Bay Oxygen Research Group (BORG)

Monday, February 10, 2025
12:00 PM – 1:30 PM

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

MINUTES

12:00 PM – 12:10PM: Introductions/announcements

Peter Tango (USGS)

Overview: Peter gave an introduction to today's agenda. Peter mentioned the continued discussion on criteria assessment protocols and options surrounding how to apply and interpret the information that will come out of the interpolator. He also made the group aware of the Beyond 2025 outcome assessment, specifically on the Water Quality Standards Attainment and Monitoring (WQSAM) outcome. There are public meetings on the outcome assessment that can be found on the Chesapeake Bay Program's calendar page. Peter also let the group know about the preliminary work on non-tidal load and trends coming out this week. Lastly, he shared that the Hypoxia Collaborative Leadership Team is thinking about how their data can be used for interpolator development, calibration, and verification. The group expects to be monitoring again in April.

12:10 PM – 12:30PM: 4-Dimensional Interpolator Development Updates

Rebecca Murphy (UMCES)

Overview: Rebecca's presentation served as an introduction and update to the 4-dimensional interpolator tool and development process. As a reminder, the BORG team is developing a spatial and temporal interpolation method to be used in the criteria assessment process. This roots back to the "Ambient Water Quality Criteria" document, which recommends the use of cumulative frequency distributions (CFDs) for the water quality criteria assessment process. This method can help reduce some of the limitations found in other criteria assessment processes.

The current approach for criteria assessments is to collect monitoring data collected throughout the tidal waterways and aggregate that to a bi-weekly basis, interpolate spatially for a single cruise, aggerate those results over a three-year period and compare to criterion value, and use to build a cumulative frequency distribution. Those results compared to reference curves are used to quantify attainment or not for a certain segment over a three-year period. This is a complicated process with a lot of parts that can be looked at, especially if

evaluating higher frequency data is considered. This meeting is focusing on the interpolation approach.

The current interpolation approach has a few weaknesses. It cannot evaluate high frequency data, like the 15-minute values. It isn't truly a 3-dimensional interpolation. The horizontal layers are stacked and put together to make a 3-dimensional picture. One cruise, a two-week period, is evaluated at one time and made static. Lastly, the process is not statistical. The new interpolator will use and interpolate all of the high frequency data to create hourly oxygen data points. All the data will be interpolated together, not in layers. Interpolations will also be done in time, so they don't have to be artificially split. Time is the fourth dimension of this interpolator, adding onto the spatial interpolation. This process will be statistical, so uncertainty bounds can be worked into the process if deemed useful.

Comment: *Matthew Stover:* Back on slide two, in integrated reporting we've used these other assessment methods pretty successfully, so I'd hate to undersell them. I know there are some limitations and bias with where you choose monitoring locations. I think a lot of those biases are true whether you use a 3-d interpolator or a 4-d interpolator, because they all depend on where the data is collected – they all should be based on the monitoring data that is on the ground. I think that isn't just a limitation of station-by-station monitoring; it's a limitation of all monitoring. Any interpolation method is just that. It's creating synthesized data where we don't have any. On slide 5, one concern both Virginia and Maryland have where it kind of touts the benefit of the new interpolation is using all high frequency data. One of the assumptions that goes into is in using the ConMon data. We're basically taking the ConMon data, that has measured data every 15 minutes and we're instead rarifying it down to one measurement per day to interpolate with those mid-channel stations. So, it's not really using all frequency data. The only way it is is using those diel cycles to help judge what's happening at that mid-channel station that we only have once or twice monthly. I don't want to oversell this new method because I think we still run into some of the same challenges. I also don't want to undersell some of the alternative methods we appreciate that the Bay Program is considering. I know one of the things that Gary and others have offered to help us with is reviewing our Fishing Bay data and running some of the ConMon data from that through the current 3-d interpolator. This should be telling because it has taken something we are familiar with and trying to force fit it a bit. It should be interesting to see what that says versus our station-by-station assessment which we're doing. Ultimately, we'd hopefully be able to look at that through the 4-d interpolator. I just don't want us to gloss over some of these other methods, which I think still have validity and could be helpful to consider.

- ***Response:*** *Rebecca Murphy:* I agree that no interpolation method is going to be accurate to the true conditions if there is no data in a certain area that has some different conditions. It's all very dependent on the monitoring. That's what we're doing; we're interpolating monitoring data. It's super important to plan data collection locations in a smart way. I hope it becomes clear as we go through this, and as Jon goes through his results, that we truly are interpolating the high frequency data. I think some of our earlier presentations made it seem like we were just doing a daily interpolation, but that's definitely not the point. We will be using the 15-minute data and interpolating between the values. We don't have that data everywhere, but we are using it to its full extent.

