Chesapeake Bay Program

Science. Restoration. Partnership.

Criteria Assessment Protocol (CAP) Workgroup Meeting

February 9th, 2026
9:00 AM – 3:00 PM
Visit the meeting webpage for meeting materials and additional information.

Purpose: In this meeting, the group heard from Leah Ettema (EPA) with the second part of her presentation on Dissolved Oxygen Assessment Methodologies. The first part of this presentation was given at the December CAP WG Meeting. Matt Stover (MDE) shared the findings from his survey of Clean Water Act Administrators from other states. Then, Peter Tango (USGS) and Qian Zhang (UMCES) shared the history of the Stoplight Chart and explained the Water Quality Standards Attainment Indicator. Next, Breck Sullivan (USGS) dove into the revised Watershed Agreement, specifically the changes to the Water Quality, Standards Attainment and Monitoring Outcome and next steps. Lastly, the group participated in a discussion, diving further into the points discussed earlier.

Minutes

I. **Welcome, Introductions & Announcements**
   *Lead: **Peter Tango** (U.S. Geological Survey, USGS)*

   Upcoming Conferences, Meetings, Workshops and Webinars

   - Choose Clean Water Conference – May 18-20, 2026. Lancaster, Pennsylvania.
   - Chesapeake Community Research Symposium – June 1-3, 2026. Annapolis, Maryland.
   - Restore America's Estuaries' 2026 Coastal & Estuarine Summit – September 22-25, 2026. San Francisco, California.

II. **Background and Approaches to Dissolved Oxygen Assessment Methodologies, Part 1: Instantaneous Criteria Review**
   *Lead: **Leah Ettema** (U.S. Environmental Protection Agency, EPA)*

   In this presentation, Leah synthesized information from the various forms of Bay criteria documentation to bring cumulative understanding and posed questions on topics that need to be further discussed. She began by reviewing the conclusions found and questions posed in part 1, which was presented at the December CAP WG Meeting.

   Next, Leah explores a question she thought about after the December CAP WG Meeting, "why can't instantaneous minimum be assessed using the cumulative frequency diagram (CFD), but the deep channel can?" This is explained in the 2007 Tech Addendum.

   The next question asked is "when did the concept of the 4-d interpolator originate?" Leah explains that it was first found in the 2003 Criteria. Next, she moves into the 2010 document, where she revisits the question "why are exceedances allowed of the instantaneous minimum?"

Then, she focuses on the 2017 document, which talks about assessment methodology in detail. There are three approaches that are covered. These are direct assessment with enhanced monitoring, conditional attainment, and continuous data assessment methodology, which Leah describes. As she explains state methods for continuous data assessment, she also covers the three approaches to exceedance frequency. Then, she shows examples of what the region three states are doing for this. She also asks questions that are crucial to answer when calculating exceedance frequencies. These must be documented to ensure replicability. She explored existing state methods for critical periods of monitoring, combining data types, overwhelming evidence factor, and spatial considerations.

As Leah continues to work through the 2017 document, she explains the three zones, where they came from, how they were delineated, and how they are used to analyze anoxic conditions. She also explores how they impact assessment, their assessment procedures, and analyzing assessment mean. Like Tish mentioned in the discussion, there are limitations with the assessment mean method of Rule 2-Alt and may not be acceptable. Leah continued to explore the recommendations laid out in the 2017 document and she poses questions for the group to consider moving forward.

Actions:

1. Revisit the 2017 document's continuous data recommendation of Rule 2-Alt: "No more than two consecutive days with 10 percent time (>2.5 hours) exceedance during a single season. This translates into 6 or more hours or about 0.2 percent of the summer season." There are concerns about allowing high levels of exceedance if they aren't on consecutive days and there are concerns that this doesn't apply to discrete monitoring. When updating and finalizing protocol in the target of the new Watershed Agreement, revisit this topic and decide on the best path forward.

Discussion Notes (during presentation):

Q: Tish Robertson: Why wouldn't we make the same assumption about the instantaneous minimum and deep channel for the deep-water use?

- ***A:** Leah Ettema:* I don't know.
- ***Response:** Tish Robertson:* If the issue was not being in the photic zone, then we would apply the same rationale to deep water.
- ***Comment:** Joe Morina:* Could it be physical isolation? There's a geomorphology to that versus deep water and open water.
- ***Response:** Tish Robertson:* Maybe. We have deep water that is technically in deep trenches.

Q: Leah Ettema: Are there continuous monitors in the deep water?

- ***A:** Peter Tango:* Yep.
- ***Response:** Leah Ettema:* I was also curious how we know that the variability is the same.
- ***Comment:** Mark Trice:* We also have our own profiler that takes deep water measurements. In the past, we operated NOAA's as well. I guess that's not technically deep water. Gooses is the deepest.

Q: Tish Robertson: When you say it's not technically deep water, do you mean it's not in the deep water use?

- ***A:** Mark Trice:* Right. We did run the vertical profiler in the Bay at 10 meters, but it was so hard to operate that we didn't leave it there more than a year.

Comment: Rebecca Murphy: Some of the recent NOAA vertical arrays go through the deep water and deep channel. That might be what you're mentioning with Gooses, but this is the since 2022 data in the Lower Potomac, Mainstem off the Choptank, and Gooses Reef area. We do have some of that data, but it was definitely not there when the 2007 Tech Addendum was written. There was a desire to assess the deep channel instantaneous criteria, and this was a pretty good justification, but deep water also has the 30-day mean. It could be assessed with the method of averaging two samples. It wasn't necessary to justify doing the instantaneous for deep water because we also have the 30-day mean. We don't have a 30-day mean for deep channel.

- *Comment (from chat): Jim Hagy:* I think that University of Maryland at one time operated a bottom sensor at "R-64" ... mid Bay. In the 1990s.

Comment: Tish Robertson: In the summer, we do monitoring runs every other week. When you build the CFD for the deep channel, you can have a two-week period where you have really low dissolved oxygen (DO), because that's what a cruise represents. That's the duration that would be allowable for DO < 1. It's not hours, it's weeks because that's what the monitoring was.

Comment: Tish Robertson: The visual, basic form of the interpolator has a data limit too. I can't remember what it is, but you can't put in tens of thousands of observations. It crashes. Maybe the Bay Program's copy doesn't do that, but the copy I have does.

- *Response: Leah Ettema:* This was interesting to me because I've heard that it's impossible to assess all of the criteria at once. We could but we'd have to force something into the tool that it wasn't designed for. According to the 2017 document, we could theoretically get it if we wanted to. I think resources have been spent developing a better tool for this data, rather than forcing it into something it wasn't designed for.

Comment: Tish Robertson: If we were to do this method (conditional attainment), we'd want to do it on a segment by segment basis.

- *Response: Leah Ettema:* It'd be a lot of work to develop that for every segment.

Q: Tish Robertson: To be considered an exceedance day, is that where at least one observation was below the standard?

- *A: Leah Ettema:* Yes, like a daily minimum assessment. You could change it to be three hours or something else.

Q: Joe Morina: Are they calculating exceedance day based on one instance of an exceedance?

- *A: Leah Ettema:* Yes. For them, I think it is usually an hour of monitoring.

Q: Peter Tango: Given that criteria has seasonal components, does the concept of focusing on only the critical period hold true when you have a statement that says it must be met from January to December?

- *A: Tish Robertson:* Presumably you'd have other data that would represent the whole year. In Department of Environmental Quality (DEQ) guidance, it is just a suggestion and it's so people can't just collect and use winter data. We would have other data because our monitoring programs are sampling throughout the whole year and we would use that data. We need to prioritize it with the continuous monitoring (conmon) data.
- *Response: Amanda Shaver:* Our inland DO criteria don't have the seasonal components. We want to make sure we are capturing that critical period.

- **Response:** *Tish Robertson:* In our partnership with Virginia Institute of Marine Science (VIMS), they have deployed conmons as early as March or April. In the Bay waters with our current conmon data sets, we'd be able to get much of the year.
- **Response:** *Dave Parrish:* And they're year-round now, except they're pulled for ice at the moment.
- **Response:** *Tish Robertson:* Our methodology was designed for inland waters more than the Bay.
- **Comment:** *Leah Ettema:* Other states are similar. PA usually deploys for multiple months, but it's common to only have a month of deployment because they want to maximize the number of water bodies they're monitoring. Having more than 30 days of data at a station is not always common.
- **Response:** *Joe Morina:* For our inland waters, essentially all of our conmon assessments are coming from USGS stations. We deploy some conmon for our TMDL (total maximum daily loads) studies, but that's very small compared to the stations we have from USGS.

Comment: *Joe Morina:* When I combined the data types, I see that the discrete data is generally more outside of the conmon window. It would be difficult to say it's diluting it because it's a similar value. You have 100 conmon and one discrete at approximately the same magnitude.

- **Response:** *Tish Robertson:* The grab samples would be associated with swapping out the sensor. You'd expect them to be pretty similar.

Comment: *Tish Robertson:* I like the overwhelming evidence factor. We should think about something like this in our methodology. I think of these worst case scenarios where you can have 3-4 weeks of low DO but still be meeting with the 10% rule. The duration of that would be unacceptable.

- **Response:** *Leah Ettema:* The 10% is arbitrary. It's not tied to achieving aquatic life use. It's a consideration of data quality, artifact, and representativeness of monitoring. We don't know what the exceedance frequency should be for wildlife, so having leeway around that can be helpful.
- **Response:** *Tish Robertson:* It requires the person who is analyzing the data to visualize and dig into it rather than just running the script.

Q: *Joe Morina:* On the spatial considerations slide, isn't option four how we do the Bay assessment? Because we take all the data from the station and combine it at the Bay segment level.

- **A:** *Leah Ettema:* Yeah. The main difference between option three and four is that the Bay has spatial modeling to extrapolate condition in unmonitored areas. In option four, you take what you have and combine them.
- **Response:** *Tish Robertson:* I see what Joe is saying because with option three, we don't know whether a segment had hotspots or the exceedances are happening in time, when we get the DO results back. We don't know if it's spatial or temporal. It could be that we're exceeding in the temporal dimension rather than the spatial.
- **Q:** *Joe Morina:* You wouldn't get that all based on the curve?
- **A:** *Tish Robertson:* No. All you know from the exceedance rate is that the curve went outside what's acceptable. You don't know what's driving it. You'd have to visualize it.
- **Response:** *Peter Tango:* We do see that. If you plot the CFD, we can look at how the magnitude of exceedances is distributed relative to the default curve.
- **Response:** *Joe Morina:* Then, you could infer whether that is spatially or temporally driven based on the curve. I mean I'm not looking at all of the curves.

- ***Response:** Amanda Shaver:* Sometimes they're borderline and you can't see it.
- ***Response:** Peter Tango:* Yeah. One point could pop out in 12 months of data.
- ***Response:** Tish Robertson:* If we wanted to diagnose a segment, we could get that information if we plotted the CFD.

***Comment:** Tish Robertson:* We do use dataflow for water clarity. Dave is the expert on that. We do it based on the whole segment. It's spatially explicit. We don't break it out by where the impairment is, but we say the segment is impaired at the whole acreage.

- ***Response:** Dave Parrish:* There's an acreage goal for attainment for the segment as opposed to DO, which would be a threshold. It's the step before the acreage assessment. You'd need to have something similar.
- ***Response:** Tish Robertson:* The goal for water clarity is sediment specific, but we don't have something like that for DO.

***Q:** Tish Robertson:* For the slide "Common Elements for a Continuous Data Assessment Methodology," would you have these as recommendations or best practices?

