Data usage in 4-dimensional (4-D) interpolator development and application

Rebecca Murphy (UMCES/CBP)
Bay Oxygen Research Large Group
May 20, 2024

Key input on presentation from:
Elgin Perry (statistics consultant) and Jon Harcum (Tetra Tech)

Purpose: Build a tool for more complete criteria assessment

DO criteria that currently can be evaluated with existing approaches and data

Table 1. Chesapeake Bay dissolved oxygen criteria.

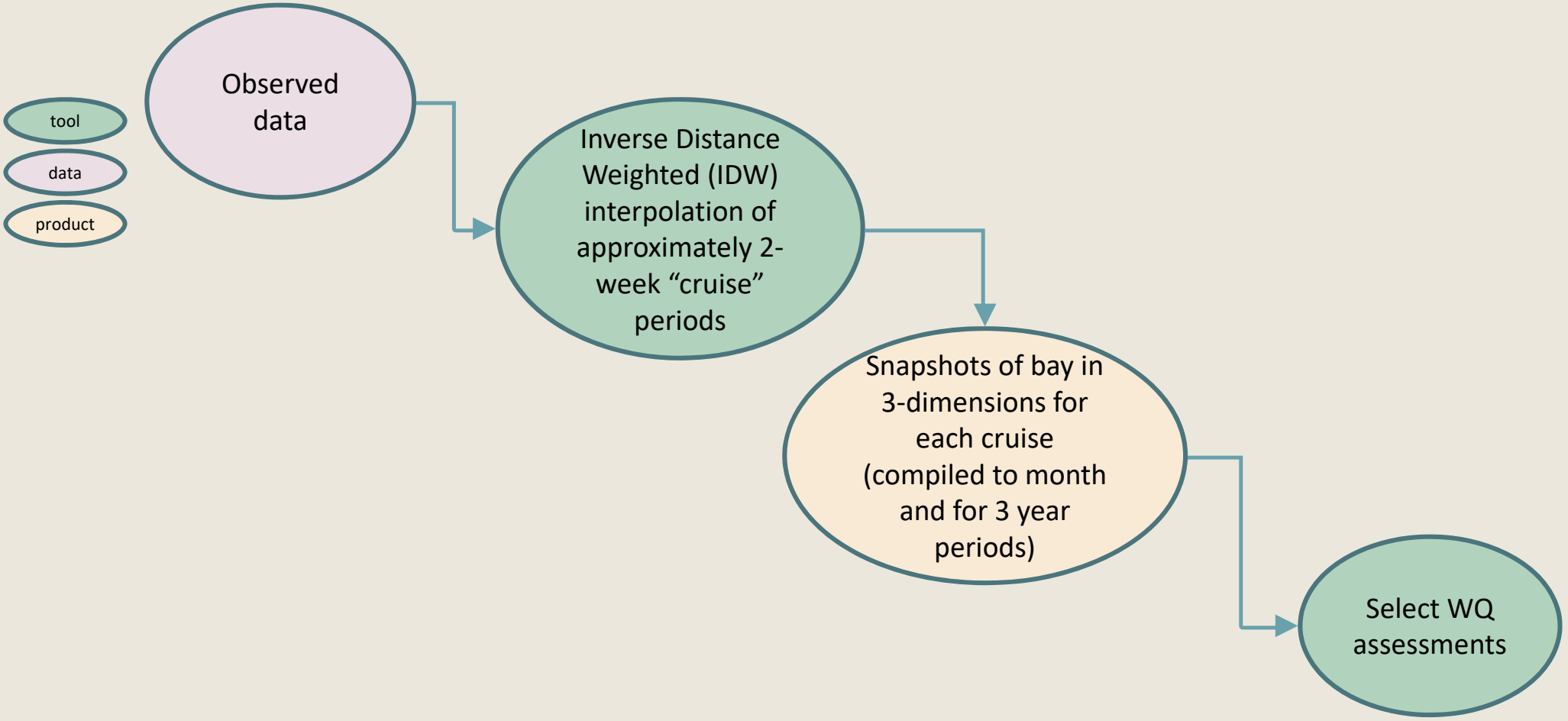
Designated Use	Criteria Concentration/Duration	Protection Provided	Temporal Application
Migratory fish spawning and nursery use *	7-day mean ≥ 6 mg liter^{-1} (tidal habitats with 0-0.5 ppt salinity)	Survival/growth of larval/juvenile tidal-fresh resident fish; protective of threatened/endangered species.	February 1 - May 31
	Instantaneous minimum ≥ 5 mg liter^{-1}	Survival and growth of larval/juvenile migratory fish; protective of threatened/endangered species.	
	Open-water fish and shellfish designated use criteria apply		June 1 - January 31
Shallow-water bay grass use	Open-water fish and shellfish designated use criteria apply		Year-round
Open-water fish and shellfish use	30-day mean ≥ 5.5 mg liter^{-1} (tidal habitats with 0-0.5 ppt salinity)	Growth of tidal-fresh juvenile and adult fish; protective of threatened/endangered species.	Year-round
	30-day mean ≥ 5 mg liter^{-1} (tidal habitats with >0.5 ppt salinity)	Growth of larval, juvenile and adult fish and shellfish; protective of threatened/endangered species.	
	7-day mean ≥ 4 mg liter^{-1}	Survival of open-water fish larvae.	
	Instantaneous minimum ≥ 3.2 mg liter^{-1}	Survival of threatened/endangered sturgeon species.[1]	
Deep-water seasonal fish and shellfish use	30-day mean ≥ 3 mg liter^{-1}	Survival and recruitment of bay anchovy eggs and larvae.	June 1 - September 30
	1-day mean ≥ 2.3 mg liter^{-1}	Survival of open-water juvenile and adult fish.	
	Instantaneous minimum ≥ 1.7 mg liter^{-1}	Survival of bay anchovy eggs and larvae.	
	Open-water fish and shellfish designated-use criteria apply		October 1 - May 31
Deep-channel seasonal refuge use	Instantaneous minimum ≥ 1 mg liter^{-1}	Survival of bottom-dwelling worms and clams.	June 1 - September 30
	Open-water fish and shellfish designated use criteria apply		October 1 - May 31