Comment: *Tish Robertson:* I wanted to lend support to what Matt was saying, but I also want to mention the statistical bullet there. I think the benefit of the new interpolation technique along statistical lines will be that we are not using the same algorithm everywhere. With the 3-d interpolator, we're using inverse distance weighting (IDW) which assumes we have the same relationship throughout the whole Bay. We know that it's spatially variable as well as temporally variable. Going forward, I think when you are presenting the benefits, you can mention the weakness of the 3-d interpolator is that it presumes that everything is the same. There is a static relationship in the predictability of dissolved oxygen. With a statistical technique, you can throw it a lot of different variables to understand what is going on in a particular region, without assuming that dissolved oxygen can be predicted the same in CB4MH as it can in CB8PH. That's something that has always gotten under my skin.

- ***Response:*** *Rebecca Murphy:* Thank you. That's a great comment. I will add that to my talking points. Great point.
- ***Response:*** *Tish Robertson:* I know that uncertainty can be a scary thing, when it comes to resource managers. I think you are right to point it out, but it may not be seen as a benefit to everyone.
- ***Response:*** *Rebecca Murphy:* Thanks. That also helps explain and respond to some of your comments, Matt. What we're doing is building a relationship between the observations and spatial and temporal features, depth and distance from the deep channel – things like that. Looking at the relationship between the dissolved oxygen (DO) and those fixed features that we know at every point in the Bay, how far it is from the mid-channel and how deep it is. To build a smart model of interpolating the oxygen, not just this inverse distance weighting relationship. It will be validated with the data to see how well it does. If it doesn't do well, we'll change that.

Comment: *Matt Stover:* Thank you, Rebecca. There is probably much more to learn on the 4-d interpolator. I know one of the things we have talked about having is a 1on1 with the Bay Program to help MD better understand what's going on behind the curtain on this project. I think that will be helpful to give us a higher comfort level. I also want to make the point that I don't want this to be the only method that we consider for assessment. I hope the Bay Program is willing to help us explore other methods of assessment, including those that Virginia and Maryland are working on with weight of evidence approach. Those hold some validity and are a lot simpler. One of the questions we have is "how long is this going to take to run and who's going to do this in perpetuity?" There are a variety of logistical questions there and sometimes simple can be good. I don't want us to automatically discount simple methods that have worked for us. Not to say we are closed off to the 4-d interpolator, that's not accurate. I also don't want to talk down these other methods that could be useful, and I want to keep them on the table.

- ***Comment:*** *Gary Shenk:* I think there are some takeaways from the conversations I have had with those in the Chesapeake Bay Program office. Number 1, which is more so coming from me, is the way I think about it, backing way out, and the reason why I love the deep-water bio-reference curve that we have now, is that all we are trying to do here is take some measure of dissolved oxygen that is indicative of that segment, and we're trying to correlate with whether or not it's a healthy habitat or ecosystem. No matter how we came up with that signature, which happens to be the 3-d interpolator with a CFD. Jeni Keisman found this really clear delineation, once you massage data in that way, between areas that were fair or better and not healthy. Whatever we end up doing, that's what we're trying to do here. We're just trying to have "here is a signature that indicates a healthy ecosystem." The second point is that whatever we do we want to do together. We have this whole series of publications. They're mainly the CFD but also in that 2017 document there are those different ways when you have an individual station. Matt, I think that's what we're talking about doing with your Fishing Bay data; following those partnership published methods. That's what we intend on doing, doing it together, and finding a way to say, "is this area healthy?"
- ***Comment from chat:*** *Peter Tango:* Matt, The earlier work was proof of concept at higher temporal resolution than we have ever had in Elgin's pilot efforts while not needing extensive supercomputing powers for working out the approach at the full temporal data density we have - also, we have never interpolated 7 day mean and 1 day mean even so that is a massive leap in improvement to match

the environmental conditions over a static 2 week representation. To Rebecca's point, I think Jon will show the next level of intensity in the interpolation.

