

Update on the 4-dimensional interpolator progress

Rebecca Murphy (UMCES) and Peter Tango (USGS)
Water Quality GIT
Feb. 26, 2024

Key input on presentation from:
Elgin Perry (statistics consultant) and Jon Harcum (Tetra Tech)
Breck Sullivan (USGS) and August Goldfischer (CRC)

Presentation outline

- Reminder – Why a 4D interpolator?
- Data sources being used
- Daily DO interpolation
- Hourly DO evaluation
- Thoughts on feeding into criteria evaluation
- Guidance documentation

Why a 4D interpolator?

- USEPA (2003) published Ambient Water Quality Criteria for Chesapeake Bay, need for criteria assessment protocols.
- 2007 STAC Workshop: Assessing the feasibility of developing a four-dimensional (4-D) interpolator for use in impaired waters listing assessment

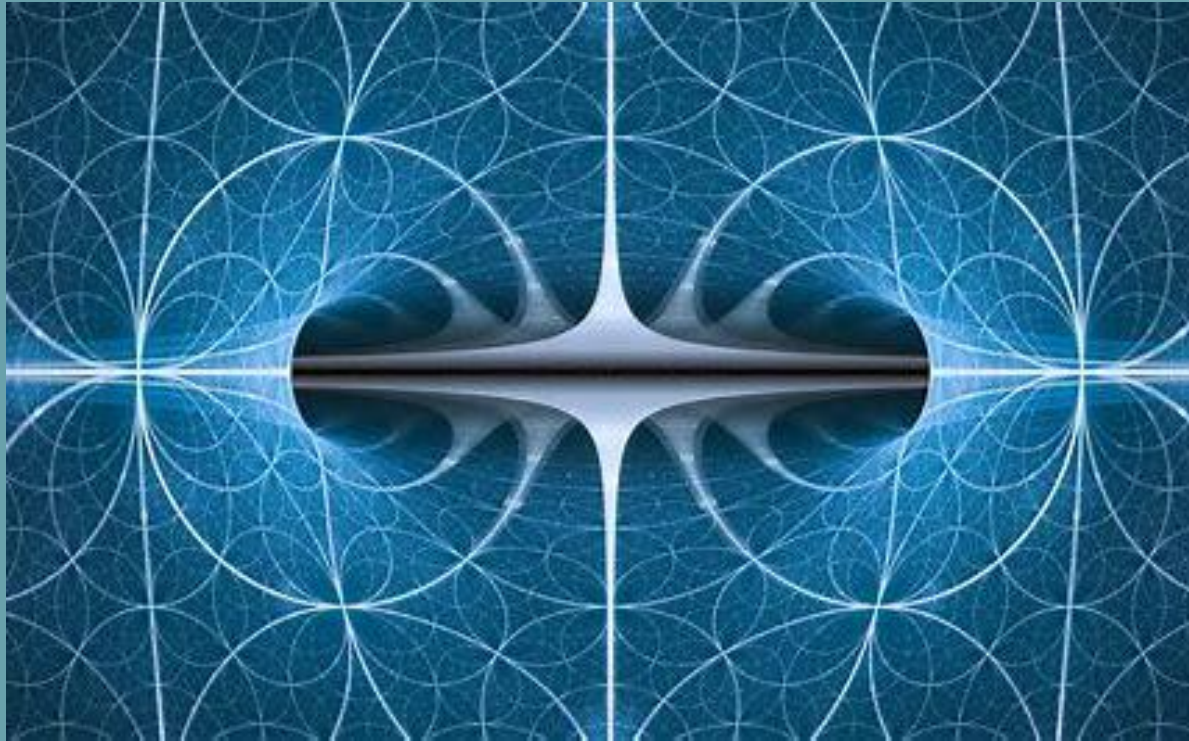
Why is the Chesapeake Bay Program interested in 4d interpolation?

- *To allow combining and integrating measurements from numerous disparate datasets*
- *To generate a more complete interpolation of available data in space and time.*
- *To improve the ability to evaluate water quality for the 303d listing process.*

2007 STAC Workshop consensus

- Sampling frequency (biweekly to monthly) is insufficient
- Spatial resolution of the existing Chesapeake Bay datasets is insufficient
 - *for successful extrapolation to four dimensions.*

“However, there is an on-going effort among Chesapeake Bay partners to acquire funding to deploy continuous monitoring buoys, which are equipped with vertical profilers in deep water areas of the Chesapeake Bay and tidal tributaries. If these efforts succeed, then the shortcomings of existing datasets will be greatly alleviated.”



17 YEARS
LATER...WHERE
ARE WE TODAY?

Welcome to the
Chesapeake Bay 4D
water quality interpolator
development program!

Requirements for an Updated Chesapeake Bay Interpolator

■ New features:

- *Temporal and spatial interpolation of water quality parameters in Chesapeake Bay*
- *Statistical estimates of uncertainty in the estimates*
- *Reproduce the short-term variability in the data*
- *Integrated vertical interpolation technique*

■ Features to retain and update:

- *Usability by partner analysts with automation for routine analyses*
- *Visualization of the results*
- *Analysis of dissolved oxygen, clarity, and chlorophyll a*
- *Post-processing to identify regions for each designated use (DU)*
- *Reasonable spatial extents for interpolation (e.g., not interpolating across land)*
- *Incorporation of new data streams & types, as available*

Bay Oxygen Research Group – Development Team

- Coordinators - Rebecca Murphy (UMCES) and Peter Tango (USGS)
- STAR Staffer - August Goldfischer (CRC)
- Development to-date - Elgin Perry (statistician); Jon Harcum and Erik Leppo (Tetra Tech); Angie Wei and Rebecca Murphy (UMCES/CBP)
- Advisory Team
 - Richard Tian (UMCES), Gary Shenk (USGS), Isabella Bertani (UMCES), Jim Hagy (EPA), Breck Sullivan (USGS), Kaylyn Gootman (EPA)
- ORISE Fellow – Wes Slaughter, U of MD Ph.D. candidate

4-D interpolator development timeline

| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|--|--|--|
| Draft January 2024 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Priority categories for 2024 are in red | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Calendar Year | 2022 | | | | 2023 | | | | 2024 | | | | 2025 | | | | 2026 | | | | 2027 | | | | | | |
| Calendar Quarter | Q1 | Q2 | Q3 | Q4 | Q1 | Q2 | Q3 | Q4 | Q1 | Q2 | Q3 | Q4 | Q1 | Q2 | Q3 | Q4 | Q1 | Q2 | Q3 | Q4 | Q1 | Q2 | Q3 | Q4 | | | |
| | Jan-Mar | Apr-Jun | Jul-Sep | Oct-Dec | Jan-Mar | Apr-Jun | Jul-Sep | Oct-Dec | Jan-Mar | Apr-Jun | Jul-Sep | Oct-Dec | Jan-Mar | Apr-Jun | Jul-Sep | Oct-Dec | Jan-Mar | Apr-Jun | Jul-Sep | Oct-Dec | Jan-Mar | Apr-Jun | Jul-Sep | Oct-Dec | | | |
| Project Year | Year 1 | | | | Year 2 | | | | Year 3 | | | | Year 4 | | | | Year 5 | | | | Year 6 | | | | | | |
| 1. Development-daily estimates | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 2. Develoment-hourly estimates | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 3. Development - shallow water | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 4. Development - GIS tasks | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 5. Development -combined daily & hourly | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 6. Development-criteria evaluation | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 7. Software | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 8. Documenting | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 9. Training | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 10. Year of Review | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 11. Operational | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Future meeting considerations