[1] At temperatures considered stressful to shortnose sturgeon (>29°C), dissolved oxygen concentrations above an instantaneous minimum of 4.3 mg liter^{-1} will protect survival of this listed sturgeon species.

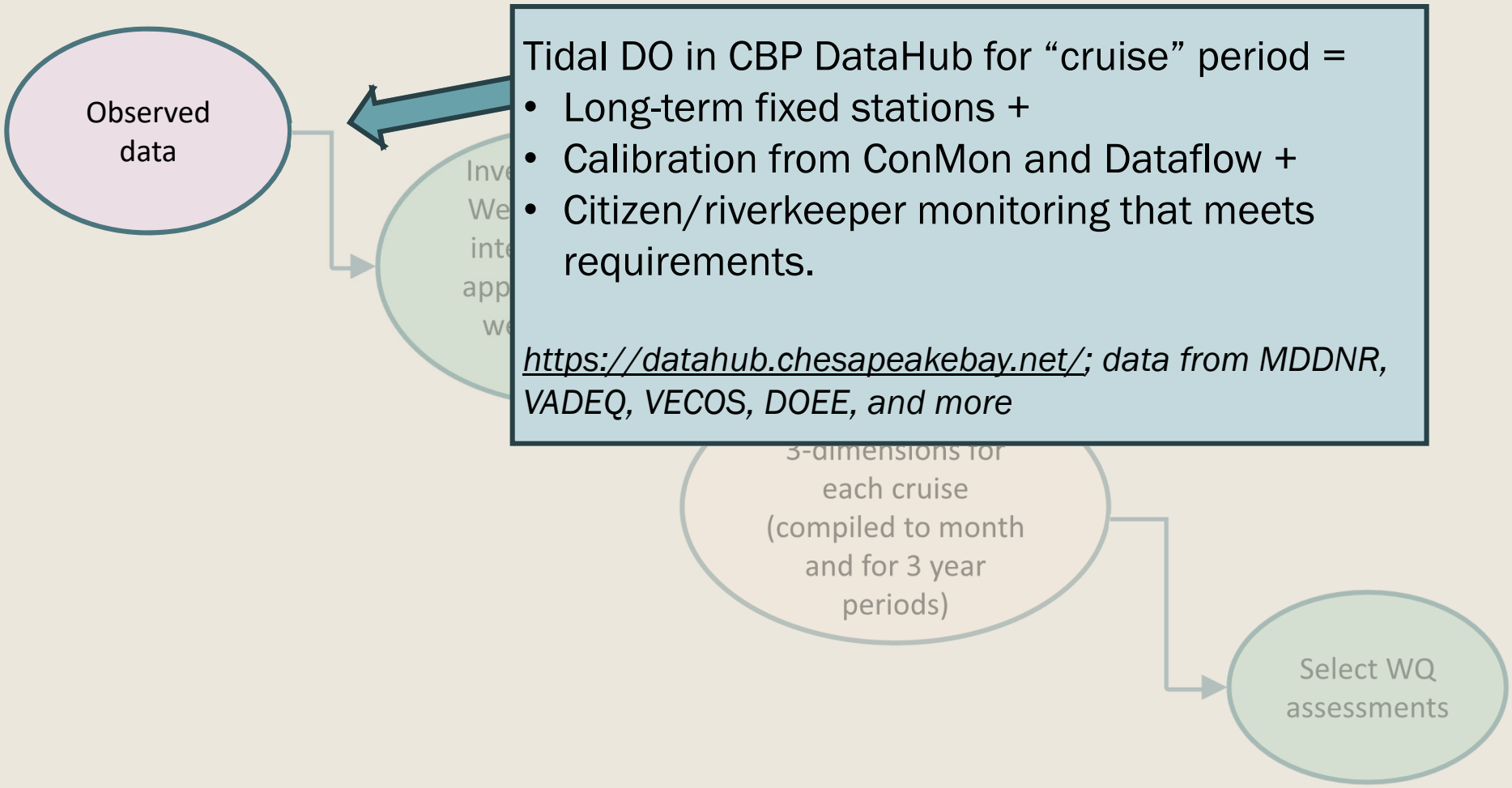
*Note a 30-day mean 6 mg/L MSN value is evaluated for purpose of the WQ indicator.

From EPA 2003 Ambient Water Quality Criteria