- ***Comment:*** *Richard Tian:* Following up on all the comments here. At the end of the day, we have to figure out how to do criteria assessment using the 4-d interpolator. I'm thinking at a later point, as Matt said, MDE did a lot of data collection in Fishing Bay. I wonder whether we can use Fishing Bay as an example and start from there with how to apply the 4-d interpolator and complete the criteria assessment. We can also think about how we can do it in the 3-d interpolator tool, but I think there's also a more stable method approach compared to how it would work. I also think there is always the issue of how the ConMon data can represent the whole segment. There is some thought in the addendum from 2017 where they are trying to handle that. They have divided it into three zones, zone 1, zone 2, and zone 3. My understanding is that the ConMon station shouldn't be used for zone 2, but there is the concept for condition containment, which is that if your zone 2 is attained, your zone 3 is attained as well. What I am trying to say to that is there is some kind of consideration and you're trying to better understand the data issues.
- ***Comment:*** *Matt Stover:* I just want to say to Gary's point; we want to move together on these things. What we would really like is whether it's the BORG, the Criteria Assessment Protocol (CAP) Workgroup, or any of the other workgroups we have, is to not close our minds to other methods in addition to the 4-d interpolator. I think there are simpler methods that could get to the same answer. Maybe I'm wrong on that, but I would hate to close our minds to other options we think are viable, and I think Virginia does too. We're not closing our minds to the 4-d interpolator because I do see a lot of benefits from it. We don't fully understand it or feel comfortable with it or explaining it. We're curious about how long it will take to meet our deadlines when it comes to the integrated report every two years. I love Elgin but I don't think he's going to work forever, so there are turn-over considerations there. It's probably going to be something we'll never run internally. Not that we've run the 3-d interpolator in Maryland but we could. That's where a lot of our concerns come from and is the point I wanted to bring across to the group.
- ***Response:*** *Rebecca Murphy:* Thanks. We're definitely taking all of your considerations for sure. I hope you'll be able to run it. Hopefully we can make things clearer as we go through this, and we hope to have some time for questions at the end too.
- ***Comment from chat:*** *Peter Tango:* Also - Ad hoc sample placement is non-statistical typically. Sample design is key to applying appropriate assessments

based on sampling theory. There is a handshake between how sample sites were located in any study and the appropriate statistics to support interpretation. We continue to strive for that appropriate handshake of statistics theory and methods of assessment underpinned by sampling design, or, lack of design considerations. We continue to strive toward robust statistical support behind providing the assessments resting on a firm foundation of sampling, design and evaluation methods.

- ***Response from chat:*** *Matt Stover:* I agree Peter. Thanks for those thoughts. I just ask that we all keep an open mind to other methods and make the good faith effort to explore them as a partnership.
- ***Response from chat:*** *Peter Tango:* Thank you Matt - those are excellent guiding concerns for how the method sees its development and user-friendliness incorporated, just like the Generalized Additive Models (GAMs) trend assessment software that shifted us from nonparametric trend assessment of the early Chesapeake Bay Program (CBP) to modern CBP, and the Chesapeake Assessment Scenario Tool (CAST) software States use for Watershed Implementation Planning as two major developments used by the States. Please keep the guiding concerns and thoughts shared, keys to our application development considerations 🧐
- ***Response from chat:*** *Matt Stover:* Thanks Peter. I think the difference with those tools, i.e., GAMS trend software, CAST, is that they, to my knowledge, don't allow State staff to understand all of the underlying code and assumptions that are baked into them. When it comes to running assessments for the 303d list, we have pretty much always had access to those details and been able to run them and adjust those assumptions and alter our assessment methodologies if we so desire. My fear with the 4D Interpolator is that realistically we will never have that level of familiarity to be able to properly vet the assumptions incorporated and change them if we find them to be inaccurate.
- ***Response from chat:*** *Matt Stover:* Peter, ultimately what we are asking is for the Bay Program's analytical and research support to explore these other non-4D Interpolator methods of assessment. And thanks to Gary for offering and helping with that!
- ***Response from chat:*** *Rebecca Murphy:* Matt – that makes sense as a need. We worked closely with the Department of Natural Resouces (DNR) and the Department of Environmental Quality (DEQ)/Old Dominion University (ODU) staff for the tidal trends method development and those trends are currently run by partners at DNR, ODU/DEQ, and the Department of Energy and the Environment (DOEE)/ Metropolitan Washington Council of Governments (COG).

Those folks frequently come to us with suggestions for method changes, which we all work on together. For this tool, we'll hopefully get more into hands-on work for partners who want to do it very soon. Certainly, I agree every part of the method has to be well documented, so everyone understands it.