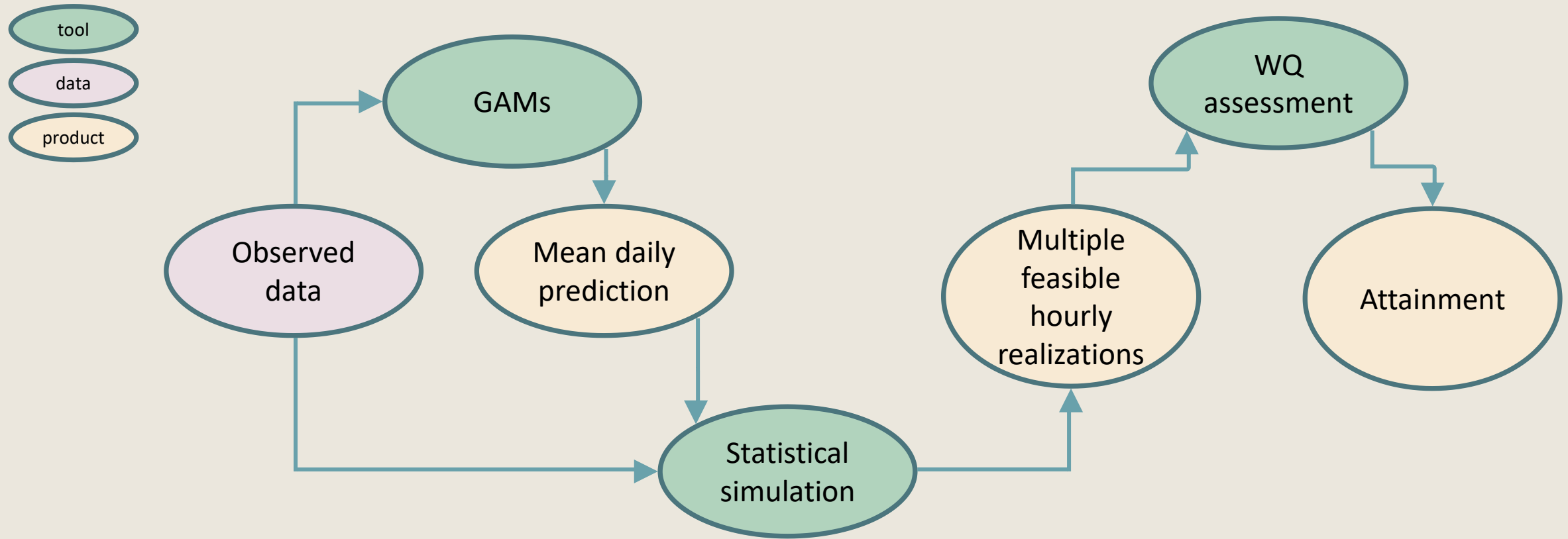
Through 2024-2025 development years:

1. **Bay Oxygen Research Group:** This large group of developers, stakeholders, future users, researchers
 - *We will meet approximately every 3 months*

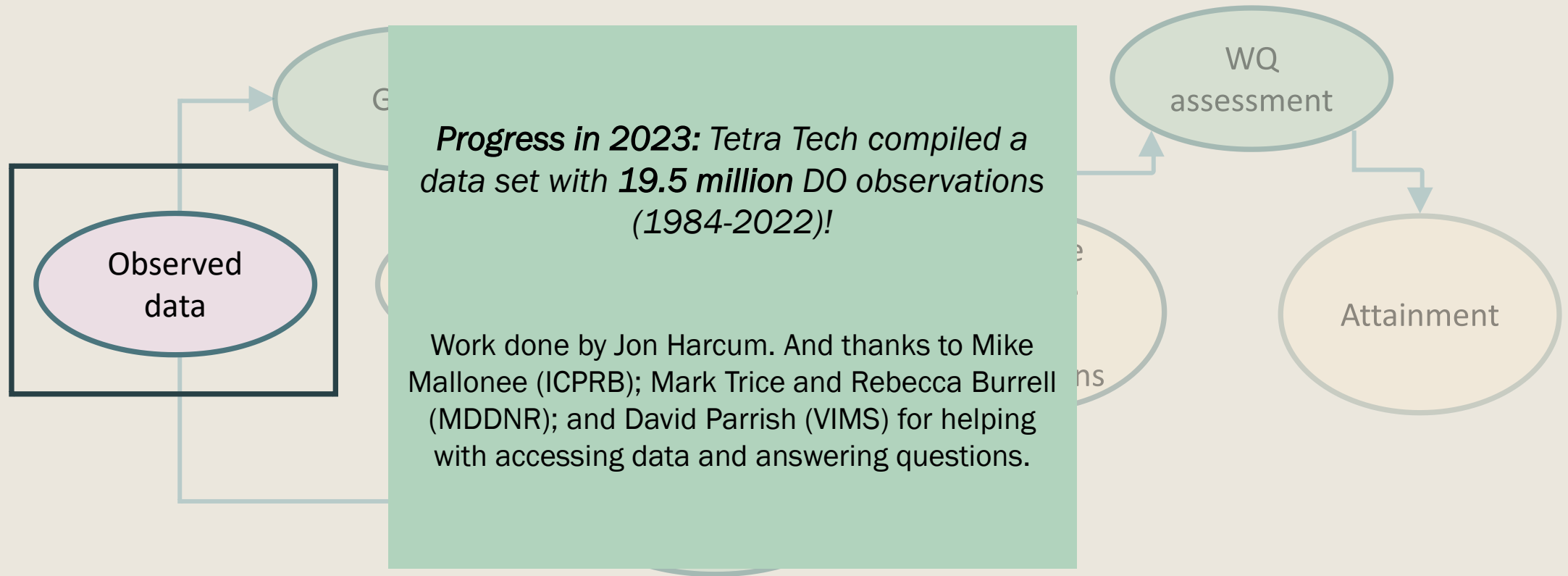
2. **Smaller BORG development group –**
 - *meets monthly 3rd Monday 12:00-12:50*
 - *Currently: CBP staff, contractors, EPA researchers*
 - *We'd like any future users to join the monthly meetings who want to know more details about the technical development*

Based on what you hear today, let us know if you or someone in your group should be on the Development Team invite. We'll send a reminder with the minutes.