Current process (simplified)



Current process (simplified)



Goal for BORG team & 4-D tool

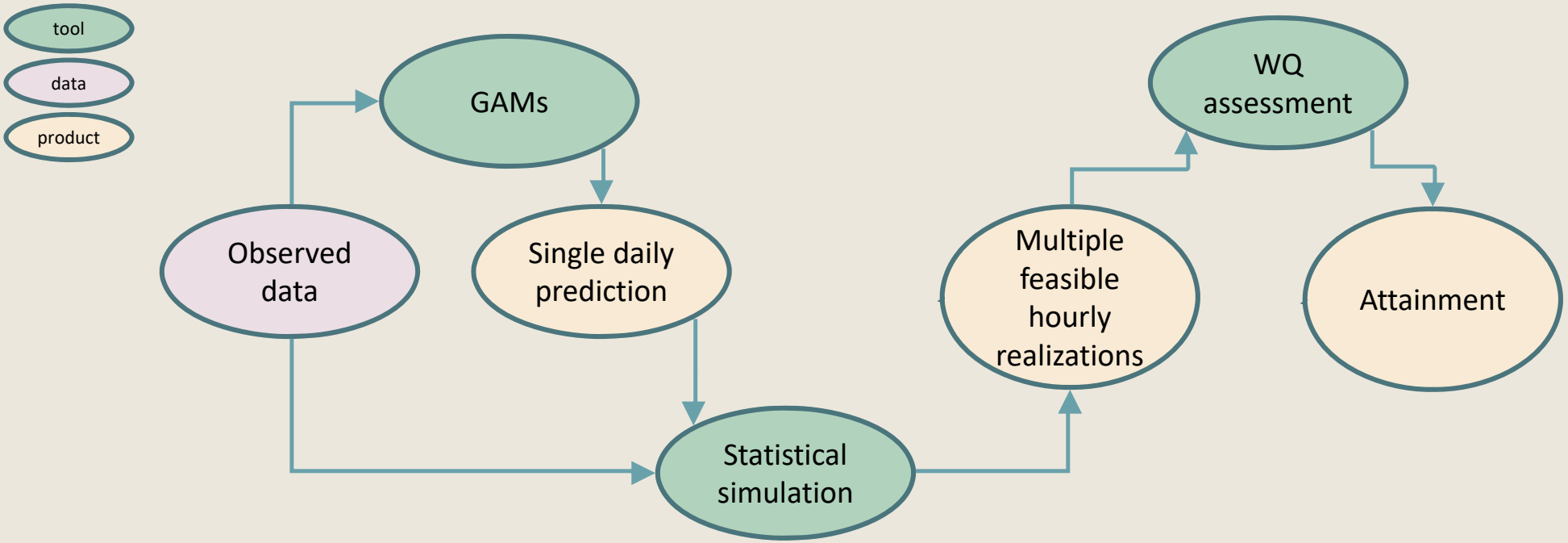
To develop a spatial-and-temporal interpolation tool for water quality monitoring data collected in the tidal waters of the Chesapeake Bay, thus enabling the evaluation of both long- and short-duration water quality criteria.

