Criteria Assessment Protocol (CAP) Workgroup Meeting

Monday, June 9, 2025
1:30 - 3:30 PM

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

Decisions Needed: None

ACTION ITEMS

- ✓ Continued conversations within the CAP WG on pycnocline data and how that can be averaged with the designated uses.
- ✓ Share any feedback or questions on Elgin's presentation or his more detailed summary (available on the meeting website) to Elgin (eperry@chesapeake.net) and Rebecca Murphy (rmurphy@chesapeake.net) in the coming months, especially if more analysis on applying the CFD approach or 10% rule using the 4-d interpolator output would be helpful to CAP decisions.

PARTICIPANTS

Peter Tango (USGS, *Chair*), Tish Robertson (VA DEQ, *Chair*), Matt Stover (MDE, *Chair*), Allison Welch (CRC, *Staffer*), Becky Monahan (MDE), Carl Friedrichs (VIMS), Claire Buchanan (ICPRB), Clifton Bell (Brown and Caldwell), Jillian Cudnik (EPA), Rebecca Diehl (DOEE), Elgin Perry, Leah Ettema (EPA), Gabriel Duran (CRC), Jon Harcum (TetraTech), Lewis Linker (EPA), Joseph Morina (VA DEQ), Rebecca Murphy (UMCE), Richard Tian (UMCES), Amanda Shaver (VA DEQ), and Tom Parham (MD DNR).

MINUTES

1:30 PM **Welcome, Introductions & Announcements –** *Peter Tango (USGS), Chair*

Upcoming Conferences, Meetings, Workshops and Webinars:

- One Water Summit 2025 – July 8-11, 2025, Pittsburgh, Pennsylvania.
- 2025 Capital Area Natural Resource Management Symposium – August 22, 2025, Washington, D.C.
- Chesapeake Watershed Forum – November 7-9, 2025, Shepherdstown, West Virginia.
- Coastal and Estuarine Research Federation Conference – November 9-13, 2025, Richmond, Virginia.

Summary: Peter gave a brief introduction on the agenda and the addition of the new co-chairs, Tish Robertson and Matt Stover.

1:40 PM **Examples of criteria assessment using 4-D interpolation results for the entire CB4 segment** – *Elgin Perry*

Description: *Elgin and the 4-D interpolator team have been working to link the 4-D interpolator with criteria assessment. Elgin has partially completed this work in the CB4 segment and would like some feedback from CAP WG.*

Summary: Elgin presented the updated work for the 4-D interpolator and criteria development. Elgin mentioned that this presentation is similar to previous presentations given to CAP WG, but this new information focuses on the entire CB4 segment.

Elgin used the first few slides to give an overview of dissolved oxygen criteria, the work he will be presenting, and the data that was used in the interpolation. In this talk, Elgin focuses on June 1 – August 31 data because that is when the vertical array data is available. There are a few methodologies that Elgin has been using, and touched on in his last presentation, which is shown in slide 3. In slide 4, he shows the scope developed for interpolation data. The reduced dataset is what he used for today's demonstration. On slide 5, Elgin went over the two different instantaneous minimum methods, CFD assessment and 10% rule.

Next, Elgin goes over the results for the instantaneous minimum methods using the selected data set. For the instantaneous minimum CFD assessment, it appears that there is a small amount of space that almost always violates this criteria over this summer period. For the 10% rule, the fraction violations were calculated and used to create the distribution function. This was used to calculate the probability of exceeding the 10% threshold, which is very small in this example, meaning it is likely in compliance.

Weekly means have a similar method. In this method, the first step is to calculate the weekly mean over a week for each cell.

Q: *Claire Buchanan:* For the weekly mean, you have done a hundred runs for each of the hourly steps. How are you calculating the mean? Are you averaging the 100 runs and then across time for the mean?

