Bay Oxygen Research Group (BORG)

Monday, May 19, 2025
12:00 PM – 1:30 PM

Meeting Materials Link
This meeting was recorded for internal use to assure the accuracy of meeting notes.

MINUTES

12:00 PM: Introductions/announcements
Peter Tango (USGS), Chair

Overview:

- Colin Hawes took a moment to introduce himself to the group. He is a PhD student at Virginia Institution of Marine Sciences (VIMS), working with Marjy Friedrichs.
- Peter Tango introduced the meeting agenda and connected today's conversation with the CAP WG and the new Watershed Agreement Outcome language.

12:05 PM: 4-dimensional (4-D) Interpolator Development Overview
Kaylyn Gootman (EPA)

Overview: Kaylyn Gootman gave a brief overview of the 4-D interpolator project and progress. She summarized the purpose of the Bay Oxygen Research Group (BORG). Kaylyn also dove into the background of interpolation, specifically in the Chesapeake Bay. She introduced the current interpolator, developed over 20 years ago, and explained the need for an updated interpolator. Lastly, she shared the goals of the interpolator and the project timeline.

Questions and Discussion

Q: *Marjy Friedrichs:* Hourly DO throughout the Bay is the output, you said? Is this in volume or actual concentration?

- ***A:*** *Elgin Perry:* It is concentration. The spatial grid on the surface is 1kmx1km. In the depth dimension, we are going in meters. In the time dimension, we are going in hours. We are predicting a concentration at each point in that 4-D grid. We are doing this in a stochastic way, so that we can do this repeatedly and get

an idea of the uncertainty of our predictions. That's based on the residuals we observed from calibrating the smoothing curves in the beginning along with spatial and temporal correlation. It's concentration and it is done repeatedly to get an idea of variability.

- ***Q:*** *Marjy Friedrichs:* The vertical resolution would be every meter?
- ***A:*** *Elgin Perry:* Correct. We are having a discussion right now about how deep the Bay is at every location. We are working on that.

Q: *Marjy Friedrichs:* Would this go back to 1985 and extend to the most recent data?

- ***A:*** *Elgin Perry:* We could do that, but we haven't at this point. We have been focused on the most recent decade. We have only implemented this for one segment and one year on an hourly scale.

Q: *Marjy Friedrichs:* When you are done with this process would you use that information to create the hypoxia volume estimates too?

- ***A:*** *Elgin Perry:* We have been focused on criteria assessment, which looks at each cell for a pass or fail and aggregates that over segments in time. I don't see why we couldn't use this with post-processing to achieve that. Jon has worked on a system of storing these simulations and post-processing them as needed.
- ***Comment:*** *Marjy Friedrichs:* If you could post this in a cloud or server, then this could be a huge resource and dataset for students and researchers.
- ***Response:*** *Kaylyn Gootman:* That is part of the plan. We want all of this to be accessible.

Q: *Matt Stover:* Have there been any drafts of documentation that folks could review?

- ***A:*** *Rebecca Murphy:* Not yet, we have some memos of decisions that have been made and descriptions.
- ***Response:*** *Breck Sullivan:* Peter and I set a date for us to meet to draft an outline in a format that others could review. It is on the way.
- ***Response:*** *Matt Stover:* I think those decisions are most important to us. I appreciate the fact that the 4-D is taking all of the data and looking at the data hourly. That is important to us. We want to be party to those decisions.
- ***Response:*** *Breck Sullivan:* Most of the decisions have been brought to the larger group. Maybe it would be important to identify what those decisions were. We recognize that it is a partnership decision and so we have been bringing those decisions forward to the group.

- ***Comment:*** *Matt Stover:* For the CAP WG meetings, we are starting to include whether there are decisions being made at the start of the agenda. This way people know that those meetings are very important to attend. That could be an idea for this group too.
- ***Response:*** *Elgin Perry:* Of course, we want to hear from you. The best way to do that is to attend these meetings because we bring pretty much everything here.

12:30 PM: Progress Update: Recent Work
Rebecca Murphy (UMCES)

Overview: Rebecca's presentation gave an update on the recent work. Rebecca divided the recent work into four categories.

