Leveraging the power of big data to model landscape-scale metabolism in Chesapeake Bay

Update to the Modeling Workgroup January 2025
1/8/2025

Project background

- Funded by Merrill Foundation through Chesapeake Global Collaboratory (CGC)
- Responsive to new initiative - CGC - with emphasis on stakeholder engagement

Project Goal

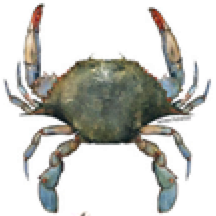
- Leverage 1) availability of WQ predictions from CBP CH3D-ICM model, 2) previously published metabolic models → model metabolic habitat conditions for organisms under management-relevant scenarios

Focus on higher trophic levels (HTLs)

- Linking HTLs to water quality a focus of regional management agencies
- Historical challenges to linking HTLs to water quality
- Metabolic approaches may provide more nuanced response than abundance/biomass

Target species & CBP Model scenarios

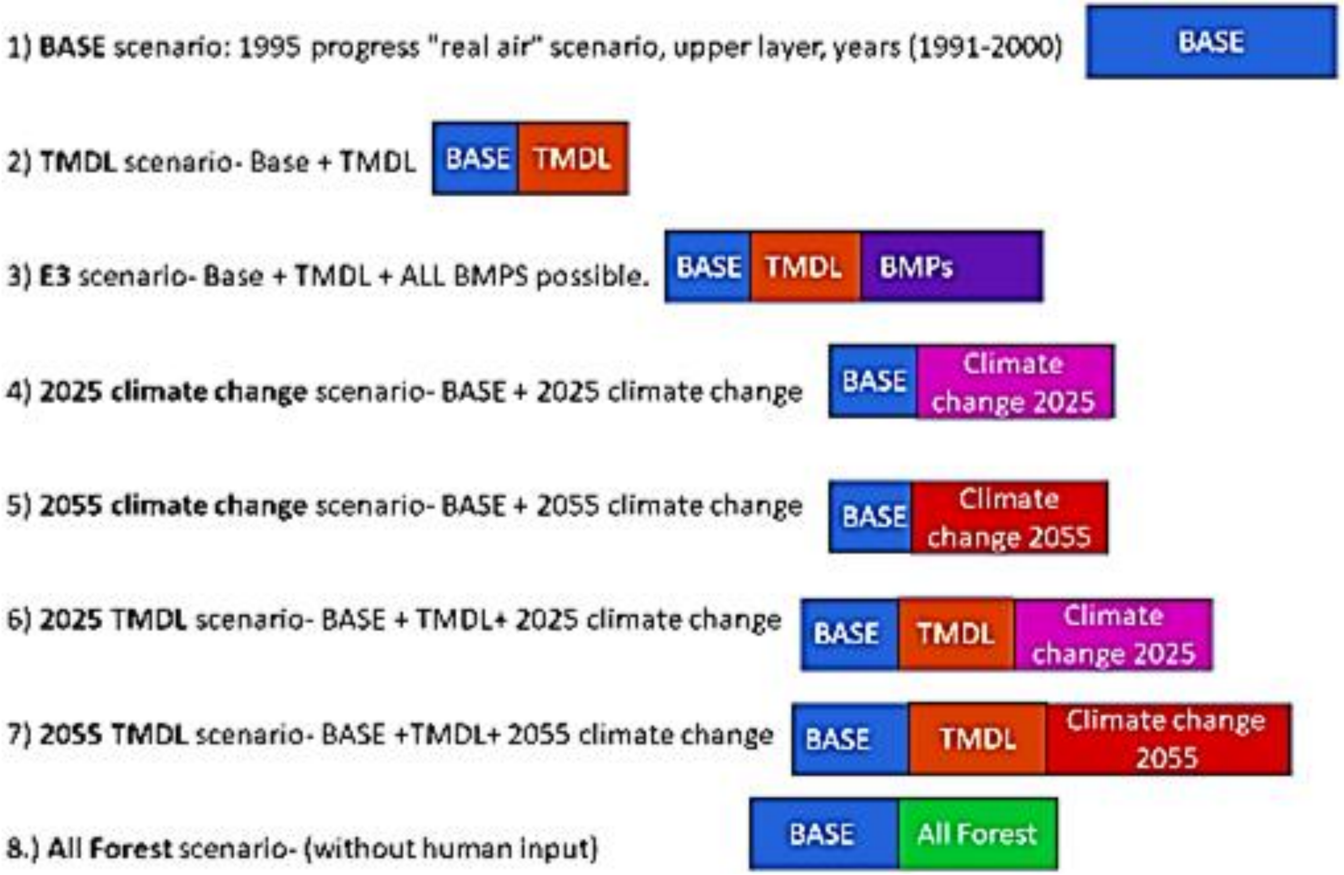
Eastern oyster (*Crassostrea virginica*): iconic species, ecosystem engineer, shallow habitat



Blue crab (*Callinectes sapidus*): cosmopolitan & iconic species

White perch (*Morone americana*): tidal fresh to oligohaline fishery species, highly abundant in tributaries

