A space-time interpolation tool for Chesapeake Bay dissolved oxygen

CERF presentation excerpts: data use focus

Rebecca Murphy[1], Jon Harcum[2], Elgin Perry[3],
Breck Sullivan[4], and Peter Tango[4],

[1]University of Maryland Center for Environmental Science at the Chesapeake Bay Program (CBP),
[2]Tetra Tech,[3]Statistics Consultant, [4]USGS at the CBP,

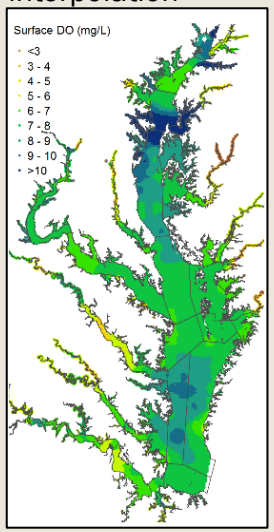
BORG

Dec. 15, 2025

Current interpolation:

Inverse distance weighting

Example spatial interpolation



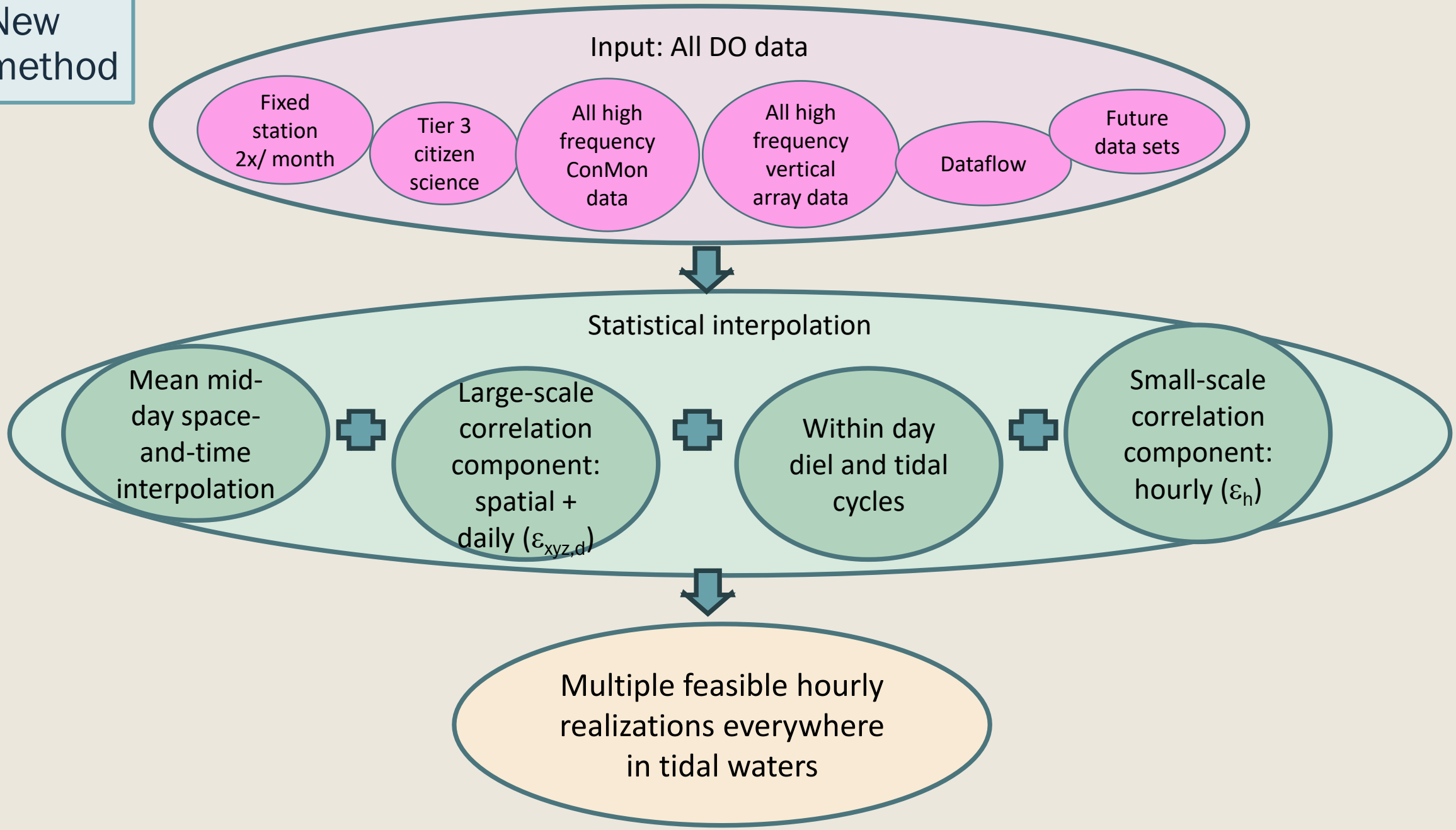
Problems with current interpolation

- Does not use the high frequency data
- Vertical layers interpolated horizontally and stacked;
- One cruise at a time, meaning a 2-week period assumed static; and
- Assumes same, fixed relationship between DO observations (not statistical).

This NEW "4-D" interpolation will:

- Use all high frequency data