Specifically, the tool should be able to:

- *Interpolate observed dissolved oxygen in space and time ("4D"),**
- *Provide statistical estimates of uncertainty,*
- *Reproduce daily and hourly variability of the data, and*
- *Allow for post-processing of the interpolation output into designated uses (DU).*

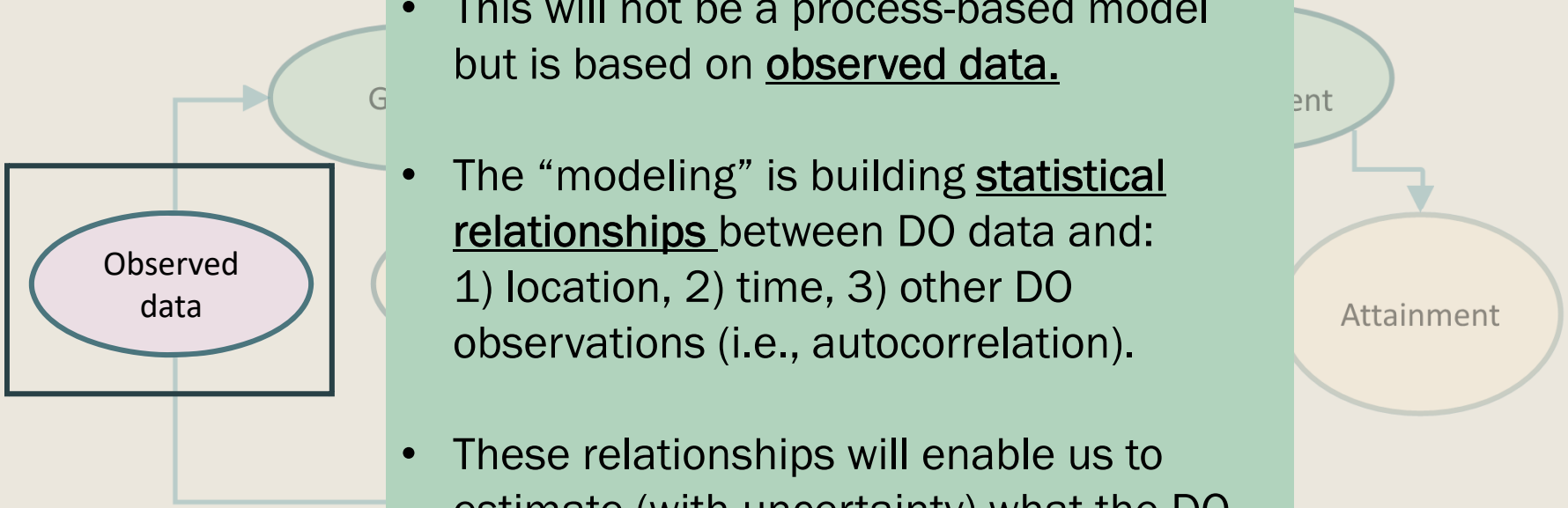
*Note: Focus on development so far has been on dissolved oxygen, but ultimately chlorophyll *a* and clarity may be evaluated as well.