Blue catfish (*Ictalurus furcatus*): invasive species, rapidly naturalizing with a burgeoning fishery

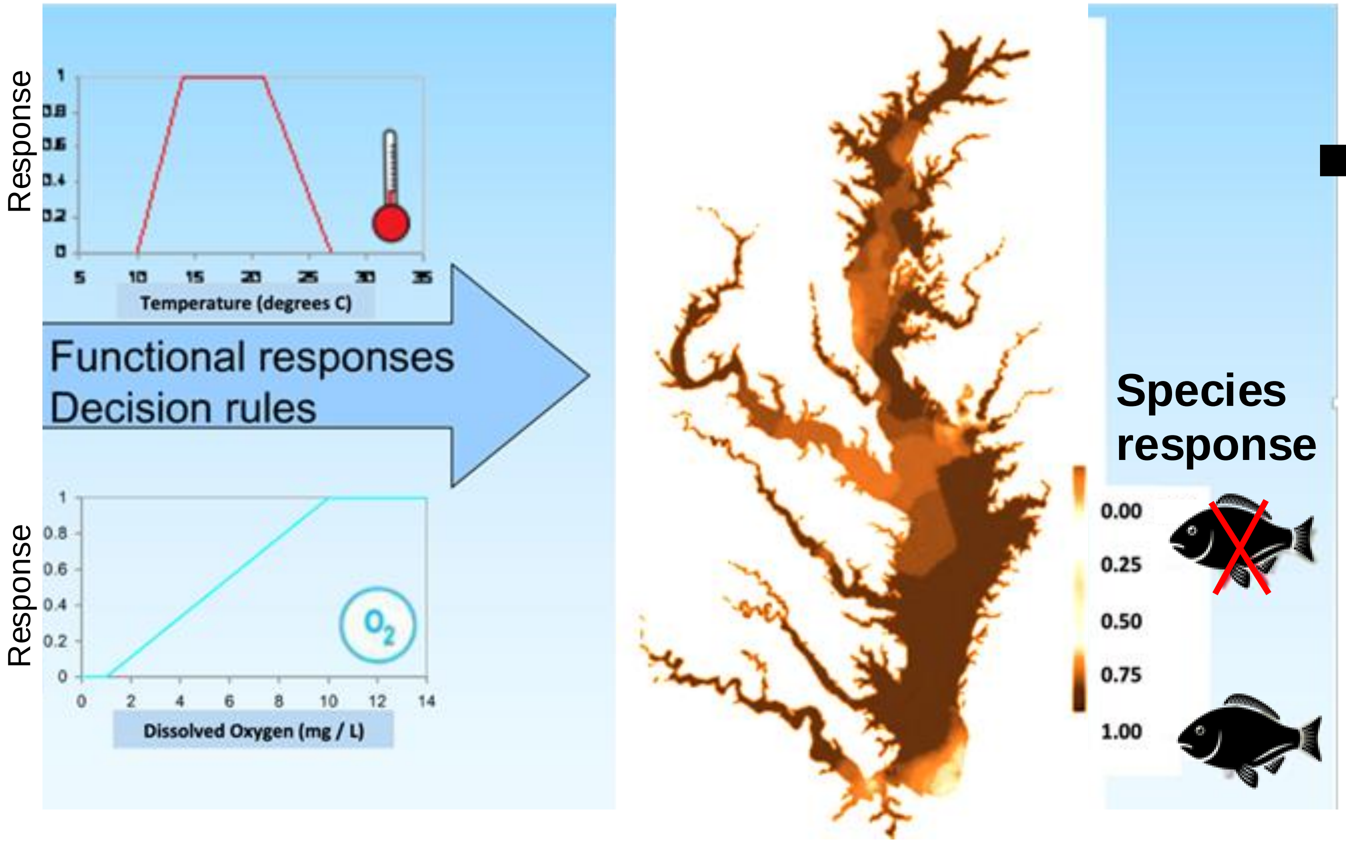


Generalized approach: Metaboscape development and analysis

Scenario 1

Temp

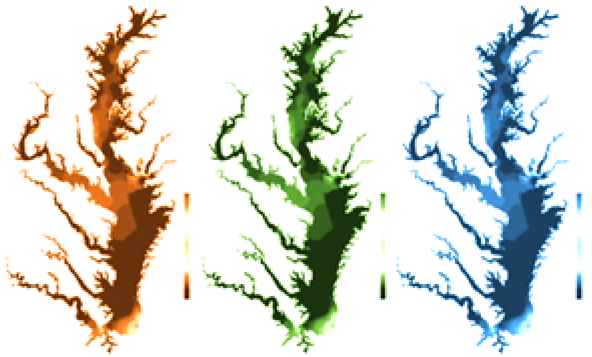
O_2



Response ~ α(Temp) + β(O2)