- Interpolate all data together, not in layers.
- Interpolate in time, so that we do not have to artificially split time periods.
- Statistical and flexible– allowing for different DO relationships with space and time depending on the data.

New method

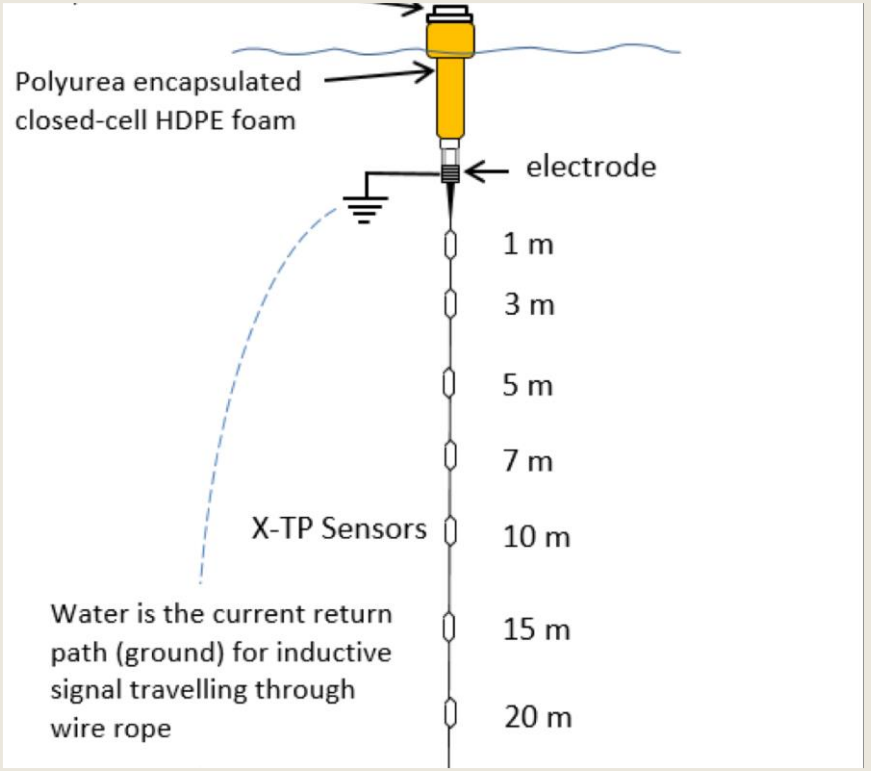

Input: All DO data
Fixed station 2x/ month
Tier 3 citizen science
All high frequency ConMon data
All high frequency vertical array data
Dataflow
Future data sets
Statistical interpolation
Mean mid-day space-and-time interpolation
Large-scale correlation component: spatial + daily ($\varepsilon_{xyz,d}$)
Within day diel and tidal cycles
Small-scale correlation component: hourly (ε_h)
Multiple feasible hourly realizations everywhere in tidal waters

Reasons for the space-time interpolator design

- We need **hourly, depth-resolved DO** for full criteria assessment
  = horizontal + vertical + seasonal + within-day structure together.