- ***A:*** *Elgin Perry:* I am only using 10 simulations for this example, but I am keeping them separate. Then, you have a space time domain where your dimensions in space and time are latitude, longitude, depth and hour. On my next slide, I put together this table which is a time series of dissolved oxygen (DO). I have dates on the left-hand side because I wouldn't have been able to fit hours. Each number represents 24 hours. When I do this, I am averaging hourly values from the beginning to the end of the week. I am calling that the sequential mean average for the week, which I am doing for every interpolator cell. Then, I look at the moving window weeks which lose a day from the previous week and add a day from the next

week. In the right-hand column, I use that information and get more estimates for weekly mean.

Continued Presentation Summary: Next, Elgin shows the violation fractions for the two methods. The red dots represent the sequential weeks, and the moving window weeks are the black dots. There are multiple red dots per week, because there are multiple simulations. Next, on slide 11, Elgin shows the CFD results for the weekly mean CFD and moving window weekly mean CFD. Last time Elgin presented this data, there was a large difference between the two, which he now believes was a programming error. Now, the two methods give similar results. The data that falls to the right of the blue curve is out of compliance, which is a small fraction of space over a large fraction of time for these graphs.

For the 10% rule, both methods show similar results and have a small probability of exceeding 10%.

Comment from chat: *Tish Robertson:* I also have found that you get very similar continuous monitoring exceedance rates when you use sequential versus moving windows.

Q: *Leah Ettema:* Could you explain what the CFD versus simple 10% are telling us? They are both telling us a fraction of space-time exceedance in different ways, correct?

- ***A:*** *Elgin Perry:* Yes, it is expressed in different ways. The CFD makes an allowance for small amounts of space to be out of compliance for a small amount of time. If you have a large amount of space out of compliance for a small amount of time, that is unacceptable. If you have a small amount of space out of compliance for a large amount of time, that is unacceptable. Of course, if it's large and large, that's totally unacceptable. It gives you a finer resolution of how space and time are interacting with the compliance structure. With the 10% rule, it is all thrown together into one space-time domain. The CFD tends to be more protective. It shows more violations than the 10% rule would, but maybe that means 10% is too high and we should make it 5%. I like the simplicity of the 10% rule and that we can make probability estimates of exceeding that threshold. Those statements are harder to create with the CFD. There was a review of the CFD done by STAC in 2006. It is comprehensive in explaining what it does and what problems it has. (Link to the 2006 STAC review of the CFD: https://www.chesapeake.org/pubs/cfd_stac_final.pdf)
- ***A:*** *Tish Robertson:* With the CFD, we are presuming that we can regulate the temporal distribution of the exceedances. With the simple 10%, we are not making a statement of how the exceedances are spread out over time. It is kind of flattening out the temporal distribution.

- ***Response from chat:*** *Peter Tango:* 10% rule doesn't assume the distribution of violations, the CFD does, and a bioreference curve was desired in order to not make such assumptions about the distribution of violations such as the symmetry of the 10% CFD hyperbolic curve. I believe that complements what Tish and Elgin said.

Continued Presentation Summary: Starting at slide 13, Elgin compares the vertical array data to the interpolator data. The map shows the CB4 segment. The red dots show the fixed network stations, and the blue-green dots show the vertical array. The matrix of grey dots show the surface cells of the 4-D interpolator. Elgin will be using the two grey dots in the blue boxes to compare the interpolator data to the observed data from the corresponding station. In this comparison, Elgin only used the vertical array data. On slide 14, Elgin shows the time series graphs that compare the interpolator data (in blue and yellow) and the observed data (in black). If you'd like to continue comparing this data, Elgin encourages you to visit the report that goes along with this presentation. (The report that goes with this presentation can be found on the meeting webpage and here: https://www.chesapeakebay.net/files/documents/CriteriaAssessmentUsing4DInterpolator-Document_ElginPerry_06.09.25.pdf)

Elgin noticed that the observed data at West Gooses has a lot more violations than East Gooses (table 5). The table on slide 16 shows the rate of violations for each of the 10 4-d interpolator simulations for both of the cells associated with West and East Gooses. Elgin was concerned that the interpolator wouldn't be able to pick up these extra violations in West Gooses, but it has, and that can be seen in this table in comparison to the ranges below it. Slide 17 shows the same data but in a distribution function. This shows the for East Gooses, the probability of exceeding 10% is very small, but for West Gooses, there is a much higher chance that the true mean is not less than 10%.

