



Bay Oxygen Research Group (BORG) Meeting

July 1st, 2026
10:00 AM – 11:30 AM

[Visit the meeting webpage for meeting materials and additional information.](#)

Purpose: This was the June monthly meeting for the Bay Oxygen Research Group (BORG). This meeting focused on addressing questions about the 4-D interpolator. Presenters provided an overview of the tool and focused on one component of the tool that the 4-D Development Team has been working on for the past few months.

Minutes

I. Welcome, Introductions & Announcements

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

Upcoming Conferences, Meetings, Workshops and Webinars

- [Restore America's Estuaries' 2026 Coastal & Estuarine Summit](#) – September 22-25, 2026. San Francisco, California.
- [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](#).

II. Documentation Questions and Review

Lead: Breck Sullivan (USGS)

Action Items:

- The Development Team will make the 4-d interpolator documentation available via Sharepoint. This will include the diagram Rebecca shared.
 - This action was completed. The Sharepoint can be reached [here](#). If you have trouble accessing it, please contact Rebecca (rmurphy@chesapeakebay.net).

Discussion Notes:

Q: *Tish Robertson:* I haven't been able to look at the documentation in detail, but I've skimmed it and it seems comprehensive. I'm looking forward to doing a deeper dive. Was there a deadline for feedback on it?

- **A:** *Breck Sullivan:* There is no due date. We are trying to leave time during these meetings for questions, but you are also able to email them once you have time to review it.

Q: *Leah Ettema:* Was that documentation shared with everyone?

- *A: Breck Sullivan:* Yes, and I'd be happy to forward it to you. (Feel free to email Rebecca (rmurphy@chesapeakebay.net) if you need a copy.)

Q: Breck Sullivan: I am wondering how you'd like to receive more documentation. We have a few more things that we could send next week. I don't want to overwhelm you with documentation. Would you like me to send it when it's ready? Or wait until I hear from you about reviewing the first part of the documentation and being ready for the second? We are planning to save time in these meetings for questions around documentation. I could send you a new piece after each BORG meeting.

- *Q: Amanda Shaver:* Could we collaborate and leave comments in a SharePoint file?
 - *A: Breck Sullivan:* I have not set it up on SharePoint, but I can. That could be easier, because then I could upload things when they're ready and all in one place.
 - *Comment: Amanda Shaver:* Multiple versions with different comments could be difficult, but if we were all in the same document, we could see answers to other people's questions.
 - *Comment from chat: Becky Monahan:* That would be great on SharePoint. That way it's all there and if there are updates, they will just be there!
 - *A: Breck Sullivan:* I will create a SharePoint. That will probably be ready next week because of the holiday. I will send it out and you can let me know if you have trouble accessing it.

III. Overview of Current Work

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

In this presentation, Rebecca shared a new workflow diagram that the Development Team has created. This flows through the data, mean mid-day interpolation, parameterization, and then into the simulation. She used the old workflow diagram to provide context with the new diagram. The items within dashed boxes are still under development and the items highlighted in pink are computations.

First, Rebecca described the "Organize Data" section. This data includes all of the sources that have been mentioned previously including the continuous monitoring stations, the fixed stations, and other data sources. The dataset has been recompiled including up to 2024 data and code has been developed for annual data additions. All of this data is put together into a "Data Lake."

This is used to run the daily interpolations for the pycnocline and oxygen. The generalized additive models (GAM) are fit by a year at a time in the interpolation regions previously presented. This flows into a large residual database of dissolved oxygen (DO) GAM residuals. This includes interpolations from almost every historical dataset.

Next, Rebecca covered the parameterization step, where the residuals are fed into the different correlation computations. These parameters will be able to be updated with new data each year.

Then, everything is pulled together into the simulation to simulate error and sum results to get the hourly DO. They do multiple simulations to increase confidence in results. They are still working on the simulation testing. This will be finalized at the end of July.

Currently, the team is discussing the approaches for linking this to the CBP model scenario output and for criteria assessment computations.

Breck also mentioned that Rebecca can walk through this diagram again, if needed, and will circle back to it in future presentations.

Discussion Notes:

Q: *Tish Robertson:* Will you also be producing hourly temperature and salinity results?