Compare and analyze across scenarios

Scenario 1 *Scenario 2* *Scenario 3*

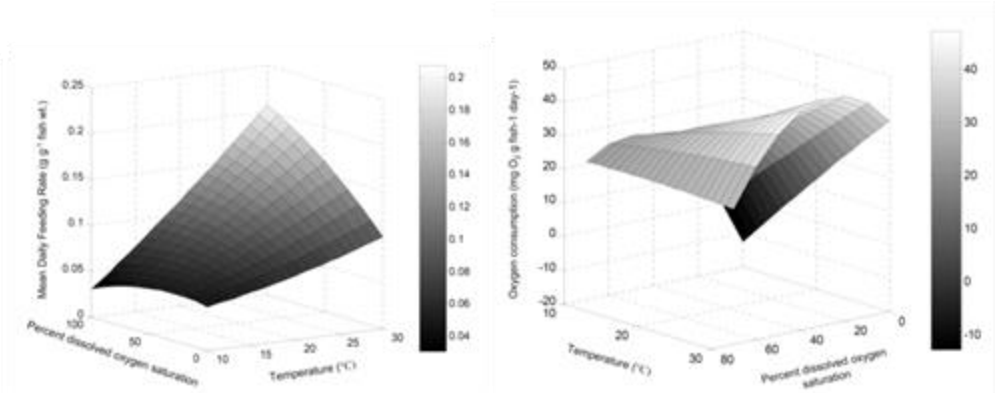


- Differences among scenarios
- Uncertainty
- Identifying knowledge gaps

Initial implementation: Metaboscape development for White Perch

White perch (*Morone americana*)

- Among top 5 most valuable finfish commercial fisheries in Chesapeake Bay
- Commercially caught using fyke/pound/gill nets, hook and line (1-2 M lbs yr^{-1})
- Important recreational fishery (~1.5 M lbs yr^{-1})



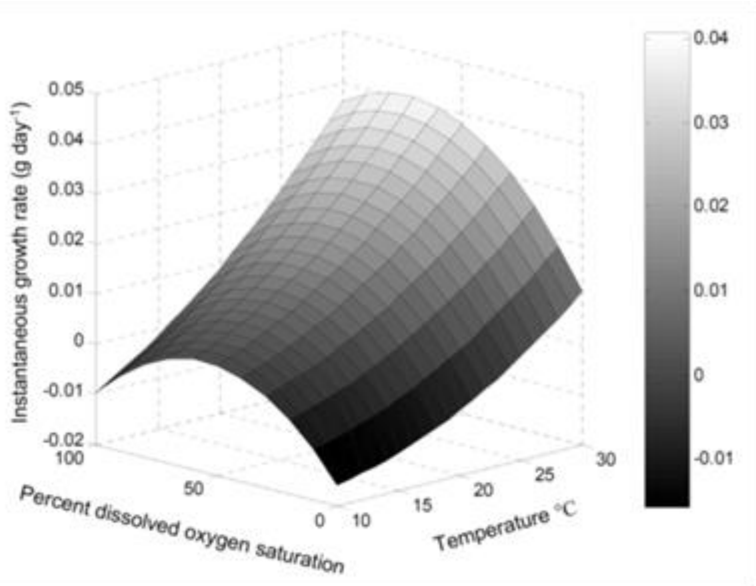
Daily feeding rate

Oxygen consumption rate

- Bioenergetic study of young-of-year during summer juvenile growth period
- Predictor variables
    - Water temperature
    - Salinity (not significant)
    - Dissolved oxygen
- Response variables:
    - Instantaneous growth rate (g day^{-1})
    - Daily feeding rate (g g^{-1} fish wt)
    - Respiration rate (mgO_2 g $fish^{-1}$ day^{-1})

Temperatre	Dissolved Oxygen	Salinity	Replicate Number
	2006		
6°C	70%	4	3
12°C	40%	4	3
12°C	70%	1	3
12°C	70%	4	3
12°C	70%	8	3
20°C	20%	4	3
20°C	40%	4	3
20°C	70%	4	3
20°C	70%	16	3
	2007		
20°C	70%	4	6
20°C	70%	1	3
28°C	20%	4	3
28°C	40%	4	3
28°C	70%	4	3

$$y = \beta_o + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1^2 + \beta_4 x_2^2 + \beta_5 x_1 x_2 + \varepsilon$$



Instantaneous growth rate (IGR)

Hanks and Secor 20011, Hanks 2009

Provided CBP Phase 6 Output:

Follow us, GitHub CGC-UMCES

- **Format**: Unformatted binary file (~47 GB)
- **Details**: 56,920 total cells, 19 layers, 10 years, and 37 variables
- **Access**: Requires Fortran code and several support files to read and explain the data.

Python Implementation: We developed a Python-based "Decipher" tool, adapted from Richard Tian's Fortran code. No additional support files required, simplified workflow with robust handling of binary data.

- **Performance**:
  - Optimized for shared-memory machines
  - Processing speed: (e.g., 32 cores, Intel Xeon Gold 6148 @ 2.40GHz). ~10 seconds per day. One year of data deciphered in ~1 hour.
  - Supports chunk-based processing for efficient memory usage!
- **Processed Output:**
- **Format**: Deciphered and saved in previous described structured NetCDF format; Missing values filled with NaN, which can be natively compressed, reducing the size from ~47 GB to ~29 GB.
- **Customizable Saving Options**

➢ **Variable-based**: Single variable for all 10 years per NetCDF file (~755 MB/file); total size for 37 variable ~29 GB.