Comment: *Peter Tango:* Elgin, I find that when we are talking about our sampling design and where we can and do drop these in the importance of our spatial variability. If we were west orientated in our samples, we have considerably more violation than if we are east orientated. Capturing and representing spatial variability becomes important when you know what you're up against. We don't want to over or underestimate. We want to be able to see the impact of regionalization on the sampling site. It is helpful to see it in these terms.

Next steps: Lastly, Elgin went over his slide on the next steps. He encourages everyone to keep thinking about these items and how they relate to assessment as the team works on them.

Q: *Clifton Bell:* Can you come to any conclusions about whether there is bias? Is that something you are planning on looking at whether the interpolator may have a consistent bias compared to the array data?

- ***A:*** *Elgin Perry:* The array data were used in the calibration of the interpolator, therefore, we're using an unbiased estimation process. It would be almost impossible for it to be biased with respect to the data that we're using to calibrate it. We want to know whether there is bias in places where we don't have data. In the future, we will try to do some bootstrapping type assessments where we take a place where we have data, kick it out, refit the model and go back to see how well it predicted the data. That will allow us to address what you want to know.
- ***Response:*** *Clifton Bell:* I know you've got an unbiased estimator, but it's been influenced by data from a lot of different locations. The array data here would seem to indicate that this is as close as we can get to the true answer because it is direct measurements for this particular location. So, isn't it the most accurate basis of comparison to what is interpolated?
- ***Response:*** *Elgin Perry:* You're right. There is the possibility that in doing our smoothing with the array data since a lot of it gets used there. It would have such a strong influence that it would probably not be biased. Even if fixed station data was showing something is, it is only once or twice a month, so it won't override the influence of the vertical array data. There is that possibility that if you had neighboring stations that were pulling the smoothing curves away from what was going on at your particular location that it could be biased. We're still early. It wouldn't be fair to pass judgement either way. There are only a couple of cases we have looked at. We need to do a lot more of that before we make a final judgement.

Q: *Matt Stover:* I don't think we have anything in our blue and white documents that mentions what to do when you have a variable pycnocline. We don't know what our assessment rule or logic is. What you're saying is when we have deep water it can exist one day but not another. That's the point you're bringing up?

- ***A:*** *Elgin Perry:* Yes, in the 3-D interpolator they interpolated the pycnocline but assumed it was constant for an entire month. The variability of the upper and lower pycnocline is a fairly new discovery. Historically, we thought the designated uses were very stable. With this new vertical array data, we can see that they are changing within a week, so we need to rethink how we do those weekly averages or whether we need a different metric, like duration below a threshold. It will be a big discussion for another time.
- ***Response:*** *Matt Stover:* I'm guessing we don't have a sense for how often this happens or if certain sites are more prone.
- ***A:*** *Rebecca Murphy:* There are probably certain places where it doesn't move as much. It is a fundamental question of how to do higher frequency interpolations. Even if it's not much in some places, the pycnocline is not stable all summer. We have to consider that a cell of water at five meters deep is sometimes open and sometimes deep water. How do we compute these averages with designated use

with that information? There are options. We will have to pull out examples and discuss it as a team.