- ***Response from chat:*** *Peter Tango:* Oh definitely understood Matt - thank you
- ***Comment from chat:*** *Peter Tango:* To your point Matt - if you go way back in the early tech documentation, Elgin and others recommended multiple assessment methods, e.g., there are suggestions for logistic regression as I recall. Either in the back of the 2003 doc or a 2004 tech doc. Further, Steve Jordan had a publication around 1996 that judged condition based on something similar to the CFD. And Walter Boynton and others have similar style evaluation approaches. Personal feeling that we have a variety of indicators (hypoxic volume, hypoxic volume days, days above or below criteria, etc.) that we could discuss for supporting information. Insightful and informative info versus what you have in your regs seems like the great bridge and has been something that evolves. So, the future always seems an opportunity to update methods considered and accepted for standards assessment as they get vetted and in your work with EPA for accepting proposed methods for assessment.
- ***Response from chat:*** *Matt Stover:* Thanks Peter. Agreed, our methods should always be open for critique and improvement. I think, as part of the CAP workgroup's next set of meetings, it may behoove us to talk about the big picture of DO assessment and to talk about the different methods that the partnership should test and consider, in addition to the 4D interpolator. VA and MD will be meeting in the coming weeks to talk about what that agenda might look like and will be sure to share with you when we have more concrete suggestions for that.
- ***Response from chat:*** *Peter Tango:* Excellent Matt. Thanks - we will coordinate for CAP, that has been especially helpful.

Continuation of presentation*:* Rebecca continued her presentation on the 4-d interpolation tool and development process. There are 4 components to developing the statistical interpolation which will be used together in the final product. In this meeting, Jon will be presenting how the team has worked through these components and brought them together. The team took the tool they have now and ran the data of one segment for one year to test the tool and answer some questions. The interpolator runs, which is a huge success. During this test run, the team tested the components by adjusting them and making sure the expected results are achieved. The team also compared the data found to other data points to ensure it lines up with the draft results of the tool. The team is still developing the tool in 2025, focusing on documentation, and assessing how the tool will be linked

to criteria assessment. In 2026, they will focus on training those who will be using the tool to ensure everyone can access and run it.

Next steps:

- ✓ Work with the CAP group to brainstorm ideas on how results will be used in criteria assessment.
- ✓ Method development. Continue necessary parametrization of all 4 parts of the 4-D tool.
- ✓ Software development. Ensure use of all data, compile data every year, determine file management and storage, and then work with tool users. (continuing into 2026).

12:30 PM – 1:10PM: Preliminary Interpolation and Simulation Approach Applied to CB4MH 2022

Jon Harcum (TetraTech)

Overview: There are four components of the interpolation tool, a mid-day space and time component, a large-scale correlation component, a daily cyclic component, and a small-scale uncertainty component. This presentation will go through each of these components. The data used for this project was Chesapeake Bay data, Eyes on the Bay data, VECOS ConMon data, Profiler data, vertical array data from NOAA, and pilot dataflow data from Wes Slaughter. The group intends on using Tier 3 citizen science data but hasn't incorporated that yet. By the end of the presentation, there should be an understanding of how that data was turned into the graph on the second slide where the black line represents the west-gooses station in comparison to realizations at a nearby grid point at depths 1, 3, 5, 7, and 9 meters. West-gooses station only went down to 7 meters.

Mid-day space and time interpolation: The data for this process was from the Chesapeake Bay DataHub and the ConMon data subsampled to hourly values to mimic what the DataHub data looked like. Part of the reasoning for that was to preserve some of the statistical properties that Tish was mentioning. CB4MH is the dark blue segment in the middle of the map. While creating this, the team included the adjacent segments, represented in tan, because the surrounding segments can inform the findings inside the segments. Rebecca has segmented the entire Bay including the adjacent boundary segments for each segment. The terms included are the hour of the day, the day of the year, the sampling depth, the depth of the station, and a water body longitude and latitude. These last terms give an indication of how far north and south you are in the Bay in the primary direction of flow and how far east and west you are in the transverse direction. The 2-way interaction

terms and 3-way interaction terms help improve the overall performance of the model. The graphs on the upper right show the observations at CB4.3W at depths 1,3, 5, 7, and 9. The graphs on the lower right show the vertical array data at west-gooses subsampled to daily values that mimic the Chesapeake Bay fixed station data.

Once that GAM is obtained, the next step would be to say, “how do we represent that when we want to interpolate across the rest of that segment?” The team uses the water quality interpolator grid. This grid is created in a rectangular area but all points on land are flagged as not water but used for some statistical properties. This is expanded into a cuboid in depth that will also have pink, non-water points. This brings in the 3-dimensional quality. The fourth dimension is looking at every day of the year. Once the four dimensions are represented, the GAM can be applied and see how it predicts the nearby grid points. The west-gooses station has two nearby grid points, one to the north and one to the south. The observed data is represented by the red circles, which is CB4.3W, and the green triangles, which is west-gooses. As a reminder, this does not concern the daily cyclic component yet. That will come later. This is only using the day of the year. Based on the graphed values, the team feels that the prediction values align well with the observed values, both in the day of the year and throughout the depths. This gives the mean prediction which is analogous to the Bay Trends tool.