➢ **Yearly-based:** All 37 variables for a single year per NetCDF file (~2.9 GB/file); total size for 10 years: ~29 GB.

➢ **Customized**: Flexible saving options

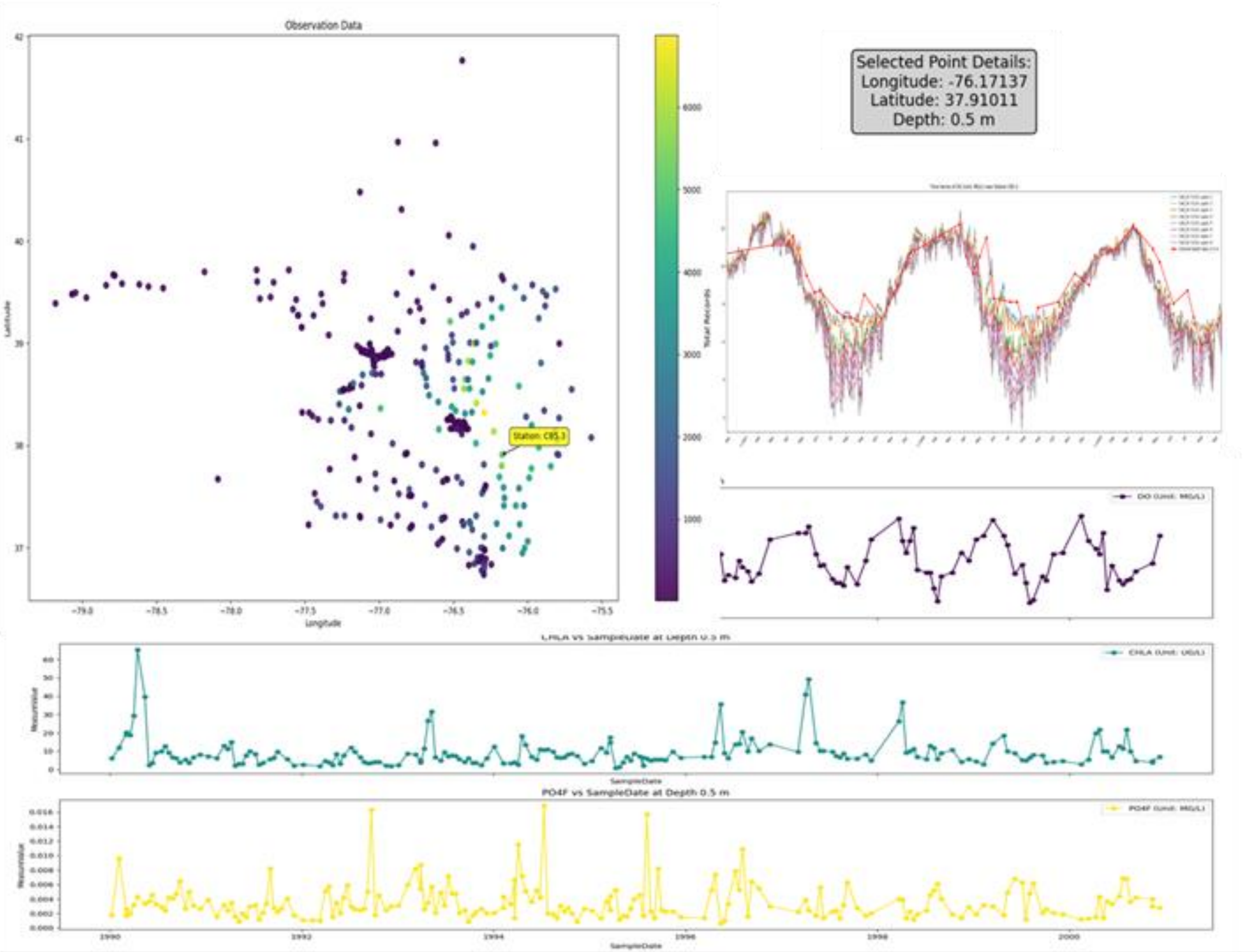
We are actively developing more public tools for Environment People, ask us about coding, HPC, and AI, we are happy to help!
https://github.com/CGC-UMCES https://scipe.umces.edu

Public Access:
Code Available: *CGC-UMCES GitHub Repository*
We also have tutorial: *Optimizing Data Processing on High-Memory Multi-Core Systems*

