Bay Oxygen Research Group Meeting 6/10/22

1:00 PM Introductions

Peter Tango (USGS) and Rebecca Murphy (UMCES) are co-coordinators of the team. Richard Tian (UMCES) is a modeler working with the modeling team with Gary Shenk, and also involved with criteria assessment. Breck Sullivan (USGS) is the Scientific, Technical Assessment and Reporting (STAR) coordinator and works with Elgin Perry and Tetra Tech on their projects. Jon Harcum works for Tetra Tech and has done a lot of General Additive Models (GAMs) work on Bay Trends with Elgin and Rebecca. Jon put together tool used by state stakeholders for recording long term trends with and without flow adjustments. Erik Leppo also works for Tetra Tech primarily with code and also worked on the Bay Trends map. Elgin Perry is a statistical consultant working on the 4-D Interpolator. Amy Goldfischer (CRC) is a STAR staffer with monitoring groups. Isabela Bertani works with Richard and Gary on the modeling side, and help with data analysis. Isabela sat a Tetra Tech office in Fairfax a few years ago. Gary Shenk (USGS) used to be involved with the details of criteria assessment from 2003-2010, though since then he's involved with that more at a distance. He's involved with tracking the overall project of updating the TMDL tools and watershed methods.

Elgin Perry commented that in the STAC workshop Jim Hagy asked some pointed questions about process models vs empirical models and what is the difference. He thought about that question and had things to say when he's back with us (for a future meeting since Jim was unable to attend today's meeting).

1:15 PM Monitoring Data Update – *Peter Tango*

- Even when talking about water quality standard attainment assessments alone there are new, high integrity datasets coming online supported by QA programs including citizen science, satellite, and new innovative approaches. These can expand our capacity and fill spatial and temporal gaps.
- There are millions more data points available than what is in the CBP data hub. There's community science with half a million datapoints. The Chesapeake Monitoring Cooperative temperature data set was utilized for a USGS temperature workshop. Community science data is categorized in tiers for different uses, with different levels of rigor. All tiers have QA standards but a different degree to which they're evaluated. Tier 3 is the most rigorous and equivalent or even more rigorous to states' reporting of water quality measurements. There are fewer groups per more rigorous tier. Most of the CMC data is more recent, but it's expanding. One of the programs, South River Federation, has had their data used in regulatory assessment of the South River. MDE also contributes to CMC database. They incorporate water quality profiles at over 100 sites.
- The hypoxia collaborative helps sync up 4D interpolator – we needed increased resolution in open water, deep water, deep channel that reflects duration of criteria we

haven't been able to monitor with biweekly to monthly datasets. Within the group are agencies and institutions interested in open water sampling such as fisheries, water quality standards, modeling. The DO, salinity and temperature is what we need to assess DO in the bay and habitat conditions that set up designated uses. The pilot test to demonstrate capacity of system to collect data during autumn turnover in the mid-bay led to insights off the mouth of the Choptank.

- Monitoring review. At River Input Monitoring sites there are 10 locations slated for continuous monitoring as boundaries to tidal zone. That would include DO, temperature and conductivity. Looking for 11 vertical arrays in main stem Bay and 10 at the boundaries.
- Other continuous water quality monitoring. Long history of eyes on the bay, VCOS. History going back to late 1990s and 2000s. Could be considered for future network integration. Additional sites for NOAA and CBIBS, offshore single depth, some sites still active. Transitioning some of these to vertical arrays for hypoxia collaborative.
- Remote sensing. SAV folks are looking at sub-daily to daily imaging at 1-4 m resolution. Revolutionizing the opportunities. Sync up Sentinel, new sensors with image processing satellite. Transition of scale in space and time to be more valuable for the monitoring we're doing. NOAA looking at sea surface salinity, temperature and more (Ron Vogel's work). Those data are publicly available, some through MODUS.
- Other resources. From 2018 fisheries workshop there was data inventory between NOAA and USGS, as part of output there were over 400 data layers recognized as potentially useful. Which ones have parameters we need or covariates that could help inform our work is something to follow up on. Along with shallow water monitoring, DATAFLOW single layer surface water quality mapping, high resolution along those tracks interpolated. This isn't done throughout the Bay just specific areas. There are water quality maps along the James River. The possibilities are access to millions more data, high quality data sets, to support the 4D interpolator assessment. I know there's fisheries data set used for habitat suitability index modeling.