The next step is when the mid-day space and time GAM was created, it creates a set of coefficients, and a vector of a variance covariance matrix associated with those coefficients. They are random variables that can be used to look at a multivariate random distribution of those coefficients and inform how confident we are in the mid-day values.

The graphs on the next slide (slide 5) are a representation of 10 mid-day realizations at the same two grid points. The observed data is now represented by the pink color. Because these grid points are close to the observed data location, the confidence in the model is high. As you get further away from the observed data points, there is more uncertainty.

Large-scale correlation: The large-scale correlation is looking at the correlation of day-to-day values, spatially, and through depth. If the dissolved oxygen is high today, it is likely to be high tomorrow. If the dissolved oxygen is high in one location, it is likely to be high 100 meters from there, or maybe it’s low there depending on the situation. As Rebecca mentioned earlier, different “knobs” within the system can be edited to calibrate the correlations. The high frequency ConMon data is used to inform the day-to-day correlation. The DATAFLOW data is used to inform the spatial aspects of correlation. The vertical array data has helped to inform the correlation of depth. A future step will be to include the profiler information from our fixed station

network to help expand that. The blue box on slide 6 shows a water quality interpolator grid at 1 kilometer. There are three points highlighted here. The red, the blue, which is a kilometer south of the red, and the green, which is a kilometer east of the red. By editing the "knobs," as in day-to-day, spatial, and depth correlation, the team can show how that impacts the interpolator and compare the results to the observed values. When the correlations are very high, like in this graph, all three of these graphs are going to be very similar. In this graph, all of the correlation values have been set to 0.99.

On the next slide (slide 7), the graph shows what happens when the north and south correlation is low. The first and second columns still track along the mean (the grey line), but they vary from each other more than in the previous examples. [Jon skipped over some of the slides on adjusted correlation values to save time for questions.]

On slide 10, Jon made the spatial correlations high, but the correlation through time is low. The red lines have more jitters than the previous examples. Due to the high spatial correlations, each of the lines look similar. This comes into play when the team runs numerous realizations to get a cloud of data.

On slide 11, Jon compares the results of one realization at high correlation values to multiple realizations, still with high correlations. This shows the mid-day GAM with correlated uncertainty in space and at the daily level.

Day-to-day correlation: The day-to-day correlation is largely driven by the ConMon data, vertical array data, and a couple of vertical array stations that are in the VIMS VECOS data set. The purpose of this is to understand how the dissolved oxygen correlates from one day at 11am to the next day at 11am. The high frequency data was subsampled to get the observation closest to 11am. This is shown in graph A with the grey lines in the back representing the raw data and the grey circles representing the values that were subsampled to represent 11am. A GAM is fit to remove the seasonality in the data. This is represented by the red line in graph B. Then, small slices of the data were taken, as seen by the blue band on graph B, with that data fit onto graph C. What they are interested in calculating is the residuals between the symbols (the raw data that was subsampled to 11am) and the fitted model (the GAM model). Then, they calculate a lag one autocorrelation coefficient. In this case the correlation was relatively high at 0.66. The process would be completed across the whole data set and all of the data sets used. A lot of data on the surface, with a few with depth. These correlations would help inform the lag one autocorrelation values used on the previous slides.

Spatial correlation: This is some of the work done by Wes, the graduate student at University of Maryland, with oversight from Elgin. The DATAFLOW data from the

Potomac River was processed in a manner similar to fitting variograms. The team wanted to remove the data from when the boat was stopped. Graph A shows boat velocity over time. The team removed the data in the black, when the boat was stopped, and treated each color in the graph as data chunks that were taken into the spatial analysis. In graph B, each data chunk is shown spatially on the river. Graph C shows the dissolved oxygen levels over the cruise. Each colored data chunk was fitted into a variogram, as seen in graph D. Those were then averaged to get the black variogram line on graph D. That can be used to calculate the correlation spatially. For example, the correlation at 1000km comes out at about 0.5 for this dataset on the Potomac. However, each data set has varying amounts of data. Elgin is working on a process that would get an individual variogram out of the total cloud of data, which would be more representative of the dataset.

Depth correlation: This is working with the East Gooses vertical array. They calculated the cross correlation between depths 1 and 3, depth 3 and 5, depth 5 and 7, etc. They created the cross-correlation matrix shown on slide 14. When looking at the yellow highlighted values, the team could calculate the lag between the 1m depths. Then, they square rooted that to get the 1m correlation, which is 0.84. That information can be used to turn the knobs and help inform the large-scale correlation.