WQ Assessment with 4D interpolator

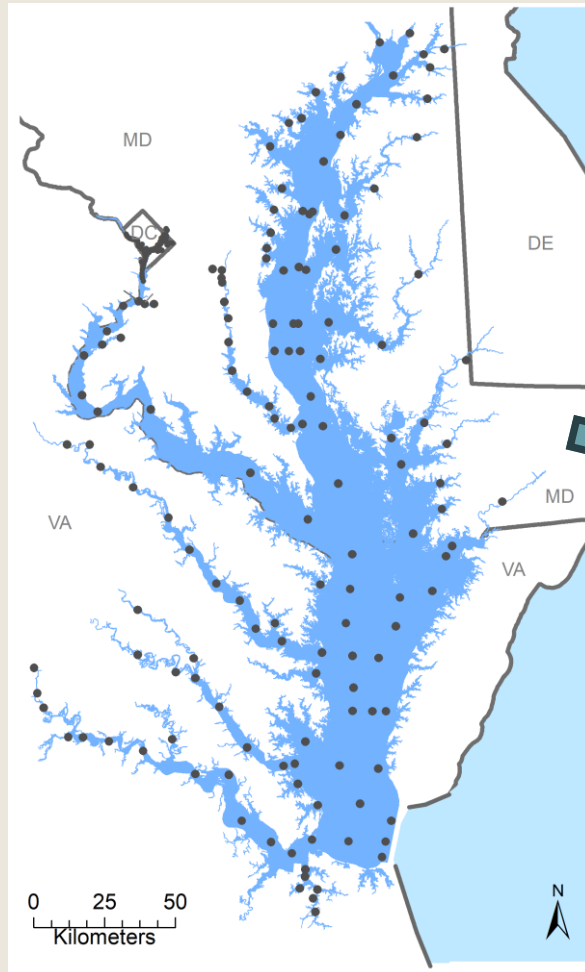


Full use of the data



Full use of the data

Bi-weekly long-term sampling
(DOEE, MDDNR, VADEQ, CBP)



Shallow water continuous monitoring
(MDDNR and VECOS)

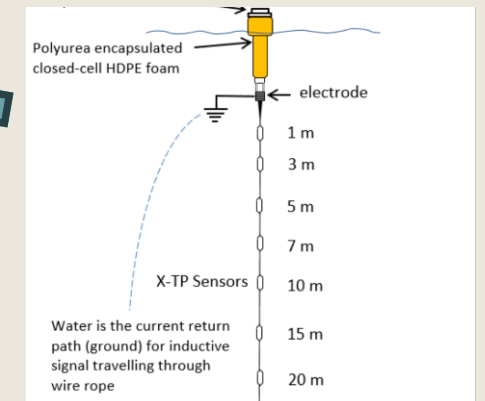


From <https://eyesonthebay.dnr.maryland.gov/>



From <http://vecos.vims.edu/>

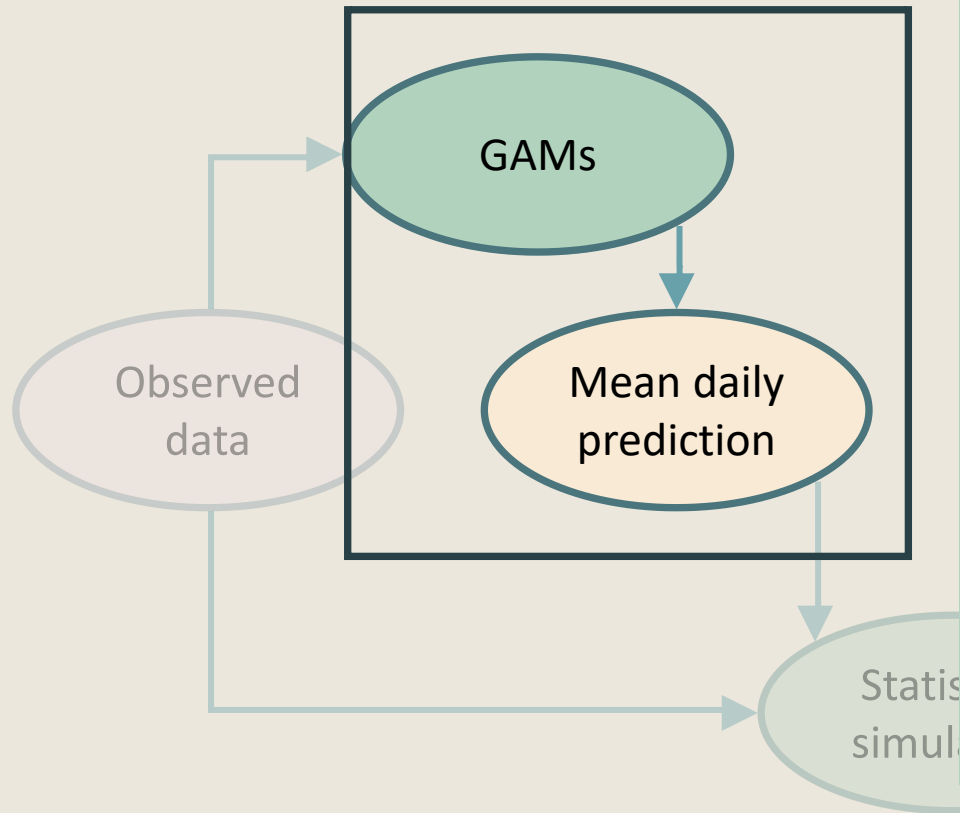
New continuous vertical
arrays (NOAA & CBP)



<https://www.chesapeakebay.net/who/group/hypoxia-collaborative-team>

And more for validation & testing: Citizen science, data collected with fisheries studies, and research data sets

Generalized additive model (GAM):