- ***A:** Leah Ettema:* Yeah, if EPA could issue a best practices recommendation, this is what I'd recommend.
- ***Q:** Tish Robertson:* Why wouldn't we have this for discrete monitoring too?
- ***A:** Leah Ettema:* Good question. Most of these are the same. Calculation details are more straightforward. Critical period is just the way discrete monitoring happens. It's not usually done every day for two weeks. That's more unique to continuous data.
- ***Comment:** Tish Robertson:* With critical periods, we could have a data collector concentrate their bacteria collection to just December and we would use it.
- ***Response:** Leah Ettema:* Yeah, and Combined Sewer Overflows (CSO) is still happening in December. I think they're all relevant for discrete data. For target sample size, people describe that for discrete data, but it can be different. Discrete usually needs 10 samples to make an assessment decision but continuous data may need 10 days or more and have different considerations for that target sample size.

***Q:** Peter Tango:* We could probably get down to the microsecond these days with the oxygen collection technology. Is that relevant? We have limitations on capacity and are also doing it biweekly and are considering that in the deep channel. The daily minimum is usually based on hourly data. When we're thinking about the instantaneous minimum and continuous data, what's the resolution that's helpful without being overly abundant? This can work into our discussions at the end of the day. When we are looking at the number of samples in the instantaneous minimum, are we doing 10, 15, 20, 30-minute data or hourly data? It will be really important for us to agree on that moving forward.

- ***A:** Leah Ettema:* The next slide clarifies that. In the 2017 Tech Addendum, it was considered an hour.

***Comment:** Tish Robertson:* We haven't come up with rules for the Bay monitoring data we've been using. We don't have a minimum data requirement for using the CFD in a particular area. Fortunately, our monitoring program has been so excellent that we haven't had to consider that. During the pandemic, we had some big holes and yet we still crunched the numbers. I love continuous monitoring data, but I feel like we have such a high bar for it, but we don't have the same scrutiny for discrete data, and it doesn't make sense that we would do that.

- ***Response:** Leah Ettema:* That's the balance of using all data or having a rationale for using all the data in discrete even when it's not great.

Q: Clifton Bell: This question applies to discrete and continuous, but I think it applies more to continuous given everything that can go wrong with these sondes. We do monitoring for our clients and they are sometimes nervous about continuous data because of biofouling or calibration derivatives or other concerns. Everything we're talking about today is assuming the values are correct. Is the quality assurance (QA) implicit that you followed the QA plan and only used the data that passed?

- *A: Leah Ettema:* Yes. This is all data that has been QAed and is valid according to QA procedures. The EPA Continuous Data Workgroup has been trying for years to give guidance. The first part is on the QA portion of continuous data. How do you do QA? How do you document that it's quality? For discrete data, that's handled by the lab, but for continuous data, it's the people who are collecting it. It's a very different QA process. It's super important to consider and have documentation on how those decisions are made. All assessment is with valid data. EPA requires a justification for not using data. That it didn't meet QA requirements is an appropriate justification for not using it in the assessment.
- *Response: Tish Robertson:* In 2012, the continuous monitoring was novel at that time and we didn't know what the QA/quality control (QC) situation was going to look like. We had VIMS data and they have a top-notch program. We came up with the 10% of 10% because we were thinking about future conmon data sets that we'd have without knowing how to evaluate. It was a hedge against saying something is impaired when it isn't. Now we can have more confidence, so it might be something to revisit.

Q: Amanda Shaver: Can you discuss the overwhelming evidence clause again? We're you saying there needs to be multiple lines of evidence?

- *A: Leah Ettema:* No. That was more so something I commonly saw because the exceedance frequency isn't necessarily tied to achieving aquatic life use. A Best Professional Judgement (BPJ) allowance could help if you have a high magnitude or duration of exceedance that indicates impairment, but the exceedance frequency wasn't triggered. It didn't meet 10% exceedance but you had DO in the twos for a whole week, but it didn't trigger the 10% and is clear that it indicates impairment.

Q: Joe Morina: The Massachusetts example said that conmon is more reflective of the environment, so they'll always hedge towards that. I know you said you were concerned about the incorporation of discrete data into the conmon and potentially diluting it. I think we're going to have that with the Bay assessment as we're mixing continuous and discrete. I have seen a lot of times where the conmon data is for three months, and the rest of the year is discrete data. It seemed like by following the Massachusetts example, you would disregard the rest of the year in favor of the initial four months of conmon data. Is that better? Is that more preferable than mixing the discrete and conmon?

- *A: Leah Ettema:* It will depend on the parameter and how the monitoring occurs. My main concern with discrete is that if you're monitoring at 3pm in the afternoon for DO, you're not capturing the minimum. That is a critical period on a diel scale. For the Massachusetts example, an appropriate rationale would be relying on continuous data, since most of the discrete data is in the afternoon and continuous data better captures the critical periods for the diel cycle. For other parameters, it will be a little different. Some don't have diel cycling. I think it depends on how the monitoring program is set up and what the specific rationale is for those.
- *Response: Joe Morina:* If we were calculating a daily average and we had 50 conmon data points and one discrete, we'd average them all together.

- ***Response:** Leah Ettema:* It goes back to how you're monitoring. You only have a few discrete compared to a longer duration of a continuous data set. It's not going to have that big of an impact.

Comment (from chat): *Rebecca Murphy:* In terms of combining types of data, a challenge in the Bay is that the location of monitoring is frequently different with discrete data being collected in deeper waters, while the continuous is in shallow waters. We need to be sure the deeper discrete samples don't get overwhelmed by the sample size of the conmon in shallow waters.

- ***Response:** Peter Tango:* That's something I recall from the early conmon discussions. From 2010-2012, we spent three years looking at the relationships between nearshore and offshore waters. There are interesting relationships out of those efforts. The 30-day means can look very similar but when you look at the diel influences that relationship breaks down. The 4-d interpolator is trying to be informed about the relationships across depths and habitats.

Comment: *Amanda Shaver:* You mentioned that we haven't realized the connection between attaining DO criteria and aquatic life in the Bay. It's important for us to consider that when were are doing our assessments and thinking about extreme events. This is a way to report on conditions periodically. We're doing it every other year. We need to marry it to other things we're doing at the Bay Program, like the living resources assessment. I think the realization is going to come in on whether we're meeting out whole program goals versus tying so much to if we meet all these DO criteria, the Bay is saved. Program wide we need to consider that this is just one piece of the puzzle that will help us achieve the bigger, broader goals. For a state's perspective, it's just a requirement we do every two years. It matters for resource allocation and directs monitoring efforts, but we're not going to meet all of our goals by meeting the criteria in these 92 segments. We put a lot on these decisions when it's just one piece of the bigger puzzle.

- ***Response:** Leah Ettema:* From a TMDL (total maximum daily load) perspective, it carries a lot of weight, but that's just one big piece of the whole.
- ***Response:** Peter Tango:* We've talked about the bio reference curve and the derivation was strengthened by the default 10% margin being similar to the creation of the benthic informed curve. If all else fails and we have no insight on whether 10% is a reasonable assessment, that gives comfort. The 10% gave a buffer that seemed within reason of what biology was showing us, at least for that deep water zone. Claire did nice work on James River with the Phytoplankton IBI (Index of Biotic Integrity) work. There were different percentiles in different segments that seemed to better match a curve that was similar to 10%. There's information to say that we're working in the realm of what the critters are telling us, but I don't think it's putting fish out there versus bugs versus phytoplankton. It's not a uniform interpretation that I can see from our work to date that could be a viable way to move forward. The criteria themselves were set to be the minimum survival expectations.
- ***Response:** Amanda Shaver:* Only a handful of species were evaluated for the criteria.
- ***Response:** Peter Tango:* Those are just the living resource connections. It's not the end all be all but there is some connection to living resources. It's helped set us up for looking at whether the Bay is looking better or worse.

Comment: *Tish Robertson:* We also have wetlands.

- ***Response:** Clifton Bell:* Yes, wetlands and submerged aquatic vegetation (SAV). There are interesting results coming out of the Phase 7 model when you put SAV in a segment.

Comment: *Leah Ettema:* There is no explanation for why discrete or continuous could be used for one zone and not the other. Maybe we didn't have data for zone 3 at the time.

- ***Response:*** *Tish Robertson:* That's not true. I don't know why we did that.
- ***Response:*** *Leah Ettema:* That was my only hypothesis.
- ***Response:*** *Peter Tango:* I'm not sure why that didn't go in that slot.
- ***Response:*** *Tish Robertson:* Maybe we just didn't add it. It was just an oversight. It wasn't like we had a reason we were excluding it.
- ***Response:*** *Leah Ettema:* It wasn't discussed at all.
- ***Response:*** *Amanda Shaver:* We probably just don't or didn't have the data.
- ***Response:*** *Tish Robertson:* We do. I am thinking the Taskinas Creek would be one. That's not in the mainstem of the York River.
- ***Response:*** *Joe Morina:* Does the Chickahominy count as a tributary of a tributary?
- ***Response:*** *Tish Robertson:* No, because it's a segment. Pohick Creek could be up for debate. We have some examples of that.
- ***Response:*** *Peter Tango:* How can we do something when we haven't been able to say anything? It's an ongoing effort.
- ***Response:*** *Leah Ettema:* Don't let perfection get in the way of progress.
- ***Response:*** *Peter Tango:* I know Tish and company have had continuous data in shallow water systems since the early 2000s.
- ***Response:*** *Tish Robertson:* When we were coming up with this, we were thinking of enhanced monitoring. If we got funding for Cadillac level monitoring, this would be the guide we'd use. If we had put continuous data for zone 3, then we could deploy conmon in this little creek. It wasn't based on what we were currently doing. It was based on what we could potentially do in the future.
- ***Response:*** *Amanda Shaver:* And to potentially prioritize.
- ***Response:*** *Tish Robertson:* The tributaries of the tributaries sounded good when we were coming up with it. We have segments like CB7 that have a million little tributaries up the Eastern Shore. Would we be monitoring those individually? How would we do that?
- ***Response:*** *Amanda Shaver:* And conmon in one isn't the same as conmon in another.
- ***Response:*** *Tish Robertson:* You did CB7 on the Eastern Shore and you had several locations where there was continuous monitoring, but you weren't monitoring all of those tributaries.
- ***Response:*** *Dave Parrish:* No, not all at once.
- ***Response:*** *Tish Robertson:* When thinking about resources, it would be impossible for us to do all of those tributaries of tributaries in an effective way.
- ***Q:*** *Matt Stover:* When you say effective, do you mean census sampling of all of them?
- ***A:*** *Tish Robertson:* For us to do an assessment, like for an integrated report, there would be some segments we could do it for, but it would be hard to do the big ones with lots of little creeks. How would we do subsegments on those? As you get more refined, you are saying that there is a different habitat in that part of the segment and you have to have a monitoring station there.
- ***Response:*** *Amanda Shaver:* It's more in line with what we've done when sub setting the Bay segments for benthic monitoring. That way we're only looking at a handful. We haven't taken into consideration point sources or anthropogenic impacts. They're smaller zones.
- ***Response:*** *Matt Stover:* What you describing is similar to what we found in Fishing Bay. We monitored a couple of small tributaries and one didn't meet. It was coming out of Blackwater Refuge, it wasn't able to flush and had issues. The other one looked similar but it was ok. It was all discrete monitoring.
- ***Response:*** *Becky Monahan:* We did get a continuous monitor there.
- ***Response:*** *Tish Robertson:* You're getting great information, but are you going to have to maintain that sub segmentation scheme and go back to those sub segments forever? The whole point of the segment level was that we wouldn't have to be that precise.

