Bay Oxygen Research Group (BORG) Meeting

March 16th, 2026
12:00 PM – 1:00 PM
Visit the meeting webpage for meeting materials and additional information.

Purpose: This was the March monthly meeting of the Bay Oxygen Research Group. In this meeting, Rebecca Murphy (UMCES) shared concepts of data visualization pieces the 4-d interpolator development team is working on. Then, Breck Sullivan (USGS) shared unanswered questions from the group that are being tracked. She asked for feedback on additional questions to add and whether certain questions should be prioritized.

Minutes

I. Welcome, Introductions & Announcements

Lead: ***Breck Sullivan*** *(U.S. Geological Survey, USGS)*

Upcoming Conferences, Meetings, Workshops and Webinars

- Environment Virginia Symposium – March 24-26, 2026. Lexington, Virginia.
- 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. Data Visualization

Lead: ***Rebecca Murphy*** *(University of Maryland Center for Environmental Sciences, UMCES)*

The interpolator team is getting closer to showcasing results. They have been able to run the whole interpolation for all of the segments for one year. The team is creating visuals to communicate these results in the most beneficial way possible. In this presentation, Rebecca shared the visual concepts produced by the 4-d interpolator team. They hope to get insight from participants on what to add or change for most effective understanding.

Some of the existing visualization ideas include maps, time series plots, and summary graphics visualized by segment and as a whole. Rebecca showed draft visuals for each of these ideas. For the map ideas, they would have to choose the time period of those, which could be the full results or summer season.

The development team welcomes thoughts after the meeting via email. Feel free to reach out to Breck Sullivan (bsullivan@usgs.gov) or Rebecca Murphy (rmurphy@chesapeakebay.net) with any feedback or additional ideas.

Discussion Notes (during presentation):

Q (from chat): *Leah Ettema:* For the time series - is that an output for the entire assessment unit or for an individual cell?

- ***A:*** *Rebecca Murphy:* This would be an individual cell. The cell closest to the monitoring station. If it was an entire assessment unit or segment, the oxygen would vary a lot through space. I don't think the graph would show us too much.

Comment: *Carl Friedrichs:* I had a comment on the time series plot. I was thinking that a statistic like the 95th percentile might be better than the full range. The full range will be dependent on the number of simulations. The more simulations you do, the larger the full range will be. It's sensitive. A statistic like 95th, as long as you have more than 20 simulations, or 99th, if you have hundreds of simulations, would be independent of the number of simulations. Tish mentioned that the tests that are done on the data would be good to do on the model as well. It was in the context of looking at assessments. If the phase 7 model results could be plotted on top of this, that would be very cool. Then you could see if the model is likewise in the 25th-75th percentile of the data. The model results could be presented like the threshold maps where the dots show the location and the fraction of time they were below the thresholds. This would be an effective way of helping folks understand how close the model results are to the observations. I really like the formats you are using. They are very intuitive.

- ***Response:*** *Rebecca Murphy:* Thank you. It's easy when I'm just drawing dots in PowerPoint. Hopefully the real results will look clear. I will note the percentile idea you had. In terms of showing the estuarine model results with this, it would be some work to figure out how to do it. We do think we will need to figure out how to show other data on it. It probably wouldn't be too hard of a stretch to show the dynamic model results as well. There are a lot of cross-pollination ideas that could be helpful for looking at those results.

Discussion Notes (after presentation):

Q: *Matt Stover:* I really like these ideas, and I think they will be very helpful. For the table you have on slide 7, it looks like you're comparing the results from the 4-d interpolator with the data. What is the data depicting?

- ***A:*** *Rebecca Murphy:* Great question. I was thinking about Fishing Bay. In a heavily monitored segment, you could take all of the data to see the fraction less than one and compare to the interpolation methods. In a segment that isn't highly monitored, like if there are 2 fixed stations, we wouldn't expect to get the same numbers because that's why we're interpolating – to fill in the gaps.
- ***Q:*** *Matt Stover:* This would take all of the data from it to get a 30-day mean and fraction and compare that. Is that what you're thinking?
- ***A:*** *Rebecca Murphy:* Yes, but there are a lot of problems with that. Like EE3.0 would get overwhelmed by the high frequency stations. If you're trying to compare to the stations, it is better to look at the time series plots. This might be better for comparing the 4-d to itself.