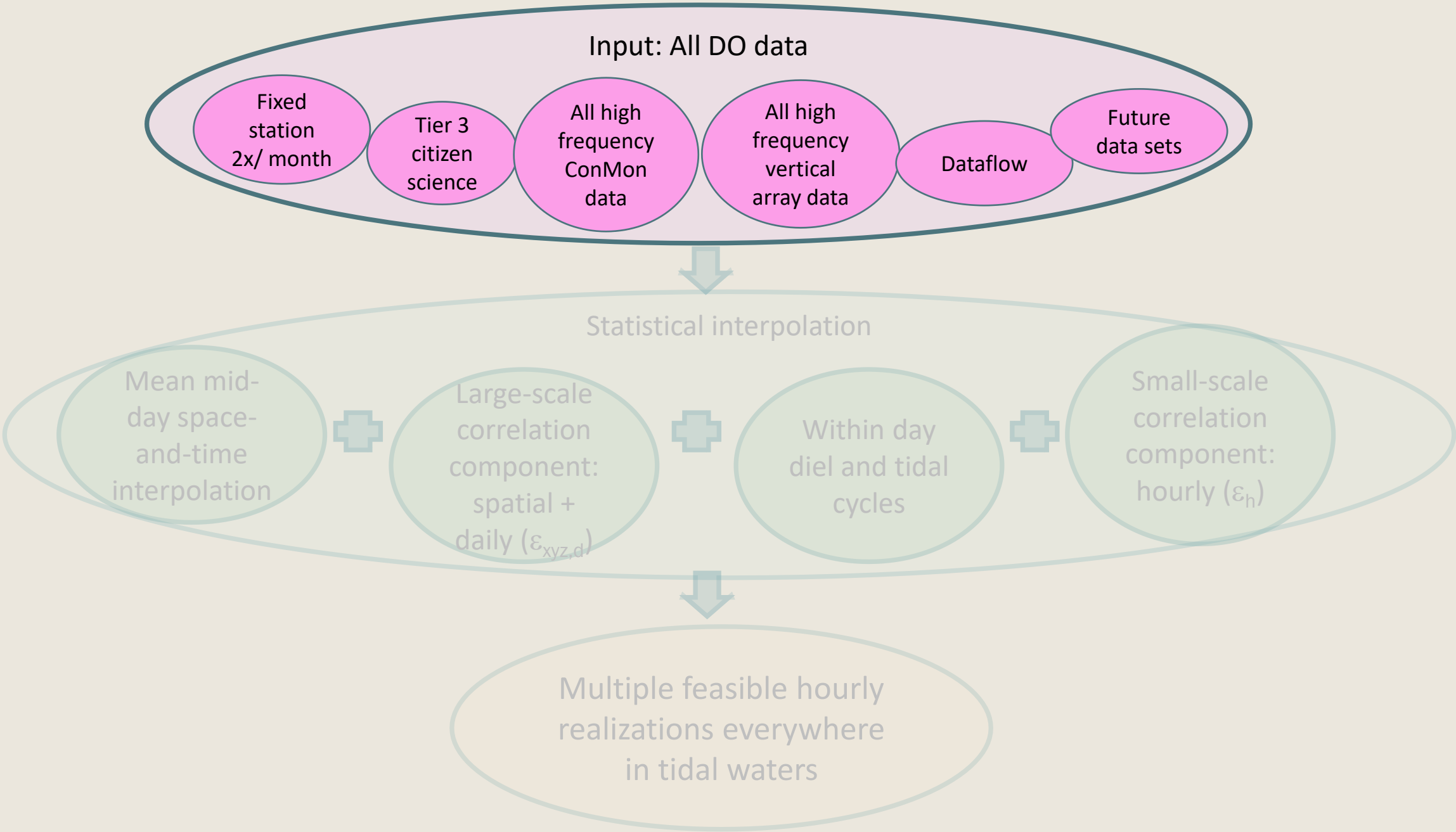
- **Ideal world:** vertical arrays every ~1 km, all depths, all days.

- **Reality:** mixed networks (fixed stations, cruises, arrays) ⟶ patchy space and irregular time.