- *Response: Matt Stover:* We took the approach that we can't monitor all of them, but we could capture a few. Don't let perfect be in the way of good. We're going to try to explain away one with the Blackwater conditions.
- *Response: Tish Robertson:* I hear you. We haven't sat down to think about it. The prospect of doing that level of enhanced monitoring without additional funding is hard to justify.
- *Response: Matt Stover:* We have a Baywide plan for where to place all the continuous monitoring devices and arrays, and even a schema for dataflow.
- *Response: Peter Tango:* For 20 years, we had the rotational nearshore conmon approach to get three-year assessments in each of the 92 segments. You all effectively carried through. The next level of that with arrays is part of what we had University of Maryland Center for Environmental Science (UMCES) looking at. They've done a couple of segments. What segments do we want more information from? That's to inform what a location provides and whether there is bias. There isn't strict guidance on monitoring protocol and where to place meters at this time. Censusing versus sub sampling to create estimates seems like something to consider. Jeremy Testa and his student, Amir, just presented at an Integrated Trends and Analysis Team (ITAT) meeting and spoke on sampling locations when considering presence of wetlands, distance to headwaters, and other characteristics. They had some 1960s data. They didn't have a continuous 50 years of data, but it was interesting to see discrete changes relative to wastewater treatment plants and efforts to reduce nutrients. It would be beautiful to have 50 years straight of data, but it's unlikely. If we're looking at condition changes that can be backed up with management actions, then maybe there's some temporal sequencing like you've pointed out. Some systems do rotations every 3-5 years. We need to think about the considerations of where they were placed and the value of continuous data in a time series.

Comment: *Tish Robertson:* If you take the Rule 2-Alternate at face value, you could have 50 days exceeding as long as they aren't consecutive.

Comment: *Tish Robertson:* Having two consecutive days below the instantaneous minimum sounds bad, but what if you have three years of annual continuous monitoring data with three days that have consecutively low DO?

- *Response: Peter Tango:* I think the Murder Kill River example had one day in the year of low DO and it was impaired. There is an 80-page technical document behind the rationale.
- *Response: Tish Robertson:* It's hardcore given how we're doing the deep channel where we allow quite a bit. For open water or migratory fish spawning uses, we say that three days is enough and it doesn't matter what the denominator is.
- *Response: Amanda Shaver:* It's one thing if you have the criteria but you have instances where it's continuously meeting or continuously over. The borderline ones need to more resources to dig into. We have these examples and can figure out what that means.
- *Response: Tish Robertson:* There are segments that are meeting the 30-mean and 7-day mean but if you look at the continuous monitoring data and apply this rule, they fail. What do we do? How do we make this a good segment? It's meeting all of the other criteria. When Will Humely and I came up with this, it sounded good at the time. In practice, I don't know how I feel about it.
- *Response: Leah Ettema:* There's a difference between consistent low DO and diel cycling. That's not considered with this either.
- *Response: Peter Tango:* The frequency and duration of low DO is related. Longer durations are associated with a higher frequency of violations. The likelihood of 50% of days exceeding without meeting this doesn't seem likely based on our conmon results.
- *Response: Tish Robertson:* Yes, 50% would be unusual but you could have 25%. You could have a whole lot of exceedances. We're applying stringency to this method that we're not

applying to discrete. There's no reason for it. It's worth having a discussion on this. The two consecutive day thing sounds great from a resource protection perspective, but there is no analogous rule for the discrete data. If we were collecting discrete data on a daily basis and there were three days of low DO, we wouldn't say it's an impairment, but we're saying it for continuous monitoring data. We are coming up with an additional frequency or duration element.

- *Response: Leah Ettema:* That makes sense because you only have discrete data every two weeks, but with continuous data you have the daily values to evaluate that.
- *Response: Tish Robertson:* There's no rule that says you can't have a daily discrete monitoring program.
- *Response: Leah Ettema:* It's worth making them compatible.
- *Response: Peter Tango:* Can we take this as an action item? Soon we are hearing from Breck on expectations with the new Watershed Agreement, where we have a couple of years to put protocol in place. This seems like we are uncomfortable with the statement in practice given your insights and this is something we should talk about in the next two years.
  - o *Action Item:* Revisit the 2017 document's continuous data recommendation of Rule 2-Alt: "No more than two consecutive days with 10 percent time (>2.5 hours) exceedance during a single season. This translates into 6 or more hours or about 0.2 percent of the summer season." There are concerns about allowing high levels of exceedance if they aren't on consecutive days and there are concerns that this doesn't apply to discrete monitoring. When updating and finalizing protocol in the target of the new Watershed Agreement, revisit this topic and decide on the best path forward.

Comment: Peter Tango: There will be paperwork for the 4-d interpolator. We're committed to addressing the needs for protocol assessment. The interpolator is a tool for helping us assess those and we need agreement on definitions and acceptable protocols. The Scientific Technical Advisory Committee (STAC) will perform a panel on the 4-d interpolator to examine the function. There will be a supplement for the 4-d interpolator, similar to the blue bible.

Comment: Breck Sullivan: I've come across a CAP WG agenda from 2013 titled "The Chesapeake Bay Instantaneous Minimum Dissolved Oxygen Criteria: Considerations for Alternate Definitions." It seems like that is similar to the discussions we're having today.

- *Response: Peter Tango:* That would be a good foundation to foster these discussions.

III. **Survey of State Assessment Methods for High Frequency Dissolved Oxygen**
*Lead: **Matt Stover** (Maryland Department of the Environment, MDE)*

Matt Stover sent a survey to several states within the Association of Clean Water Administrators to gain an understanding of the other state assessment methods for dissolved oxygen and potentially use this insight to inform assessment methodology in the Bay. He got responses from 14 states, which were Delaware, Florida, Georgia, Louisiana, Massachusetts, Minnesota, Missouri, Montana, Nebraska, Nevada, Oregon, Utah, Washington, and Wyoming.

The questions were:

1. In terms of duration and frequency, what are the different ways that your State specifies DO water quality criteria?
2. What assessment rules or methodology does your State use for assessing each of these DO criteria?
3. Does your State use an instantaneous minimum-like criteria for DO assessment?
4. For your instantaneous minimum DO criteria, do you use the 10% rule, a 'never-to-exceed' rule, or something else?

5. How does your State assess high frequency DO criteria using discrete data?
6. How does your State assess longer term DO criteria using discrete data?
7. Do you have any procedures or methodology for assessing DO criteria that were derived based on discrete monitoring data but for which you are now using conmon data to assess?

In this presentation, Matt dove into the notable responses for each of his questions. One common theme was that discrete samples were typically compared against the minimum criteria and a number of states would not use them for longer term criteria. Matt outlines some other general themes in the last slide of his presentation.

Actions:

1. Consider a presentation on how the 4-d interpolator is used in relation to the TMDL and Bay model.
2. In future discussions, consider what methodology is needed for assessing high frequency criteria with the 4-d interpolator.
3. Decide on the time period for instantaneous minimum. One hour? Two?

Discussion Notes (during presentation):

Q: Tish Robertson: Does Florida have percent saturation in its standards?

- *A: Matt Stover:* Yeah. They have it specified by region, also.

Comment: *Leah Ettema:* I think Florida has a conversion. They normalize it to a certain time of day. If you collect it at 9am, there's a calculation to get what it would be at the time of day they assess.

- **Response:** *Matt Stover:* They shared their calculator with me. They took their criteria, modeled it for an entire diel cycle, and compared the results at the certain time of day they analyze to the threshold to determine whether it was an impaired sample.
- **Response:** *Clifton Bell:* A diel cycle can be very different depending on where you are.
- **Response:** *Matt Stover:* I guess they did it by region, which helps.
- **Response:** *Leah Ettema:* I don't know what you would do for an unimpacted or impacted stream.
- **Response:** *Peter Tango:* Mark Trice did some work for us with dataflow data to remap some of it to 6am to account for those diel cycle swings, but we didn't use it as an assessment technique.
- **Response:** *Matt Stover:* It makes a lot of sense. If you have to collect the data when you can get out there, I don't think too many people are going pre-dawn. At least you know based on the diel cycle whether you are above or below.
- **Response:** *Tish Robertson:* I'm guessing that relationship would be based on a reference stream.
- **Response:** *Dave Parrish:* It's interesting that you would do it to the threshold as opposed to doing it to the data itself. If you had a continuous monitoring station, then you could infer it, but to do it to the threshold itself is tough.

Q: Tish Robertson: When Missouri is assessing instantaneous minimum, do they allow exceedances or is it a value that should never be exceeded?

- *A: Matt Stover:* I don't remember off the top of my head, but I believe it is in one of my next slides.
- *Response: Leah Ettema:* I don't think I've ever seen anyone treat it as a true never-to-exceed value.
- *Response: Matt Stover:* I can't remember which one, but one state said two or three strikes and you're out.

Q: Tish Robertson: Remind me what independent applicability means.

- *A: Leah Ettema:* It says each data type determines impairment individually. If you have biological data, like an IBI, that's attaining, doesn't mean you can say DO is attaining. You can't use biological data to trump chemical data.
- *Q: Tish Robertson:* So that is probably what Wyoming is doing?
- *A: Leah Ettema:* That's my guess. That's how people try to assess DO using biological data. That's a common comment. They are different lines of evidence and have different strengths and weaknesses therefore they must be considered independently.
- *Response: Matt Stover:* It's like the sediment quality triad where you have multiple lines of evidence, but if one fails, you have to list it.
- *Response: Leah Ettema:* Yes, or explain why that is an acceptable approach given what your water quality standards are. Without additional explanation, we wouldn't typically accept it.

Comment: Peter Tango: It sounds like we need a workshop on the 10% rule.

- *Response: Tish Robertson:* We like it because it's so easy.
- *Response: Leah Ettema:* One challenge with EPA giving out assessment methodology guidance is the 10% rule. EPA doesn't want to endorse its use because there's no real rationale. Writing something that acknowledges its being used but not endorsing it is difficult to do from a policy perspective.
- *Response: Matt Stover:* When assessing temperature, we were considering using the 10% rule or creating something biologically based. When we did that, it came out to be 10.3% and we did all this work to come back to the 10%.
- *Response: Leah Ettema:* That would be great to share if there is an actual rationale behind it.

Comment: Leah Ettema: The fact that states won't use discrete data to assess longer term criteria is likely because they only have one or two samples within a 30-day period. It's probably not enough to calculate a 30-day mean.

- *Response: Tish Robertson:* If they had daily discrete data, they probably would use it.

Q: Peter Tango: Our continuous data is technically high frequency discrete. Do people think of discrete of low resolution or once a month?

- *A: Tish Robertson:* Yeah, once a month.
- *A: Leah Ettema:* Or every two weeks. If for two weeks in row, why do that if you can put a continuous monitor there? Unless you can't afford continuous monitoring.

Comment: Tish Robertson: I have always said it's weird that we're assessing the 30-day mean criteria but not the instantaneous minimum.

- *Response: Matt Stover:* That carried through pretty much every question.

Comment: *Tish Robertson:* Oregon took their 10% of 10% method from us.

Discussion (after presentation):

Comment: *Peter Tango:* It's interesting to see the reverse in the first two general themes. We get the high frequency and people don't want to use them in the 30-day, but they're using low frequency for instantaneous. What's important is the short duration. Can you hold your breath? If you have low DO in the middle of the day, that's a real problem.

Q: *Rebecca Murphy:* Did you get any insight on spatial questions? A requirement of number of monitoring locations in a reach? Anything about interpolation?

- **A:** *Matt Stover:* No. In hindsight, I should have asked some questions related to that.
- **Comment:** *Peter Tango:* Leah, you mentioned some spatial considerations in your presentation.
- **Response:** *Leah Ettema:* I had four options for how it could be handled, if the state mentioned it at all. The most common options were splitting the assessment units if you get different results or base your results on the most limiting one – if one station is impaired, the whole unit is impaired. Most assessment units are not as big as they are in the Bay.
- **Response:** *Amanda Shaver:* We're trying to get to a one-to-one or more static version of our assessment units. It takes a lot of work to update and manipulate the assessment units each cycle. We're hoping to have one monitoring station per assessment unit.
- **Response:** *Tish Robertson:* Most assessment units aren't so big. From a resource standpoint, you probably wouldn't have a state deploying multiple continuous monitoring stations. It's only in the Bay where we have these huge segments.
- **Response:** *Leah Ettema:* Your reservoirs may be the only compatible thing.