WQ Assessment with 4D interpolator



Some general notes on data use

- The data used for the current interpolator will still be used.
- This will not be a process-based model but is based on **observed data.**
- The "modeling" is building **statistical relationships** between DO data and: 1) location, 2) time, 3) other DO observations (i.e., autocorrelation).
- These relationships will enable us to estimate (with uncertainty) what the DO was in a place and time when there was no data collected.



Uses of data, as envisioned to-date*

1. Explore patterns to inform development

2. Development & testing

3. Validation

4. Application

**Changes are very possible as we move forward in development and learn about additional data sets and how the models work.*

Uses of data

1. Explore patterns to inform development
   - Identify high frequency cycles (e.g., tidal & daily)
   - Parameterize temporal autocorrelation
   - Parameterize spatial autocorrelation
2. Development & testing
   - Fit and estimate daily DO
   - Fit and estimate hourly DO deviation
   - Fit and estimate pycnocline depths
3. Validation
   - Conduct hold-out tests
   - Conduct future validation tests using additional data sets
4. Application
   - Fit and estimate DO and pycnocline for specific 3 years

Uses of data

1. Explore patterns to inform development

2. Development & testing

3. Validation

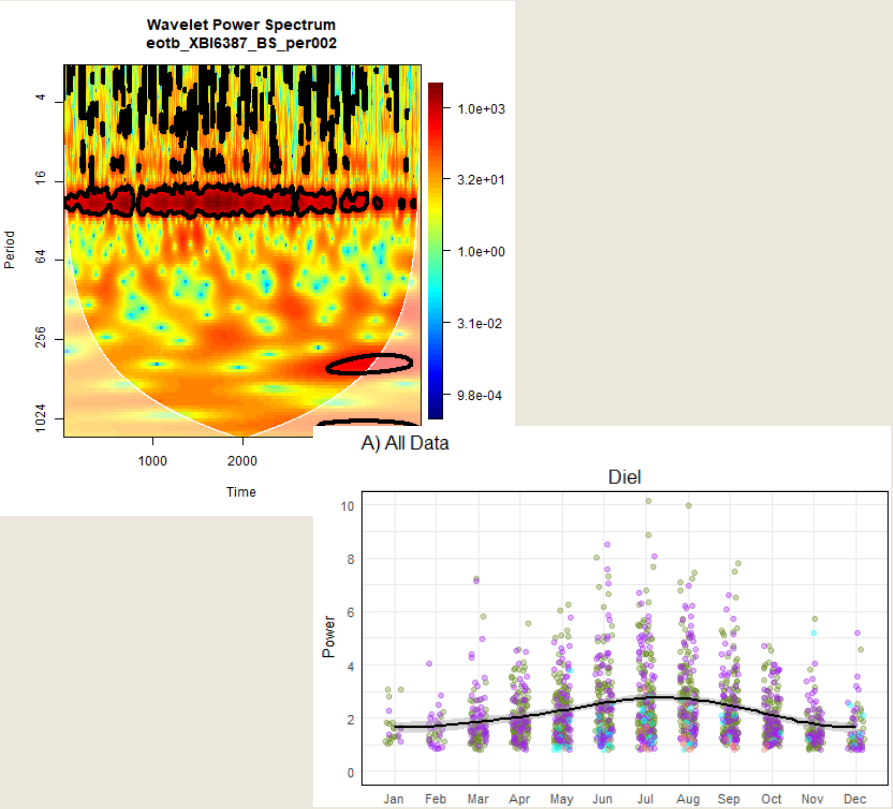
4. Application

- Identify high frequency cycles (e.g., tidal & daily)
- Parameterize temporal autocorrelation
- Parameterize spatial autocorrelation

1. Explore patterns to inform development

Identify High Frequency Cycles

When and where are there sub-daily cycles in DO?



From Jon Harcum and Erik Leppo, Tetra Tech

- EOTB: Eyes on the Bay (MDDNR)
  - *2001-2022*
  - *126 ConMon stations*
- VECOS: Virginia Estuarine and Coastal Observing System (VIMS)
  - *2003-2022*
  - *52 ConMon stations + 2 profilers*
- NOAA vertical arrays
  - *2022-2023*
  - *5 stations*
- Additional data used in Umbrella Criteria Analysis
  - *1987-1995*
  - *23 stations with sufficient record length (EMAP, plus some others)*

➢ NOTE: data >4 days duration needed for this analysis; and we can add additional data in other steps.

