Bay Oxygen Research Group

Friday, June 24th, 2022
9:00 AM – 9:50 AM

Meeting Materials Link

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

NEXT STEPS

- ✓ Continue work on the Mainstem GAM-based prototype
  - Cross Validation implementations (Elgin)
  - Experiments with increasing k-values & targeting flow improvements to times/places where we really want to capture the dynamics (Rebecca doing some work, Elgin)
  - Explore tide, temperature, wind as other variables to incorporate
  - Next: consider simulation phase for short-term temporal and spatial variability
- ✓ Bay-wide identify DO, salinity and temperature data sets beyond our fixed station monitoring (Tetra Tech?)
  - Map out time frames, locations, types of data
  - Identify which data is continuous in space and time and how we will access it/store it as needed
  - QA data as necessary/consider need for QA tools
  - Connect with fisheries group to pull in DO data they're using in assessments for Habitat Suitability Index (Peter help with contact)
- ✓ Start work on shallow-water DO variability – small tributary focus
  - Possibly target a small tributary with record of continuous data to explore how interpolation process will work there
  - Work on how interpolation in smaller system will be linked to the whole bay
- ✓ Start work on a larger tributary example and if mainstem GAM can be translated
  - Pick a tributary with intensive monitoring during some period (York or Potomac perhaps?)
  - Work on how interpolation in larger tributaries will be linked to whole bay
- ✓ Software
  - Start thinking of how this will scale up to the whole bay
- ✓ Meeting format
  - Full group public meeting at end of summer; repeat quarterly
  - Monthly smaller group working meetings

NOTES

- Jon Harcum: I created some PowerPoint slides to frame the discussion at an overview level to help navigate where we go. Hope to get feedback on what we need to be doing for next steps in terms of acquiring Dissolved Oxygen (DO) datasets and get it in an organization that's useful for model development and calibration. Downloaded data sets from CBP data hub. The data goes

from 1990-2019 and includes data from cmon (community/citizen monitoring) and DeFlow. When running up and down rivers they pause for a moment to calibrate and do a vertical profile. I grabbed data from 1990-present because that's the model period Elgin was looking at. In terms of the dataset most of the data is from 144 stations. 102,00 events. Key lets you know about number of events at each of those sites. Right hand panel shows other sites that happen to come down with the datasets (supplemental data collection). Vast majority are 2-3 year time periods, some are 8-10 years.