Q: *Joe Morina:* We're using our continuous data for all of the assessments in the interpolator, not just the long duration, correct? Could one of the data congruency issues be addressed if we use the conmon for the high frequency 30-day mean and use the discrete for the deeper waters?

- **A:** *Rebecca Murphy:* We are putting all types of data together as input into the 4-d interpolator – high frequency, discrete, and high frequency spatial (dataflow). The idea is to combine them all and output an estimate of oxygen every hour on a spatial grid. With the interpolator, we are working on accounting for the misbalance of data quantity so we can still get good interpolations in the deeper waters, even with a lot of high frequency shallow water data nearby. We don't want to take data out of the interpolator. There are approaches we are building in to help balance the weight of the data.
- **Response:** *Amanda Shaver:* It sounds like the interpolation is tied to the frequency of the criteria.
- **Response:** *Joe Morina:* I was thinking about your earlier comment about having an amount of overwhelming shallow water data compared to the deeper waters. You won't always have conmon data to assess all of those uses. I was just thinking that through.
- **Response:** *Rebecca Murphy:* We're working on that and the interpolator should be able to do that. If we didn't interpolate and we were doing a count with all the data, there could be a concern that the high frequency data would overwhelm that computation. That needs to be considered for any data-based assessment.

- ***Response:** Tish Robertson:* That's a great point. I thought about that too in evaluating the exceedance frequency of the conmon stations separately from the discrete stations. That denominator would be huge if you combined all the data.

Q: *Peter Tango:* The tool is using all of the data to allow us to evaluate the spatial and temporal distribution of conditions down to an hourly scale. If our protocol is bi-weekly data in three locations in the segment, is that a misuse of the available information or is that an application of the protocol decided by the community for criteria assessment purposes?

- ***A:** Tish Robertson:* I would say the second. There's an idea I've had about combining these disparate data sets. They have their weaknesses and strengths. I'm for leveraging their strengths individually. The advantage with discrete data is it covers the whole year, and they are spatial in a way the conmons aren't. Having them be their own thing makes sense even though we are talking about data. They could be lumped together but I don't think that would be a fair treatment. The advantage of having the interpolator is that you're infusing all of that information, but that's not the only way you can combine it.
- ***Response:** Leah Ettema:* When you have all those different stations, it depends how the final assessment unit decision is made. If you have one station that indicates impairment and others that don't, how do you make that final decision? From EPA's perspective, if you have an assessment unit that is attaining but one station is impaired, we need a scientific rationale for why you're making that decision. To me, the most defensible rationale would be that the 4-d interpolator says that less than 10% of space is exceeding.
- ***Response:** Amanda Shaver:* In Fishing Bay, they could point to that one station that was cited in Wetland Drainage that says this is what is happening here and we need information in this area to understand the impact to the broader segment.
- ***Response:** Tish Robertson:* It's a little more complicated than having a station where all your exceedances are coming from. It would be much more likely to find exceedances happening all over.
- ***Response:** Peter Tango:* That's the magic that could come out of what we're doing. Like on Lake Erie with chlorophyll, where they took their least productive situation over a time series and used it as a baseline for a summer of chlorophyll assessment. They built their assessment off of comparisons to that. That's using satellite information, but it's analogous to what we're talking about. You can imagine what the space-time of that chlorophyll map would look like for the summer. We know that a small percentage of the area and time would show up in a lake-wide assessment. It's the same approach to the question. We'd have more insight into what that one station represents. Jeremy Testa and his student were showing the relationship of nearby stations and what the range of effects were compressed through time.

Q: *Tish Robertson:* TMDL is based on segments, right?

- ***A:** Peter Tango:* Yes.
- ***Q:** Tish Robertson:* Do we disrupt the whole TMDL concept we've established in 2010 by coming up with a different assessment framework, like when we talk about sub segmentation? The allocations developed by the TMDL didn't take into account hot pockets where we could have exceedances. It's based on segments and fixed stations. It's not based on continuous monitoring or 4-d interpolations. How do we communicate that? We get totally different results in the assessment world than we are seeing in the model.

- ***Response:*** *Leah Ettema:* That's where the stoplight charts come in because they were developed based on meeting the TMDL.
- ***Response:*** *Peter Tango:* The model is our planning tool to reach our best estimate to the effects of adjusting conditions. If the sub segments fit within the segments that we can roll up, that's probably a fair marriage. If it's a completely different segmentation scheme, then that's a question relating to the TMDL world and matching up how those allocations are made.
- ***Response:*** *Tish Robertson:* We didn't have tributary of tributary data. While we're waiting for the 4-d interpolator to be ready, if we used the 2017 document to do enhanced monitoring, are we identifying impairments that don't overlay with our understanding of the system based on the modeling? Are we complicating the picture by identifying impairments in a way we didn't conceive of when the TMDL was established?
- ***Q:*** *Amanda Shaver:* Is the 4-d interpolator going to go into the model? Or inform the model somehow?
- ***A:*** *Breck Sullivan:* The results from the estuary model are going to be put into the 4-d interpolator to make sure the scenario that runs through the model helps meet water quality standards attainment. Within the Bay Program, it serves two roles – the water quality standards attainment and the Water Quality Implementation Plans (WIPs) targets. Maybe the sub segments could roll up into one segment, but moving away from the 92 segments would not be a possibility. The Bay Program is taking a new direction of trying to focus within the shallow waters. They're considering a tiered implementation of finding where in the shallow waters we can target restoration work. Doing research and analysis within these tributaries of tributaries can support that but it doesn't change our ultimate targets.
- ***Response:*** *Amanda Shaver:* I think the assessment rules are separate from the work happening with the 4-d interpolator. I don't think it's going to impact what's going to happen with the target. That's going to stay static based on the TMDL and we're going to use the interpolator as we need to with our assessment determination.
- ***Response:*** *Tish Robertson:* That makes sense to me. How we come up with the TMDL and our assessments would ideally be the same.
- ***Response:*** *Leah Ettema:* The potential is having the 4-d interpolator saying there is impairment and we can reduce the impairment, but maybe a data assessment says it is attaining so why would we need to do anything.
- ***Response:*** *Amanda Shaver:* That's where the data needs to better inform the targets. They need to do it on the modeling side. They need to better inform through data what's happening with the targets and TMDL progress versus where we are trying to use the data to make these attainment decisions.
- ***Response:*** *Clifton Bell:* And also, to have policy decisions on the Bay TMDL versus the local TMDL. I remember during phase 6 there was shallow and open water where people said they wouldn't let them control statewide or basin wide planning targets. The same question is going to come up with phase 7. How much do we let some of these open waters control the Bay TMDL? I know there is a lot more interest in local waters.
- ***Response:*** *Tish Robertson:* A more refined assessment makes sense from the assessment perspective, but it breaks down when you think of the Bay TMDL.
- ***Response:*** *Becky Monahan:* We are going to change our assessment units too, but the TMDL can stay as long as our assessment units can roll up into the existing 8 digit. Then, the TMDL

will apply to any portion of that segment that is impaired even if they are split up into however many units, we want it to be.

- **Response:** *Tish Robertson:* I'm not thinking of it in terms of the allocations. I'm thinking of it in terms of measuring performance. Delisting would be based on the segments because that's what the allocations are assigned to. We're getting into the weeds now.
- **Q:** *Matt Stover:* Are you talking about the scale assessment, like carving up a segment into little pieces because you get different assessment results at different stations versus throwing them all together?
- **A:** *Tish Robertson:* Yeah. With the 4-d interpolator it's the same scale so maybe it's not a big deal, but it's also a different method than what was used. We like to have progress in our assessment tools so it is good that we are moving towards the 4-d interpolator. It would be hard for the states to manage if there are big differences between what we found from the TMDL in 2010 and the 4-d interpolator.
- **Response:** *Peter Tango:* We're about 70% from our goal and have not gotten closer than 60%. We know that the loads aren't down significantly yet. I understand the sensitivity. The 3-d interpolator was used to interpret the output, and the 4-d should be consistent.
- **Response:** *Tish Robertson:* I'm not anticipating they are going to be similar and I would be shocked if they are. We'd be going from calculating the 30-day mean with one or two samples versus to high frequency data. Why wouldn't it be different?
- **Response:** *Clifton Bell:* The 4-d interpolator is going to predict widespread nonattainment of instantaneous minimum criteria in all of these little tributaries. It's going to blow up the map. Is that something we think is going to happen?
- **Response:** *Tish Robertson:* I don't know. Why wouldn't it?
- **Response:** *Amanda Shaver:* We just have so many borderline impairments that it fluctuates every other year. I think because the assessment is that sensitive it could blow up.
- **Response:** *Tish Robertson:* I don't know so I'm thinking why wouldn't it be different? We have different data at different frequencies with different criteria. A lot of things would be different. The 3-d interpolator is using rudimentary inverse distance weighting (IDW). It's not a sophisticated modeling technique. I'm anticipating it would be different.
- **Response:** *Leah Ettema:* It might be helpful to have someone present on how the 4-d interpolator is used. My understanding is that the TMDL and TMDL allocations are not changing. How is the 3-d interpolator used in the implementation of the TMDL, which I think is in the stoplight charts. Having an understanding of how everything is connected would be great.
  - ○ ***Action Item:*** Consider a presentation on how the 4-d interpolator is used in relation to the TMDL and Bay model.
- **Response:** *Tish Robertson:* The 3-d interpolator is used in the Bay model to test the TMDL modeling scenarios with respect to water quality standards. I'm guessing the same will be done with the 4-d interpolator. That makes sense to me. We've never had a TMDL at DEQ where we changed the assessment method that dramatically. We don't revisit TMDLs but when we do the assessment method isn't changing.
- **Response:** *Leah Ettema:* The TMDL isn't changing it's just the way it's implemented.
- **Q:** *Tish Robertson:* The targets are changing?
- **A:** *Breck Sullivan:* They potentially may be changing but probably not.
- **Response:** *Leah Ettema:* The targets aren't part of the TMDL. It's the allocations and waste load allocations. Those targets aren't something EPA approves of as part of the TMDL.

- ***Response:*** *Tish Robertson:* The states would be the end users of those targets right? That's what I'm thinking of. How do we communicate to our management what they're looking at compared to when the WIP 3 was finalized? Are we going to be in the same universe or will it be in a completely different set of numbers? What are we going to say that's causing that?
- ***Response:*** *Leah Ettema:* The targets have been revised before.
- ***Response:*** *Breck Sullivan:* We're going to be educating the Water Quality Goal Implementation Team (WQGIT) on what is being done with 3-d versus 4-d. It won't have to be you explaining to your management, but it's us. You're not by yourself.
- ***Response:*** *Leah Ettema:* It's valid, Tish. If we adopt a different approach, how will that be communicated?
- ***Response:*** *Peter Tango:* Everything being done is calibrated to the existing data. If we see something dramatically different, that will raise red flags. Things have been pretty stable considering IBI scores and hypoxic volume. If we use our previous understanding of Bay conditions, they will support the big picture staying consistent.
- ***Response:*** *Tish Robertson:* I don't know why it wouldn't be different. The tool being developed is so much better and if it's not coming up with a different result, then why are we doing it.
- ***Response:*** *Rebecca Murphy:* I think Peter said that really well. I don't anticipate we'll have a huge jump because it's based on the data. It's not a dynamic model with processes. It's data interpretation. We will be using more data, which is the new part. I bet some of those high frequency criteria will differ from the 30-day. I agree with Peter that we will not see a dramatic change. What we need is the methodology for assessing high frequency criteria with the 4-d interpolator output. Everything we discussed today was on the methods with raw data and it will be necessary to consider how we do that with the 4-d output to see what the results look like.
  - ***Action Item:*** In future discussions, consider what methodology is needed for assessing high frequency criteria with the 4-d interpolator.
- ***Response:*** *Amanda Shaver:* The assumption is that the additional data is not going to tell us a different story than we already know.
- ***Response:*** *Tish Robertson:* It's hard for me to imagine that two samples composing a 30-day mean will produce a similar result to this new method.

