Bay Oxygen Research Group (BORG) Meeting

February 23rd, 2026
12:00 PM – 1:00 PM
Visit the meeting webpage for meeting materials and additional information.

Purpose: In this meeting, Rebecca Murphy (UMCES) presented on class imbalance in the 4-d interpolator and proposed solutions. This is a continuation of her presentation at the January BORG Meeting. Then, Breck Sullivan (USGS) introduced the FAQ document and asked for feedback.

Minutes

I. Welcome, Introductions & Announcements

Lead: ***Breck Sullivan*** *(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. Class Imbalance with 4D Data Sets and Proposed Solutions

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

In this presentation, Rebecca touches on many of the topics that were discussed in this group over the last few months. The team has been working hard to improve the interpolation and make it as close to the data values as possible. This presentation focused on the things they are doing and can do to improve the model.

Rebecca started with the diagram of the 4-d interpolator that has been used many times to visualize how this tool will work. This presentation focuses primarily on the mean midday space-and-time interpolation. This section uses generalized additive models (GAMs) and smooth functions to get midday interpolation results for everywhere in the Bay using monitoring data. High frequency temporal data tends to be in shallow waters and can far exceed the number of boat collected grab samples which could lead to an over representation of shallow water patterns. Rebecca shares the methods they have explored in order to overcome this issue.

One of these methods is taking an hourly timestep of the data. She explains that taking the average of the hourly instead of subsampling would decrease variability and miss minimums. She shows an example of this with data sampled at Bush River, including the empirical density function (EDF). They are confident that subsampling this data hourly won't cause a loss in minimums. For dataflow data, they recommend subsampling spatially by 500m.

Next, Rebecca goes into the issue of class imbalance between the data used for training and prediction. She shows an example of this balance in York River. Elgin presented on this issue at the January 2026 BORG meeting. Rebecca went over this and shows how this issue impacts the interpolator results. Rebecca performed the analysis in every segment over multiple years and found that this issue doesn't affect every station but affects enough that it needs to be mitigated. The interpolator needs to be able to work with unique situations and that is what the team is aiming to do. One proposed solution would be defining an observation unit as the space and time (i.e., one interpolator grid cell in a day) that gets a weight of 1 for fitting the smooth spline functions. If there is only one data point in an observation unit, like a fixed station, it would get a weight of 1, but if there were ten samples in one unit, like a conmon, each would get a weight of 0.1. This is called "weight by observation units." With this method used to weight the data for GAM fitting, the error in the resulting GAM predictions at fixed stations decreased dramatically, while the error in the predictions at the conmon stations increased slightly.

Rebecca concluded with a short summary of these steps needed for accounting for class imbalance in the input data and why they are beneficial.

Discussion Notes (during presentation):

Q: *Matt Stover:* Most of our conmons are in the shallow water area and it makes sense that we wouldn't want bias towards the shallow water. In Fishing Bay, we had a profiler in the deeper mid channel area. If we did more of that in the future, would our approach change? If we had more conmons outside of the shallow water, maybe it wouldn't be as biased.

- ***A:*** *Rebecca Murphy:* Having more high frequency data away from the shore and representing larger areas of the water would be very helpful. This decision could be reevaluated in the future, but there are a lot of benefits of subsampling hourly that aren't just because of the misbalance in data sites.
- ***Response:*** *Elgin Perry:* When you're dealing with a class imbalance problem, it's feasible to achieve balance by increasing observations from the minority class. That would be ideal. In this interpolator, the finest unit in time is an hour. When we get to the small-scale variation component of our interpolator, we need to be able to estimate what the hour-to-hour autocorrelation is because that is the scale we are trying to predict. 15-minute data doesn't help us with that. To get the hour-to-hour autocorrelation, we'd still need to subsample that to an hourly scale. If we wanted to use 15-minute data, we would need to retool the interpolator to quarter hour increments instead of hourly. When the computing power is available, I'm sure we'll start doing that, but the current computing power constrains us.
- ***Response:*** *Matt Stover:* So, to paraphrase, we aren't doing this to just overcome the spatial bias caused by conmons primarily being in shallow water areas. Computing power is also a limiting factor that makes it reasonable to use an hourly timestep. Rebecca showed that we don't think we're losing anything with that hourly timestep. Is that the rationale?
- ***Response:*** *Elgin Perry:* Yes, that is a good synopsis.
- ***Response:*** *Peter Tango:* This question about spatial representation is exactly why we had the UMCES team working on the project looking at different segments and the distribution of near shore and offshore sites. They are going to see if there are places that we can put monitors to reduce bias. Hopefully that will inform our segments going forward.

Q: *Leah Ettema:* Does GAM use least squares?