Rebecca: always good to be reminded of the data. Liked that you touched on satellite info, especially temperature measurements on surface water. The potential for including that is good, especially looking at these oxygen anomalies linked to high or low temperature events (or at least correlated with). Something to look at in the future. Peter: people are looking at this more. That data will be available.

Elgin: thinking what we should do is compile a list of where we have information gaps. Examples: we have almost no data that gives us information about spatial autocorrelation of DO in deep water. When we go into this interpolation we don't have data to inform that. The scan fish device that VIMS has that goes into deep water or Andrew Muller at the Naval Academy has unmanned mine sweepers. That tool should get that data for us. Let's think about where sources are to fill our weaknesses. We're thin on data that tells us about diel patterns of

DO in mid channel. The buoy profilers help. Back in 90s there were some EMAP deployments that we could draw on, but we're still pretty thin on that data. We could make a laundry list of what we need.

Peter commented: part of our monitoring review was identifying new gaps. Someone in USGS mentioned talking with Navy about having some work done using AUVs. I'll let them know we have some types of data we'd like to glean from surveys they'd like to do. They're looking to do a transect from top to bottom of bay maybe, can keep you posted on that.

Richard: can you put a link to a data hub/central location where we can data source? Because we depend a lot on the data hub. If a link is available, this would facilitate a lot to get the data from other sources. Is citizen science considered as a data source? Peter: yes. They're in their second round of funding. Typically coordinating with STAR on where and when we need info. For a particular tributary where there's work being done we're trying to give them guidance on where we can use additional information. And if they want their data used in the way we use it, we offer them protocols so we can align them up.

1:30 PM Proposal to BORG for implementing Cross-Validation of 4-D estimation GAMs – *Elgin Perry and Rebecca Murphy*

Rebecca went over the cross-validation document with Jim Hagy's comments. Elgin worked on setting up a prototype for using GAMs in the mid-part of the bay, CB4, CB5 and CB-3. Using the oxygen collected year round and fitting one GAM to all of it. Using variables like date and day of year and recently added in flow as well. Based on r squared it performs well and he has great videos watching the oxygen going through time. It captures the average patterns well and some of the interannual variability. As he's added more variables and flow terms it doesn't adjust these large scale parameters very much because there's 120,000 or more observations. The next step for improving this approach is to implement a cross-validation technique where some samples can be removed and use those as validation samples and see how well the model performs for those locations we know the data values for.

Jim put some comments here and seems to agree with the approach. One of the main drivers also was it's a daunting task to test all of the flow average periods that flow could be impacting across the whole year for oxygen. There were some interesting results with a long lag time for flow that would influence current predictions of oxygen. This cross-validation would be targeted to testing flow periods and trying to find good selection of what to use for flow fitting. We thought about what to use as training data and what to use as testing data. One of the first approaches we're proposing is not to randomly split up data. Sometimes more than one observation in a month at each station. This tends to be in the spring and summer. The approach we'd try is to take out second data point in the month, and hold that second monthly data set as testing set. Another possibility would be to use a jack-knifing approach where we take out some data and then put it back in and do another set where we take out more. This

would be done for certain stations that we would take out every other station going down the Bay.

Elgin: my thought is that we would do both of those approaches. Pulling out the second monthly data set is testing how well the model is interpolating over time. Second approach is test of how well its interpolating through space. We have fewer stations than dates though so that's why I thought maybe jack knifing with stations would be good idea.