On slide 15, the results calculated for the West Gooses station are used to create 10 realizations, as shown by the green lines on the graphs. The purple points are the observed data points. The team feels good about the accuracy of these preliminary results.

Within day (cyclic): The ConMon, high frequency, vertical array data is used to pull out hourly data, as seen in graph A. This data is taken out by a day and fit to a daily cyclic model, as seen in graph B. Sine and cosine functions are used to fit the red line to the observed data. The coefficients of the red line are saved for each repetition of similar data. That is used to calculate the within day cyclic patterns. The uncertainty in this red line compared to the observed data gives the small-scale correlation component, seen in graph C. This correlation value calculated was pretty small and not statistically significant. Then, the daily cyclic component is overlayed, as seen on slide 17. If you looked at the daily cyclic realization for a grid point and graphed it over the year, you'd get graph A. The sinusoidal pattern gets stronger in the summer but is weaker in the winter. In graph B, you can see that the pattern changes throughout depths, getting weaker as it gets deeper. This is just one realization.

Small-scale correlation: Once looking at the small-scale correlation, the magnitude of the scale of the graph is very small. Also, the graph looks like white noise. This means that there is a very small correlation. However, it's important to have that

because as it moves across the Bay, the level of uncertainty may change. Not every segment or grid point behaves the same way.

On slide 19, all pieces of the statistical interpolation are brought together, and the steps are shown through the four graphs. Adding all of these together brings the full simulation, as shown on slide 20. There are 10 realizations on the two grid points near the "west-gooses" station. As a first go, the team is feeling good about these results. The graph on slide 21 shows the same data but for just July 1-11.

Next steps (same as before):

- ✓ Criteria Assessment link: work with CAP team on how results would be used in criteria assessment (starting now)
- ✓ Method development: continue necessary parametrization of all 4 parts of the 4-D tool (continuing through entire 2025)
- ✓ Software development (continuing through entire 2025 into 2026)

Q from chat: *Tish Robertson:* I am curious how data that falls outside of the grid but is representative of tidal waters are going to be handled. Currently if a data point is within a segment's data region, it is used to inform the interpolation in that segment…even if the data point was collected in a tidal creek not represented by the grid. Will this be the case in the 4D interpolator?

- ***A from chat:*** *Rebecca Murphy:* Tish - we were just looking at some data like this actually. If outside the grid, but tidal nearby, it should be included in this interpolation as well.

Q from chat: *Becky Monahan:* Can someone help me understand what else the interpolator outputs and CFDs are used for besides MD and VA's Integrated Reports?

- ***A from chat:*** *Tish Robertson:* Becky, it is my understanding the CBP still doing its reporting of water quality standards (WQS) attainment, which for DO relies on the interpolator/CFD. Also, the total maximum daily load (TMDL) modeling output is post-processed using the interpolator/CFD, to determine with the TMDL loadings are predicted to meet WQS.
- ***Response from chat:*** *Becky Monahan:* Thank you, Tish, for answering my question!

1:10 PM – 1:30PM: Questions/Discussion

Q: *Tish Robertson:* I just had a thought while looking at those last few slides. Forgive me if I sound incoherent. I'm really worried about our ability to capture the instantaneous minimum exceedances. Those lowest concentrations, right? I'm just throwing this idea out. As we test the performance of these models, can we get some indication of how off we are we the exceedance frequencies of the instantaneous minimum criteria with respect to the array data actual observations. I think that would help us understand whether it's a big deal that we're not capturing all the low values that we're observing or if isn't a big deal.

- ***A:*** *Rebecca Murphy:* I completely agree, Tish. As we've tried to caveat here, these aren't the final results, and we want to be capturing the data. One thing I think will be a very good test, that we have done preliminarily, is in certain spots like this compute what the exceedances are from the vertical array raw data. Then, from the interpolation, what do we get? What percentage of time are we going below the different criteria? That will be a type of check that we need to do. We're going to keep parameterizing too. Jon could maybe add to that answer.
- ***A:*** *Jon Harcum:* That's a great question, Tish. You're exactly right. We do need to make those comparisons where we have the opportunity to do that. The vertical array data is a great place for us to take that opportunity to do so. What Elgin is working on right now is now that we have a portion of a simulation, how do we take that and want to look at the criteria? The shorter duration intervals and instantaneous minimums and see how those best fit. We get ourselves into a best apples and apples comparison because here we had these multiple realizations where the line fits pretty well. With more than 10 realizations it would cover it up pretty darn well. I think that that's a great point but given Elgin's late winter task to try to look at that from a statistical standpoint to make sure we have a solid analysis. He's still well into the development aspects of that. It's definitely on our mind, but we don't have anything to show you today.
- ***Response from chat:*** *Becky Monahan:* I would be really interested in seeing those kinds of checks, Rebecca.
- ***Response:*** *Tish Robertson:* I just wanted to throw that out there. It's good to know you guys have been thinking about that. Thank you.
- ***Response:*** *Elgin Perry:* I'll chime in. We do worry about that too, Tish. That's foremost on all of our minds. I think one thing that is important to remember is with this simulation process, we're going to get a distribution of estimates for what the violation rate is for any point in the space time continuum. What we hope is that the distribution of points that we generate is going to give us a confidence interval that covers the true value we observe in those areas, like where we have the profilers and can actually observe true value. That's going to be our first benchmark about how well this thing performs. We're just getting to the point where we're able to start doing some benchmarking and those are the questions we hope to answer.