Comment: *Peter Tango:* We have protocols for the 30-day mean and that is something the 4-d interpolator could use now. We never approved a seven-day mean assessment. We could always test it with just the fundamentals until we decide on a reference.

- ***Q:*** *Tish Robertson:* You mean we wouldn't assume the 10% distribution would work?
- ***A:*** *Peter Tango:* I imagine that the 4-d interpolator will give us a picture. If we don't have the protocols in place, can we give them that rule? That's something we should work with the 4-d group to decide.
- ***Response:*** *Breck Sullivan:* Even in 2013, we had the conversation of whether instantaneous minimum is an hour. It would be helpful to understand that part for us to be able to make these comparisons.
- ***Response:*** *Tish Robertson:* I think an hour makes sense. That's the standard for other minimum criteria.
- ***Response:*** *Clifton Bell:* It's somewhat conservative because a lot of the studies point to two hours with exposure in the labs.

- *Comment: Peter Tango:* Let's bring this decision back to the next meeting.
  - *Action item:* Decide the time period for instantaneous minimum. One hour? Two?

IV. Multi-metric Water Quality Standards Attainment Indicator Overview
Lead: **Peter Tango** *(USGS) and* **Qian Zhang** *(University of Maryland Center for Environmental Science, UMCES)*

Peter Tango's presentation gave the "story" of the Water Quality Standards Attainment Indicator walking the group through the history and use of the indicator. First, he described what the Chesapeake Bay Program (CBP) was measuring before the indicator, which was an overall water quality achievement metric in the Bay Barometer. He also goes over the "stoplight chart," which was explicitly described in the 2010 Bay TMDL and serves as a communication tool. Peter explains how the results for each of the 92 segments get presented and evaluated on a Bay wide scale. Next, Peter shared what the indicator does to account for missing criteria assessments. He also explained how segments are accounted for when determining attainment. Peter walks through the review and publishing process for the indicator. In 2016, the "Blue Bible" was published. Then, Peter shared the uses of the indicator for public consumption, the annual updates, and published assessment results.

Actions:

1. Talk to the Clean Water Goal Team about communicating the difference between the Water Quality Standards Attainment Indicator results and state 303d results.

Discussion Notes (during presentation):

Q: Tish Robertson: Is that stoplight chart (the first one presented) in Appendix M?

- *A: Peter Tango:* This one is not but the next one is.
- *Response: Tish Robertson:* The first one has values in the migratory fish spawning nursery use and we haven't been assessing that one.
- *Response: Peter Tango:* This has been going back 15 years. This is the precursor to the current indicator and stoplight chart.

Q: Leah Ettema: You're inputting different loads into the model. Does it output the values for fixed discrete monitoring stations for DO?

- *A: Peter Tango:* Yes.
- *Response: Leah Ettema:* The output values for the fixed stations are outputted into the 3-d interpolator.
- *Response: Peter Tango:* Yes. You would run the base scenario and modified loads, which get you a new dataset about expectations. Then, go back to the monitoring sites and pull those out as if it was Tuesday at 11am. The way we do the interpolator on our monitoring data is the same as how they sampled the model to run for the scenario outputs to evaluate attainment.
- *Response: Leah Ettema:* So, the model may produce outputs for every cell but we're picking them out where the fixed stations are to run it through the interpolator.
- *Response: Tish Robertson:* It's the relationship between the calibration model and the scenario model. It's not the actual model output. It's the change.

Discussion Notes (after presentation):

Q: *Joe Morina:* This came into being because the WQGIT required it and it's a way to track how close we are to attaining these goals. Is this a communications tool or is there any other aspect it's used for, like planning?

- *A: Peter Tango:* I hope this is used for planning. All 92 segments are represented. Qian has championed the R Shiny App that lets you look at the trends through time for any segment, criteria, or period of time. It's set up for those insights. In 2012, Liza gave a presentation that stated if they have no method, they will just call it a zero because they don't know. We've improved upon that and said that we should be giving credit where reasonable credit can be applied for attaining or non-attaining. Going into 2017, we wanted to use the best of what was available. We can do better than x-ing it out when there is no data at that resolution. That was our effort to use the best available information in practice. The R Shiny and the various publications try to give guidance that is hopefully insightful to managers. There are some graphs and tables that show the trends and patterns of attainment.
  - *Comment (from chat): Qian Zhang:* R shiny apps that Peter mentioned: https://wqs.chesapeakebay.net/wqs_attainment_deficit/ and https://wqs.chesapeakebay.net/wqs_attainment_indicator
- *Response: Joe Morina:* The way the attainment is determined in this approach is testing our criteria and if they pass, it's attainment. This is different from our approach of testing the criteria that we can test, but if we can't test all criteria, it's not attained even if they're fully supported. It's more relaxed attainment in the indicator analysis versus the 303d listing.
- *Response: Peter Tango:* It's based on the analysis that says it's reasonable to stretch the available data one way or the other to get the results.
- *Response: Leah Ettema:* For 303d listing, it's impaired now. When having the ability to delist, you need all the criteria for the regulatory decision.
- *Response: Amanda Shaver:* For our consent decree, it's a whole thing.
- *Response: Peter Tango:* We never adopted that conditional report. It's in the indicator but the regulating community has never adopted it.

Q: *Amanda Shaver:* Does the indicator feed the Barometer for standards attainment and trends or does the Barometer get trends from somewhere else?

- *A: Peter Tango:* For the outcome, we use the attainment indicator and its trends. For the Barometer, it's usually only one trend for the Bay as a whole.
- *A: Breck Sullivan:* The tidal trends for each individual station are not included in the Bay Barometer because it was never part of the outcome language, but now it is.
- *Q: Amanda Shaver:* We just saw a fact sheet from Virginia that did it by river. So the Barometer just does the whole Bay or the indicator only does the whole Bay?
- *A: Peter Tango:* The indicator does the whole Bay and individual segments, but we've typically only been asked to give the high-level single graph that shows the Bay as a whole and links to all those other analyses about segments and criteria.
- *A: Qian Zhang:* For the Bay Barometer, they typically just grab the information from the ChesapeakeProgress website. That is the whole Bay and the binary attainment indicator. The single number for the entire Bay.
- *Q: Amanda Shaver:* Recently, they shared with us that it is done by state, are the trends by basin?
- *A: Breck Sullivan:* They have recently started dividing information up by state. I don't know how they do that.

- *Response: Qian Zhang:* That is new to me. In the past, they just used the single number.

Comment: Becky Monahan: In the recent Scientific, Technical Assessment and Reporting (STAR) Team meeting, Alex Fries was talking about indicators. Maybe the best explanation of the indicator is that it is using the same data for a different purpose. The indicators are almost like a report card. I am concerned about the use of the word attainment and water quality standards attainment because we are not able to assess all of the criteria. I'm worried that we will have different stories. In our world, nothing is attaining, but your map shows that lots of things are attaining.

- *Response: Peter Tango:* We usually put the word estimated in front of it to be the differentiator as it is not the explicit accounting.
- *Response: Breck Sullivan:* When it is presented, it's normally speaking to the outcome in the Watershed Agreement, which is water quality and not the regulatory aspect.
- *Response: Peter Tango:* About every two years, we go over the difference between an indicator and a census. I understand that for some people a census may be the indicator, but the indicator doesn't have to be 100%. We are not counting every molecule in the 18 trillion gallons. It is our estimate of conditions. Now we have 64,000 square miles of habitat mapped to 1m scale. That's their indicator that they're using to show change over time in habitats that we can't match on the water side. That is a census of habitat while we're looking at an indicator of conditions and change over time with the best available information.
- *Response: Amanda Shaver:* For toxics, they just pull from our 303d information when this is creating a new piece of information. It's challenging because the public doesn't understand how they compare. We talked about this when designing the outcome. We need to streamline these messages because it's not an efficient use of time to be explaining this. I think there is streamlining that we should consider to make it a clear picture for everyone.
- *Response: Breck Sullivan:* We want to tell more of a story with the indicator because there are so many questions. Hopefully we can work together to help explain the differences. I think we do say that we created this method because we don't have enough data, but we could explain that better and mention the 303d reports.
- *Comment (from chat): Qian Zhang:* I searched for the Bay Barometer on CBP's website and found a couple of state-specific fact sheets, e.g., https://www.chesapeakebay.net/files/Bay-Barometer-Factsheet-2023_Virginia.pdf. Its "Water Quality Standards and Attainment" section talks about the overall estimated attainment indicator only and has no further breakdown by region. In the same section, it does have a breakdown of the water quality trends at the nontidal stations for Virginia. And here is the fact sheet for Maryland: https://www.chesapeakebay.net/files/Bay-Barometer-Factsheet-2023_Maryland.pdf
- *Response: Becky Monahan:* Ours all say it's impaired so that is why I'm concerned.
- *Response: Peter Tango:* 29% is saying it's impaired. The 29% is not saying the Bay is healthy and you are saying it's not. I think we are communicating similar results. That might be something to raise to the Clean Water Goal Team to see if we can do better communication and messaging.
  - *Action Item:* Talk to the Clean Water Goal Team about communicating the difference between Water Quality Standards Attainment Indicator results and state 303d results.
- *Response: Melinda Cutler:* In some communication tools, it gets lost that we're not assessing all criteria. It's assumed that everything is being assessed to the average user. There are no

unknowns there. In our office, we've had to explain that the Bay isn't meeting because we haven't assessed all of the criteria. Making things clearer would be helpful.

- **Response:** *Joe Morina:* I would take Melinda's point a step further and encourage all of the dielogue across the partnership to explain what the indicator is. The purpose, why we are mandated to do it, what it does, and why it's different from the maps produced by Virginia/Maryland/Chesapeake Bay Foundation (CBF). That way when CBP and VA DEQ maps are showing different results it's communicated. We should have a resource to show up as a partnership and address what the 303d list does versus a communication tool versus what is required for the TMDL.

Q: Breck Sullivan: We are trying to have conversations about where we address progress for the Watershed Agreement. When people say that your results are different than what they're seeing from us, are they looking at ChesapeakeProgress?

- *A: Becky Monahan:* Normally they just say, "I've seen" or "I've heard." It is confusing with not only your website, but all the different report cards. You could find a report card that says exactly opposite of what we're saying. All of the report cards are saying different things. It almost looks like we are lagging behind you because you are able to show all this attainment. I want to make sure that this doesn't take away from the fact that we all want to assess all criteria and that should be a focus too, so we are able to tell stories.
- **Response:** *Amanda Shaver:* For us, it's the trends. We have the non-tidal trends, tidal trends, and trends for state code. The trends are also problematic, and we need to do a better job explaining. The biggest thing that is missing is mentioning that there are other sources saying different things and highlighting the different uses. We are going to continue to put tools out there. In the STAR conversation, I think that's where we need to assess the need for these things. We're bound by regulation, so we must put them out. If it's just about coming out with the best communication tools for the new Agreement, maybe it's been re-thought.
- **Response:** *Breck Sullivan:* We didn't rethink it, but it's required of us to report on our progress.
- **Response:** *Amanda Shaver:* In the context of everything else out there, how can we complement something and make sure we're not putting out redundant but different results that can be confusing.
- **Response:** *Joe Morina:* Here is an example from our 2024 Integrated Report.

  "Water quality trends analysis would be improved with the addition of flow normalized plots and references to similar water quality evaluations. Example: the Chesapeake Bay indicator tool (from partner organizations such the USGS weighted trends)."

