Bay Oxygen Research Group (BORG) Meeting

April 20th, 2026
12:00 PM – 1:30 PM
Visit the meeting webpage for meeting materials and additional information.

Purpose: This was the monthly meeting of the Bay Oxygen Research Group. In this meeting, participants heard from Rebecca Murphy (UMCES) on the thinning of segment grids. This presentation followed a similar one from November 2025. This was followed by a presentation from Jon Harcum (TetraTech) on the simulation count for the interpolator, where the group discussed how many simulations should be used in the tool. Lastly, Breck Sullivan (USGS) shared the timeline for the Phase 7 model and how the interpolator development and review fits into that.

Minutes

I. Welcome, Introductions & Announcements

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

Upcoming Conferences, Meetings, Workshops and Webinars

1. Choose Clean Water Conference – May 18-20, 2026. Lancaster, Pennsylvania.
2. Chesapeake Community Research Symposium – June 1-3, 2026. Annapolis, Maryland.
3. Restore America's Estuaries' 2026 Coastal & Estuarine Summit – September 22-25, 2026. San Francisco, California.
4. Chesapeake Watershed Forum – November 13-15, 2026. Shepherdstown, West Virginia. *Session proposals are currently open and due May 27. More information can be found here.*

Announcements

1. The 4-d interpolator development team is starting to work on a case study with Fishing Bay.
2. The 4-d interpolator development team has received scenario output examples from Richard to help with the comparisons of the 3-d versus the 4-d interpolators.

These tasks will be shared with the BORG Team when they are ready to be shared. The development team is hoping to have the Fishing Bay results ready in the summer.

II. Grid Thinning

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

This presentation looked at the possibility of thinning the resolution for some of the segments in the interpolator grid. Firstly, Rebecca provided background on what the interpolator grid is and how it is divided into segments which have varying horizontal resolution based on segment area. This is the same grid as the 3-d interpolator.

At the November 2025 BORG Meeting, the development team recommended thinning five segments so the interpolator would run smoothly on a normal laptop. The group agreed that these

should be thinned as it improved runtime and didn't have an impact on the results. Participants also recommended thinning the rest of the 50m segments to 100m. The development team has been working on this to understand what impact there would be on the results if these segments were thinned. They also looked at thinning some 100m resolution segments to 200m.

Rebecca shared a list of these segments. On the webpage for this meeting, the grid maps for each segment are available in two zip files. In each of these maps, you can see the original and thinned grids using checkboxes in the upper right corner and the depth of each grid cell displayed by color. The development team has looked through each of these and provided notes and recommendations for each segment. For example, EBEMH has good coverage throughout the segment when thinned, but ANATF_MD does not.

Within these grids, they want to accurately represent the three dimensions, including depth. They did an analysis to see how the vertical representation changed when thinned. Rebecca showed examples of plots where the depth distribution is compared. When comparing for the 50m to 100m segments, there are very similar results for the original and thinning depth distribution. A few examples, like ANATF_MD, have a minor difference, but these are usually segments with such few grid points that they shouldn't be thinned. For the 100m to 200m, there were similar results.

Next, Rebecca shows the dissolved oxygen (DO) interpolation results from the 4-d interpolator for each segment before and after being thinned. These were displayed as empirical distribution functions (EDF). The graphs are labeled where the thinned and non-thinned results are most different as well as how different they are at approximately the 10% point of the distribution. They found that the differences at the 10% point is almost always less than the maximum. Next, they were able to summarize the differences in a graph by segment. This can help determine whether each segment should be thinned or not. They also looked at which direction the differences were going to ensure there wasn't a bias in DO distribution and found that there isn't. The distribution is symmetrical.

Lastly, Rebecca shared the development team's conclusions given each of these tests and suggestions for which segments to thin. For the 50m grid resolution segments, there are four segments that possibly shouldn't be thinned to 100m. For the 100m grid resolution segments, about half of the segments can probably be thinned and the other half possibly shouldn't. The team hopes that BORG members can look at this list (slides 19 and 20) and confirm or make suggestions of changes. In general, none of the impacts are large, so if state or other partners prefer changes, the development team would be happy to discuss.

Discussion Notes:

Q: *Joe Morina:* I have a question about the repercussions of this. In one of your examples, you were dropping the last meter based on this thinning and Elgin pointed out that you'd be interpolating that meter unless you were collecting data there. In the event that some of the bottom depths aren't being covered in this thinning, what does that mean assuming there was data collected at the bottom depth? Would it not be in the interpolation?

- ***A:*** *Rebecca Murphy:* We would not cut any data no matter what happens with grid thinning. This is focused on where the results would be. All the data would still be used in the interpolation. The interpolation output simply wouldn't include that depth if it was thinned. There are only a handful of these so we can check but I don't think it is likely that there would be a sample in that interpolator grid. We could check that and make sure. We still wouldn't be cutting the data even if we cut the grid.

- ***Q:*** *Joe Morina:* The results of the interpolation wouldn't extend to the depth that was thinned, correct?
- ***A:*** *Rebecca Murphy:* Yes.