Comment: *Melinda Cutler:* If there is one station that has a higher fraction of dissolved oxygen (DO) exceeding, but everything else looks good. If all of the data is together, you might not get that nuance as much.

- ***Response:*** *Rebecca Murphy:* Yeah, those are great points. Our approaches for criteria assessment have more decisions and take more steps, so this table isn't applicable to that. This could be more for comparing the interpolator to itself – year to year or segment to segment.

Comment: *Matt Stover:* It would be interesting to compare the 3-d and 4-d interpolators. I like the idea because we keep talking about using an alternate assessment methodology. We haven't settled on anything yet. It might be helpful to use some of the segments that are heavily sampled. We know Fishing Bay well and have a good sense of what is going on there. Essentially the whole segment is meeting DO criteria for all designated uses with the exception of the black water creek sample. We would expect some level of exceedance in that creek, but it should be spatially limited. I think it's temporally limited to the summer. We would hope that the 4-d interpolator would show something similar. If it doesn't, we'd have a lot of questions to understand how it used that data to come to a different conclusion. Maybe using that data, we were coming to inappropriate assumptions.

Q: *Leah Ettema:* Isn't that what your time series plots get at since you're showing the data on top of what it predicted?

- ***A:*** *Rebecca Murphy:* Yes. That would be the best way to look at the data compared to the interpolation location by location.
- ***Q:*** *Leah Ettema:* Would you do that for discrete monitoring points?
- ***A:*** *Rebecca Murphy:* Yes. For discrete monitoring, the blue line would be dots in the graph on slide 4. It would be more common that we'd have those. This compared to the station you are talking about with the low DO could be interesting. We'd be able to do that.

Comment: *Matt Stover:* Melinda and I were just saying that all of these look like they'd be very useful and interesting. Our minds always go to Fishing Bay because that is probably the most perfect data set that we will ever be able to have. Viewing it this way is very helpful for validation purposes.

Q: *Leah Ettema:* In the time series maps, have we set criteria for what is acceptable? If your data fell outside of the range, that wouldn't be acceptable. Are we trying to get it within a certain percentile of the simulations? If that does exist, it would be good to add to the graph. What do we consider acceptable?

- ***A:*** *Rebecca Murphy:* That's good to think about. We will probably see a variety. I think the answer is no, we haven't decided what is acceptable. Early in the development, we would see results that were problematic, which is why we made some of the method adjustments we brought to this group.
- ***A:*** *Breck Sullivan:* We have been trying to communicate that with this tool we are not trying to get the correct DO in every place at every time but making sure we have the frequency of data meeting the criteria correct. We haven't set a full understanding of that frequency.

Comment: *Jim Hagy:* When thinking about whether we look at a point and say whether the interpolation is acceptable based on a rule. I think you'd want to evaluate the interpolator from a broad perspective instead of going to a particular spot and deciding that the data is bad. I'm thinking that a lot of data sit on the edge. If you are right above the pycnocline on the Western Shore of the Chesapeake Bay, there are times when the bottom water will come up and the data will be a lot lower than an interpolation would be. It's real but it's not absent having a profiling continuous monitor everywhere. There may be some processes we don't pick up. I am concerned for that being ammunition to say that the whole thing is wrong.

- ***Response:*** *Rebecca Murphy:* I agree that looking at in aggregate is what we want to get right. We don't know the aggregate. That's what we're interpolating to. There are tradeoffs and we will have to look at a lot of summaries.