- **A:** *Rebecca Murphy:* We've been very focused on oxygen. We know that in order to do some of the criteria analysis, you need to split things by temperature and salinity. Ultimately, yes, but not by the end of July. That will be in our linking to criteria part of the project. The Data Lake includes temperature and salinity and the code allows you to run GAMs on those instead but we haven't done quality analysis (QA) yet.
- **Response:** *Tish Robertson:* So, the long-term plan is to have the interpolator give us those parameters too.
- **Response:** *Rebecca Murphy:* Yes, we might do those on the daily level.
- **Response:** *Tish Robertson:* I don't know about that since we have the instantaneous minimum (IM) criteria that hinge on the temperature. The critical temperature for sturgeon is 29 degrees. I don't know if we'll hit that with a daily mean.
- **Response:** *Jon Harcum:* We look at that in the combined pycnocline. Code has been adopted to compute the pycnocline. We look at the temperature and salinity there, but we don't have a direct GAM for those things at this point in the development.
- **Response:** *Rebecca Murphy:* Thank you for your comment. That's a great point. I understand the high temperature needs and how that will vary throughout the day.
- **Response from chat:** *Elgin Perry:* Tish, my thinking on other parameters is that DO is one of the more complex parameters to interpolate and that if we can get this tool working for DO, then it should not be difficult to make it work for temperature and salinity which are typically better behaved than DO.
- **Response from chat:** *Jim Hagy:* Elgin, I agree very much with water temperature. You might find some challenges with salinity because it can change abruptly in some places with rain events.
- **Response from chat:** *Tish Robertson:* Elgin, thanks for the assurances! High-frequency temperature is going to be essential for the Open Water IM criteria assessment. The salinity is important too for the Migratory Fish Spawning Nursery 7-day mean. But it doesn't have to be hourly. A daily mean would be sufficient.

Q: *Becky Monahan:* Where is the data coming in and being evaluated? Is it in the parameterization box?

- **A:** *Rebecca Murphy:* We are using all of the hourly data in the mean mid-day interpolation. The Data Lake is the hourly data, which is pulled in a region and put into a GAM. Those fit the hourly data in space and time to different relationships and give us a daily estimate for each grid cell. The data is used in the mean mid-day interpolation step, but the result is daily. To get an hourly estimate we're using the data in the parameterization step.
- **Response:** *Leah Ettema:* Could you clarify where the three-year assessment data go in?
- **A:** *Rebecca Murphy:* You'd do this three times. When we were testing, we tried a three-year period, but it was better to interpolate the data one year at a time. You'd pull one year out of the Data Lake and fit an interpolation to get the results. Then, you'll do it again for the next two years.
- **Response:** *Becky Monahan:* Thank you. That is a helpful explanation. The data is being used in the blue box (mean mid-day interpolation) to get the daily results. Then, the grey box (parameterization) is to get the hourly prediction. We are using the data in real time year to year. The model isn't made with 10-year data, and we are using the data for the model. The data is being updated every year.
- **Response:** *Rebecca Murphy:* Yes. We will definitely update that every year. The blue part will only use the data from the current year. The grey box uses historical data as well. The

parameterization will use the current years data and all of the data before because we don't always have high-frequency data everywhere.

Comment from chat: *Amanda Shaver:* It may help to add a definition of "Data Lake" to the slide.

- **Response:** *Breck Sullivan:* Thank you for the suggestion. We can do that in this visual and the documentation.
- **Comment:** *Jon Harcum:* The documentation Breck sent out includes how we screen the data and look at the different remark and comment codes that are processed in the left-hand box. They are in the Data Lake and organized by year and segment. All of the flags related to data qualifiers have been removed. All of the data in the Data Lake is good to use.
- **Response:** *Becky Monahan:* Most people will probably go to this document over the documentation so it would be nice to have a one-sentence summary in the legend.
- **Response in chat:** *Leah Ettema:* and/or adding more description to the top of the blue box describing something like "each year of 3-year assessment window"

Q from chat: *Amanda Shaver:* Breck can you share the diagram in the SharePoint so we can provide any additional thoughts/comments?

- **A:** *Breck Sullivan:* Will do!

IV. Day-to-day Correlation

Lead: Jon Harcum (TetraTech)

This interpolation is assembled from additive components, one of which is the large-scale correlation component. Jon showed the observed data from West Gooses station in 2022 and points to the "memory" of the data from one day to the next highlighting the idea that each day is similar to those near it and how similar patterns can be observed between the adjacent depths.