Q: *Becky Monahan:* Thank you guys so much. This is an incredible amount of work. I must admit that deep into the weeds here I was only getting half of it so I will have to review these presentations again. My question was along with Tish's point as well. Have you guys looked at how the interpolation would differ from just looking at the raw data? If they are different, we would want to know why, and if they're not different, then what is the point of all the interpolation? That's what we're getting at with our Fishing Bay project is seeing if we can put all of those sampling stations out there and they're capturing the real time data, and I'm listening to Gary here a little bit that anytime you are assessing data, you are using some sort of interpolation, right? So, as we're looking at your screen here (slide 21), you have the west-gooses station outline and you have your pink one to the east of it. If those are just real data, interpolating them without the interpolator would just be either averaging them together or seeing what the number between them would be. I think this is what Matt is getting at too, but there are other ways of doing this without these complex models. I so appreciate your work on this and I know we need some of these models for TMDLs and that's my question, where else is this getting used? Because certainly for IRs for Maryland and Virginia, and Tish knows a lot more about this than I do, but I'm not going to be able to explain this to anybody and I'm not going to be able to run this. I think for IR assessments we need something so much simpler. I would love your help and expertise, and the Bay Program's help on figuring out what else we could do, because I just keep seeing presentations about the 4-D interpolator and I feel like we're not all headed that way. So, I'll just leave you guys with that.

- ***A:*** *Jon Harcum:* I think the answer to your statement is we hear you. Rebecca, Peter, and Gary mentioned earlier in response to Matt's comments earlier that we hear you. I'm not a Bay person, well I work for them as a consultant, but I think that from earlier conversations that the Chesapeake Bay Program folks have heard what's being said. So, thank you for that.

Q: *Jim Hagy:* I have a question that's more technical. Looking at this graph here (slide 21) particularly about halfway down the depth, you see some jumping out and down in the oxygen concentrations. This suggests to me that what we're observing is a physical transport process, maybe like a periodic titling of the pycnocline in some way. So, what you're seeing is a mixture of a vertical and temporal correlation. This raises in my mind a broader question, which is you are using DATAFLOW data to characterize a long track autocorrelation. That data applies to surface water. We have the sensor data that shows if you're on the bottom and clearly away from an interface, we know there is tidal components to it. In that case, you are mixing spatial and temporal autocorrelation. I know you've thought about these things. How do you feel that the inability to purely separate the different components of autocorrelation affects the analysis?

- ***A:*** *Elgin Perry:* This is something we are just starting to look at, Jim. When I look at the graphs on the screen. The black curve looks like it's subject to episodic events that you don't see in our smoother looking simulations. There is

something going on that the model is not capturing. At this point, my hope is that we're catching enough of the noise that our confidence intervals are going to look realistic, even if we are missing some of these episodic events that are clearly happening at the real time scale that our model is not catching. If our confidence intervals don't have good coverage probabilities, I'm not sure how we would go about trying to remedy that problem.

- ***Response:*** *Rebecca Murphy:* Thanks, Elgin. We might have to talk about that more at a development meeting – zooming in on some of these results and thinking about that. Thank you, Jim.

Comment: *Kaylyn Gootman:* Thank you all so much for your work and effort. These questions are really important. Becky, I want to say something. I also really appreciate your comments and feedback. Being able to talk about this for people like me in my office, or for our director or our partners is something that is at top of mind for us. That is very important feedback. Even with usability aside, simply being able to talk about this in a less detailed way, where the common person can understand as it relates to IR or other uses is at the top of mind. I really appreciate your feedback. That is something we've been talking about and have some ideas about making it clearer and making sure everyone is on the same page in understanding. I wanted to let you know that and be appreciative of your feedback.