1. Explore patterns to inform development

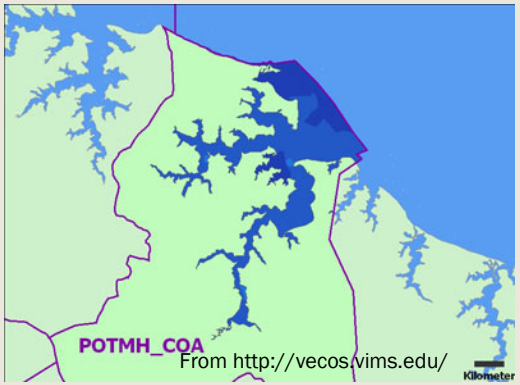
Parameterize Autocorrelation

Fill in the correlation matrix to represent similarity of data points close to each other

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

From http://vecos.vims.edu/

10/13/2009 VECOS



From http://vecos.vims.edu/

Temporal autocorrelation

- All ConMon and vertical array data (previous slide)
  - *Used in Arima analysis to get autocorrelation parameters (Elgin Perry, Jon Harcum)*

Spatial autocorrelation

- Dataflow – all years of data from MDDNR and VECOS
  - *Focus on horizontal dimension*
  - *Initial work is starting in Potomac fitting variograms (Wes Slaughter, UMd)*
  - *Future work: test and compare elsewhere*
- High frequency vertical arrays & fixed stations with multiple depths
  - *To identify spatial correlation in the vertical*
  - *Not started*

Uses of data

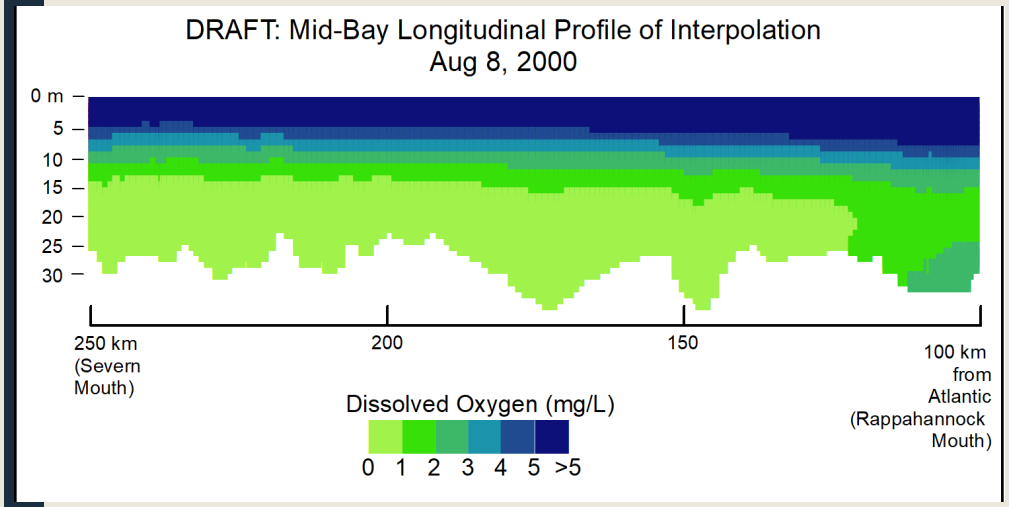
1. Explore patterns to inform development
2. Development & testing
3. Validation
4. Application

- Fit and estimate daily DO
- Fit and estimate hourly DO deviation
- Fit and estimate pycnocline depths

2. Development & testing

Estimate Daily DO (and pycnocline)

Uses a Generalized Additive Model (GAM) for mid-day estimates, everywhere



- DataHub DO, salinity and temperature data
  - *This is MD, VA, and DC fixed station data; ConMon & dataflow calibration data; and additional approved data*
- EOTB, VECOS, and NOAA high frequency ConMon and Vertical array data, daily subset
  - *Subset to a single daily value from each station to be consistent with sampling time of fixed station data*