On slide 5, Jon showed a seasonal GAM pattern and daily departure patterns and combined them into the simulated mid-day dissolved oxygen. This is based on a correlation value of 0.4. On the next slide, he showed what the different correlation values look like with a correlation value of 0 having no correlation. When there are higher correlation values, the data departs from the mean in a similar manner for longer. The figures on the next slide show how the correlation value looks throughout a four-day period. Elgin and Jon believe that higher correlation values produce results that look more similar to observed dissolved oxygen data.

Next, he discussed some of the steps within the mean mid-day correlation, which help get the day-to-day correlation. On slide 10, he showed an interpolation region and explained how the interpolation uses the target and boundary segment stations. They begin with the available DO data which may be subsampled to ensure there is no bias from different data availability. Then, they fit the mid-day GAM to this data, which is shown as black dots. On the next slide (13), Jon removed the expected mid-day DO to get the DO residuals. These residuals are developed during the mean mid-day interpolation and used in the parameterization, including the day-to-day correlation.

This data is hourly but needs to be subsampled to get a daily value. They are looking for a residual near 11:00. They select one residual in a four-hour window around 11:00. On slide 16, he plots the residuals for West Gooses at 1m and can calculate the day-to-day correlation for July 1 using the data from June 1-August 1. Then, he plotted current day against previous day for each day in this time period to estimate the local lag-1 correlation. Next, he plotted the record availability on a heat map with each interpolation region. This allows them to see where and when there was continuous monitoring (common) data available.

Jon calculated the day-to-day residual correlation for each interpolation region and found that the correlation is usually positive. The variance between each segment supported the need for regional parameterization.

On slide 18, Jon shows a graph that represents day of year versus correlation value. The colors represent the type of monitoring data with blue being vertical profiles and green being commons. Using this, Jon can see the change in correlation over the seasons. He found that there is a strong relationship with day of the year, but not as strong with depth. They used the Fisher's transformation on the regional day-to-day lag-1 correlations. Then fit that to a common model based on day of the year. They didn't include depth in the equation because they found little correlation in most examples, but they are unsure about the results they got for the Lower Rappahannock, which was shown on slide 19. They didn't see it often enough to move it forward in the model equation. From this, they can calculate the simulation that is representative in space and time. They can also use this correlation value in future simulations, if there isn't new common data available.

Discussion Notes:

Q: *Tish Robertson:* Jon, you mentioned that the Lower Rappahannock was an oddball because the depth seemed like an important variable for the correlation, whereas other segments don't have that. Was that the takeaway?

- **A:** *Jon Harcum:* Yes. That's exactly the takeaway. I believe the station's name is RPP 01.26. That is the blue in these charts. It's a profiler so it is taking observations at various depths throughout the water column. It usually has at least one observation to each depth in an hour before it starts back over. That's why we have estimates for the lag-1 correlation through the depth. The top of the graph is the deepest and has the highest correlation. As you move towards the bottom of the graph or the surface of the water, the correlation decreases. The green shows the common stations.
- **Response:** *Tish Robertson:* I've always said that the Rappahannock is the weird one but I don't know why it's different than the other tributaries.
- **Response:** *Jon Harcum:* That's why we chose to not include this correlation. If there was a physical reasoning that could provide context, we might be able to bring it forward, but we don't feel like we have that.

Comment: *Jon Harcum:* In the beginning, Becky asked about where the data was being used. The mid-day model fits the middle of the data, but you still have the deviations from long or multi-day runs below or above that mean. That's what we're trying to get at with this day-to-day correlation. When we're looking at the 1m data on slide 12, we can see these runs of data where they move together. That's where the day-to-day correlation comes in. We also have the depth correlation where 1m and 3m would be similar. That's another component, as well as space. So, stations that are close together are correlated based on how far apart they are. That is spatial correlation using dataflow data. We can present on the depth correlation if people are interested. Those three together form the large-scale correlation component.