Rebecca: one thing that's important is hopefully we can get out of this technique is target in on some places and times that are performing poorer than others. We really want to predict low oxygen values in summer well when they start to occur in the southern edge of where the lowest oxygen concentrations are. By taking out certain times and locations hopefully we'll get some. Be able to show error spatially and temporally on some different graphics to target in on improving the model in places where it's important for us to get it right for purposes of criteria assessment. We pulled this together to show what we're hoping to have for next steps for the interpolator development. To commit some time to implementing this cross-validation make sense at this point.

Elgin: I had proposed looking at places where the model performs poorly in space and time. One place we're concerned about is how well the model performs when we're close to our criteria. Looking at the DO range and after we do this validation try to isolate the space within the DO range that's close to our criteria and see how well it's performing there is another strategy for trying to assess a local performance that doesn't look at whole prediction state. Jim seemed to be indicating this in his comments. Rebecca agreed. Was excited to spend some time working with models. Having these error results will help us, especially tweaking the flow. We have a lot of knowledge about what causes bottom low summer DO, those cutoffs around 1 and 3 as well. We can target that we know likely be the spring flow, that's deepest for middle depths. There is potential for improving in those areas.

Elgin: at the Chesapeake modeling conference (Chesapeake Community Research Symposium) Marjy Friedrich demonstrated student data. When you look at DO in an aggregation like hypoxic volume days, get relationship between hypoxic volume days and flow. But when you get down to taking flow and predicting just the concentration of oxygen at individual stations, that relationship is hard to find. I spoke with Marjy and she agreed it doesn't work well at individual station concentration level. Would be good to talk with her and see if she has ideas about how we can do this better. She agreed to be available to us.

Richard: Given that we are looking at the higher frequency time series, the reason that the paper from VIMS illustrated that temperature is a predictor for DO as Rebecca mentioned. Also temperature data would be much more available than DO. Maybe this can fill some gaps in between. Our data is two weeks apart, we don't know what happens in between. Some kind of higher frequency data as a predictor. I know you started out this work with older data, but in recent twenty years there is more data. Can look at starting with most recent data rather than

older data because this methodology is more valuable in the future than going back. Rebecca mentioned the trend at the transition zone of southern and northern border where DO changes a lot seasonally. For the present interpolation, we do it segment by segment. Wonder whether we can adopt that kind of approach, meaning that you fit a GAM model for each segment and do the prediction.

Elgin: those are all good ideas. The idea that we might be able to make better sense of the flow data if we put in another continuous flow variable like temperature, I can see that. At the conference there was an interesting paper by Dong Liang where he was trying to reconstruct mid channel salinity profile on a more continuous basis by using cumulative info from stations as predictor variable and interpolate in what was going on in the mid-channel. That's an interesting approach we might benefit from. If we shifted our test window 10 years into a more current period, we'd have a lot more con-mons to work with if we adopted an approach like that.

Gary Shenk: Trying to connect it to spatial autocorrelation being the biggest challenge we have. I don't know if it addresses that directly. I don't have a feel for the scales we'd have to understand that spatial autocorrelation on. If we're dropping some stations or certain months out using these methods is the distance between the existing stations sufficient such that these stations can help us assess whether that spatial autocorrelation is adequately simulated?

Elgin: in terms of scale, I've never done Kriging of DO but have done Kriging of other con-mon parameters. For instance with chlorophyll, when you do Kriging you get a variogram where it shows points that are close to each other have a high correlation and when you move out in space they become less correlated. Eventually you get to what's called the sill, which is a range where they look independent. With chlorophyll that sill in the James River, it happens at 1-2 km. Our fixed station network is way further apart than 1-2 km. From a statistical point of view our fixed station network are almost independent in terms of the kind of spatial correlation we need to be capturing from somewhere to do estimation. Right now we're shooting for km by km horizontally. In vertical direction, we're shooting for every meter. Because we do profiles at every station we have a lot of info about correlation in the vertical data. We do have data flow data that gives us information about autocorrelation at the surface on a horizontal direction, but we don't have anything that gives us autocorrelation at a horizontal direction at the bottom. That's a big void.