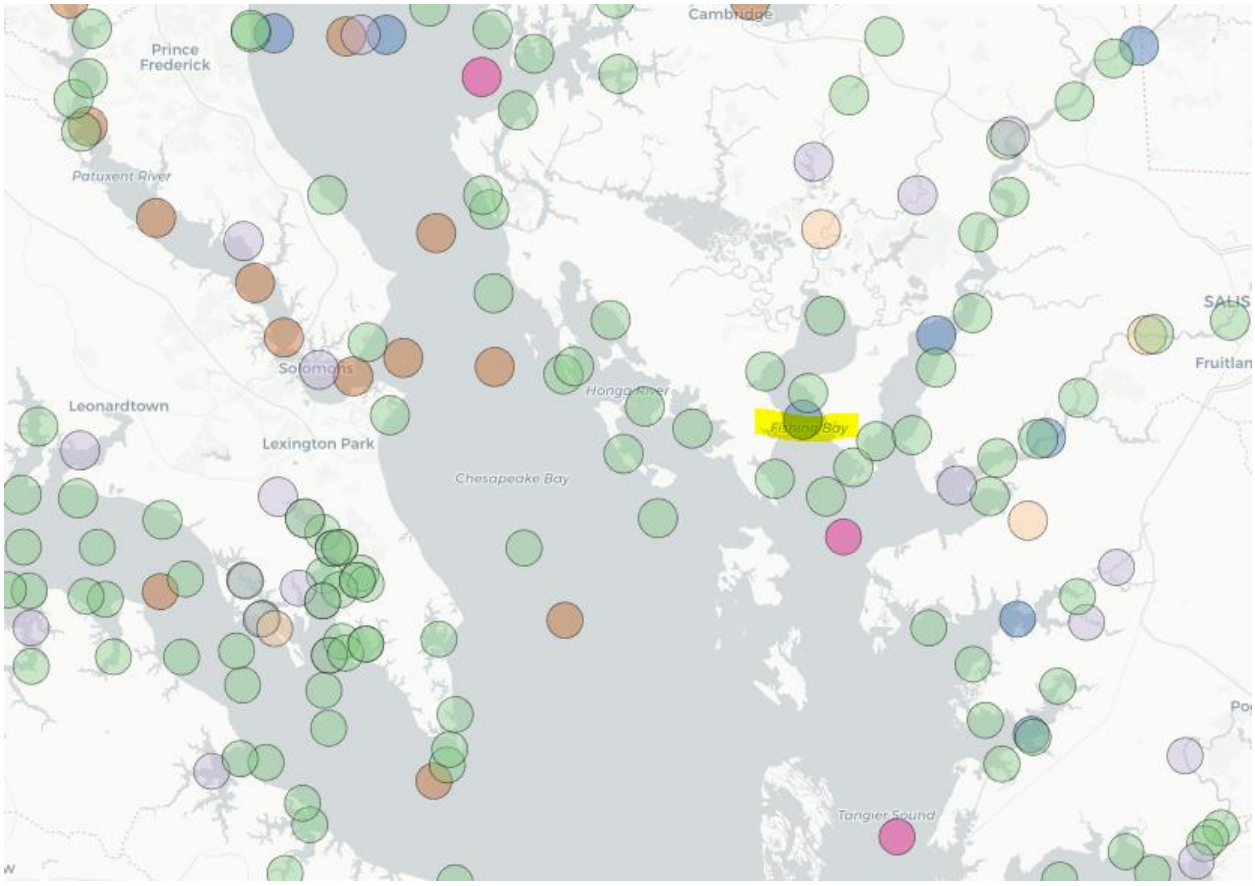
- Looking at CB4.3C and CB4.3E. In the middle is a supplemental site XEF3551 at Dominion Gooses Reef. There's a bit of a shoal there. Are we able to use that site as a calibration, validation type site? We do have some sites in the main channel.
- EE2.2 Mouth of Little Choptank. Had concerns about different depth.
- Peter asked how we got profiles at Gooses Reef given that it's a fixed bottom site. Jon did not know, but can follow up.
- Jon: Discussion point 1 (160,000 DO observations). Recognizing that they'll typically be shorter duration, many have 3 year time period. Maybe need to find those sites that match up with what Elgin has done thus far in terms of using them for additional calibration or in validation.
- Discussion point 2 "clear errors". There's not a whole lot; probably won't affect the model whether they're in the dataset or not, but can flag for taking out, quick sensitivity analysis. Small effort to confirm they don't matter. My sense is with data going into data hub they've already gone through some level of screening but we've maybe not interrogated the data this way so maybe something to be aware of.
- Discussion point 3 total depths. These are based on relatively few number of cases where looked at more data, bathymetry data. There's some in later records of 2010-20 that's also occurring. It's not been completely eliminated but think data we're getting is reasonable. Given 1km by 1km by 1m depth my sense is we need to be using that data.
- Other thing I looked at is identifying outlying profiles. Example mouth of Potomac. Lower DO environment started earlier in that particular year but when we look at DO and salinity and other profiles that happen earlier in the month. Those things are real data. Maybe second phase of model talk about statistical simulation. At the moment don't necessarily need to spend time. Identify some of those sites and look at that for phase 2.
- Jim Hagy: big picture question. Chesapeake Bay monitoring data has industry leading data screening procedures and there's a ton of data such that the kind of things where there's actually a data problem is a small fraction of the total amount of data. I'll give you that this 4.9 is probably a wrong number, but for this effort I wonder if we should proceed with the assumption that the data we have are good data. There's a slippery slope for data that doesn't fit. If it's real and you take it out then your statistics on how well you're predicting or describing data with the model are biased, probably not much. With the 4.9 send it back to data people and see if they can track it down and remove it for everyone's benefit not just us. For people who use sonds, a lot of the sonds report percent saturation and mg/l, and we write down both. If there's an error in either DO, temperature and salinity you can calculate the mg/l from the percent saturation and if they don't agree you know one of those numbers is wrong. You could find out that the percent saturation combined with the DO and salinity is 4.9, then you can confidently change that number rather than throw it out. However, it seems like a tempest in a teapot when you have so much good data it swamps these issues.

