are interested (agoldfischer@chesapeakebay.net).

Action and research need: Investigate and potentially develop new bioreference curves.

Decision: In the interim of new bioreference curves being developed, use the 10% curve.

Action: Category 3 – talk with EPA regarding using this category. Leah Ettema (EPA) and Peter will talk about this next week.