Gary: Richard and Marjy have models of Chesapeake Bay that are completely different top to bottom. Everything about them is different except they're both process models. Is there any value in the fact that these two models exist and if you're using info from one model about its calculated spatial autocorrelation, there isn't any reason to believe it would have the same as the other model. They don't have the same parameters. Is there some knowledge to be gained from one tested or the other?

Elgin: what a model is hoping to do is tell you about what the mean of the process is doing. To get a parameter that we can use as a spatial autocorrelation parameter it would be used in next layer of this model. Use GAM to give estimate of what mean process is doing and then add on random component on top of that. It's that random component which needs to know about spatial and temporal autocorrelation and things like rare events and how often those occur. Hard for me to see where we'll get all the info we need to make that work. Don't think it will come from process model because that's designed to estimate mean, but we need tools to estimate the noise. Maybe the difference between 2 models represent noise.

Gary: process models are pretty noisy. I think of statistical models as representing means and process models…watershed process model was too noisy and that was the problem with it. We've gotten away from it because we're interested in the mean. Elgin: let's keep that on the table then.

Jon Harcum: echo that I appreciate what Richard had to say about fitting GAMs to segment specific GAMs and looking at additional parameters. Re Gary's points about auto correlation and stations we have: the stations are too far apart to see significant correlation. They would be independent sites. When we're at the de-flow sites, when they pause when doing runs up and down river, they will pause and take some readings with depth. Some of that data perhaps is in the data hub as well as some additional data that states will submit. May also have some profile data. We may be in a situation where it's not necessarily jackknifing stations out but if we have a chance to characterize these additional data where we have sites, maybe we have an opportunity to use spatially independent sites without having to go through the jackknifing process. I don't know because we haven't looked through the data to characterize how rich it is.

Richard: Elgin you talk a lot about spatial correlation. Can we find any relationship vertically at each station? If we can, can that help you to do the interpolation? If we have a lot of data for let's say temperature but we don't have that in the bottom layer of water column, if we can get a relationship vertically we may extend signal down the water column.

Elgin: yes. We have a lot of information about autocorrelation in a vertical direction. Even though we only get profiles once a month, we have a lot of months of them. I don't know of any way to translate that vertical autocorrelation information into horizontal autocorrelation to the bottom but maybe some one can work that out.

Rebecca: I did a ton of variograms of oxygen in the Chesapeake Bay. I think there is spatial autocorrelation in fixed station samples throughout mainstem of Bay. Chlorophyll is much lower. But there is a pretty regular pattern of oxygen in the summer. I don't think our nearby stations are completely independent. They're definitely spatially autocorrelated.

1:50 PM Tetra Tech Dissolved Oxygen (DO) data review – *all discuss, provide feedback*

Jon Harcum reviewed the memo from Tetra Tech. They went through Chesapeake Bay Data Hub and pulled DO data from 1990-2019. About 515,000 observations are related to DO

observations from 144 stations we use for tidal trend analysis. Another 150,000 from other stations. We looked at a couple stations in more detail to identify data errors to screen out and see if refinement is needed. We went through a couple of strategies. Erik's done a lot of looking at continuous data and looking for spikes, flatlines and things like that to clean up data. We said since we have that tool let's throw it at this data set. Something that was interesting at the third STAC meeting on DO, Elgin pointed out a few cases where he found observations that didn't match up well with the preliminary models. Is it good data, anomalous, rare event? We wanted to see if we could find those and make sense of them.

Jon showed graphs with all DO data from station CB5.3 1990-2019. Picked a few observations to look at in detail. In April, the DO was lower than typically seen in April but not all that different from what's typically seen in May. Is this an early transitioning to low DO environment or is it an anomaly? In September, there's an observation that's probably a data error and should be removed. The tools we typically see in real time monitoring networks to screen data is really good at picking up things like this. Notice in the June observation we have reported data down to 28 m when most of the other data only goes down to 26 m, or 21 m for November observation. These are the things that we did note we need advice on what to do about it – cases where depths are different from typical depth. Maybe the boat drifted to deeper or more shallow portion.