- ***A:*** *Rebecca Murphy:* Yes.

Discussion Notes (after presentation):

Comment: *Breck Sullivan:* If there is an area where there isn't an imbalance, this weighting proposal doesn't impact the results. It provides a solution where needed and doesn't affect where there isn't an issue. It also gives a method that can be used if imbalances are found in future data sets.

Q: *Leah Ettema:* On slide 17, we are trying to get the mean midday space and time interpolation at each cell so we will use that data. When we are interpolating for the other cells without data, we are weighting the data that's available to do the interpolation. Is that correct?

- ***A:*** *Rebecca Murphy:* The first step is not to get a mean midday average in each cell and then interpolate. When we fit a line through all of these points, the multiple samples in one unit are getting a lower weight than one lone data point. We're not aggregating in a pre-processing step, but it's within the system. One of our independent variables is time. Each of these are collected at a different time in the day. We'll still get the part of the model that shows how the oxygen differs at different times, even though they would be weighted less for getting the whole day estimate.
- ***Q:*** *Leah Ettema:* There's one spline per assessment unit. Is that correct?
- ***A:*** *Rebecca Murphy:* No. It's fit throughout the group of segments in a year, but it's lots of spline functions because there's one for time of day, day of year, bottom depth, xy location and interactions between them. It's one model fit to the entire year of data in a segment.
- ***Q:*** *Leah Ettema:* In the beginning, when you were talking about mean midday space and time, does the spline include all of those other components as well?
- ***A:*** *Rebecca Murphy:* Yes. The mean midday is at the output. We don't take a mean of each sample to get a daily value first. That's on the output side, which is the middle of the day prediction. We will be trying to add things along these lines to the frequently asked questions document.

Q: *Carl Friedrichs:* This is great. I'm very pleased with how you found a way to include the conmon data and weigh it appropriately. I was confused about the spline fitting too. Thank you for asking. There is hourly data at the conmon everyday in the spline that is weighted by $1/24^{th}$. In the deep channel, there is only an observation every two weeks. Does that cause an issue in the spline fit?

- ***A:*** *Rebecca Murphy:* With this approach, we will still have data everyday from the conmon while the fixed station is 14-16 times a year. That's the level I think the GAM can handle. When we put our dissolved oxygen data in, it has variables like spatial location, depth, bottom depth, date, and time. We fit a model to that and use that to predict everywhere. There is enough flexibility in the GAM to handle the variance in the frequency of the data. It is possible it wouldn't have worked, but empirically we can see that predictions are doing OK at this level of weighting.

Q: *Carl Friedrichs:* Does it compensate? At the location of long-term monitoring in the channel, you're using many depths in the vertical. Does that add more information which helps compensate for the fact that there is only one position in the vertical at the conmon?

- ***A:*** *Rebecca Murphy:* Maybe it helps in terms of the overall horizontal picture, but that depth has even less data, because there isn't high frequency data there. That's a place that could use more high frequency data. It probably helps a little bit with how different open water is compared to shallow water, but that's another challenge.

Q: *Carl Friedrichs:* Is there a vertical spline?

- ***A:*** *Rebecca Murphy:* Definitely. When it's a standard stratified profile of oxygen, that is what our spline would look like.

- ***Response:*** *Carl Friedrichs:* Because of that vertical spline, the data lower in the central channel observation is constraining the fit at the top. Maybe it's adding more power to the influence of the data in the central channel, even though there's deep channel. I'll have to listen to more presentations. I'm out of my understanding. I think it's hard to come up with a theoretical reason that's perfectly predicted to why you should subsample at different places in exact intervals. I think the one-hour sampling approach sounds like a really reasonable approach. I'm very comfortable with the path you're taking. Kudos to how well this is working. I suggest everyone listen to Rebecca's presentations as often as possible.
- ***Response:*** *Elgin Perry:* I think Carl's intuition is correct. When you're sampling in the mid-channel every two weeks, you could make a case that you should be weighing that data by 14 at the same time we're weighing high frequency data by $1/24^{th}$. When we plugged in the weights of $1/24^{th}$, it seemed to fix the problem. We decided to quit messing with it to prevent introducing more complexity than was needed to get reasonable estimates. You're also correct that the depth profile of the fixed station adds more weight without introducing a weight to the fixed station data. That's helped by the fact that the correlation over depth tends to be pretty strong. When I looked at a predicted depth profile compared to an observed depth profile, it's amazing how often the predicted profile is shifted ±1 of the observed data. Those factors all help to compensate for the paucity of data we have at the fixed stations.