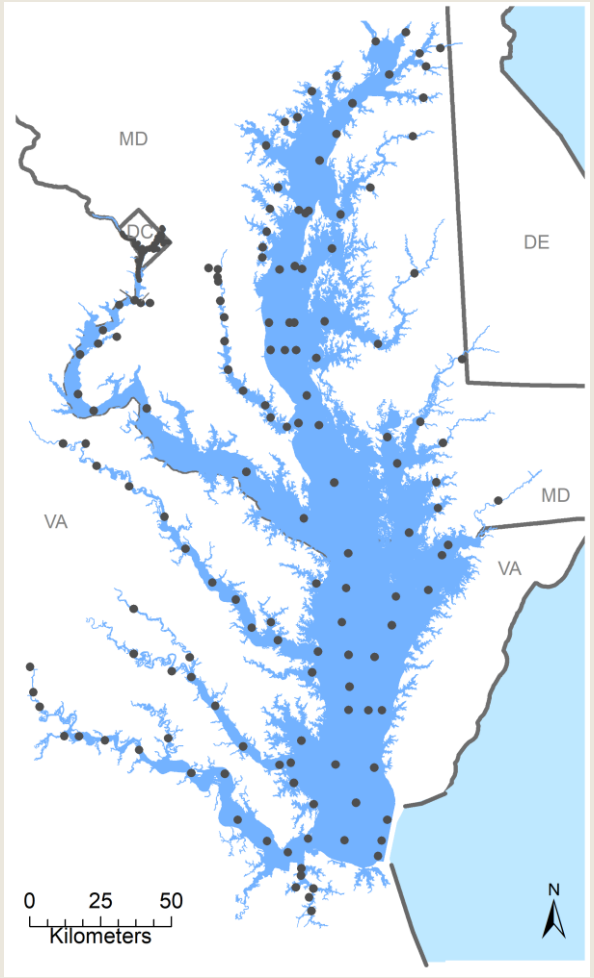
Our approach:

General additive model (GAM) for the **mean mid-day** DO
+ Empirically fitted **horizontal, vertical, and temporal** correlations (MVN errors)
⟶ a **4-D interpolator.**


Input: All DO data
Fixed station 2x/ month
Tier 3 citizen science
All high frequency ConMon data
All high frequency vertical array data
Dataflow
Future data sets
Statistical interpolation
Mean mid-day space-and-time interpolation
Large-scale correlation component: spatial + daily ($\varepsilon_{xyz,d}$)
Within day diel and tidal cycles
Small-scale correlation component: hourly (ε_h)
Multiple feasible hourly realizations everywhere in tidal waters

Multiple types of data being used

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

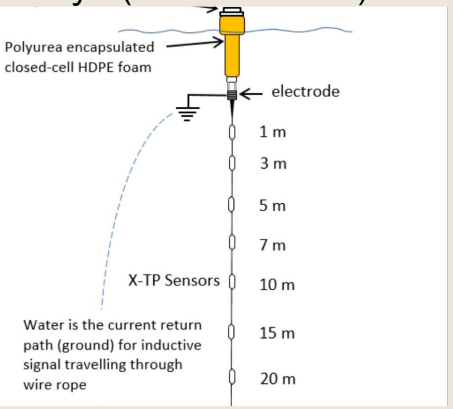


Shallow water continuous monitoring (MDDNR and VECOS)

From http://vecos.vims.edu/

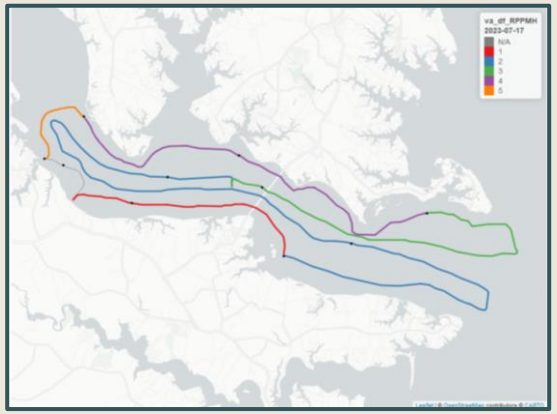
From https://eyesonthebay.dnr.maryland.gov/

New continuous vertical arrays (NOAA & CBP)