- ***Response:*** *Matt Stover:* It seems like you brought up an action item for the CAP WG for a decision that has to be made in the future with perhaps a bit of background work.
- ✓ **Action Item:** Continued conversations within the CAP WG on pycnocline data and how that can be averaged with the designated uses.
- ***Response:*** *Rebecca Murphy:* We definitely need to look at some graphs so you can visualize what is going on here.
- ***A:*** *Richard Tian:* I wanted to mention that the pycnocline is also averaged. Is the frequency for computation the same as the data? Is it averaged the same way as concentration?
- ***Response:*** *Rebecca Murphy:* Yes, currently it is the same frequency as oxygen interpolations, and we would be doing that with the new interpolator too.
- ***Comment:*** *Peter Tango:* Back in 2003/4/5, there was a fixed pycnocline in 18 segments. The dynamic pycnocline was acknowledged in the 2010 document. There was a step there but getting to the higher frequency is presenting a new issue.
- ***Comment:*** *Tish Robertson:* I was going to throw into consideration that maybe we don't do the averaging of the pycnocline that we do now, and we find a shorter time duration or scale. We could parse the data on that time interval and then do the average for the seven-day-mean or one-day-mean. There is no reason that we have to continue with the way we are doing it now.
- ***Response:*** *Peter Tango:* The ability to see what we've seen in the data and to try to match that to what we call our designated uses and apply this – I'm glad the team is already working on that.
- ***Response:*** *Rebecca Murphy:* I think you are right on Tish. I don't think we need to do the same thing. It might not be advisable to do so.

Comment: *Matt Stover:* Are you looking for comments and feedback by a specific date or just a general review thus far?

- ***A:*** *Rebecca Murphy:* I don't think we have a specific date. There weren't any real decision points here. Eventually, we will need to make some decisions, like how to do weekly means, but he is going to run through some more examples to help flesh out the discussion. If you have questions, thoughts, or want clarification, let us know. We are aiming to get a lot of this done by the end of the year.

Comment from chat: *Leah Ettema*: potential conversation for another meeting. it seems that locations/cells with more temporal and spatial monitoring (continuous and array vs bimonthly discrete) will have more influence on the interpolator results for an unmonitored cell? To understand potential bias due to monitoring locations (and frequency), when

assessment results are generated, it would be really helpful to see a map of the data that goes into each one. I think those would be helpful to see for even these CAP discussions.

- ***Response:*** *Peter Tango:* That is something we are looking at with one of our projects. We're looking at sample site distribution relative to influence in order to inform bias in space. It is a consideration we are accounting for and look to place resources to minimize.
- ***Response from chat:*** *Rebecca Murphy:* Great feedback, thanks!

Q from chat: Tom Parham: Elgin - Nice presentation. Will the monthly attainment values be calculated similar to the weekly process using the hourly predictions?

- ***A from chat:*** *Rebecca Murphy:* Tom - I think Elgin had to leave. But yes, we would calculate a monthly average for each interpolator cells from the hourly estimates.

2:20 PM **State Approval Pathways –** *Tish Robertson and Joseph Morina (DEQ), Becky Monahan and Melinda Cutler (MDE), and Rebecca Diehl (DOEE).*

Description: *Presentations from each jurisdiction (MDE, DEQ, DOEE) on their approval processes followed by discussion to best inform the CAP WG decisional processes.*

- *Overview of state approval pathways. – DEQ*
- *Approval process for integrated report and steps for adopting changes to water quality standards. – MDE*
- *Water quality standards rule-making and approval process. – DOEE*

Presentation: **Virginia's Approval Pathways** **–** *Tish Robertson and Joseph Morina (Virginia Department of Environmental Quality)*

Summary: Tish gave an overview of the water quality standards rulemaking. Every 3 years, states have to open their water quality regulations in a public hearing for feedback to seek improvement. This is to follow federal and state legislation. This goes through many organizations, finishing with EPA's approval.

Next, Joe shared the water quality assessment process. This is a two-year process that is approved by EPA. This is the process that creates the biannual integrated report and uses monitoring data.

These two processes are closely related and work within each other. When changing the water quality standards, Virginia must create a Notice of Intended Regulatory Action (NOIRA) which is subject to approval by many state offices. This document is presented to the public for a public comment period. Then, a regulatory advisory panel is convened. After that, they must go to the state Water Control Board with proposed regulations for approval. Then, it goes to public comment, again. The timeline for this process is 18

months, but it rarely meets that because the proposal can spend a long time on the governor's desk. There is no time limit for the governor.