Q: *Becky Monahan:* It seems like we're now weighting the fixed station higher than the conmon. I understand it's because the fixed station has multiple depths as we've just talked about, but the conmon is a more accurate predictor of dissolved oxygen. Should we weight that a little higher? I agree we want to weight it because there are 24 times more data points, but maybe it should be weighted higher than the fixed station.

- ***A:*** *Rebecca Murphy:* The conmon gets more weight because of the frequency of sampling. Let's say that in our observation units, we are considering it as one a day. We would still have every day for conmon, while we might only have once every two weeks for fixed station. I don't think we need to be concerned about the conmon having too little of an influence because it still has a lot of influence. It's the only thing that influences daily cycles. Whether that's from the GAM part of it to help get the best prediction we can at 11am or it's from the further parts of the model where we do the daily cycles and the variance every day. That's all coming from high frequency data. It gives a lot of information to the interpolation. We wouldn't be able to do any of this without the high frequency data.
- ***Response:*** *Becky Monahan:* I'm thinking that this graph (slide 17) will be the way the data will be represented at the end output, which you explained is not what it's going to be. This is the way it's organized for the spline units, which I'm understanding. I think I need to dive into that a little more.
- ***Response:*** *Rebecca Murphy:* We are embarking on documentation this year and hopefully we can make sure things are being explained well. Your feedback would be very valuable for that.

Q (from chat): *Amanda Shaver:* Would a similar approach be taken with the array or mid-channel sampling absent of any shallow water conmon? Maybe there is discrete volunteer monitoring in a tributary.

- ***A:*** *Rebecca Murphy:* The vertical arrays from Virginia, Maryland Department of Natural Resources (DNR), and National Oceanic and Atmospheric Association (NOAA) would have multiple depths at high frequencies. That data is great and we really need it. It would be the same thing where you can imagine 24 dots in a deeper cell. It would be weighted 1/24 to fit

the function because we still want to make sure it's not overwhelming where it exists. Any high frequency data would use the same method.

- ***Response:*** *Amanda Shaver:* I was thinking that we could have the scenario where surface water sampling is happening in a segment, either ambient or volunteer monitoring. Those shallow water discrete samples would be weighted as one in the shallows, which would be the opposite of what you're showing.
- ***Response:*** *Rebecca Murphy:* Yeah. I could see a world where there were multiple samples at depth and there is only one in the shallows, then it would be the opposite. We want to make sure shallow water conditions are interpolated correctly too.

Comment: *Breck Sullivan:* We can give everyone time to digest this information and come back with any additional questions. We are diving into documentation. Today's questions were helpful for us to know how we can explain this better. This approach allows us to use the data that's available and also be able to weight it appropriately to get the correct interpolation results.

Q (from chat): *Matt Stover:* So, would the weighting be done on a segment-specific basis?

- ***Q:*** *Matt Stover:* I thought it was an interesting point that was brought up where you have the opposite temporal data density. Would it be done on a segment specific basis – on the basis of how and where different frequencies of monitoring were completed in that segment?
- ***A:*** *Rebecca Murphy:* No, what I described is a universal rule for the 4-d interpolator. I don't see it that would be different with a different data distribution. Our idea is that the observational unit is one station, one depth, and one day. That's not different for a vertical array versus a conmon versus a fixed station. We have a big data set and if you want to interpolate Fishing Bay, you'd select that segment and the segments it's grouped with would be included – the boundary segments and any it's grouped with. The data is pulled for those segments from the larger database and then the processing is done where the observational units are identified. Every sample in one observational unit is grouped together and the weights are applied. Then the GAM function is fit and the predictions are made. It would be done individually for a segment or group of segments, but the definition of the units will stay the same. We don't want to be changing the rules for each segment. It should be universal.

Q: *Matt Stover:* If you had the opposite situation with a conmon in the mid-channel and fixed station on the nearshore, then the function you'd use would recognize where the greater data density is and weight it appropriately. Correct?

- ***A:*** *Rebecca Murphy:* Yes, exactly. You could imagine a "groupby" function, where it groups by depth, station, and day. It doesn't matter if that group contains 1 or 24 observations. Then, we'd weight by however many are in the group to get to the unit.
- ***Comment:*** *Matt Stover:* That makes sense because you would still have the same concerns of bias for that segment if you had a ton of samples in the main channel and not a lot in the shallow water areas. Even though, in that case, you would be representing a greater proportion of the volume of that segment with the continuous monitoring data.
- ***Response:*** *Rebecca Murphy:* I like this example and think I should make a visual for it because it does happen. One example is our east goose NOAA vertical array in 2022. If for some reason a station was abnormal, this would be an important step.
- ***Response:*** *Matt Stover:* In your first few slides, you showed the proportion of volume represented by a certain station. In this case, it would be flipped because most of the volume is in that deep portion of a segment. If we had a conmon in that area, maybe we wouldn't want to de-weight it as much. I'm not sure. It would be interesting to see what the numbers look like for Fishing Bay or one of these types of segments because it may yield a different result.