2. Development & testing

Estimate the Hourly DO Deviations

Uses Fourier analysis to estimate the expected deviation from the mid-day DO at each hour of the day

Example: Fourier analysis with just daily cycle to fit hourly DO (DO_h)

$$DO_h = lc * h + sc * sin\left(\frac{2\pi * h}{24}\right) + cc * cos\left(\frac{2\pi * h}{24}\right) + \tau$$

coefficients

h = hour 1:24

- EOTB, VECOS, and NOAA high frequency ConMon and Vertical array data, hourly subset
  - *Subset to hourly only if collected at higher frequency*

Uses of data

1. Explore patterns to inform development

2. Development & testing

3. Validation

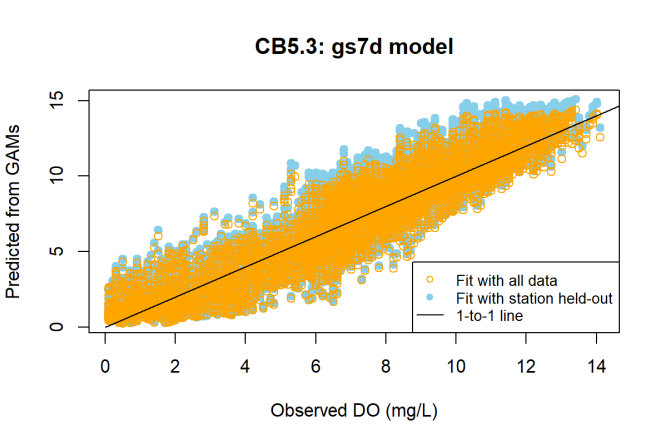
4. Application

- Conduct hold-out tests
- Conduct future validation tests using additional data sets

3. Validation

Evaluating the Tool

Validation tests



Hold out tests

- All same data as daily GAM fitting

Additional validation tests

Not started. The extent of this work to be determined by time, funding, and other evaluations of performance

- Water quality collected along with fisheries data
- Other citizen science data
- **Location-specific:** Perhaps hold-out tests in some high-density sampling program segments (e.g., MD's work in Fishing Bay)

Uses of data

1. Explore patterns to inform development

2. Development & testing

3. Validation

4. Application

- Fit and estimate DO and pycnocline for specific 3 years

Interpolation for 3-year Assessment Periods

- **DataHub data**: for the same time period, as done currently
- **EOTB, VECOS, and NOAA high frequency ConMon and vertical array data,** used in 2 ways:
  - *Subset to daily for daily component, and*
  - *Used at hourly level for hourly deviation estimates*.*

- *We may use some high frequency data outside the 3 years for predicting hourly deviations.
- Decisions on when & where to model tidal and daily cycles and autocorrelation parameters will be built from more than just 3 years of data.

Other possible data sets & uses

- Shallow water explorations
  - *NOAA daily products for salinity and temperature based on satellite data (Wes Slaughter's work)*
- Pycnocline:
  - *Possibly use freshwater flow measured at USGS gages*
  - *Maybe meteorological data as well*
- Bathymetry:
  - *We need to know the bottom depth everywhere, and want to use any updated information if possible.*

Sources and thank you!

- Data compilation thanks: Mike Mallonee (ICPRB); Mark Trice and Rebecca Burrell (MDDNR); David Parrish (VIMS) and Carl Friedrichs (VIMS); Jay Lazar and CJ Pellerin (NOAA); Erik Leppo (Tetra Tech).
- MDDNR EOTB: https://eyesonthebay.dnr.maryland.gov/
- VECOS: http://vecos.vims.edu/
- CBP DataHub: https://datahub.chesapeakebay.net/
- Umbrella Criteria Analysis data citations:
  - *Tidal Monitoring and Assessment Workgroup. (2011). Retrieved from https://d18lev1ok5leia.cloudfront.net/chesapeakebay/documents/2011_umbrcrit_wrkshp_rprt_dec2011pt.pdf*
  - *Tidal Monitoring and Analysis Workgroup (TMAW). (2012). STAC Publication 12-02. Retrieved from https://www.chesapeake.org/pubs/289_UmbrellaCriterionActionTeamTidalMonitoringandAnalysisWorkgroup2012.pdf*