Parameterizing spatial correlation: Jon has been working on this and now it has been mostly wrapped up. The graph Rebecca displays on slide 3 shows the middle of each DataFlow data set and a mean correlation was computed from the different times the DataFlow was collected in that spot or region. The dark blue are places that are more correlated, and the oranges and reds are less correlated. There is a lot of variability across the Bay that should be represented in the tool. It is important to note that this is surface data, and it is missing spatial correlation at depth. That is a future improvement to be made.

Q from chat: *Leah Ettema:* Can you remind me, what is dataflow data?

- ***A:*** *Rebecca Murphy:* DataFlow is a type of data collected in the Chesapeake Bay. A boat goes through the water, and a sensor measures multiple water quality parameters. It is spatially dense for surface water information.
- ***A from chat:*** *Tish Robertson:* Dataflow = "underway" monitoring
- ***A from chat:*** *Matt Stover*: I like Rich Batuik's description of it, calling it the "drunken sailor survey".
- ***A from chat:*** *Peter Tango:* Multi-probe WQ sensors capturing measurements every 3-4 seconds. As Tish said, underway monitoring. Oh yes "drunken sailor" survey

Q: *Carl Friederichs:* Why aren't the York River and Rappahannock River displayed on this map?

- ***A:*** *Rebecca Murphy:* This is just a background map; it isn't our grid. We should try to change our background map for our final report.

Dataset structure and compilation: Jon is working on this as well. This is being done to address the concern that the high-frequency data is not being used at its high-frequency. Once this is finished, the team plans on going back to the generalized additive model to use the high-frequency data at the frequency it was collected at. This structure will be used for yearly data inputs, which will need to be a regular process for the 4-D interpolator, just like it is for other Bay Program data tools.

Pycnocline: This step is being done by Rebecca and is well documented in the criteria documents. On slide 5, Rebecca shows data for one part of the Bay where this analysis was performed.

Linking the tool to criteria assessment: Elgin has been working on this step. This will directly relate to the Criteria Assessment Protocol WG.

Questions and Discussion

Q: *Tish Robertson:* Elgin said that the simulations are stochastic. Elgin, can you explain what you mean by that for generating the CFDs?

- ***A:*** *Elgin Perry:* When we fit the smoothing curves, there are deviations between the smoothing curves and the observed data which gives us an idea of the variability that the observed data have around the smoothing curves. Once we do our mean projection for a cell in the 4-D grid, we add noise to that based on deviations of the observed data and smoothing curves. In that, we try to reflect the spatial and temporal dependence observed in those residuals. The smoothing curves represent the changes in the mean behavior of dissolved oxygen. That's not going to capture the extremes you want when you're assessing the instantaneous minimum criteria. That's why we think it's important to put that stochasticity into the data. We do that repeatedly and go through the CFD calculation for each simulation. That is where we get the multiple black lines from the graph on slide 6.
- ***Q:*** *Tish Roberston:* When you say stochastic, are you describing a random process and are you sub-sampling the distribution?
- ***A:*** *Elgin Perry:* We are not doing subsampling at this point. We are generating random numbers from a normal distribution. We're pumping them in the reverse direction through our beta-logit transformation, which gives them a distribution that seems to fit the observed data well.
- ***Q:*** *Tish Robertson:* You're doing that for each interpolator cell?
- ***A:*** *Elgin Perry:* That's correct.
- ***Q:*** *Tish Robertson:* Is that process happening independently for each cell?