Next, Joe shared a timeline for the integrated report (IR). One thing Joe highlighted is that Virginia sends their data to their regional assessment staff for manual review. This is likely to differ from some other states because some states do their whole assessment through scripting and don't have a manual review.

Q: *Richard Tian:* For this timeline, what years of data are you analyzing?

- ***A:*** *Joseph Morina:* The example I am showing is for the 2026 IR and you can see under "Data Solicitation Deadline" the six-year data window is 2019-2024 for this example.
- ***A:*** *Tish Robertson:* For Bay DO, we would use the three most recent years within those six years. For this example, we would use 2022-2024 period for Bay DO.

Q: *Richard Tian:* My understanding is that you would need federal approval to do an amendment, correct?

- ***A:*** *Tish Robertson:* Yes. Water quality standards are provisions, but you can't call any provision a water quality standard unless EPA says it's a water quality standard. It is a big deal to get EPA approval.

Q: *Rebecca Murphy:* If a change happens to a method, not the standard, but the interpolation method, where does that fall into this process?

- ***A:*** *Tish Robertson:* We would consider that to be assessment guidance. It wouldn't be a standards change because in our standards all we say is using the CFD for Bay DO. We don't specifically say what interpolation method we use. For our upcoming triannual review, we are going to propose some changes to the implementation language in the Bay criteria section. Last year, we were going to do fast track rulemaking on that language because right now it says we must use the CFD. We would like to keep the CFD reference but include other approved methods which are in the guidance. These would be worked out in CAP WG and will be reviewed by the public.

Comment: *Leah Ettema:* I wanted to add EPA's role in the IR. The IR includes water bodies that are impaired for a pollutant that needs a TMDL (total maximum daily load) in category 5, as well as water bodies that are impaired but have a TMDL, and water bodies that are attaining. Technically, category 5 is the 303D list, waters that are impaired for a pollutant without a TMDL. EPA reviews and approves the 303D list. In regenerating that approval, we review the assessment methodologies and that supports our approval of that list. We don't actually approve or disapprove an assessment methodology; we approve or disapprove the final decision.

Presentation: **Maryland's Water Quality Standards Regulation Timeline** **–** *Matt Stover, Becky Monahan and Melinda Cutler (Maryland Department of the Environment)*

Summary: Matt began the presentation on Maryland's regulation. He mentioned that Maryland's process is similar to Virginia's, but they may be running on different timelines. Matt mentioned that Maryland's process is slightly shorter than Virginia's because Maryland doesn't have to go through a water quality board. He also reviewed the considerations for the water quality standards and the approval pathway.

Next, Becky presented the integrated report process and timeline. As a new step, they are trying to make their timeline dates regular for each integrated report so that partners will know what date they can expect to be reached out to or presented for public comment. Becky added the star in August to emphasize that they need all final data and information by this point to submit to EPA through ATTAINS. Becky mentioned that they are stricter with the timeline but more flexible with the years of data they are asking for. The deadline for the IR is April 1st but sometimes it delays into the following months.

Rebecca Diehl agreed that the water quality approval process was very similar for DOEE.

Q: *Richard Tian:* Becky, your 2026 IR covers the data from 2021-2023. Virginia would be 2022-2024. Is that correct?

- ***A:*** *Becky Monahan:* You understand correctly. That is something we talked about with Virginia. We should be using the same data. What happened in the past is that I wasn't willing to hold up the whole IR process. Since our year requirements are not as strict, we were able to move back a year. Virginia has stricter data calls and they are willing to push back their IR dates to get that new data. We need to talk with Virginia to align our timelines better.
- ***Q:*** *Matt Stover:* Virginia, you run the data yourself, correct? QAQC and 3-d interpolator? In Maryland, we rely more on others to assist with that.
- ***A:*** *Amanda Shaver:* Exactly correct. We are able to run analysis in house. Becky, you are right. We seem to be pushed to use more recent data. We do have datasets that aren't available in that timeline, but that is more so than our biological data, fish tissue analysis or benthic workups. Those typically follow a five-year time frame.