Q: *Jim Hagy:* The results seem to show that for the most part thinning isn't a problem especially if you have a lot of cells to begin with. Is there a reason that the larger cells, like 1 km, wouldn't go to 2 kms for the same reason we do the thinning? If that also had no effect, why maintain 1000m in open waters?

- ***A:*** *Rebecca Murphy:* That is something worth thinking about. The 1 km grid makes more sense with the common sampling resolution between some of the closest stations. These 50 and 100m resolution grid segments are so thin. They were designed to make sure we cover the area of the water well, but we're not getting enough information from the samples to make a difference. I think some of the coarser ones, with 500m and 1 km, would likely start to have more impact on interpolation results. That's a good point though.
- ***Comment:*** *Breck Sullivan:* We were initially trying to keep the same grid as the 3-d interpolator but ran into issues with the 50m. That's when we wanted to change those but keep the rest the same. That is something to consider.

Q: *Jim Hagy:* Was the computational time for the main Bay segments not a problem?

- ***A:*** *Jon Harcum:* The short answer is no. I can look up the details of that but it didn't seem prohibitive.
- ***Comment:*** *Rebecca Murphy:* It's ironic that these small areas are the ones that take the longest time because the grids are so fine.
- ***Comment:*** *Elgin Perry:* It wasn't computation time. It was that the five problematic segments required a bigger machine. You couldn't run them on your laptop. You had to get a work station or something to run those. Our motivation to thin those was to make the code accessible to more people. The rest of the segments, including the 1 km segments in Main Bay, run fine on a standard laptop. That's why we weren't considering thinning those and stuck with the resolution from the 3-d interpolator.
- ***Comment:*** *Jon Harcum:* When we fill out the grid cells in some of those thin diagonal segments, we fill in a lot of points in order to create a large-scale correlation matrix. What seems like a small body of water can magnify quickly.

Comment: *Rebecca Murphy:* The teams can think about this for the next month as we approach our next meeting. If you have any thoughts or want to see anything specific, we'd be happy to point to the right materials and give our insights. Please email me (rmurphy@chesapeakebay.net). I'll be reaching out to those who weren't on this call to make sure they have everything they need and understand it. We will be bringing this back in May after people have time to look over it.

III. Testing Simulation Count

Lead: ***Jon Harcum*** *(TetraTech)*

When running the interpolator, the goal is to run the smallest number of simulations to yield a scientifically stable answer. This depends on what metric is being looked at, like the mean simulation value or a 10^{th} percentile. This would change how many simulations are needed. Jon shared a graph that can visualize how many simulations the group would be going for.

Next, Jon explained how they study the different number of simulations and what they look at when determining this. This slide (3) shows a graph of how the different metrics change the number of samples needed. Jon presented some questions for the group to guide discussion and help the development team in determining the number of simulations needed.

Discussion Notes:

Q: *Jim Hagy:* I noticed on your graphs that the x-axis was "few, more, many." Why don't you have the actual number of simulations?

- ***A:*** *Jon Harcum:* It was meant to be qualitative today. On slide 1, if I put the numbers back on it would go from 5-80. I wanted to keep our conversation more qualitative today.
- ***Q:*** *Jim Hagy:* The point is just that there is a number where more simulations gives us relatively the same results and that is what you are trying to find. Eventually, you'll have to know what that number is, but you don't want to dwell on it now. Is that what you're saying?
- ***A:*** *Jon Harcum:* Yes, that's exactly right. We can do this type of analysis when we run a simulation on a segment-by-segment basis. It may be that some segments look like the graph on slide 3 while others might take longer to converge to the reference point we're using. When we look at a graph like this, what characteristics would you look at to the point of feeling comfortable? Then, we could keep that in mind as we look across the 92 segments. Then, we can be informed when we come back with the actual results.

Comment: *Rebecca Murphy:* It might help this group better understand if we talk about what our summary statistic might be. Say we're talking about a segment of Fishing Bay, which only has open water. We might want to aggregate the fraction of DO that is less than 5mg/L from each simulation. For that situation, using the mean makes sense to me. It might help to describe what the summary statistic might be.