- Rebecca: when Jon and I were talking yesterday we were close to getting there. Especially with this method, if there's a handful of truly erroneous data it won't matter.
- Jim: there's some data sets I've seen where low oxygen water near the bottom breaks off with tidal straining along an isohaline and you end up with this thin stream of low oxygen water in the middle of a profile. I've seen it repeated so it's real. There's weird stuff in the world and we should collect enough data so we don't have to second guess our data based on it looking weird.
- Richard: when they set up the data hub, they do serious work of quality control from the beginning a lot of time. Jim made a good point if we think of something wrong and correct it for everyone, not only for us, at one time I saw a salinity number in the DO column, obviously it's wrong. It's inevitable that there are mistakes, but better to work together with Mark to get more information and correct dataset at datahub rather than in our spreadsheet.
- Jim: similarly when the sample data are from deeper than the total depth, the ship gets to the station, the person looks at the depth sounder and fills out time of arrival depth, then you drift a little into deeper water, to me we should ignore that stuff. The depth comes from the instrument that measures the DO, it wasn't lying.
- Peter commented in the chat: mentioned seeing something similar in data sets Jim. Have also read about thin layer hypoxia observations in other systems. Just saying it can be real. But good to have some background checking about the event to continue to sustain data integrity. He added that it's good to interrogate the data from time to time. Not let it stop where 99.9% of the data gives us the momentum to keep going.
- Jim: the situation where you have more ad hoc data than you know what to do with is the norm. How do GAMs deal with little bits of one year of data sprinkled into datasets that have 20 years of data? My guess is these data are better suited for validation.
- Elgin: it is a good point to raise. The GAM we're using right now that has a long term trend term and space in the model and season in the model. My feeling is if a new site had at least a complete year of data so getting some info about seasonality at the site, it could be used as part of the training data. Sites that don't meet those criteria I'd be very skeptical about using those as training data. Any smattering of data could be used for validation.
- Jim: what if it's unbalanced? Short term data sets towards end of your long term rather than beginning? Does the GAM deal with that?
- Elgin: it tries to. The GAM we have has a tensor smooth term. It'll give you lower standard errors at later part with more data and greater ones at start with less data. My understanding is an assessment of criteria we only do that assessment for a 3 year period. We may use 20-30 years of data to calibrate the model, but only apply it for a few years. If we have an increasing frequency of datasets over time for the purposes of this model it's the 3 years we have the most data we'll be applying it and I feel comfortable it will be ok.
- Jim: what if the snippet of data is May-September, why shouldn't we use it?
- Elgin: we're expecting violations to occur May-September, if we have a piece of data that occurs over that time especially for 3-4 years at a time that's an argument for using it. That will be true of many continuous monitoring (con-mon) sites, they get pulled out in winter when there's a chance of ice damage.
- Peter: a lot of the nearshore ones. As Jon mentioned some sites have been longer term. We target critical most sensitive time for hypoxia but there's a need for both.