3:00 PM **Chesapeake Bay Program (CBP) Decision-making: An Overview** **–** *Lucinda Power and Bo Williams (EPA)*

Description: *An overview of the process and role of key partners and partnership groups in CBP decision-making. The Water Quality Goal Implementation Team's (WQGIT) decision-making framework will be provided as a case study example of how decisions are made, the timing, and determination of which decisions get elevated to the next decision-making body.*

Summary: Lucinda and Bo reviewed the structure of the Chesapeake Bay Program (CBP) and the key decision makers. They also emphasized the importance of making consensus decision-making using the consensus continuum. Next, they went through the flow of decision-making. Sometimes, these decisions stay at the workgroup level, while others move up. Sometimes, STAR decisions move to the WQGIT, but most of them move directly to the Management Board. Next, Bo went through an example decision-making process as used in the WQGIT.

Q: *Matt Stover:* I think this subject came up in the CAP WG because this group looks at assessment methodologies that are typically the purview of state regulatory agencies. We try to get on the same page so that we can assess water quality the same. If we are making decisions in this workgroup, who has primacy? Ultimately those decisions would be made by the states but here we are making decisions as a partnership so that changes some things. As a regulatory agency representative, I think it is still our decision on a lot of those things. I don't know if you have any insight of that, specifically with the CAP WG. It seems like a lot of our more impactful decisions would be taken to the Management Board.

- ***A:*** *Bo Williams:* If it's a partnership decision, then I think it affects how the partners are going to assess waters and they have agreed to it, I think that would go up to the Management Board level and it wouldn't just be informing. I think it could be decision by decision and there would have to be communication with the Management Board on that topic and a debrief ahead of time. When it is a big partnership decision that affects jurisdictional resource allocations, it would be a Management Board level decision.
- ***A:*** *Lucinda Power:* Also, we'd have to include a director level or decision-maker level within those regulatory agencies. Many times, those decision-makers refer to the technical experts because they are in the weeds and can provide scientific data. There is often a shared membership of the WQGIT and MB because of the TMDL. Then the MB could decide whether they want to formally weigh in on that.
- ***Comment:*** *Bo Williams:* Yes, I think there is that awareness factor and then the purview of the MB to decide if they want to weigh in on the issue.
- ***Comment:*** *Peter Tango:* I know the language from the WQGIT ties in, not only implementation, but checking on the data and outputs of the standards attainment assessment. That has connections with updating the methods, and they are likely going to want to understand what the methods are that are feeding the results that will be evaluated. In those moments, the modeling and monitoring data connect with the partnership. It would depend on the significance of particular topics.

Q: *Matt Stover:* Trying to decide on the criteria and the work in the CAP WG is part of the Clean Water Act requirements that states are delegated authority on from the EPA. Within other workgroups and their structures, the primacy isn't established clearly because we are required under the Clean Water Act to develop criteria and report on them. I know the

partnership is a big beast and there are a lot of different interests and perspectives. One of my goals is to make sure the regulatory perspective gets a substantial say in some of those decisions, like where monitoring resources go. From a regulatory perspective, we're having difficulty assessing all of our segments because monitoring resources are not coordinated in a way that would help us do that. I am thinking outside the box and typical org structure to be making goals in an integrated fashion so we can meet our goals too. I keep hearing from my leadership that we are pouring millions of dollars into monitoring but still cannot assess our high frequency DO criteria and show the progress that's been made by years of implementation. I think we could work together to align our goals with those of the decision makers. We want to be able to show the progress of our monitoring efforts in our integrated reports.

- ***A:*** *Lucinda Power:* Integrated decision making is something that the partnership will dive into in our Beyond 2025 efforts through the structure and governance work.

3:30 PM Adjourn