  We can't do weighted trends in the Bay because it's tidal but this is mixing all of this together. How do your trends compare to the indicator tool? How do they also compare to the weighted analysis USGS does? We do ours on concentration so it's not comparable to USGS or the indicator tool. It doesn't seem appropriate to compare. There could maybe be comparisons with the weighted trends for USGS. This question illustrates someone with CBF who is involved in this and still trying to piece things together. We need to do a better job at communicating what we're doing and not doing. If an engaged partner is questioning what we're doing, how is a member of the public supposed to know?
- **Response:** *Peter Tango:* The Bay Program has been talking about focusing on social science. The communication piece is very important. That is a great point.

Comment: Amanda Shaver: We're subject to public comments.

- *Response: Breck Sullivan:* We are too.

Comment: Tish Robertson: Amanda and I just got a question from our division director about the attainment results and how they compare to our assessments. The first thing I said is that we aren't using the same data sets. You guys are using data from the Data Hub, but we are using data sets that aren't in the Data Hub, in addition to the Data Hub data. The reporting done on the James River chlorophyll is the perfect example of that. We're using DataFlow data for that assessment. The number of criteria being assessed isn't the only difference. There is also the difference in datasets, and I had to explain that to her. I don't know if we're getting a lot of comments about the discrepancies, but it is something we notice a lot in our management notices. It's something we must respond to.

- *Response: Amanda Shaver:* When the public sees discrepancies, they disengage because they don't understand.
- *Response: Peter Tango:* This will be important to discuss. We have local interest in the folks on the ground who want to see progress. We're also answering the big Bay question of the partnership wanting one number to describe Bay condition.

V. **Revised Chesapeake Bay Watershed Agreement and Water Quality, Standards Attainment, and Monitoring (WQSAM) Outcome**
Lead: Breck Sullivan (USGS)

In this presentation, Breck shared an overview of revisions in the revised Chesapeake Bay Watershed Agreement with a focus on the Water Quality, Standards Attainment, and Monitoring (WQSAM) Outcome. Before the revision process, there were several reports published to guide revisions. Breck walks through the initial structure of the Agreement and how outcomes were edited for their current structure.

Next, Breck shared the revised Watershed Agreement and walked through it. She highlighted the Vision and how it came directly from a public comment. Breck also walked through some of the changes with the WQSAM Outcome, which is most closely related to the CAP WG.

Breck mentioned how the development of management strategies are the next steps and members of the workgroup can be involved. She also shared that there is interest in sending the CAP WG under the Clean Water Goal in the new structure and that they would be a good match due to the regulatory nature of both groups. Each goal will have its own management strategy which will be broken down by goal. She also walks through the next steps for the CAP WG.

Actions:

1. How do we establish membership in the CAP WG? How will membership be adapted to fit into the Clean Water Goal? Note: The group may want to wait for the Governance and Structure document to be released before deciding on recommendations.

Discussion Notes (during presentation):

Q: Tish Robertson: Can you explain what reclassify means?

- *A: Breck Sullivan:* In the previous Agreement, there weren't targets, but now that we have targets some existing outcomes were reclassified as targets under another outcome. An example of this is the Waterbirds target. There used to be a Black Duck outcome, but now it's a target within the Wetlands outcome that also expanded to all waterbirds. This is because the amount of wetlands impacts the number of Black Ducks. A target is still part of the outcome We had some high-level outcomes without clear numeric goals. For the targets, we can say we are improving upon the outcome when we reach the target.

Q: Joe Morina: Are they doing the process with indicators and management strategies across all of the goals and outcomes?

- *A: Breck Sullivan:* Yes.

Q: Melinda Cutler: Is this different than the indicators we were talking about earlier?

- *A: Breck Sullivan:* No. This is the same.

Comment: *Joe Morina:* Earlier we were talking about the attainment indicator as a communications tool but if it's built into the next steps like this, then it's more than a communications tool. It's like a metric. It's part of the Agreement and what we are trying to achieve.

- *Response: Breck Sullivan:* I would say that it's always been more of a communication tool because it's always been a metric to show our progress towards the Agreement.
- *Response: Joe Morina:* To me, this seems further than communication because it's an indicator of achieving progress. It's not just information for the public. It's dictating whether we've achieved what we set out to.
- *Response: Leah Ettema:* Theoretically there'd be consequences.
- *Response: Amanda Shaver:* We'll potentially be changing our methodology of what is going into the indicator based on what we develop for the first bullet. We should. If we come up with assessment methodologies and decide to assess it differently, then the indicator will change based on what we decide.
- *Response: Matt Stover:* If we try to make sure it is consistent, the assessment is our integrated reports.
- *Response: Joe Morina:* We still use the indicator to make the assumption that if you pass these five criteria, then all of the designated uses for the DO is met. Even if we keep that, this is more than displaying our progress. Like Leah said, this could have consequences, even if it's just pass or fail.
- *Response: Peter Tango:* It's a commitment with a use rather than an indicator that we created to show progress.
- *Response: Joe Morina:* We would explain this to the public by saying "this is a communication tool to say x, y, z, but we've also used it to determine if we're meeting our targets under the Bay Agreement." Maryland and Virginia's 303d listing and maps don't go towards this, but they go to the TMDLs.
- *Response: Breck Sullivan:* I want to specify that the indicator we are talking about here is for water clarity, DO, and chlorophyll *a*. The 4-d interpolator is only DO.

Discussion Notes (after presentation):

Q: Amanda Shaver: What's the date for the next revision?

- **A:** *Breck Sullivan:* 2040. 2033 is our 50th anniversary. That would be the half-way point between now and 2040 where we will do an overall evaluation but not a revision to the Agreement.

Q: *Amanda Shaver:* It says, "maintain or exceed." Are we re-evaluating that?

- **A:** *Breck Sullivan:* Yes. We will be re-evaluating that. We will see if we are on course or off course. In the past, the Bay Program has done re-evaluations. In one, they noticed that they were very behind on Forest Buffers and Wetlands. The Bay Program worked with each of the states to create more specific actions for those outcomes. At the mid-point, if we realize we're not maintaining water quality standards at all, it will raise the flag to say we need to pinpoint efforts and resources into this. We would be doing that on a semi-annual basis, but it will be a larger get-together.

Q: *Amanda Shaver:* Will there be a management strategy for STAR?

- **Response:** *Breck Sullivan:* People are considering that. We have an overall scope and function document so maybe something within that can be revised. The Bay Program has proposed having the four goal teams with a program support branch made up of STAR and the Strategic Engagement Team (SET). The program support will support the partnership and all outcomes. Our work doesn't function if we're not supporting everyone. I don't know if there could be specific actions because it depends on the needs of the outcomes. I foresee there being something.
- **Response:** *Amanda Shaver:* I'm envisioning an operations plan, or something so there is some revisiting of the work that is being done to support the outcomes so it can better align with the goal.
- **Response:** *Breck Sullivan:* I support that too because I think it would help provide better collaboration across the goals. I've also been trying to think within the management strategies so STAR can be involved in each of those.
- **Response:** *Joe Morina:* That would also be a lot of work and a lot of different types of work. It would be helpful to have someone to organize that.

Q: *Joe Morina:* We are under STAR, but we are moving to Clean Water. Where does that go in the hierarchy?

- **A:** *Breck Sullivan:* It would be under Clean Water because Clean Water is responsible for WQSAM. This workgroup has a regulatory framework and supports that outcome. The goal chairs are being considered right now. The overall structure of Clean Water is also looking to change, including bringing in the CAP WG.
- **Response:** *Amanda Shaver:* The initial six Goal Implementation Teams (GITs) will now be moving into the four new goals.

Q: *Melinda Cutler:* Are other workgroups like the Bay Oxygen Research Group (BORG) moving as well, or are they staying under STAR?

- **A:** *Breck Sullivan:* It is looking like they will stay under STAR because of the technical analysis part of it. This would be similar to how modeling will stay under STAR even though they speak a lot to Clean Water. STAR isn't going anywhere. You'll still have to interact with us a lot and we want to. That's where the extra collaboration could come in. Some data would be done within STAR, which would go to CAP, then to Clean Water. Starting in July, we

won't have a Management Board (MB). There are conversations asking, "when do decisions stop with the GIT chairs and when do they go to the Principals' Staff Committee?"

- **Response:** *Matt Stover:* In our last meeting, Peter shared that the other groups that support CAP serve several different workgroups. To me, it would be nice to know which workgroups have a say and home in on the hierarchy. It's confusing to not know who we need to get approval from. To me, I'd like to have that stuff laid out.
- **Response:** *Breck Sullivan:* We have a Governance and Structure Team that's made of signatories. They are trying to figure that out and we are waiting to find out. At the moment, it is a very internal signatory team. I think they will bring a draft of their suggestions to this month's [Management Board meeting](). I could share with you the Maryland representative that sits on the MB. I know she is really involved with management strategies and governance. The governance part is a very internal signatory decision.
- **Response:** *Peter Tango:* From the workgroup level, it's hard to know who is on the MB and not.
- **Response:** *Matt Stover:* In this day and age, there is a reluctance to have hierarchical organizational structures and decentralized decision-making. There are good reasons for doing that. I think it behooves us to follow Robert's rules or something like that. These decisions have big consequences like how we assess waters or develop models. I like knowing exactly how the decision gets made and who I need to talk with to enact change.

Comment: *Amanda Shaver:* The Toxics Workgroup has voting members and that tends to happen when you're under the GITS. We haven't had that.

- **Response:** *Breck Sullivan:* It depends. I think you could open it up to having voting. I would wait for governance and structure to be laid out. When you go to voting, is it consensus or majority? I think they'll provide flexibility knowing the decisions in each workgroup are different. It's a possibility, but it hasn't happened before.
- **Response:** *Peter Tango:* If you look at how our indicator came along, it's credit to the analytical power of those who come to the meetings. We've been more freeform with our harvesting and harnessing of those energies rather than being decision focused. In the WQGIT, they're making decisions on model scenarios and BMPs, which call for a more membership-based structure. I don't know, but if we're moving under the Clean Water Goal will we be making decisions or creating science and recommendations to bring to the GIT? That is something for us to consider in what our established scope and role is. Is it bringing decisions already made so therefore it's very important who is sitting at the table or is it the mind mill of great minds that brings the recommendations forward for them to make a decision?
- **Response:** *Amanda Shaver:* I would support some kind of voting representative to speak for each signatory to keep things moving forward, especially now that we have timebound targets.
- **Response:** *Peter Tango:* In the past, there was a strong connection between monitoring data and modeling. Then, there was a strong emphasis on the modeling world, but now there is more connectivity. This might be a good time to have discrete membership or representation. We can make this an action item and important discussion. I don't know if we can do that until the GITs are firmly in place, but we could recommend a structure.
- **Response:** *Breck Sullivan:* You might want to wait for the Governance document too.

- ○ ***Action Item:*** How do we establish membership in the CAP WG? How will membership be adapted to fit into the Clean Water Goal? Note: The group may want to wait for the Governance and Structure document to be released before deciding on recommendations.

Q: Matt Stover: Is the Governance document being reworked?

- *A: Breck Sullivan:* Very much so. Multiple people have explained the responsibilities and roles of each group, chairs, and coordinators. They're trying to get to that, but some flexibility will be involved.

Q: Matt Stover: Does moving from STAR to Clean Water change anything in how we operate?

- *A: Breck Sullivan:* Yeah, that's what will be decided based on the guidance provided by the Governance document. The Goal Team chairs will influence the decision.

Comment: *Breck Sullivan:* For each Goal Team, they are looking for a federal and state representative as chairs. Currently, there are four proposed chairs for the Clean Water Goal.

Q: Amanda Shaver: Are they targeting for June/July to get everything settled and begin working on management strategies?