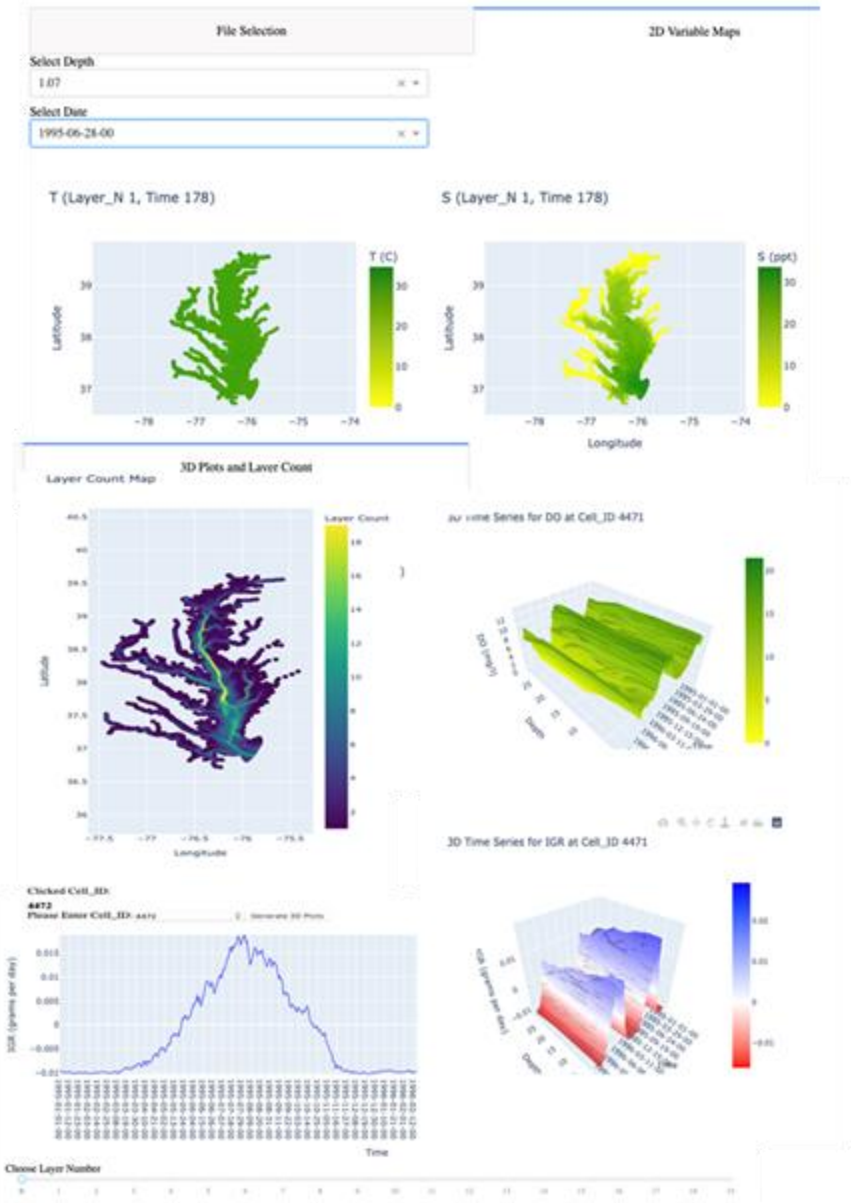
Other Tools Available:

Follow us, GitHub **CGC-UMCES**

Visualization of Time Series Data: Comparing the Selected Monitoring Station with the Nearby Simulation Cell