- ***Response:*** *Tish Robertson:* I think we need to care more about the overall assessment than getting the accuracy of those time series.
- ***Comment (from chat):*** *Matt Stover:* Point taken Jim. I don't disagree. But I do think it's good to use real data to compare to the interpolated data in a variety of segments. In the end, the interpolator should model reality.
- ***Response:*** *Elgin Perry:* I want to echo what Jim was saying. In terms of whether or not we're doing well enough with the interpolator, I come back to the idea that we have taken a step in the right direction if we are doing better than we did before. If we can see that the 4-d has better agreement with the data than the 3-d did, then we can say we're moving in the right direction. Even if we don't have a perfect tool, it's still better than what we were doing and an argument for implementation.
- ***Response:*** *Breck Sullivan:* I want to reiterate that we don't need to wait another 20 years to make improvements. In the future we can work towards making it better if we get new analysis or data. As for Matt's comment, hopefully the time series will be a way to visualize that. The development team ran the time series and saw that the tool wasn't capturing the variability of the conmon, so they fixed it. We are looking at that to make sure we're not misrepresenting the data.
- ***Comment (from chat):*** *Joe Morina:* I think we'll need to look at the DO results directly to address Elgin's point (i.e. how do we know we are doing "better"), but I do appreciate and agree with Jim's point.
- ***Response (from chat):*** *Jim Hagy:* I agree with using real data. The term that grabbed my attention was devising "acceptance criteria." Would we "reject" the interpolation for one location is the data didn't fit there? I would agree that one could look to understand why the interpolator and data might have diverged at the location at that moment in time to understand if there is a larger important issue. Writ large, however, we quantify the model performance by comparing the model to real data statistically, rather than anecdotally.
- ***Response (from chat):*** *Matt Stover:* Great explanation Jim. We agree that that's super important for evaluating the model. One concern for us as a regulatory agency is that we want to make sure that we're not assessing our Bay segments incorrectly since that has real world implications for implementing the total maximum daily load (TMDL) in National Pollutant Discharge Elimination System (NPDES) permits.

Comment: *Leah Ettema:* Thank you, Jim. That was a helpful clarification. I thought the purpose of these time series plots was to show how good the model is doing compared to the data. If this isn't the only thing we're looking at, I don't know if there is another visualization to capture the overall model performance with the data. I don't know what that would be. If this isn't the only thing we are looking at, we can show other things to show the model is performing well. Maybe that's not something we evaluate with every assessment either.

- ***Response:*** *Rebecca Murphy:* That reminds me of the conversation we just got into about the table. We tried to think of other ways to compare to the aggregated data, but we have to be careful that we're representing everything correctly. That fraction less than one might not be a fair comparison. We're scratching our heads to think of how to compare to the data besides the time series plots. Any ideas are welcome.

Q: *Tish Robertson:* Joe Morina put this idea in my ear. What if you restricted the 4-d interpolator output to just the bottom waters and compared bottom observations to bottom interpolations?

- ***Q:*** *Rebecca Murphy:* The dot graphs I showed have been tweaked to show just bottom. Do you mean the DO itself, not just the fraction below a threshold?

- ***A:*** *Tish Robertson:* Yeah. You could still use the threshold and look at the fraction. You could restrict your view to bottom and compare the bottom data to the bottom interpolation estimate for a segment. Maybe in a deep channel segment, you use 1 mg/L as the threshold you're looking at. If you're comparing the whole data set with the whole interpolator estimate, you may not be getting a fair comparison. I'm thinking about Fishing Bay where most of your low DO is in that one marshy creek. It's not in the mainstem of the segment. Another way to make a fair comparison would be to just compare the mainstem interpolator estimates with the mainstem datasets. That way you won't be biasing the dataset towards enhanced monitoring, like what Maryland Department of the Environment (MDE) and Maryland Department of Natural Resources (MD DNR) did in Fishing Bay.
- ***A:*** *Rebecca Murphy:* I think we could do that.

Comment: *Jeremy Testa:* I've been thinking about where oxygen is failing and what we'd consider to be criteria. It's in an environment where I don't know if the monitoring or TMDL motivation ever targeted. I don't know much about Fishing Bay, but if it's a similar black water river, it's probably a place where oxygen has always been low. Tidal marsh creeks are connected to places where we are trying to evaluate DO, but they will have very low DO at low tide because that's how they have always been. I hadn't thought about what this new interpolation approach would mean for adding locations that we aren't necessarily targeting with the oxygen criteria.