- ***A:*** *Elgin Perry:* It is a little misleading to say that we're doing it for each cell because we look at this huge 4-D matrix of observations. We first build in temporal correlation using an AR1 type process within each cell. That temporal vector through time gets the spatial correlation built onto it. We're using matrix algebra. It's happening with these big arrays. If you have a vector of multivariate normal observations and you take the Cholesky decomposition of the variance/covariance matrix for that vector. If you multiply a set of independent normal random variables by the Cholesky decomposition, you get a vector that has the prescribed variance/covariance matrix. We start off with independent normals, but we have done studies on spatial and temporal correlation. We can use that to construct variance/covariance matrices that reflect what our observed data looks like. They're not behaving independently when we get to the full simulation.
- ***Q:*** *Tish Robertson:* Thinking of this in terms of the spatial correlation makes a lot of sense to me. You have reminded us that we are dealing with probabilities. I am trying to understand what probability we are looking at. The probability that the segment has failed according to the CFD? The reference curve? Or something else, like the probability of a cell having certain violation rate?
- ***A:*** *Elgin Perry:* That's one of the problems I am working on now. With the CFD itself, I find it hard to do probabilistic stuff. When I present my examples, I will present simpler ways of assessing them, like the simple 10% rule. That makes it pretty easy to take the simulations generated and fit a distribution curve to the violation frequencies for that segment to calculate a probability. When you get away from the CFD, we can talk about probabilities.

Q: *Jeremy Testa:* I didn't realize that the volume that's assessed in deep water habitat could change based on the stratification. Maybe that potential for temporal variation in the magnitude of the habitats. Do you see that the change is big over time? How has that happened in the past with using biweekly data and now having this tool where it can change frequently from day-to-day?

- ***A:*** *Rebecca Murphy:* Originally when the criteria was written, there was a fixed pycnocline. Then there were some findings that the pycnocline needed to move around and sometimes did not exist within a segment. Now it is allowed to move. When oxygen is interpolated in a two-week period, the pycnocline is also identified. The oxygen cells are split up into designated uses, and the number of violations is counted. Now that we have been looking at the high-frequency data, we see that the density gradient moves not only daily, but every 15 minutes. I think we will be able to do some calculations to quantify the impact of this.

There are also some discussions to be had to identify a weekly mean in a cell when that cell might be moving between designated uses within that week. That is something we will need to talk about, probably with CAP, to assess these high frequency criteria when the pycnocline is moving around. There are a lot of questions around that.

Q: *Jeremy Testa:* If you were in the main stem of the Bay in a place that has deep channel habitat and there is a storm, so it is well mixed, would that be considered open water?

- ***A:*** *Rebecca Murphy:* If it's a storm and it's totally mixed in the data, then yes, the bottom would be open water.

Comment: *Leah Ettema:* I am having a hard time keeping track of what is variable and changes with new data input and what is a one-time method development. Something that could outline those pieces, whether and how they would change, and describe them in detail would be very helpful to have. This way it could be fully transparent and reproducible.

- ***Response:*** *Rebecca Murphy:* We will definitely prioritize making that clear. I can see the disconnect between using old data to parameterize but using the model to interpolate the most recent data. We need to lay that out and make it clear. The data isn't collected everywhere every year, so if you were only looking at one year, it would be spotty. The annual data that's collected has nice spatial coverage in every segment and would be used along with the ConMon and DataFlow data from that year to give the base interpolation. Then some of the parameters, especially for correlation pieces of the tool, are filled with older data.

Q: *Amanda Shaver:* We know from the 3-D interpolator that the data regions will go outside the segment boundary. Will those regions stay the same or is that also going to be flexible based on which data or segments are being used?

- ***A:*** *Rebecca Murphy:* When the current interpolation is done, you're focusing on one segment at a time. That segment has a little boundary around it that grabs additional data outside the segment to aid in informing interpolation. We looked at those boundaries and they weren't created completely systematically. We decided to start fresh. For any segment, it is going to at least include all of the segments that touch it to inform the oxygen estimates there. We have a large series of maps that show this that we can provide.