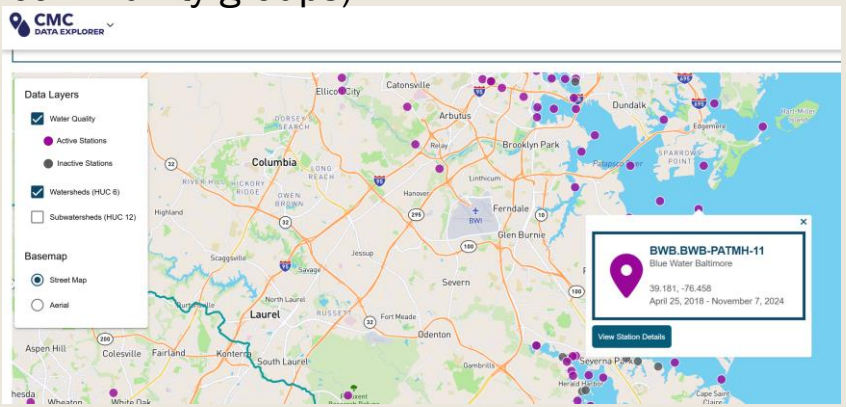
https://www.chesapeakebay.net/who/group/hypoxia-collaborative-team

Dataflow (MDDNR and VECOS)



From http://vecos.vims.edu/

Citizen monitoring (CMC and multiple community groups)



Chesapeake Monitoring Cooperative: https://www.chesapeakemonitoringcoop.org/

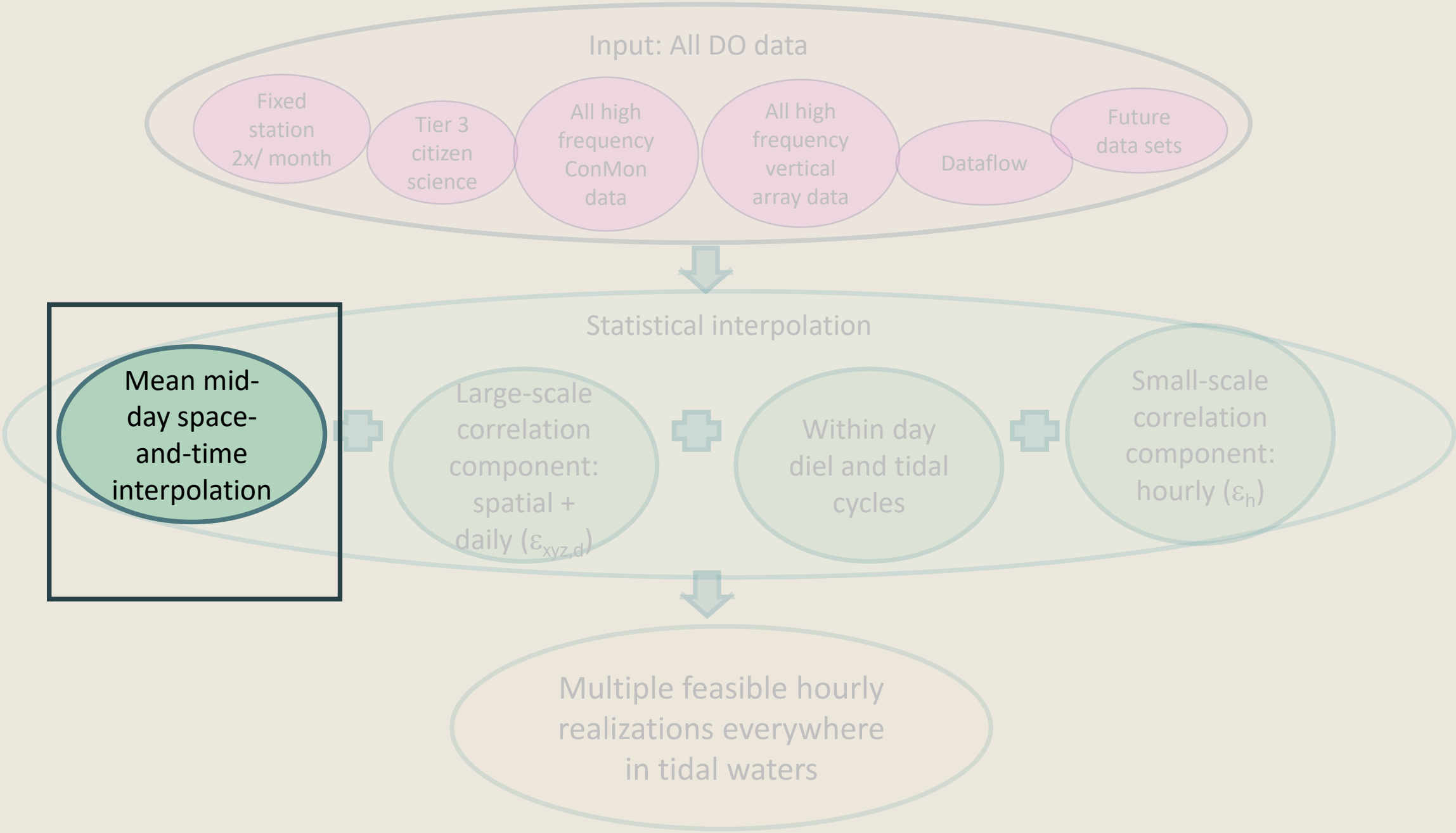
+Additional State Agency Collected Data (MDE, VADEQ)