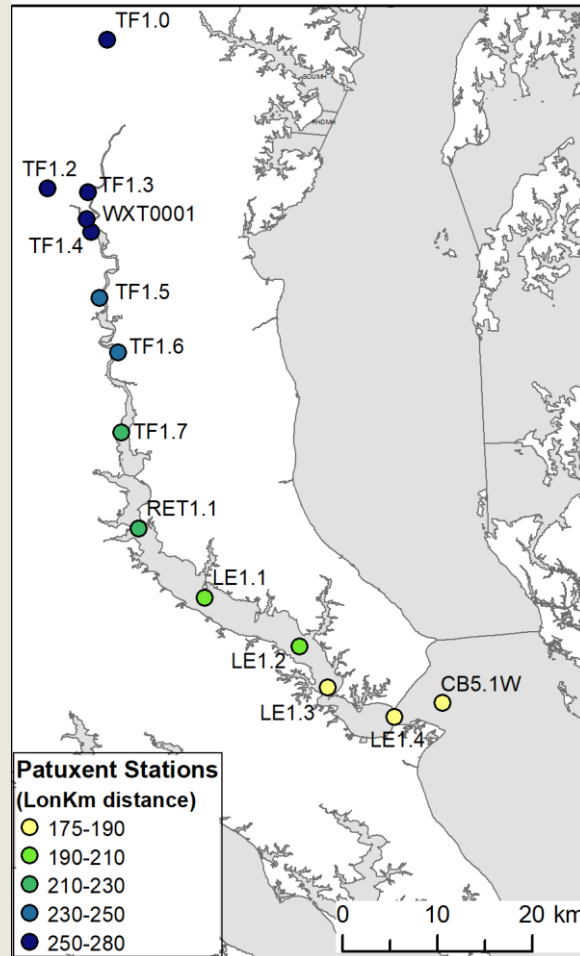


This is a statistical approach that uses data to generate a smooth function of DO in space and time.

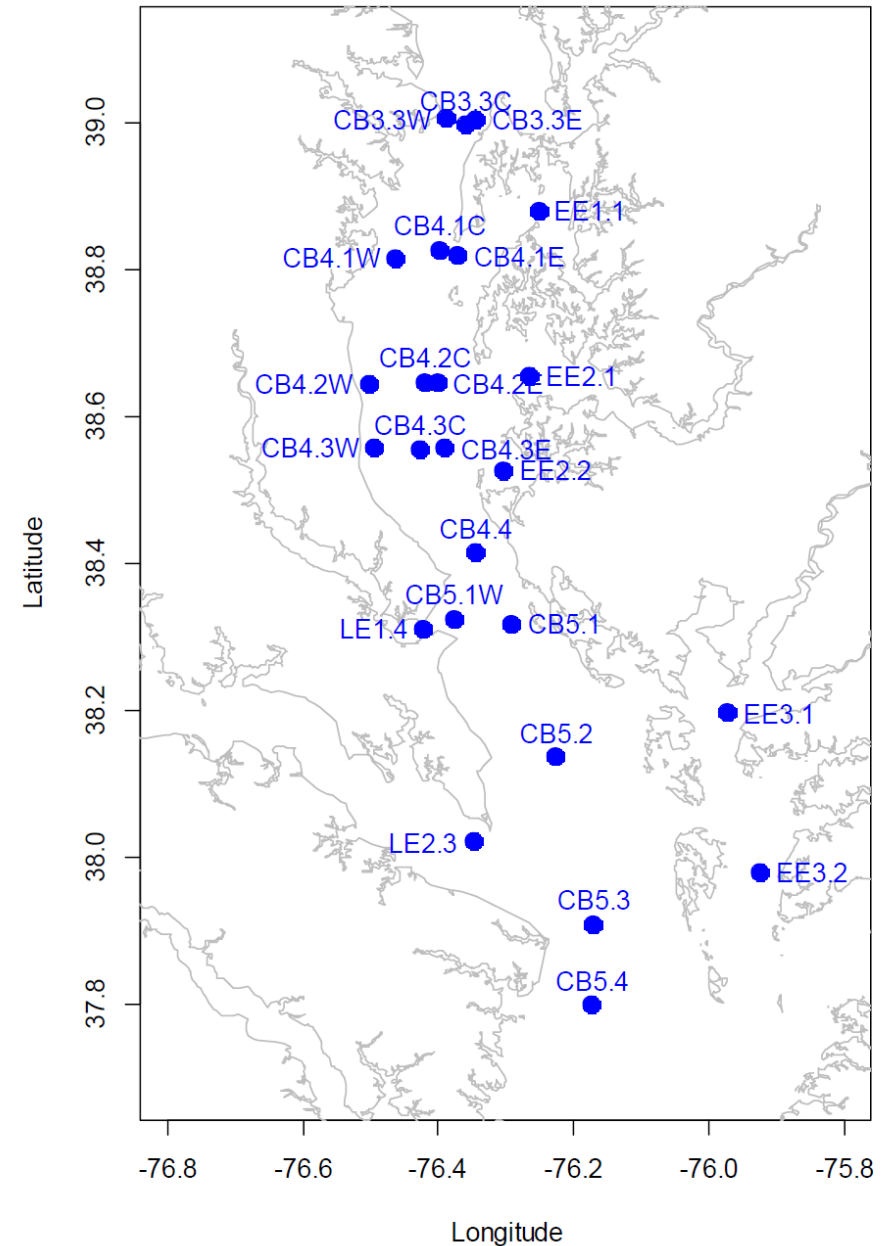
Multiple fitted GAMs will be used to estimate oxygen daily throughout the tidal waters:

$$DO \sim \text{smooth function}(\text{date}, x, y, z)$$

Daily mean: Test applications



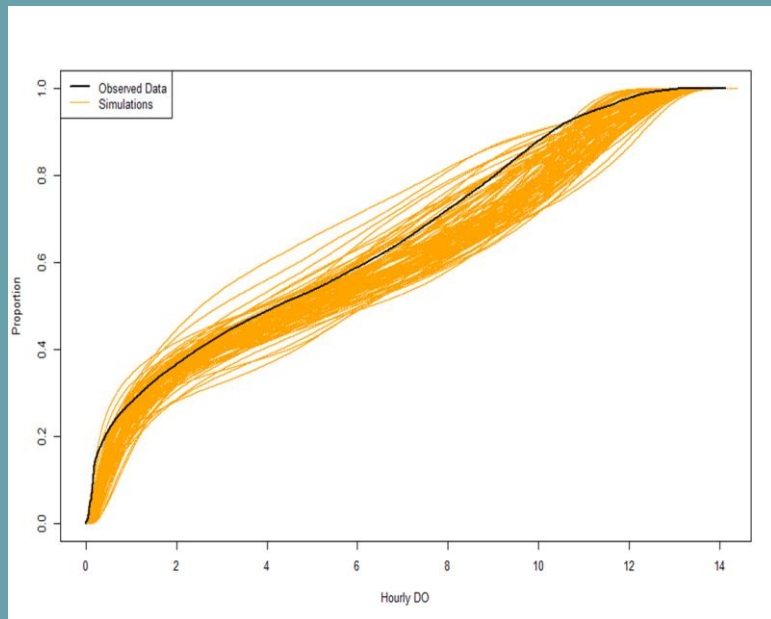
Map of stations in the test region.



Daily mean: Current work

Transforming DO

- We have challenges with both bounded DO values and a high frequency of low DO.
- Result is that a normal distribution assumption being a poor choice.
- Elgin Perry has been working on an approach (“beta-logit”) to address these.