Interactive NetCDF Visualization App

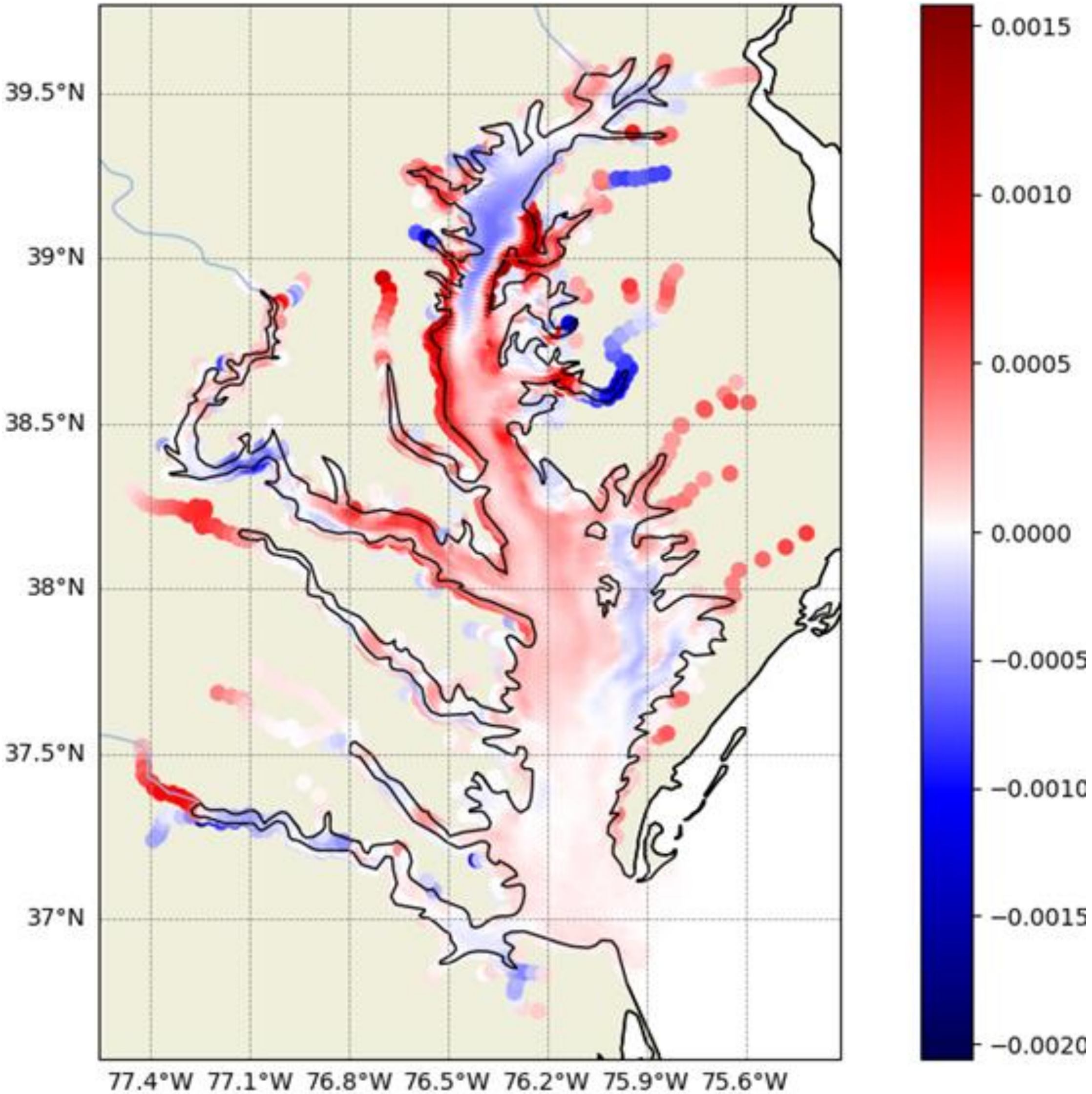


Initial implementation: Metaboscape development for White Perch

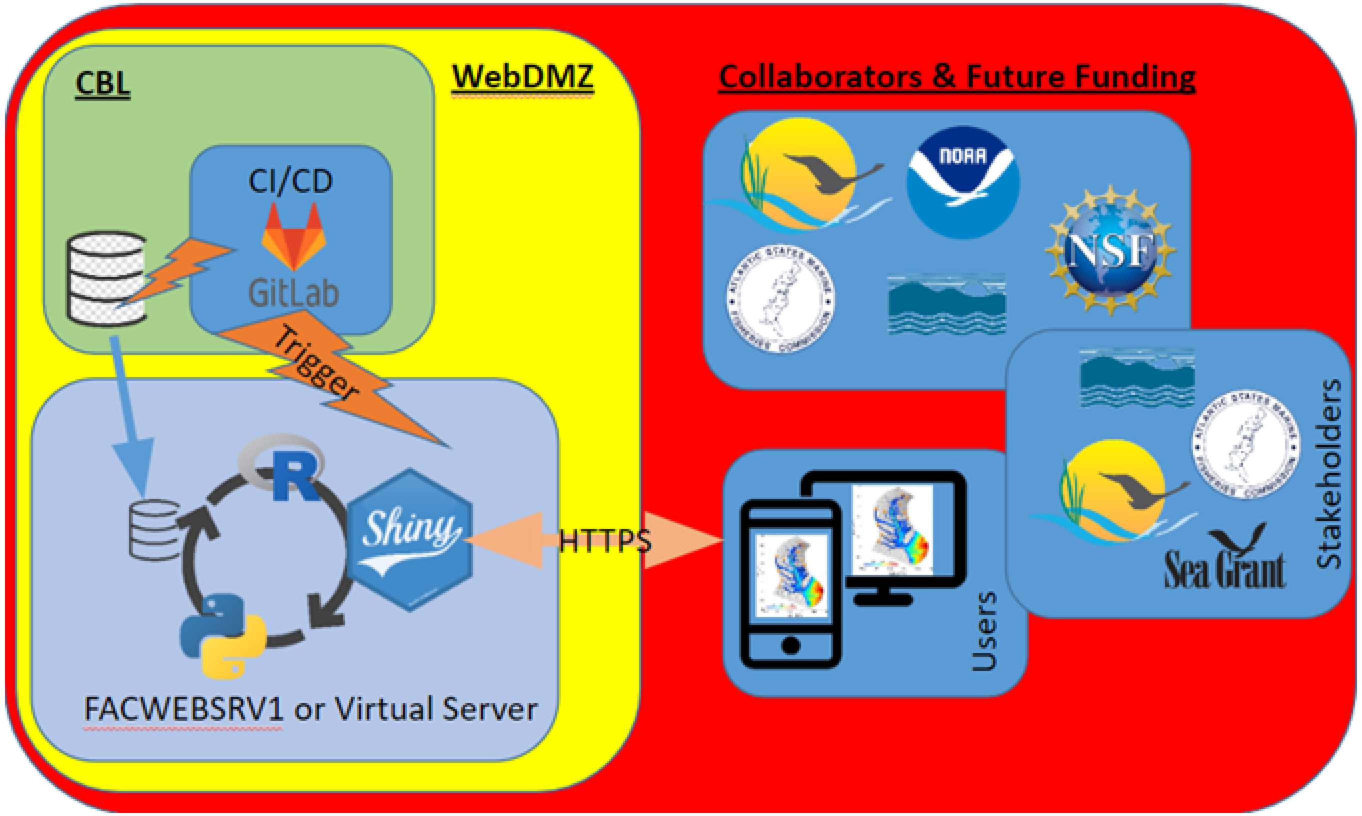
Example of model output

- Spatially explicit IGR (or other variable)
- Depth-specific (or depth-integrated) model output
- Evaluate differences (Δ) across scenario runs

- Example (1996 model year)
  - overall, increase in summer IGR in response to meeting TMDL goals in many areas
  - highlights spatially-dependent responses, some areas ↑, some areas ↓
  - ability to summarize results at various scales (e.g., HUC8/12, salinity zones, shallow vs deep habitats)