Peter: regarding the September anomaly, if there are plankton layers occasionally can get strange readings if a bloom dies and settles in the water column or mass of zooplankton has migrated down to certain depth at a particular time of day. I've seen this in lake work. Jon: Data were at 1 m intervals. Didn't consider the plankton as an issue but thank you for this comment. Wanted to look at the April profile. I plotted DO, salinity, temperature and pH. Looked across different events starting from March to October of that year. We see a fairly strong correlation between salinity gradients and DO. When we look at this data, the transition to low DO environment came earlier. It's a less normal event. This would be something when Elgin's model is running it's unlikely that his model will pick up this April 20 because it's not in the mean. But when we get into simulations of looking for extreme events we need some way to characterize these type of events.

Elgin: looking at the salinity profile for that April 20th event, it's clear the bay was stratified which gives credence to the idea that the bottom waters may have gone anoxic. My model is not going to capture that. I had about 3 of them that I identified. It happened infrequently and it happened at a time of year my model considered it to be an abnormality, but it's probably real. Because it's real when we get to that second layer of adding in anomalies, would be nice to have info about how often these happen and build into stochastic simulator, a tool that would produce these rare events at about the right frequency. Is April a month with two observations? Is that the second observation in the month? Jon: yes it is. It shows a nice transition to that low DO environment. Beginning is March-April, transitions nicely as you get

into May timeframe. Is there some characteristic to capture this is a season we'll get a low DO earlier in the year? That would be great. Thinking back to Richard's comments.

Richard: I think this is an important finding, indicating how important it is that interannual variability can happen. If we relieve the days to the GAM prediction we're doing now, I think maybe we're missing interannual variation. For example, the date of the year is the major predictor, but that can change from year to year. The only thing we have is flow to some extent. That's one issue, how can we get it on interannual scale variations.

Peter: we used to have more cruises, upwards of 20 early in the program's history, now we're down to 15-16. April's usually not a two cruises per month.

Jon: a couple of these panels included profiles or events that Elgin said when he was looking at residuals from his preliminary models didn't make sense. In particular this low DO event in January. That's not something we would tend to think would happen. We also have cases of some deeper and shallower data. This December observation – we went down to 0 and then the monitoring stopped. There are some curious profiles in the data set. One of the strategies I was thinking about was using cluster analysis to identify some of these profiles. We can look at 1-2 stations in detail but scaling up to 150 sites and all these other data sets that might become available is arduous, time consuming and subject to subjectivity. Some type of cluster analysis. There are some profiles out there by themselves; do we treat them as anomalous, rare events? And do we have an opportunity to characterize that? I went into some strategies for how to identify those profiles to look at more, or to say can we use that information to help quantify the data in phase 2 part of the modeling effort.

Jon: From the data hub, we have about 160,000 observations at stations that aren't part of the 144 core stations. Maybe this is where those stations could play a part in the cross-validation of the work rather than jackknifing stations out, maybe those sites could contribute to the process. As we go through the data, do we have minimum data requirements for data to be useful for our purpose? If we want to characterize these 160,000 observations what do we need to know about that?

There are some clear data errors. We can write a simple report on that.

Total depths and maximum sample depths varied among profiles. Some profiles went down half the typical depth. That's real data. If we are going to start to look at other sources of data those might come back up as issues we have to address. If a profile is deeper than normal, or shallower than normal, are those data that we want to use for purposes of the model or not? If they get cut off what does this mean for data usability? How does that impact the model for consistency, data integrity? Something happened in the resolution of the location of the model.

I came up with a strategy for using a cluster analysis to find those rare events. Want to see if there's opportunities to identify station profiles that help characterize the variability at a particular location or do we identify them as anomalous events.

2:15 PM Next Steps/Planning

The group did not have time to address all the discussion questions. The next meeting will address what we haven't. People will also continue the discussion over email.

2:25 PM Adjourn