Example analysis: CDF of 1 station's bottom data superimposed on CDFs of 100 simulations using resampling with Beta-Logit Transform

Bay-wide scale-up

Phase 1: Scale-up > Framework > Scale up Testing

Scale up

- Sub-estuary
- 'by' term

'do ~ s(date_d, k = 20, by = reg)'

| cbseg_92 | cb_303d_segment | cb4d_model | add_data | cb4d_by2 | cb4d_by3 | cb4d_by4 |
|----------|-------------------------------|------------|----------|----------|----------|----------|
| CB4MH | Middle Central Chesapeake Bay | CB1_3 | TRUE | | | |
| CB3MH | Upper Central Chesapeake Bay | CB1_3 | FALSE | | | |
| CB2OH | Upper Chesapeake Bay | CB1_3 | FALSE | | | |
| CB1TF | Northern Chesapeake Bay | CB1_3 | FALSE | | | |
| MAGMH | Magothy River | CB1_3 | FALSE | West | West | MAG |
| PATMH | Patapsco River | CB1_3 | FALSE | West | West | PAT |

Scale up

- GAM model

```
do ~ s(date_d, k = 20, by = reg) + #_002*  
  s(doy, bs = "cc", by = reg) + #_003*  
  s(depth, k = 4, by = reg) +  
  s(wb_lon_km, k = 4, by = reg) +  
  s(wb_lat_km, k = 4, by = reg) + #_004*  
  s(depth_b, k = 4, by = reg) +  
  ti(wb_lat_km, depth, wb_lon_km, k = c(3, 3, 3), by = reg) +  
  ti(depth, depth_b, k = c(3, 3), by = reg) + #_005*  
  ti(depth, date_d, by = reg) +  
  ti(wb_lon_km, date_d, by = reg) +  
  ti(wb_lat_km, date_d, by = reg) +  
  ti(depth_b, date_d, by = reg) +  
  ti(depth, do, bs = c("tp", "cc"), by = reg) +  
  ti(wb_lon_km, do, bs = c("tp", "cc"), by = reg) +  
  ti(wb_lat_km, do, bs = c("tp", "cc"), by = reg) +  
  ti(depth_b, do, bs = c("tp", "cc"), by = reg) +  
  ti(date_d, do, bs = c("tp", "cc"), by = reg)
```

#_002*
#_003*
#_004*
#_005*

#: 2023_10_23

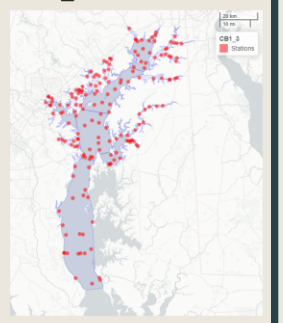
Phase 1: Data

- 1984-2022
- 844 stations
- 19.5M obs.

Filter

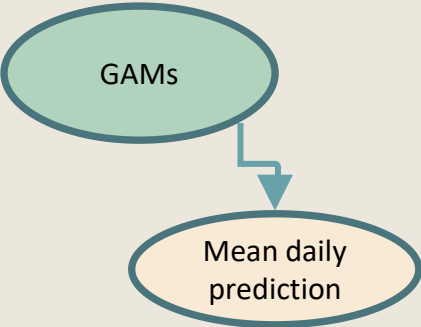
- Sub-estuary
- Date range

CB1_3

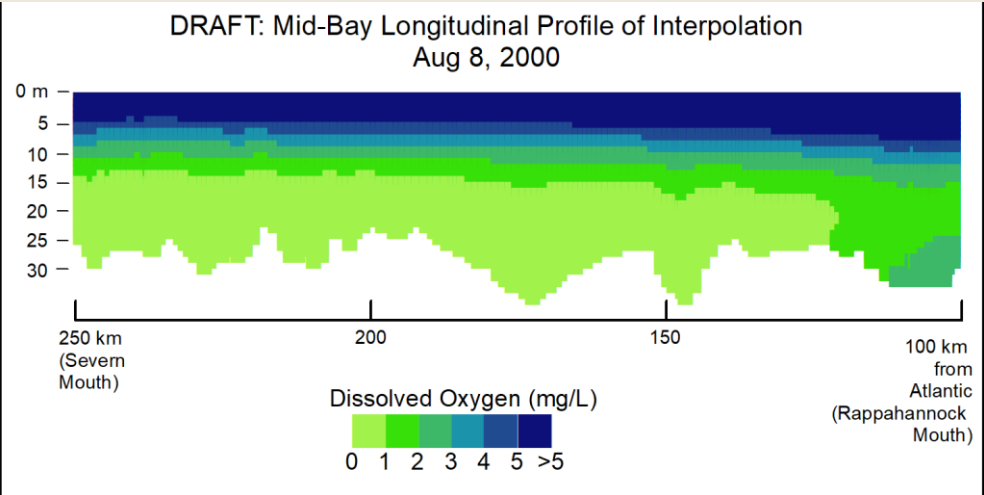


- Basic code has been built in R to apply GAM to each segment (Jon Harcum, Tetra Tech).
- Currently testing spatial and temporal options.
- Modifications will continue as we develop the method and link it to other components.

Where we are:



Example: GAM-based DO predictions along part of main channel



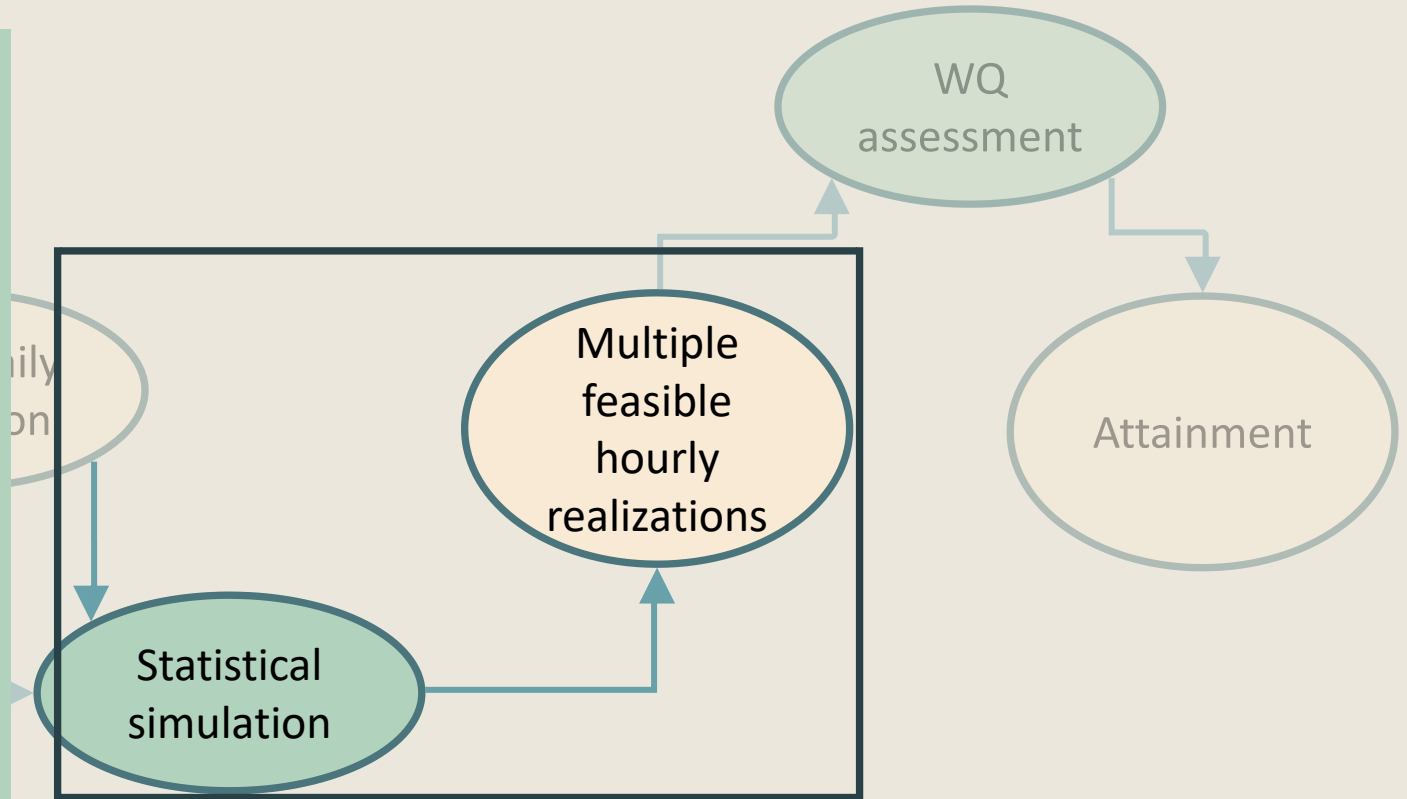
| Status and schedule | Development item |
|-----------------------|--|
| ✓ | Pilot application to mid-bay |
| ✓ | Test and modify approach in Patuxent |
| Current work, Q1 2024 | Identify appropriate statistical distribution for low DO |
| Current work, Q1 2024 | Build structure for bay-wide implementation |
| Q2 2024 | Test GAM approach on density/pycnocline |
| Q2 2024 | Draft of bay-wide implementation |
| Q3 2024 | Spatial validation of GAM estimates using additional data sets |
| Q3 2024 | Develop approach for daily error estimates |
| Q4 2024 to Q4 2025 | Continued updates as hourly approach is developed |

Q1=Jan-Mar, Q2=Apr-Jun, Q3=Jul-Sep, Q4=Oct-Dec

Hourly prediction

Involves evaluating all existing continuous monitoring DO data for within-day variability.

And then applying those patterns in a simulation to make hourly variability estimates bay-wide.



Method development with high frequency data



Have scanned through all common data from MDDNR and VECOS and recent NOAA vertical array data. Going to use this data to inform variability.

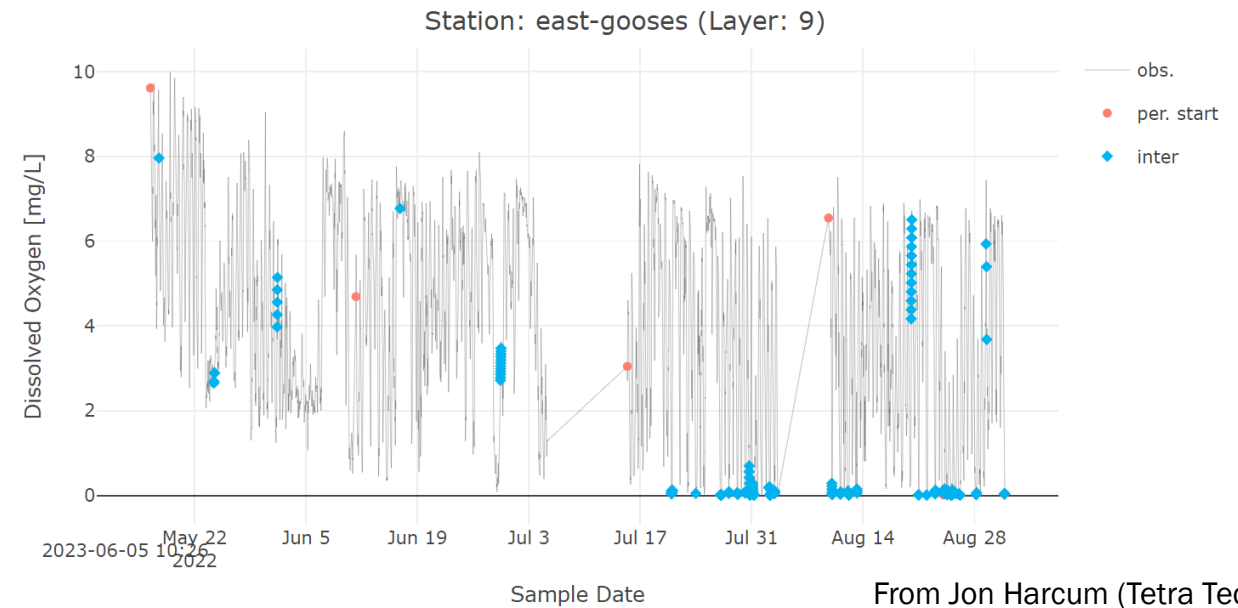
Work done by Erik Leppo and Jon Harcum at Tetra Tech.

Example: Current processing technique from Tetra Tech team using East Goose's array in 2022 at 9m

3.10.2 Period Summary

| Period Summary | | | | | | | | | | |
|----------------|---------------------|-----------------------|---------------|--------------|---------|---------------------|------------------------|----------|---------|--------------------|
| period | per_date_time_first | time_diff | rec_num_first | rec_num_last | rec_num | per_date_time_last | per_length | per_keep | num_int | rec_freq |
| 1 | 2022-05-16 11:00:00 | | 1 | 3659 | 3659 | 2022-06-10 20:40:00 | 2194800s (~3.63 weeks) | TRUE | 10 | 600s (~10 minutes) |
| 2 | 2022-06-11 06:50:00 | 36600s (~10.17 hours) | 3660 | 7123 | 3464 | 2022-07-05 08:00:00 | 2077800s (~3.44 weeks) | TRUE | 13 | 600s (~10 minutes) |
| 3 | 2022-07-15 09:10:00 | 868200s (~1.44 weeks) | 7124 | 9842 | 2719 | 2022-08-03 06:10:00 | 1630800s (~2.7 weeks) | TRUE | 98 | 600s (~10 minutes) |
| 4 | 2022-08-09 15:50:00 | 553200s (~6.4 days) | 9843 | 11897 | 2055 | 2022-08-23 22:10:00 | 1232400s (~2.04 weeks) | TRUE | 57 | 600s (~10 minutes) |
| 5 | 2022-08-24 02:30:00 | 15600s (~4.33 hours) | 11898 | 12989 | 1092 | 2022-08-31 19:20:00 | 665400s (~1.1 weeks) | TRUE | 45 | 600s (~10 minutes) |