- ***Response:*** *Jon Harcum:* Let's suppose there is a seven-day mean criteria less than 5mg/L. It might make sense to calculate a mean statistic. We'd get a mean value from each simulation. If we look across all of those simulations, we would get a range of those values. Let's suppose that the seven-day mean from simulation #1 was 4.6, #2 was 5.2, and #3 was 5.1. When we aggregate those together, we can get an overall mean and that would be a value that we would carry forward. For the same group of seven-day means, we could also look at the tails of the distributions. In this example, if I have 100 simulations, I might have a distribution of seven-day means running from 4.5 to 5.2. The middle value is halfway between the two but maybe my 10^{th} and 90^{th} percentiles might be 4.65 to 5.1. It will take a different number of simulations for a mean or percentile to become stable, with the mean becoming stable more quickly.
- ***Response:*** *Rebecca Murphy:* To me, it makes a lot of sense to find stability of the mean for all of the simulations, but I'm not sure we'd need it for the percentiles. Maybe we would need it for the 10^{th} percentile. That is part of our question here. We are asking what metrics we want to use for deciding the count of simulations.
- ***Response:*** *Jon Harcum:* Because we have the instantaneous criteria, we might want to look at percent of observations below an instantaneous threshold instead of the mean. Let's continue with the earlier example. I believe the instantaneous minimum criteria is 3.2 mg/L. Maybe the question would be, what is the percentage of hourly observations less than 3.2 mg/L and see how stable that is. That would work well where we have simulation values less than 3.2. If we have a segment that is clearly not impaired, that statistic might not be as easily evaluated. When we look at a percent less than, what are we comparing? Do we want to stick with 1 mg/L for the deep channel, 0.7 for deep water, and 3.2 for open water or do we need to customize that depending on the segment to get a more meaningful statistic.
- ***Comment (from chat):*** *Peter Tango:* I do like that thinking, John.
- ***Response:*** *Elgin Perry:* It's likely that different endpoints are going to reach stability at different rates. I think the end point we are most concerned about is the cumulative frequency diagram (CFD) because that's what we are using to assess criteria. I think we should probably

run some experiments to assess which one of those approaches stability most slowly. If we were to choose a number of simulations that gave us an acceptable degree of stability for the most sensitive endpoint, then we could call that an umbrella for the remaining criteria. We know the others will reach stability much sooner, but since we know we have to run 75 simulations, for example, to reach stability, we wouldn't go back to subset to a lower value for a remaining criterion. We've already created 75 simulations.

Q: *Jim Hagy:* If you had the most difficult to get right end-point, can you calculate that end-point while you're simulating and terminate the simulations when it converges?

- ***A:*** *Elgin Perry:* I suppose that is a possibility. You could use an adaptive approach to see how much the results are changing when you add each simulation and stop the simulation when it reaches less than 1%, for example. I think that's feasible. We could take a look at that.

Comment: *Elgin Perry:* Earlier Jon mentioned using a large number of simulations and set that as a standard to compare against. Another approach could be to conduct an experiment where you knew what the answer was. You could use expected probabilities to figure out what the exact answer will be and figure out how close you are to that. We might implement something like taking output from the water quality model and sample that according to our sampling network. We could create a dataset that was representative of a year of sampling and start running simulations to see if we could start getting close to what we know the answer was from the water quality model data.

Comment: *Breck Sullivan:* It seems like we have two different approaches to consider looking into more. This is another presentation topic that we will be bringing back to the BORG meetings in the future. If anyone has comments or suggestions, they can email us (bsullivan@usgs.gov).

IV. Phase 7 Timeline and Connection to 4-d Interpolator

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

As mentioned in previous meetings, one of the purposes for the 4-d interpolator is supporting Phase 7 to help identify planning targets. First, Breck shares what she means by the Chesapeake Bay Program suite of models or Phase 7, which includes the Land Use Change Model, the Airshed Model, the Watershed Model and the Estuary Model.

Next, Breck shares the timeline for the Phase 7 model, including a closer look into the development and review. The Scientific and Technical Advisory Committee review will be taking place from February 2027 to July 2027, and the partnership review will be from February 2027 to September 2027. The final approval will take place in December 2027. Then, she shares a closer look at what the modeling team will be working on within the model development for 2026.

Next, Breck shows a draft of how the 4-d interpolator timeline will work alongside the Phase 7 timeline. The 4-d development team plans to continue working on development through August. Previously, BORG members have asked to review the documentation for the interpolator before it is sent to STAC. This might not be possible given time constraints, but the development team is hoping to be able to send out sections as they are completed. The goal is to send the documentation to STAC and pause interpolator development by November. This is so the review can take place before the Phase 7 review begins. There are multiple factors that make the STAC review time unknown. Once the STAC review is complete in 2027, there will be time to adjust based on the STAC review feedback before it is incorporated into the Phase 7 model. In preparation for this, the development team will be providing updates and suggestions to the Water Quality Goal Implementation Team (WQGIT). The timeline in this presentation is a draft. Breck will provide updates as things are confirmed.

V. Adjourn

Next Meeting: ***May 18, 2026***

Attendees:

- Allison Welch, CRC
- Angie Wei, UMCES
- Breck Sullivan, USGS
- Carl Friedrichs, VIMS
- Cindy Johnson, VA DEQ
- Colin Hawes, VIMS
- Elgin Perry, Independent Statistician
- Jay Lazar, NOAA
- Jeremy Testa, UMCES
- Jim Hagy, EPA
- Jon Harcum, TetraTech
- Joseph Morina, VA DEQ
- Kaylyn Gootman, EPA
- Leah Ettema, EPA
- Marjy Friedrichs, VIMS
- Mark Trice, MD DNR
- Peter Tango, USGS
- Rebecca Murphy, UMCES