- *A: Breck Sullivan:* Correct. Since the Agreement was signed in December, we have 18 months to finish the management strategies, which will be June 2027. This June, the outline of the management strategies will be completed. Then, you'll have a year, which I think will be plenty of time given the amount of documentation we have for our outcome.
- ***Response:*** *Peter Tango:* The template will be our guidance.
- ***Response:*** *Breck Sullivan:* Yes, it will cover what goes in it and what will be needed for each outcome. At this February MB meeting, there will be conversations on the outline for the management strategy and there should be a lose draft of the governance document. There is a lot in flux right now. Last year, we focused on the what, this year we're focusing on the how.

Q: Matt Stover: Is a lot of that happening at the MB level before they dissolve?

- *A: Breck Sullivan:* Correct. Management strategies would happen more at the GIT level because we've already designated that the GIT chairs would be responsible for it. The MB is dissolving in July, and the management strategies won't be done until next June.
- ***Response:*** *Matt Stover:* From what I've heard the Principals' Staff Committee will go from secretary level to director level.
- ***Response:*** *Breck Sullivan:* I've heard the opposite. I've heard it's staying at the secretary level and they're trying to have a solidified person at the end. They are hoping to stick to the same acronym and change it to the Policy Steering Committee. There is some hesitation with that, like from USGS. Our whole thing is that we don't do policy, just science so having representation on a policy steering committee is iffy. They want people who can shape policy, funding, and resources.

VI. **Brainstorming Session: Planning for Next Steps within the Criteria Assessment Protocol Workgroup**
Lead: **Peter Tango** *(USGS),* **Matt Stover** *(MDE) and* **Tish Robertson** *(Virginia Department of Environmental Quality, VA DEQ)*

1. What do we want to accomplish in 2026? Beyond?
2. What do we need to discuss in our meetings in 2026?

Actions:

1. Further discuss 4-d interpolator testing and error calculations.

Discussion Notes:

Q: Peter Tango: We recognize that we are trying to operate with the understanding that our role may shift, but there are commitments in the Agreement on the protocol timeline. We need to set some fundamental definitions to set the best foot forward. It sounds like folks have some good initial feelings towards one-hour being a representative time step. That could be decision one if we want to bring that back. Should we bring that to the next meeting?

Q: Joe Morina: Is that the time interval for instantaneous?

- *A: Peter Tango:* Yes

Q: Melinda Cutler: If we see an exceedance that is an hour in length, then that would qualify for impairment?

- *A: Joe Morina:* It would be aggregated at the hour level.

Comment: Peter Tango: We saw with Tish's decision rules that it was the recommendation with two days of impairment. We can take that decision and look at it in terms of that rule and related rules to decide what we should do.

- ***Response:*** *Tish Robertson:* I'd like us to explore other rules. From what Leah and Matt presented, there are other ways we could do it. It wouldn't hurt to sit down with some methods.
- ***Response:*** *Clifton Bell:* It'd also be nice to test those methods with real datasets and sites. Maybe there are some representative sites to test on.

Comment: Tish Robertson: Clifton is raising something that I've thought about for the 4-d interpolator. Are we comfortable with coming up with rules without running the analysis? That would probably be most objective to come up with methods without seeing what the results look like. The practitioner in me doesn't want to do that either.

- ***Q:*** *Amanda Shaver:* Could we start with a baseline of what the 3-d interpolator is putting out now with the 30-day mean and instantaneous deep channel compared to the 4-d?
- *A: Tish Robertson:* That's something we could talk about. Are we going to assume the 3-d interpolator is the correct assessment?
- ***Response:*** *Amanda Shaver:* There's some worry about change and what it will look like when we go to the 4-d interpolator. The 3-d interpolator is what we know.
- ***Response:*** *Tish Robertson:* That goes into the discussion. How are we going to evaluate the 4-d interpolator?

Comment: Rebecca Murphy: I thought that what we were talking about with Leah's presentation was more like assessing the criteria without interpolation. We'd like the group to talk about how the 4-d output would be used in criteria assessment. We can still use the CFD set-up. Elgin's done

draft testing of putting hourly output into the CFD computation and it's technically possible. We could lay out some of the decision points that need to be made if we decided that would be the approach or a choice. Were we thinking about ways to do it without the CFD?

- **Response:** *Tish Robertson:* I think that we need to unpack everything. I know Elgin has expressed trepidation about using the CFD with the interpolator output. It may be good that we look into that. If we decide to keep using a CFD, do we want to use a 10% reference curve? Is there another distribution we'd like to use? I think anything is up for discussion. Why wouldn't we? We're at a time where we should be evaluating all aspects of the assessment protocol, not just the interpolation technique.
- **Response:** *Peter Tango:* I firmly appreciate the exploring there. We have some commitment timelines that will bound the number of explorations we can do. What can we do strategically if there are some paths that look better or more effective than others?
- **Response:** *Clifton Bell:* I think it is important to look at methods before rolling out, especially if we are replacing the CFD method because it's a regulatory thing. Looking back at the Bay TMDL document, it used the 30-day mean for open water and deep water and referred to Appendix D, which concluded that the 30-day mean was protective of the instantaneous minimum. They evaluated that by running the model directly through the CFD. I suppose they used the default 10% reference curve. That's where those little tributaries of tributaries are so small in area that you get benefit from using the CFD method. If you use a new method that's using conmon data there, it seems like it could be a lot more stringent than the CFD, at least for that sub-segment. That would change the TMDL. I think we want to know those answers before we roll it out.
- **Response:** *Breck Sullivan:* I don't think it changes the TMDL.
- **Response:** *Amanda Shaver:* It changes the status of the water. It's looking at things at a finer scale. It could bring up how well the tool is doing on a segment wide basis. We are potentially masking problems when we have finer scale data.
- **Response:** *Tish Robertson:* When we say, "change the TMDL," we don't mean the allocations. We mean that the conclusions we would draw from testing the simulations might change.
- **Response:** *Leah Ettema:* It would provide evidence that maybe the TMDL needs to be revised but it wouldn't change the TMDL itself.
- **Response:** *Clifton Bell:* It was intended to have spatial allowance and take into account hydrologically isolated areas that may not attain due to natural conditions. Maybe that's why we approved that method in the first place.
- **Response:** *Tish Robertson:* What you said is why a difference is inevitable to me. The resolution of data we used to do the scenario modification wasn't on a high frequency dataset. The 4-d interpolator is injecting a layer of complexity that wasn't available when we were developing the TMDL. Do we want to come up with the methods without running the results? I think we need to look at all the different outcomes of the rules we come up with but how do we evaluate what those results mean? Say we use the 4-d interpolator to evaluate the instantaneous minimum criteria for the open water use, and we found that all 92 segments fail using the 10% distribution. Will we decide that it is too stringent? What would be the basis for saying that? Maybe that is reality. What are the rules we'd use for determining that was an unacceptable outcome.
- **Response:** *Amanda Shaver:* There's no aquatic life check. The 10% is the default we've used except for some other pieces like the bio reference curve.

Q: Melinda Cutler: Is that something we could explore? Is there data that we could collect that could try to figure out what percentage is best for the aquatic life in these areas?

- *A: Tish Robertson:* The only thing I could think of that would be for us to identify the best Bay segments we have.
- **Response:** *Matt Stover:* We could use a reference segment approach. It would be a small sample size.

Q: Breck Sullivan: Now that we have better data available for open water, would it be developing a bio reference curve for that?

- *A: Tish Robertson:* I don't know how we'd develop a bio reference curve for open water because that is about fish. We use the IBI for deep water and deep channel. Open water is not benthos, it's the nekton – fish, crab, swimming critters. What index do we have that would allow us to look at reference from that standpoint? I don't think we have anything like that.

Q: Amanda Shaver: Could the shallow water/living resources inform a migratory use curve?

- *A: Peter Tango:* It seems like Jim Uphoff, Margaret McGinty, and company have done significant work on fish richness and maybe survivorship. I'm picturing curves they presented at the Maryland Water Monitoring Council about habitat conditions and influences on fish communities by taking data associated with watershed characteristics. I assume you could use the same data. I don't know if that's seine survey or seine and trawl. Their work would offer a bio reference curve style test of characteristics of related waters. We have 92 segments of conmon data for three years that should be matched with seine data. The downside to seine data is it's usually limited to sandy habitats. There are probably limitations on information there. The benefit of evaluating that use is that we haven't evaluated that. I'm curious why you think the benthic data is a bad match for open water. Benthos are important for shallow and deep water. That's where the oyster work is focused.
- **Response:** *Tish Robertson:* Benthos are not the main player of the open water criteria. The short-nosed sturgeon is the signature critter for the open water use.
- **Response:** *Leah Ettema:* I don't know if there is more literature on sturgeon, but do we know how far they'd travel to avoid low DO? That could inform a space component. How long can they tolerate different durations in hypoxia? There may be something to learn that can inform what percent to use.
- **Response:** *Tish Robertson:* Like the area of avoidance or temporal tolerance?
- **Response:** *Leah Ettema:* Yeah. Since we're primarily driven by sturgeon, that would be a rationale for that use.

Comment: *Melinda Cutler:* I'm also thinking about our temperature for nontidal streams for the assessment methodology used for the integrated report. We have continuous monitoring data and are trying to figure out how to assess it. Department of Natural Resources (DNR) did a study to compare where we have brook trout with continuous temperature monitoring data to see what amount percent exceedance of our criteria occurred where we had brook trout. That came out to 10.3%, so we use 10%. I'm wondering if sturgeon presence or richness could be paired with our dissolved oxygen data.

- **Response:** *Peter Tango:* That's exactly what they did in San Francisco Bay and the upper part of the Suisun Marsh. They looked at three decades of fish presence, absence and catch data relative to DO measures to inform DO criteria.
- **Response:** *Leah Ettema:* Jim Hagy was key to developing the Delaware River criteria, which is based on larval recruitment of sturgeon. I don't know if we have the same type of dataset available here though. That could be an option too.
- **Response:** *Peter Tango:* He's doing a webinar. Maybe we can recommend attendance. February 25th, 2-3pm. Maybe Jim would come back and talk to us about that.

Comment: *Joe Morina:* I feel like the last part of the conversation was about the criteria that we're testing against to make sure it's protective of the resource or designated use. Earlier we were talking about how we know whether the interpolator results are valid. If we do scenarios where we run the 4-d interpolator, but we also have data we didn't put into the interpolator, we could compare the cells. If it's within 10-20%, then we could be confident that it's giving a DO value we may have measured. Once we know that the interpolator is functioning correctly, it's up to us to figure out if we want to adapt our assessment methods.

- **Response:** *Leah Ettema:* Any difference in that could also inform what your allowable percent exceedance is. If we're saying 5% difference, then that could be our margin of error that informs the 10% reference curve.

Q: *Dave Parrish:* Is the 4-d interpolator output going to be a single value of DO per cell or is a distribution of DO values per cell? Is there uncertainty attached to the estimate?

- **A:** *Rebecca Murphy:* We are going to run multiple simulations. For any one cell in one hour, there might be 100 estimates out of the 4-d interpolator. That is to account for variability in high frequency data and uncertainties. We can do multiple things with that, like a distribution. We can talk about that further.
- **Q:** *Leah Ettema:* Would that be done for each hour in each cell?
- **A:** *Peter Tango:* Correct.

Comment: *Breck Sullivan:* We can show the 4-d results and we can put the observed data on top of it to compare. I am hoping there isn't data we haven't used.

Comment: *Rebecca Murphy:* Joe's suggestion was excellent and that's what I hope to do. I'd like to do some holdout tests to see what the error is in the predictions. We've kind of been doing that as we're testing and adjusting.