Developing a web-presence for the public

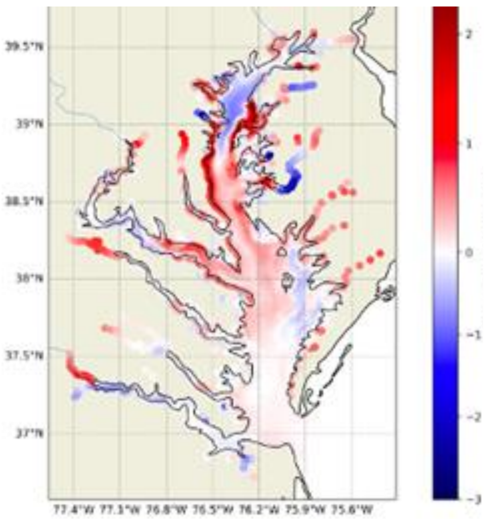


Interactive website

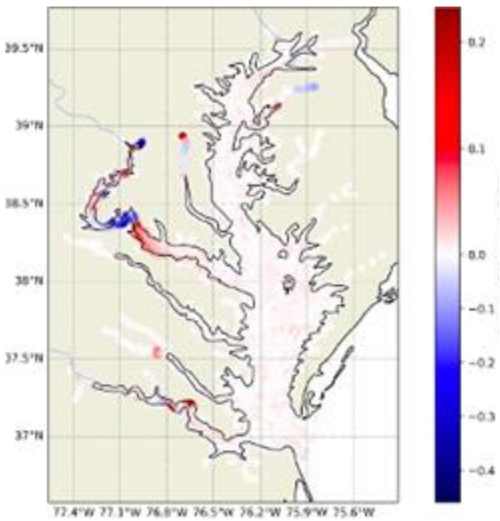
- Engage public through place-based science tool
- Foster communication with agencies and management personnel
- Visible product to highlight CGC capabilities

Next steps and Discussion questions

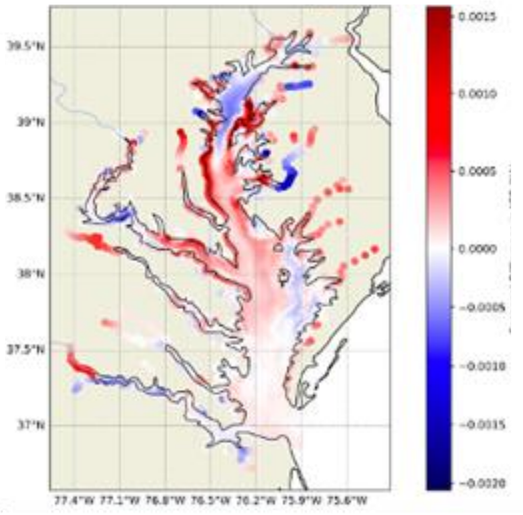
Dissolved oxygen



Temperature



IGR



Modeling goals

- Obtain and model full suite of CBP scenario runs
- Implement realistic ecological constraints for white perch
- Assess environmental drivers underlying metabolism patterns (right)
- Apply available models to other species (e.g., oyster DEB models)