Data Landscape

Informs mean mid-day space–time interpolation

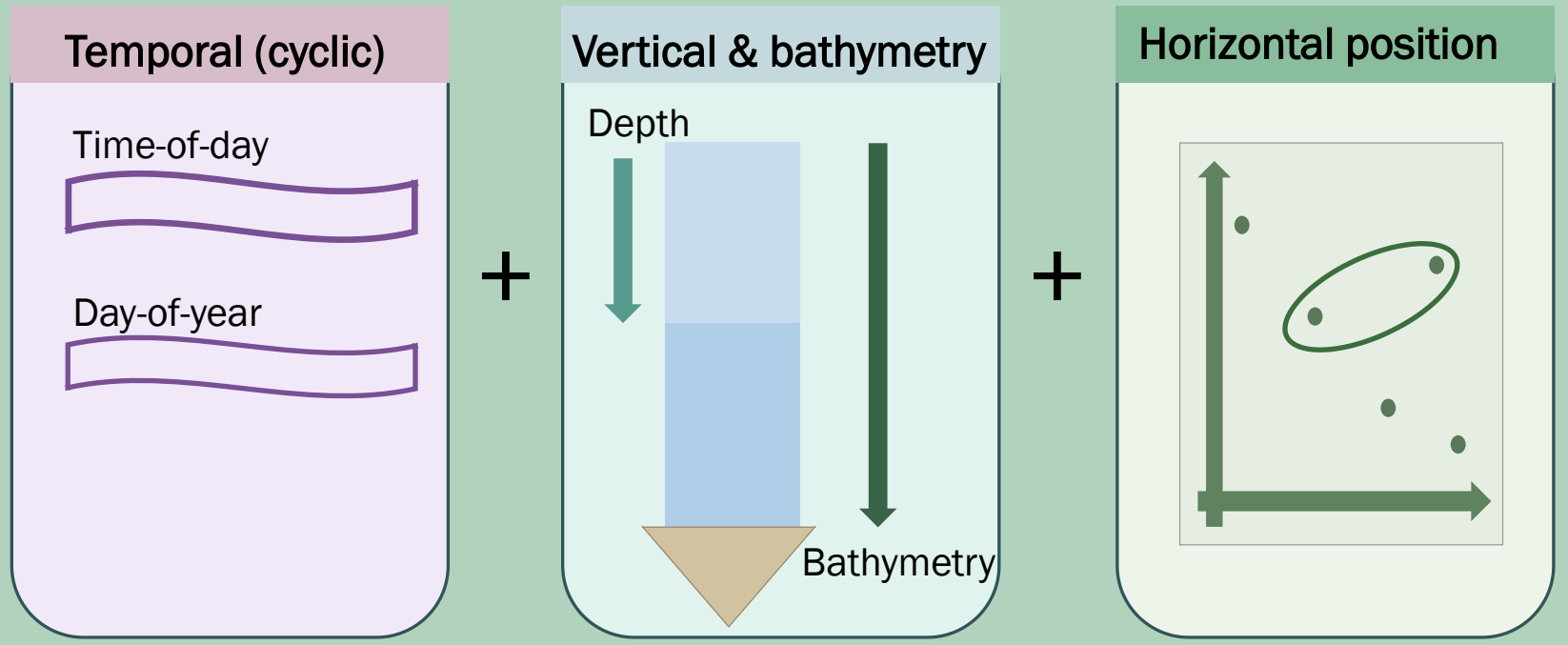
Fixed-station network	Cruise-track monitoring (**DataFlow***)	Continuous monitoring (**ConMon***)	Vertical arrays (**NOAA**)	Additional State Agency Collected Data
✓Fixed location; broad spatial coverage ✓Long-term consistency ✓Profiles (every 1-2 meter to bottom) ✓150+ sites ✓1-2x/month	✓Surface mapping continuous data (~0.5 m) ≈ 8-10 sites/yr* ≈ 4-7 cruises per year (Apr-Oct)	✓Fixed location ✓15-minute sampling ✓Fixed depth near surface or bottom ≈ 25-30 sites/yr*; typically, 6-9 months/yr; some year round	✓Fixed location; multi-depth ✓10-minute sampling ✓New since 2022 ≈ 2-3 sites/yr ≈ 5-9 month deployments	✓Expands monitoring breadth ✓1-2x/month **Citizen monitoring (Tier 3)** ✓Expands monitoring breadth ✓1-2x/month
Vertical correlation	Horizontal correlation	Daily correlation; Daily cycle; Hourly correlation	Vertical correlation; Daily correlation; Daily cycle; Hourly correlation	