3.10.3 Final Values



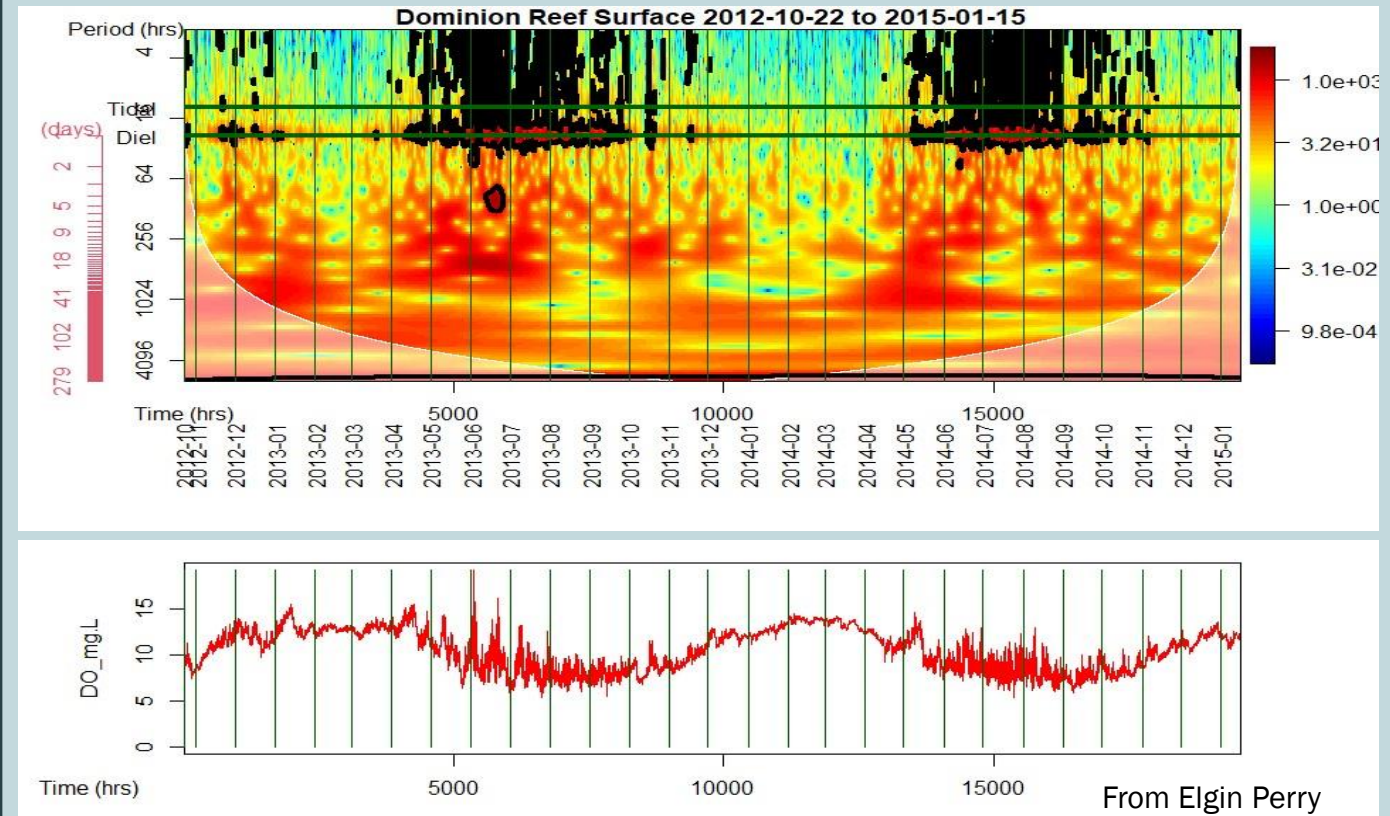
Method development with high frequency data

Conducted preliminary work with wavelet analysis on high frequency data and identifying sub-daily cycles.

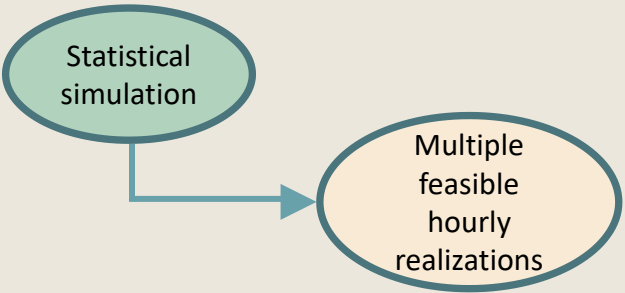
Processing of the results is underway.

Concept worked on by Elgin Perry with analysis of all high frequency by Erik Leppo and Jon Harcum.

Example: Wavelet analysis of gap-filled data to identify cycles



Where we are:



| Status and schedule | Development item |
|-----------------------|---|
| ✓ | Investigate method for hourly prediction at 1 to 2 continuous monitoring stations |
| ✓ | Standardize continuous data processing |
| Current work, Q2 2024 | Compare (and update) approach at additional continuous monitoring stations |
| Q3 2024 | Investigate approach for predicting hourly DO |
| Q4 2024 | Evaluate need for small scale spatial error term |
| Q4 2024 | Implement approach for modeling hourly predictions |
| Q4 2024 to Q4 2025 | Continued development and integration of daily & hourly tools |

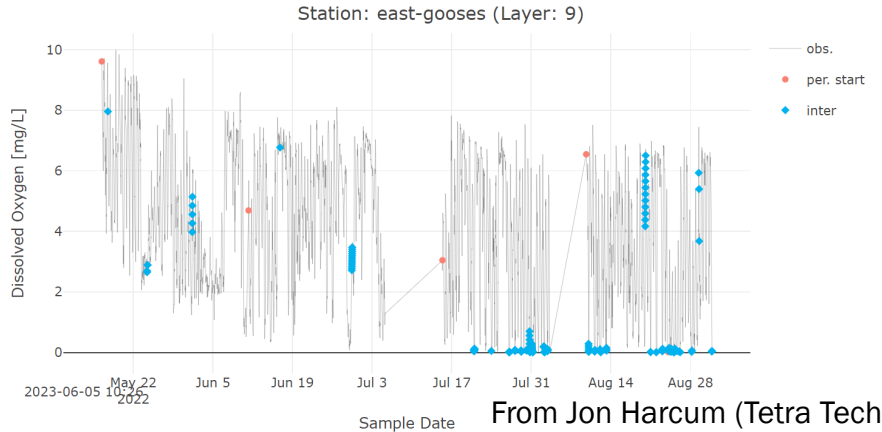
Q1=Jan-Mar, Q2=Apr-Jun, Q3=Jul-Sep, Q4=Oct-Dec