Comment: *Elgin Perry:* I want to go back to Becky's question on where the data is used. I see that the data permeates this diagram and wanted to offer that perspective. To me, the Data Lake is taking the data and splitting it into two parts. There is the part we can predict based on location, depth, and the other variables. Then there is the part we can't predict, which are the residuals. Those two things together still represent the entire dataset. We haven't thrown anything away. The predictable part is moving down the right-hand side of the figure and the residuals are moving into the parameterization box, where we do more analysis to study the behavior. If we had intensive data everywhere all of the time, we wouldn't need that box. We would have a set of

residuals to go with every piece of predictable data. Because we need to take the historical information and move it into places where we don't have data, we have to have the parameterization step. We study the behavior of the residual data in the places we have it so we can use it in years when it isn't available. To me, the data moves through this diagram. It isn't used in a particular part. Every deviation from the model has an explanation but we don't know why yet, but we can study it, see if there are patterns, and use them in parts of the Bay where we have no data.

- **Response:** *Becky Monahan:* Thank you, Elgin. I think my question was more about the three-year window. I thought we were making one model for the three-year window for the Bay, and I was confused on how that would fit the data. Understanding this workflow has helped. I understand where the data is coming in.

Q: *Leah Ettema:* Was there much variability at a site between the day-to-day correlation observed between years or were all of these one year of correlation data at a site?

- **A:** *Jon Harcum:* Yes, but it's a generally positive value. We looked to make that window ± 10 then it gets quite noisy. If I extend it out, the value tends to calm down and be a little less noisy. Ultimately, we're looking for a single representative value for the segment. It's ok that there is a little bit of noise because we boil it down to a mid-value. If we took all of the individual lag-1 correlation coefficients and pulled it together in a segment interpolation region, you'd get a probability density function like on slide 18. This is usually between 0.5 and 0.8. It is highly variable at a station but once you look at the central tendency there is some windowing in onto certain areas.
- **Response:** *Leah Ettema:* This includes multiple years of data at a site too. All of the lag-1 components for whatever data is available.
- **Response:** *Jon Harcum:* Correct. This graphic on slide 17 represents all of the data. Some points are reused as part of the boundary segments.

July Meeting Discussion

Comment: *Breck Sullivan:* Our next meeting is scheduled for July 20th. Multiple people on the Development Team will be out that day. We could reschedule for the following week or the first week of August. I know the CAP AT meeting on July 13th will focus on the criteria assessment rules using the 4-d outputs. It will have a lot to do with the tool and connecting it to criteria assessment. That might be able to supplement having a meeting on the 20th. I would love to get feedback from everyone on what we should do about that.

Comment: *Amanda Shaver:* I was thinking about the timeline we had for the 4-d interpolator and the decision rules. When those are presented on the 13th, I don't know if there would be any immediate red flags that we could figure out then. Then we could see an output on that at the August 17th BORG Meeting.

Comment: *Tish Robertson:* I'm not saying we have to have a full-fledged version of all of the criteria. It would be good to see what some of the criteria might look like. Maybe especially the ones that don't hinge on temperature and salinity.

Comment: *Breck Sullivan:* The 4-d Development Team's initial goal is to finish the tool by the end of July. Then, August and September would be for showing general results, comparisons with the 3-d and then taking in the criteria assessment decisions. I'm not sure how much could get done by August 17th. If we wait until then, it would provide more time to show more.

Comment: *Amanda Shaver:* If there are other key components not related to criteria assessment, it'd be good to have another meeting in between.

Comment: *Breck Sullivan:* We are cancelling July 20th. I will talk with the Development Team to see if we can set something up within this timeframe to go over one of these additional components. If we aren't at that stage, we will convene on August 17th. Also, this will probably be my last BORG meeting until December. Peter Tango will be taking over the coordination for these meetings until I'm back.

Note: The meeting has been rescheduled for Tuesday, July 28, 2026. More information is available on the calendar webpage linked below.

V. Adjourn

Next Meeting: [July 28, 2026](#)

Attendees:

- Allison Welch, CRC
- Amanda Shaver, VA DEQ
- Andrew Keppel, MD DNR
- Angie Wei, UMCES
- Becky Monahan, MDE
- Bella Konchar, F&M
- Breck Sullivan, USGS
- Cindy Johnson, VA DEQ
- Elgin Perry
- Jim Hagy, EPA
- Jon Harcum, Tetra Tech
- Gabriel Duran, CRC
- Leah Ettema, EPA
- Marjy Friedrichs, VIMS
- Mark Trice, MD DNR
- Melinda Cutler, MDE
- Rebecca Murphy, UMCES
- Sophia Grossweiler, MDE
- Tish Robertson, VA DEQ