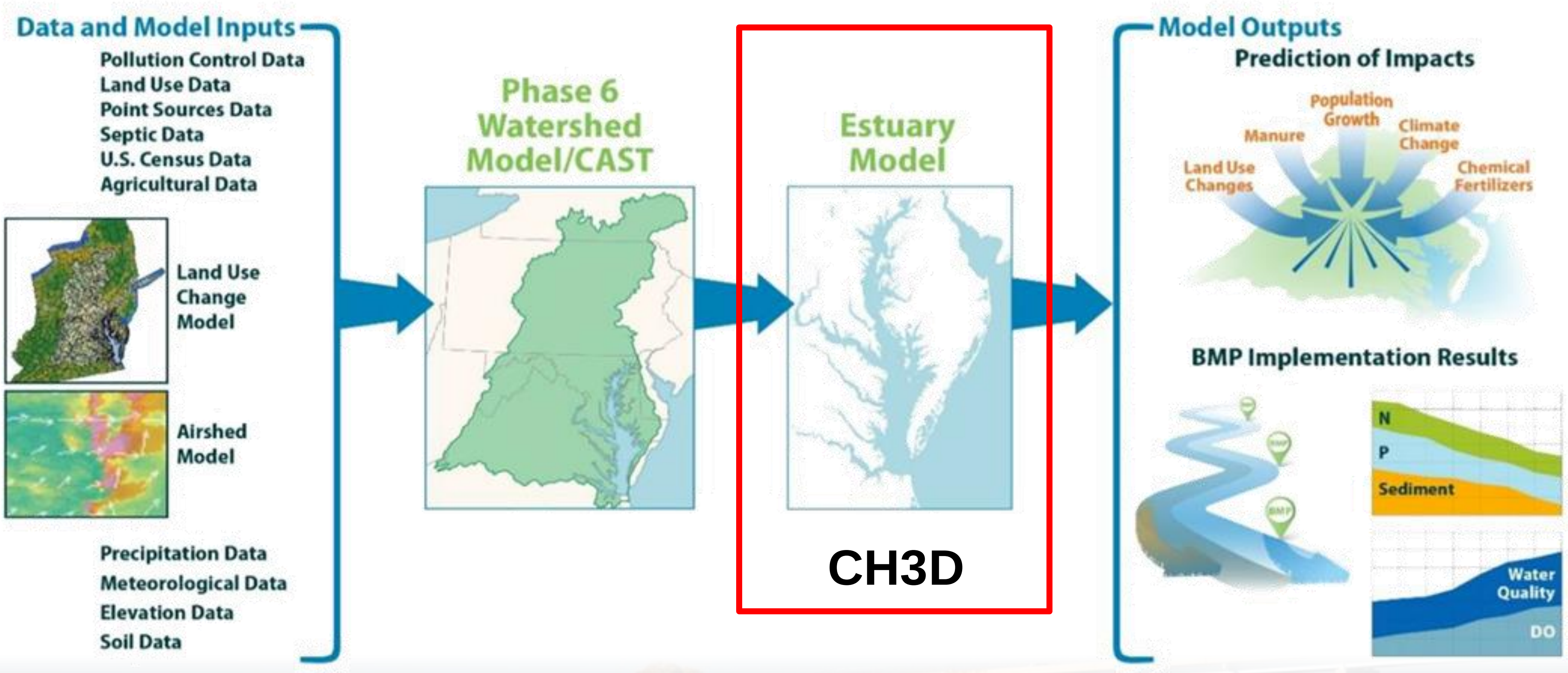
Discussion questions

- How would managers like to see results for species such as white perch?
  - Are there optimal level(s) of spatial aggregation of results?
  - Temporal aggregation?
- Are there specific, extant models of the focal species that we should consider?
- Are there other species of interest that have published bioenergetic models that we should consider if time allows (e.g., SAV, Striped Bass, sturgeon)?

Please reach out with any questions, suggestions or feedback to:

Ryan Woodland: woodland@umces.edu

Water quality predictions: Chesapeake Bay water quality model



NetCDF Structure:

Dimensions: (Cell_ID: 11064, Layer_N: 19, Time: Flex)

Coordinates:

Default:	* Cell_ID	(Cell_ID)	int32	1 2 3 ... 11062 11063 11064
	* Layer_N	(Layer_N)	float32	1.067 2.896 4.42 ... 25.76 27.28 28.8
	* Time	(Time)	datetime	Y-M-D-H

Data:

Default:	Latitude	(Cell_ID, Layer_N)	float32	Units: degrees_north
	Longitude	(Cell_ID, Layer_N)	float32	Units: degrees_east
	Date	(Time)		
	Depth	(Layer_N)	float32	Units: meter
	nwcbox	(Cell_ID, Layer_N)	int32	
	Area	(Cell_ID, Layer_N)	float32	
Flexible:	DO	(Cell_ID, Layer_N, Time)	float32	Units: mg/l
	T	(Cell_ID, Layer_N, Time)	float32	Units: C
	S	(Cell_ID, Layer_N, Time)	float32	Units: ppt
	...	*Total 37 Variables Available*		...

Follow us, GitHub CGC-UMCES

Chesapeake Global Collaboratory (CGC)
University of Maryland-
Center for Environmental Science (UMCES)

Example:

xarray.isel() to probe one datapoint in Year_1991.nc, which slice numbers are:
(Cell_ID=694, Layer_N=3, Time=1)

Latitude:37.04 (degrees_north)
Longitude: -76.052 (degrees_east)
Date: 1991-01-02-00 (days y-m-d-h)
Depth: 5.943 (meters)
nwcbox: 28287 (unknown unit)
Area: 631342.6
T: 6.5791 (C)
S: 29.370 (ppt)
DO: 10.063 (mg/l)

...

PO4: 0.005 (mg/l)
Chl: 7.2908 (ug/l)
Orgsed: 10.791 (mg/l)