- ***Response:*** *Becky:* Thank you.

Comment from chat: *Amanda Shaver:* It would be great to plan for more discussion time for these types of presentations in the future. Looking forward to continuing to review at future CAP and other workgroup meetings.

Comment from chat: *Peter Tango:* Becky - thanks for the comments. The Bay Oxygen Research Group was developed around the interpolator development. That is the focus of this group. More broadly, I do hope that our continued expanded interests you point out as options and multiple methods seems best suited to CAP WG. I know, I know, so many groups to track covering specialist focused needs versus more general needs. I really appreciate the concerns. And Becky, I think we have heard from Leah (EPA) that there is no single method applied globally. Multiple methods are norm for consideration.

- ***Response from chat:*** *Becky Monahan:* Thank you, Peter! That makes sense. So many workgroups indeed! I know work on the 4-D has been ongoing, but it seems like MD and VA haven't been as included in the decisions here. Which is confusing since it seems like the interpolators are mainly for us (MD and VA)? But maybe I'm misunderstanding all of this.
- **Response from chat:** *Peter Tango:* Let's follow up Becky, we want to continue to coordinate best as needed for everyone's needs here.

- ***Response from chat:*** *Matt Stover:* And understanding all of the assumptions that go into these assessments.

Q: *Richard Tian:* I have a quick technical question, Rebecca. Concerning the water stations included for specific segments which is polygon defined as that. Recently Tish did some revision on the data file so it would be able to include the citizen science station. My understanding is to use it as a modification tool. My question is can we use a single polygon file for both the Bay Program 3-D or 4-D interpolators? Given that, can you share or find a way to combine all of the modifications? Thanks.

- ***A:*** *Rebecca Murphy:* Yeah, we can talk about that, Richard. For everyone's information, at the moment if we are focusing on one section, we're using the data from all touching segments in at least the first part of the tool. Especially if the 3-D continues to be used, we can talk about that, Richard. Thank you.

Comment: *Breck Sullivan:* Thanks all. Thank you, Rebecca, Elgin, and Jon for all this work you've put together. Thank you to everyone who has brought their comments, and rightly so, concerns forward. I want to emphasize again that this meeting was to help show that we actually have a tool that can function and work. I'm not saying it's the final one. We still need to refine it, make sure it connects well with current monitoring data, and help inform what information we have in areas without monitoring. Because we are now at that point, I am thinking maybe in future meetings instead of us coming forward and presenting on the development, a future meeting could be bringing forward questions and conversations once you have taken the time to digest and understand this. This can be a chance to dig into "what are these other methods that we want to consider? How can we utilize the work being done in the 4-D to complement these other methods? What methods to move forward with in conjunction with this work?" Since we are at this point where we have a tool that actually works, we can try to ask more questions about how it works and get those answers, rather than the presentation and limited time for questions and feedback. Those are some thoughts for us to consider so we can make sure that we answer these questions that continue to come up. Thank you all. We will be meeting again within CAP and BORG and maybe having some more intermit meetings to help address those questions.

- ***Response from chat:*** *Becky Monahan:* Yes, Breck. Having time to review ahead of time would be really helpful!

Comment: *Rebecca Murphy:* Thank you, Jon for all that content and thank you everybody for these great discussions.

Comment: *Peter Tango:* I think for the final points here. Great presentations. It was great to hear from folks about where the concerns are and the ongoing process,

because we are building this as a community tool for community needs and Bay Program needs. We certainly want to do that as a handshake going forwards. Thank you all for the continued insight and help. Great work from the team. I really appreciate everyone taking the time today to help with the guidance and insights that are needed going forward. Good job, everyone. Thank you.

1:30 PM: Adjourn

Allison Welch (CRC), Rebecca Murphy (UMCES), Jon Harcum (TetraTech), Jim Hagy (EPA), Cindy Johnson (VA DEQ), Richard Tian (UMCES), Becky Monahan (MDE), Peter Tango (USGS), Andrew Keppel (MD DNR), Elgin Perry, Tish Robertson (VA DEQ), Matt Stover (MDE), Isabella Bertani (UMCES), Marjy Friedrichs (VIMS), Sabine Miller, Melinda Cutler (MDE), Colin Hawes (VIMS), Angie Wei (UMCES), Breck Sullivan (USGS), Sophia Grossweiler (MDE), Amanda Shaver (VA DEQ), Tom Parham (MD DNR), Gary Shenk (USGS), Jay Lazar (NOAA), Carl Friedrichs (VIMS), Kaylyn Gootman (EPA), Joseph Morina (VA DEQ), Jeremy Testa (UMCES), Leah Ettema (EPA).