- **Response:** *Clifton Bell:* If you are using that kind of output with the CFD, you might say that because we already have a spatial and temporal integration, some sites will be on one side of the uncertainty, and some will be on the other side. Therefore, we'd just take the results and apply the CFD. If you were applying that result at a single site, you might require more than a 51% probability that it's impaired.
- **Response:** *Leah Ettema:* You're saying the errors might wash out on a big scale, but not at a single site.
- **Response:** *Rebecca Murphy:* We'll probably have to lay out some examples. In initial results, Elgin has done it by having multiple lines on the CFD curve. Each simulation would result in a different assessment line on the curve. Maybe they'll all be way below the 10% curve and

it'd be obvious, but when they are close to the reference curve, some decisions will have to be made.

Q: Leah Ettema: Will we be able to extract the result for an individual cell?

- **Response:** *Peter Tango:* My understanding is that you'd be able to store the 100 simulation output and the starting conditions.

Q: Dave Parrish: Are there distances in time and space from where we have an observed datapoint where the uncertainty is so high that we exclude it from the output of the interpolator? Are we interpolating everywhere?

- **Response:** *Leah Ettema:* There should be.
- **Response:** *Dave Parrish:* I run interpolations of dataflow data on the same day. We don't exclude areas that are too far away to interpolate data, but we probably should. You are outside the domain of the data. I am curious how the interpolator might handle that and how you would integrate that into a standards assessment.
- **Response:** *Joe Morina:* In the 3-d interpolator, there is no limit on distance. It searches for the next one, even if it's far away.
- **Response:** *Rebecca Murphy:* I believe there are some limits because I've seen blank cells. This is a great decision point. If a segment doesn't have data in a year, will you decide not to use that interpolation result? Rules can be set here like I think they have in the 3-d interpolator.
- **Response:** *Tish Robertson:* Joe is right. We don't have a rule in the 3-d interpolator that you only interpolate in the domain of the data. It will go to the shoreline where you don't have data points. There is a spatial limit though.
- **Response:** *Peter Tango:* Is the 4-d down to 11 interpolation regions? I know we went down from one interpolation for the whole Bay, but not as far down as segment level.
- **Response:** *Rebecca Murphy:* There is a part where we combine segments. The structure we're building is by 92 segments. Sometimes you'll need data from other segments, but for results it will be set up to be selected by segment.
- **Response:** *Peter Tango:* I didn't mean that the analysis wouldn't be on a segment-by-segment basis. I meant it along with the data region idea.
- **Response:** *Rebecca Murphy:* To run the baseline midday mean generalized additive model (GAM), we have 27 segments. For example, the whole Patuxent is one.
- **Response:** *Jon Harcum:* We purposely include boundary segments. If we want to create an interpolation for CB4MH, CB3MH and CB5MD would also be included in that region. We include the data of the boundary segments as part of the interpolation. Data will get reused in another interpolation region. Once those GAMs have been created, the simulations would be on a segment-by-segment basis. You could run them all one morning.

Comment: Peter Tango: We started with some clean questions and shifted over to the criteria method testing. When it comes to the 4-d interpolator, we should think about the testing structure first for its applicability in protecting fish and shellfish. Within the holdout tests, we should also be able to understand the behavior of error rates and how that would inform output expectations.

- ***Action Item:*** Further discuss 4-d interpolator testing and error calculations.

Comment: *Leah Ettema:* It would be helpful if someone could summarize decision points for the 4-d interpolator and bring those to an additional meeting.

Comment: *Tish Robertson:* Dave mentioned a good point we need to work out. There's uncertainty with the distribution produced for each cell and the domain of the interpolation. If we have cells that fall outside of our comfort level uncertainty, how do we handle that? There are different kinds of uncertainty we need to tease apart, which we never did with the 3-d interpolator. How do we handle the fact that we were producing estimates that we probably shouldn't have been?

Comment: *Peter Tango:* These are essential questions that are coincident with our deliverable timelines. If we come up with a list, we could put them into phases.

Comment: *Tish Robertson:* How will we handle the different simulations for the CFD? That is something Elgin raised. We need to come up with a rule for determining when a cell is in noncompliance. The rule we have now is pretty hardcore where any excursion across the reference line means it's out. Does that mean if any of the 100 simulations cross the reference line, that is enough to say it's out? Do we need to come up with another rule?

- **Response:** *Leah Ettema:* This seems like there are two categories. What data goes into producing one CFD curve? If there are spatial areas that have too much uncertainty, are we going to exclude those from the CFD curve?

Comment: *Becky Monahan:* It seems like there are two trains of thought here. I love how Leah's presentation laid out all of the things that should go in an assessment methodology. Maybe we should start making an assessment methodology portion for the interpolator and those decisions. Then, we'd have another assessment methodology for the impaired decision. They are two separate things. There is what the interpolator does and that output and there is using that output in further decision making. There would be some overlap. I think it would help us to organize them the same way with the same topics and decision points. I feel like every time we come up with a decision, we get more confused about something else. Maybe looking at the assessment methodology topics would help us organize our thoughts.

- **Q:** *Tish Robertson:* Do you mean the slide that had best practices?
- **A:** *Becky Monahan:* Yes. I feel like you, I and Leah have read and synthesized the blue and white documents, and I've missed things. The last thing we need is another blue and white document, but we do need one synthesizing everything we've known from the beginning with the decision point and why. This new assessment methodology would describe how we are defining things now and document why. It would help synthesize all the documents that may be conflicting. That way our decisions will be documented.
- **Response:** *Leah Ettema:* Would that be everything, including the rationale, in one place?
- **Response:** *Becky Monahan:* Yes. This way we can stop going back and re-reading a million documents and interpreting them differently. I know there will be documentation for the 4-d interpolator but maybe documenting it in a familiar way would help us understand it. I know there will be documentation, but I still don't understand the decision points and where the data is going in.

Q: *Breck Sullivan:* I'm confused. We're going to have a method documentation of the interpolator, including how it was built and the rationale for the methods. You want a separate document that's for the results?

- *A: Amanda Shaver:* Yes. The implementation part.
- ***Response:*** *Leah Ettema:* There are pieces of implementation that will likely go into the 4-d interpolator technical document. For example, the data regions. That is an implementation piece. There is a little bit of overlap.

Comment: *Becky Monahan:* While they are working on the 4-d interpolator, maybe we should pretend we're not using the interpolator and decide what decision is important to us. Then, when we bring in the interpolator, we can fit the interpolator into our decisions, like only using discrete data for instantaneous. The 4-d interpolator output will fit into the decisions. There are too many unknowns and we're circling.

Comment: *Tish Robertson:* I think in Virginia, we are anticipating that there are going to be Virginia segments where the 4-d interpolator confidence is not high because we don't have the array data. We are trying to figure out a way to keep the assessment train going, which is where we might have to use alternatives. We would still consider using the 4-d interpolator, but we may not be able to use it exclusively.

- ***Response:*** *Peter Tango:* I think that's a super viable strategy. It's complementary to what was laid out in 2003. We're making progress and will be using the 4-d interpolator, but if we can't then we have other options.
- ***Response:*** *Dave Parrish:* That confidence piece is still there in the discrete data. That is what is being reflected in the uncertainty from the 4-d interpolator. I wonder if we could use the 4-d interpolator information to speak about confidence in the assessment. As opposed to throwing it out, we could say it meets or doesn't exceed, but we're not super confident in the assessment. I don't know if that fits standards assessment.
- ***Response:*** *Tish Robertson:* If we had a way to have a 4-d interpolator result that says, "we think it's meeting but we're only 45% sure," then that would give us permission to look at another line of evidence. You can use the discrete data and analyze it in a different way. I like the idea of using the 4-d interpolator to communicate the uncertainty in the approach. That would be information that we could send to the EPA.
- ***Response:*** *Leah Ettema:* As a counter, there is a lot of uncertainty in the 4-d interpolator. That uncertainty exists because we're missing data, but you'd still be missing that in an alternative assessment. Why would that be any more certain?
- ***Response:*** *Tish Robertson:* In my mind, the 4-d interpolator is making an assumption of hourly time steps based on the available information. You could look at the monitoring data without using a statistical modeling approach. Then, you could combine those two results. The uncertainty isn't in the monitoring data, but the analysis of the 4-d interpolator.
- ***Response:*** *Dave Parrish:* Whatever uncertainty is coming out of the 4-d interpolator is speaking to the lack of data.
- ***Response:*** *Tish Robertson:* Our tentative plan is that we would report whatever line of evidence indicates impairment. Even if the 4-d interpolator has high uncertainty but the discrete data reports the same thing, then we'd report that evidence. When we're fully supporting, we'd want to look at as much information as we can. We don't want to get into a situation where we only have the 4-d interpolator with a high uncertainty without having anything else to support that. Then we would run into the issue of not being able to evaluate all designated uses.

- ***Response:*** *Amanda Shaver:* We need to be careful because we haven't set a minimum sample size. We'd have to come up with rules for uncertainty based on how much data we have. We'd also have to determine how much uncertainty we are comfortable with.
- ***Response:*** *Tish Robertson:* We would need to come up with that. What's the confidence level? We've just been reporting whatever the 3-d interpolator spits out and assume it's ok.
- ***Response:*** *Peter Tango:* We have 37 million data points that hopefully give us more confidence.

Q: Jon Harcum: We just assume the 3-d interpolator is giving an appropriate answer but are asking the 4-d interpolator to tell us whether there is uncertainty. I am wondering if there is an opportunity to look at the 3-d and 4-d results with a comparable set of data and ask what uncertainty we have and have been blind to that in the 3-d interpolator. This way we wouldn't be setting a bar that's very high. We've been making assessments with the 3-d interpolator for decades. Should we be making similar levels of assessment with uncertainty and how would we quantify that?

- ***A:*** *Tish Robertson:* That's a great point. Elgin would be the best person to talk to about that. We had Elgin do a white paper for us for James River chlorophyll. We used to use the CFD and 3-d interpolator for James River chlorophyll. Elgin did an analysis using the same technique you are using for the 4-d interpolator. He found that the 3-d interpolator when applied with the CFD was no better than a coin toss. That's why we moved away from the interpolator CFD and came up with an in-house assessment method. I don't know if DO would be the same. Elgin would be the best to talk about it because he found that we were dealing with squishy results with chlorophyll. It's fair to say that we're accepting squishy results now, so why get worked up over the 4-d interpolator output? That is because we don't have a way to quantify the squishiness in the 3-d interpolator results. For the 4-d interpolator, we will. If we get that output, we should look at it and use it. Otherwise, why are we producing it?
- ***Response:*** *Leah Ettema:* There is more significance to this decision from the regulatory standpoint.
- ***Response:*** *Tish Robertson:* When we find that out, we do decide to ditch tools. I'm not saying we'd do that with the 4-d interpolator, but it is important to the states.

VII. Adjourn

Next Meeting: March 9, 2026

Attendees:

In-person:
- Allison Welch, CRC
- Amanda Shaver, VE DEQ
- Becky Monahan, MDE
- Breck Sullivan, USGS
- Clifton Bell, Brown and Caldwell
- Dave Parrish, VIMS

- Joe Morina, VA DEQ
- Leah Ettema, EPA
- Matt Stover, MDE
- Melinda Cutler, MDE
- Peter Tango, USGS
- Tish Robertson, VA DEQ

Online:
- Carl Friedrichs, VIMS

- Carol Cain, MD DNR

- Cindy Johnson, VA DEQ
- Guido Yactayo, MDE
- Jim Hagy, EPA
- Jon Harcum, TetraTech
- Joseph Wood, CBF
- Juan Vicenty-Gonzalez, EPA
- KC Filippino, HRPDC
- Kelly Gable, EPA
- Lew Linker, EPA
- Mark Trice, MD DNR
- Michael Echevarria, HRSD
- Rebecca Murphy, UMCES
- Renee Karrh, MD DNR
- Sophia Grossweiler, MDE
- Tom Parham, MD DNR
- Qian Zhang, UMCES