- ***Response:*** *Rebecca Murphy:* We had an example last meeting where some of these creeks that have naturally low DO need to be represented and interpolated correctly, but the impacts don't expand beyond where those conditions exist. We did have a problem with some of that high frequency data being over-represented. That discussion might fall more under the Criteria Assessment Protocol Workgroup (CAP WG). For those regions that fall naturally low, I don't know enough. There have to be some exceptions.
- ***Response:*** *Tish Robertson:* We have special DO criteria for our marshland/wetland influenced segments. We might have little tidal creeks flowing into a major segment that have lower DO levels, but for the most part, we've accounted for those segments where we consider low DO as a natural condition. Those are already accounted for in the goals we are trying to meet for those segments.

III. Discussion to Address Stakeholder Questions

Lead: ***Breck Sullivan*** *(USGS)*

In this discussion, Breck shared questions that she has received through emails and discussion that are being tracked. The team is planning on answering them when they have a chance. She opened the conversation to hear if there are certain questions that should be addressed first or additional questions that need to be added to the list.

Using the supporting document, Breck walked through the questions they have tracked and included a status update and when they will be addressed, if available. If participants have feedback or additional questions, they are welcome to email them to Breck (bsullivan@usgs.gov).

Discussion Notes:

Q (from chat): *Amanda Shaver:* For the schedule could you include details on how and when the 4D will be incorporated into the Phase 7 model effort? If you could review that again now that would be helpful leading up to the Clean Water Goal Team (CWGT) [as referred to as the Water Quality Goal Implementation Team (WQGIT)] meeting next Monday.

- ***A:*** *Breck Sullivan:* We are still working with the Modeling Workgroup, 4-d interpolator development team, and the WQGIT to understand the exact timeline. We are hoping to help

make that connection and incorporate the 4-d interpolator into the phase 7 model in the summer of 2027. The part of the 4-d interpolator that supports the phase 7 model is with the planning targets, which won't be worked on until 2028. We foresee that in 2027, after we've done the STAC and partnership review for phase 7 and the 4-d interpolator, we will incorporate the two models and run some results. I don't have specific dates, but I can provide the phase 7 timeline and our 4-d timeline. In the coming months, I see myself helping to merge those timelines with the Modeling WG and WQGIT. That's something we are still working on.

- ***A (from chat):*** *Kaylyn Gootman:* Planning target work start in earnest on January 1, 2028.
- ***A:*** *Kaylyn Gootman:* Thank you for the question, Amanda. Breck has been flagging the timelines and expectations with leadership at the office. Breck, some other folks, and I have been working to get that set up. We'll be happy to share that information as soon as we have the exact timing.
- ***Response:*** *Breck Sullivan:* I think this question along with the discussion will possibly come up during the WQGIT meeting on Monday. For those who are interested, the WQGIT meeting is from 1pm-4pm on Monday March 23rd. Information is available here. This meeting will include the start of our educational journey with the WQGIT. This will help them understand water quality criteria, standards attainment, bioreference curves, 10% reference curves and the 4-d interpolator. We will hopefully be going back to the group multiple times.

Comment (from chat): *Amanda Shaver:* Thanks! It's also good to understand how the updates to the 4D would affect the planning targets.

- ***Response:*** *Breck Sullivan:* We are going through a phase development for the model. Not only do we have new data, but we are also changing the process of the model. Planning targets would have already been impacted by the phase change, but we are also including a new interpolator for the assessment. There will be impacts across the updates of our various models.

IV. Adjourn

Next Meeting: ***April 20, 2026***

Attendees:

- Allison Welch, CRC
- Amanda Shaver, VA DEQ
- Angie Wei, UMCES
- Becky Monahan, MDE
- Breck Sullivan, USGS
- Carl Friedrichs, VIMS
- Cindy Johnson, VA DEQ
- Elgin Perry
- Gabriel Duran, CRC
- Jay Lazar, NOAA
- Jeremy Testa, UMCES
- Jim Hagy, EPA
- Joe Morina, VA DEQ
- Jon Harcum, TetraTech
- Kaylyn Gootman, EPA
- Leah Ettema, EPA
- Mark Trice, MD DNR
- Matt Stover, MDE
- Melinda Cutler, MDE
- Peter Tango, USGS
- Rebecca Murphy, UMCES
- Tish Robertson, VA DEQ