- ***Response:*** *Rebecca Murphy:* This reminds me of a challenge we had before we had conmon. Our fixed station network was all collected by boat, so it is more representative of deep waters. Before the early 2000s, when we only had fixed station data, there's no way the interpolations were representing the shallow water dissolved oxygen condition appropriately. We want to get that area right too. This example is based on the fact that the conmons are more frequently in shallow waters.

Q: *Carl Friedrichs:* A lot of the conmons have been in for three years and not more. If this were expanded to retrospective splines that fit multiple years, the influence of conmons may be a small period of that location in record. Right now, we're working at one year at a time. Is the full simulation of a spline limited to one year?

- ***A:*** *Rebecca Murphy:* Yes. When we started this a couple of years ago, our initial idea was to combine years – maybe 3 or 10. We settled on one as we tested through that because it helps get the interpolation for that year better. Expanding beyond a year didn't work as well. We do one year at a time. The structure will be the same with the observation units and variables, but we do it separately by year.
- ***Response:*** *Carl Friedrichs:* I think that makes a lot of sense for now. The long-term models tend to be relatively independent on each year. Wet years and dry years can behave differently. That makes a lot of sense. If you were forcing one spline to fit all that, you might get some communication between years that you don't want. When you do a long-term fit for nitrogen or Secchi at an individual station, you do go through all the years. If it were all the time, you would run into issues that some places had data intensely for a year and not for the rest of the record.
- ***Comment (from chat):*** *Peter Tango:* The first 6 phases of the Bay model underpinning the TMDL (total maximum daily load) did not have shallow water either. New frontiers here in Phase 7.
- ***Response:*** *Rebecca Murphy:* You're right. It was our trends work that made us try to fit multiple years at a time initially. Those GAMs [for trends] are fit year-to-year. We are currently set up for the other parts of the interpolator tool to use multiple years of data to inform it. We've talked about that with this group before and we will continue to try to make that clear. There is such rich information in the high frequency data, but it's not always collected everywhere. We use the entire high frequency data set to inform what the daily cycles look like.

III. 4-D Interpolator Frequently Asked Questions (FAQ) Document

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

Breck introduced the draft FAQ document and asked for feedback from the group. Participants are also welcome to send feedback or questions they'd like added to Breck over email (bsullivan@chesapeakebay.net).

Discussion Notes:

Comment (from chat): *Matt Stover:* Thanks so much to the folks who worked on the FAQ!! I've already found it to be very helpful!!

Comment: *Leah Ettema:* I love the idea of this FAQ. If I understand right, when you show the flowchart, the green is an output of the statistics. I'm not sure how they all fit together. Maybe there should be a flowchart that describes in more detail how you get from the pink chart to the green part.

- ***Response:*** *Rebecca Murphy:* That's a good point. Each of these is the output. We need to talk about the methods used to get to this, which we will include in the documentation.
- ***Response:*** *Breck Sullivan:* That's helpful to explain that these are outputs of the components. That would be a great idea to go from input data to the statistical part.

Comment: *Melinda Cutler:* I want to echo that and this FAQ is great. A flowchart of the process would be really helpful. It could highlight the specific types of input data and which part of the model aspects each of them goes into.

- ***Response:*** *Rebecca Murphy:* We did have an earlier flowchart that we will probably have to edit. It might be a starting point for that idea. We changed to this one to make it clear that all of the data was being used. That other one would be more along the lines of the process.
- ***Response:*** *Breck Sullivan:* I want to give a shoutout to Elgin because he's been commenting that we should make a flowchart of the process. That's something we can put some time into.

IV. Adjourn

Next Meeting: ***March 16, 2026***

Attendees:

- Amanda Shaver, VA DEQ
- Andrew Keppel, MD DNR
- Angie Wei, UMCES
- Becky Monahan, MDE
- Breck Sullivan, USGS
- Carl Friedrichs, VIMS
- Cindy Johnson, VA DEQ
- Efeturi Oghenekaro, DOEE
- Elgin Perry
- Gabriel Duran, CRC
- Jay Lazar, NOAA
- Jon Harcum, TetraTech
- Joseph Morina, VA DEQ
- Leah Ettema, EPA
- Marjy Friedrichs, VIMS
- Mark Trice, MD DNR
- Matthew Stover, MDE
- Melinda Cutler, MDE
- Peter Tango, USGS
- Rebecca Murphy, UMCES
- Sophia Grossweiler, MDE
- Tish Robertson, VA DEQ