- Elgin: in winter you can have an imprecise prediction that still tells you things are ok. Being weighted toward summer would be a good thing.
- Jon: are there other datasets Tetra Tech should engage with to bring in for validation training process in the near-term or more further down the road?
- Rebecca: any continuous monitoring datasets that states have collected, the calibration info should be in the datahub, right? Because we want to use continuous data eventually but that's not in there.
- Peter: the bulk should be covered, yes.
- Richard: yes, there is a lot of con-mon data flow specified based on that we can pull out violation.
- Peter: one of the datasets I'm less clear about that seems to have a long history – the fisheries folks when they're doing their surveys, they do some kind of DO measurements and some of that data is being used for their habitat suitability modeling. It seems very challenging to get access to, but we might be able to. I probed fisheries folks a few times but don't have clear answer yet on accessibility of that data or where it resides. Apparently it's a long-term data set that relates to how they look at and connect their many fisheries surveys with habitat information to make those associations that they use to calibrate their HSIs. The hypoxia network development - those stations are coming online and expected to come online more so. Those are datasets that are vertical sensor arrays that will be collecting data at real time out in deeper waters of main channel in Chesapeake Bay and maybe lower tributaries with 10 min observations of salinity, temperature and DO. These are profiles. Expectation of having 11 operating in the not too distant future. These are string of sensors. Virginia may have a climber or one that goes up and down. But the bulk are fixed. There's the citizen science ones, some on weekly to biweekly timetables on tributaries that may be of use in the training data set. If those are approved at Tier 3 (meaning doing equal to or greater than type of quality assurance as state programs), those may become a higher value dataset as time goes on. We talked about satellite information to inform what we have but a couple of efforts both on Aaron Bever's side and the presentation during the workshop suggest there's real challenges in only having some of the efforts trying to look at the surface DO and infer anything underneath that so didn't seem profitable to put a lot of effort into using that to drive DO patterns right now.
- Jon: is there something you want us to do this month, next month, in regards to going out and reaching this additional data beyond what I've talked about today?
- Peter: I'll see if I can connect you with the fisheries dataset because that's the largest unknown potentially helpful dataset that we don't now have ready access to.
- Elgin: Surprised when Jon showed me how much data there is available that's not part of the fixed station network. I was only using data from fixed station network. I am about to embark on cross validation stuff and will be retooling some of my software. It might be a good opportunity to have Jon pull together a bigger dataset. Right now the dataset I'm using doesn't have much information about profiles in shoal waters. Some of the data he's presented today would be capable of giving us this information. I propose we ask him to put together a new test dataset. I can work on developing a tool, but when it's time to start next phase of applying that tool and producing those results might be nice to have a more complete dataset to try it with.
- Rebecca: Those are good ideas. Following up with fisheries group is a great idea, if you can give us a contact, Peter. Someone always mentions that data. That would be great to track that

down and decide if it's useful to us. And also what Elgin just suggested. And looking at maps Jon made, wonder if it's smart to target in an area in 3 year period that has intensive shallow water monitoring to find a place where we could try to pick a different region and start prototyping this approach. Would be great to think about reasonable next steps. Elgin's continuing with cross-validation, Tetra Tech looking at more data, Breck has shallow water data. Would like input on if we should identify a different area to explore the approach.

- Peter: remind folks that MD DNR is focusing on Fishing Bay as a place they want to do their pilot 3 year all out monitoring blitz to make an assessment at a higher data resolution than we typically have had in a smaller system. Fishing Bay is on the Eastern Shore north of the Pocomoke, south of the Choptank.
- Jim: It would be interesting with something that's at a fundamentally different scale with variability having a different signature. Fishing Bay might meet that. I was going to say Mattaponi.
- Meeting plans: proposal is that larger more public meetings on a more quarterly scale, kind of like the modeling workgroup. If folks are comfortable with that we can target that. If there is another frequency you think that would be helpful?
- Richard: suggest you contact Lew Linker.
- Rebecca: we have a BORG website and list of members. Folks like Mark Trice might be wondering what's going on, though we did have a larger group update at the STAC meeting. Maybe at end of summer have larger group meeting. Then this small group and anyone else that we need to talk about the work to continue once a month. Plus Angie Wei. Continue once a month to keep us moving forward.
- Breck commented in the chat that she can help share shallow water stations with more than 10 years of data.
- Jon shared a map in the chat showing where Fishing Bay is along with a number of sites. Where I've been drawing from, I'll share that as well. We want to build that out more to help identify in Elgin's model, is it a 2003-2005 the perfect data time to do that drill in. We'll work on that more.


Cambridge
Prince Frederick
Patuxent River
Solomons
Leonardtown
Lexington Park
Chesapeake Bay
Hooper Island
Fishing Bay
Tangier Sound