- ***Q:*** *Amanda Shaver:* Would it only be the segments with the same designated uses? Are there any adverse effects to assess one designated use over another?
- ***A:*** *Rebecca Murphy:* We did not restrict it to being only segments with the same designated uses because the water there is connected even if the designated use cuts off. However, we don't use all of the Bay together because that would be hard to interpolate accurately with the high amount of spatial variability. It seems to be ok to use adjacent boundaries for smaller regions. We don't want sharp boundaries between segments and data sets.
- ***Response:*** *Tish Robertson:* To provide some context for Amanda's comment, in the 3-D interpolator, we noticed that the data region for Tangier Sound extends into CD5 and CD6 and when we don't have any data for Tangier Sound, the interpolator uses data from the deep channel and deep water habitats to inform Tangier Sound. In a technical sense, there is nothing wrong with that, but if we measured 3mg/L for CB5, that's what is represented for Tangier Sound. Ideally, we'd have a station in that segment, so it wouldn't have to reach all the way to CB5. Really the problem isn't the model but that we don't have the data. That shouldn't be a problem with the 4-D interpolator because we will have the ability to generate estimates even when we don't have a monitoring station, right?
- ***Response:*** *Rebecca Murphy:* I have looked at all of these segment groups that we created for oxygen interpolation and compared them to the fixed station data. Jon built a tool that can look at each of these, and Tangier is a great example of one of the places where the estimates were getting a bit weird. This is when we realized we need additional interaction in the generative additive model (GAM) model to adjust for bottom depth versus distance. This helped tweak the tool to get a good spatial picture. Within our next steps, we are looking at the places that are resulting in poor interpolation results, especially where we don't have any nearby data points. We realized we need to restrict our output, so it doesn't extrapolate where we don't have data but uses the best information we have nearby.
- ***Comment:*** *Amanda Shaver:* From a criteria assessment perspective, I think it'll be important to know where we have data within the segments versus not. Potentially, we could set some rules around confidence in the assessment or consider what information has gone into making that decision.
- ***Response:*** *Rebecca Murphy:* Yes. To me, it seems like if there is no data in a segment, we might not want to use it. Some count of the data used would be great.

1:05 PM: Progress Update: Next Steps

Rebecca Murphy (UMCES)

Overview: Rebecca went over the next steps, displayed at the end of the slideshow.

1:15 PM: Open Discussion

All

Q: *Guido Yactayo:* What were the limitations for not including time in the current interpolation model? Was that the absence of continuous information or a computational limitation? I am curious about the history of the tool. My second question is whether you will be able to extrapolate in time as you add more parameters and more information?

- ***A:*** *Rebecca Murphy:* For your first question, I am not positive, but I think it's because the data has historically been taken by boat about every two weeks or a month. I bet it wasn't a thought to interpolate through time. Also, data analysis has been getting better every year, so it may have been computationally feasible as well.
- ***A:*** *Elgin Perry:* When we were moving from 2-D to 3-D, that was a huge accomplishment, and we didn't start thinking about 4-dimensional interpolation until a few decades later. Temporal resolution was also much sparser than what we have today. We also had limitations in computing power.
- ***A:*** *Rebecca Murphy:* On your other question about extrapolating into the future and making future predictions, that's not something we need to do with this for the criteria assessment. I could see a researcher taking it and figuring out how to do that, which would be possible, but it wouldn't be done for criteria assessment because it wouldn't be bounded by data.

Comment: *Kaylyn Gootman:* The team is always open to feedback. You do not have to wait for the meetings.

- ***Comment:*** *Breck Sullivan:* For this presentation you saw the blue box that highlighted how our users would be using that information. Let us know if that was helpful or what other things might be helpful. If an idea comes up, you can always reach out to us.

Participants appreciated having the time within the meeting to discuss each presentation and absorb the information. Kaylyn also offered to meet individually with people if they had further questions, ideas, or discussion topics.

1:30 PM Adjourn

Attendance: Allison Welch (CRC), Rebecca Murphy (UMCES), Jon Harcum (TetraTech), Peter Tango (USGS), Tish Robertson (VA DEQ), Becky Monahan (MDE), Kaylyn Gootman (EPA), Matt Stover (MDE), Marjy Friedrichs (VIMS), Angie Wei (UMCES), Colin Hawes (VIMS), Breck Sullivan (USGS), Elgin Perry (independent statistician), Jeremy Testa (UMCES), Melinda Culter (MDE), Guido Yactayo (MDE), Carl Friedrichs (VIMS), Joseph Morina (VA DEQ), Jim Hagy (EPA), Andrew Keppel (MD DNR), Leah Ettema (EPA), Isabella Bertani (UMCES), Amanda Shaver (VA DEQ), and Gabriel Duran (CRC).