Example: High frequency data processing

3.10.2 Period Summary

| Period Summary | | | | | | | | | | |
|----------------|---------------------|-----------------------|---------|---------------|--------------|---------------------|------------------------|----------|---------|--------------------|
| period | per_date_time_first | time_diff | rec_num | first_rec_num | last_rec_num | per_date_time_last | per_length | per_keep | num_int | rec_freq |
| 1 | 2022-05-16 11:00:00 | | 1 | 3659 | 3659 | 2022-06-10 20:40:00 | 2194800s (~3.63 weeks) | TRUE | 10 | 600s (~10 minutes) |
| 2 | 2022-06-11 06:50:00 | 36600s (~10.17 hours) | 3660 | 7123 | 3464 | 2022-07-05 08:00:00 | 2077800s (~3.44 weeks) | TRUE | 13 | 600s (~10 minutes) |
| 3 | 2022-07-15 09:10:00 | 868200s (~1.44 weeks) | 7124 | 9842 | 2719 | 2022-08-03 06:10:00 | 1630800s (~2.7 weeks) | TRUE | 98 | 600s (~10 minutes) |
| 4 | 2022-08-09 15:50:00 | 553200s (~6.4 days) | 9843 | 11897 | 2055 | 2022-08-23 22:10:00 | 1232400s (~2.04 weeks) | TRUE | 57 | 600s (~10 minutes) |
| 5 | 2022-08-24 02:30:00 | 15600s (~4.33 hours) | 11898 | 12989 | 1092 | 2022-08-31 19:20:00 | 665400s (~1.1 weeks) | TRUE | 45 | 600s (~10 minutes) |

3.10.3 Final Values



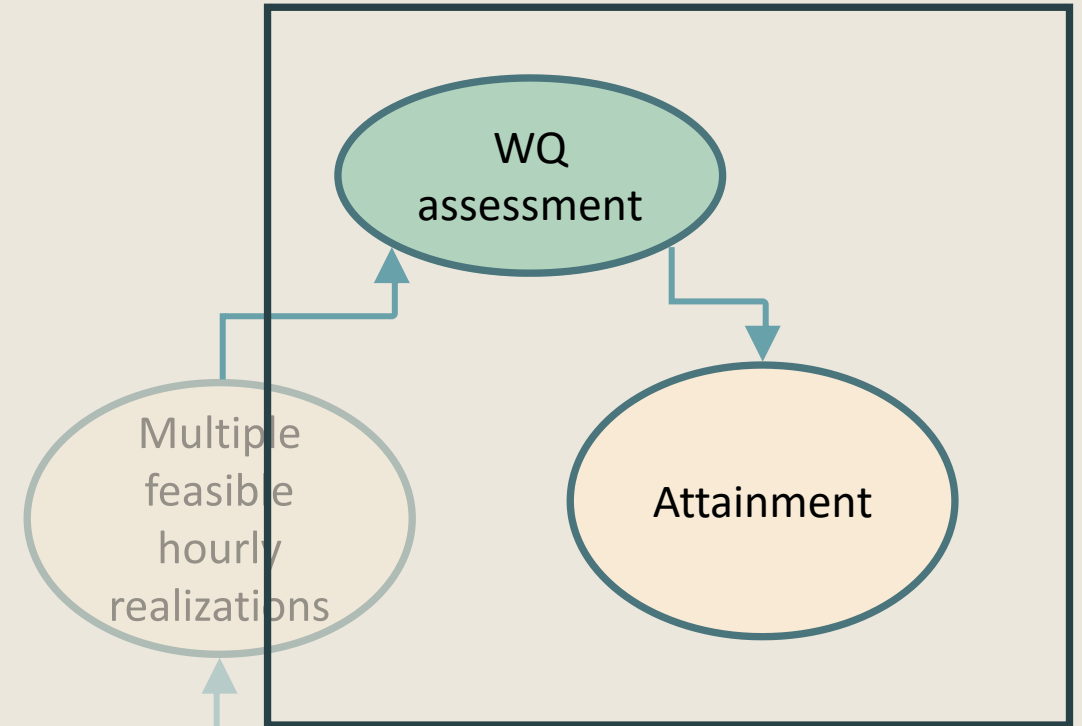
Link to WQ assessment

4d interpolator outputs will not be precise hourly DO estimates.

The goal is for them to provide a measure of segment impairment over the appropriate temporal and spatial scale for a criterion.

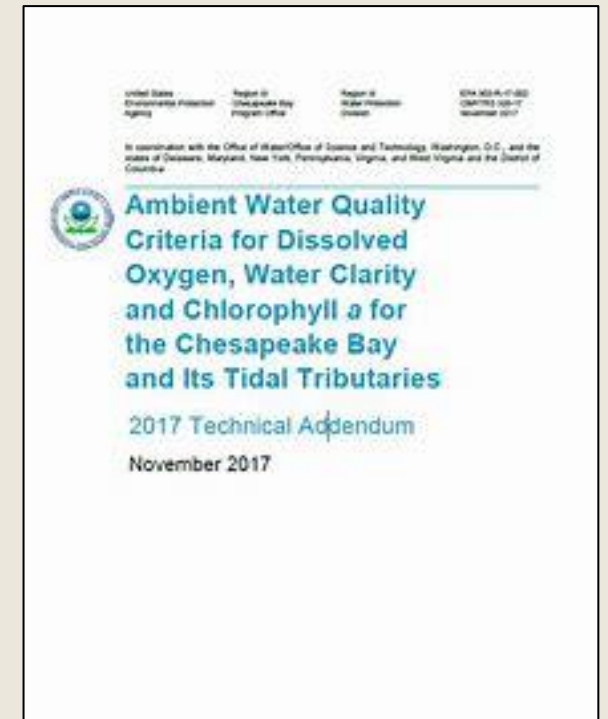
Timeline for development: ongoing through 2024-2025

Decision protocols: CAP WG, Modeling WG, EPA



Documentation

- Creating a new EPA Technical Document supporting
 - *Dissolved oxygen chapters needed include*
 - Development underpinnings of the new 4D tool
 - Updates on dissolved oxygen assessment protocols
- Outline shared among Monitoring team for review
- First outline review CAP WG March 7, 2024
- To be presented by CAP WG, future review by STAC, approval by EPA



Thank you!