*Sites commonly rotated about every 3 years to broaden coverage


Input: All DO data
Fixed station 2x/ month
Tier 3 citizen science
All high frequency ConMon data
All high frequency vertical array data
Dataflow
Future data sets
Statistical interpolation
Mean mid-day space-and-time interpolation
Large-scale correlation component: spatial + daily ($\varepsilon_{xyz,d}$)
Within day diel and tidal cycles
Small-scale correlation component: hourly (ε_h)
Multiple feasible hourly realizations everywhere in tidal waters

Mean mid-day space-and-time interpolation

- **Goal:** Capture the central tendency of <u>all the data</u> in a year. Output an estimate every day, everywhere, to add to hourly variability.



Uses Generalized Additive Models (GAMs) with:

- **Cyclic smooths (cc)** for time-of-day and day-of-year.
- **Tensor-product smooths (ti)** to model interactions among depth, bathymetry, and horizontal position.

Mean mid-day space-and-time interpolation

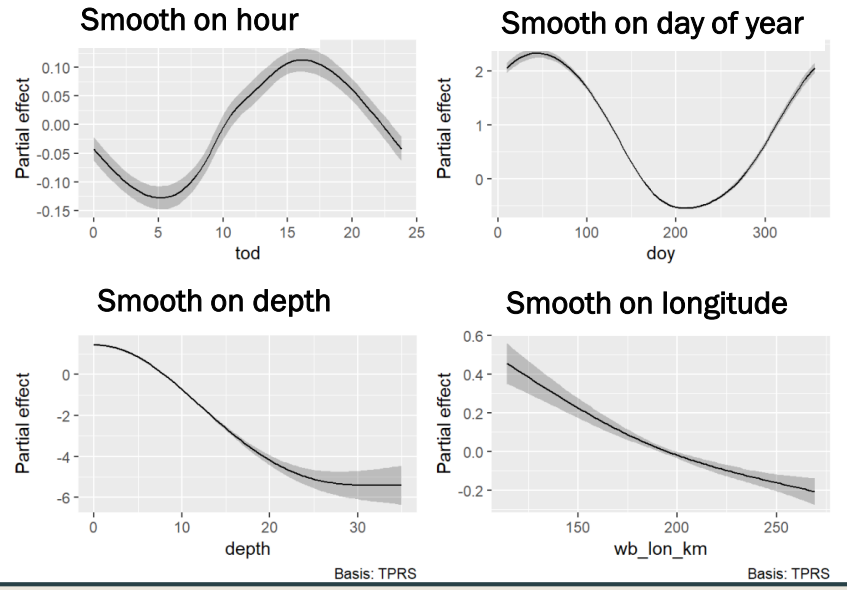
Multiple smooth terms are fit to the data

Tabular output

Approximate significance of smooth terms:

type	source	df	F	p.value
parametric terms	NA			-
smoothed terms	s(tod)	3.47	102.7910	<0.0001
" "	s(doy)	4.00	2164.2063	<0.0001
" "	s(depth)	2.46	2229.1807	<0.0001
" "	s(wb_lon_km)	2.15	36.8366	<0.0001

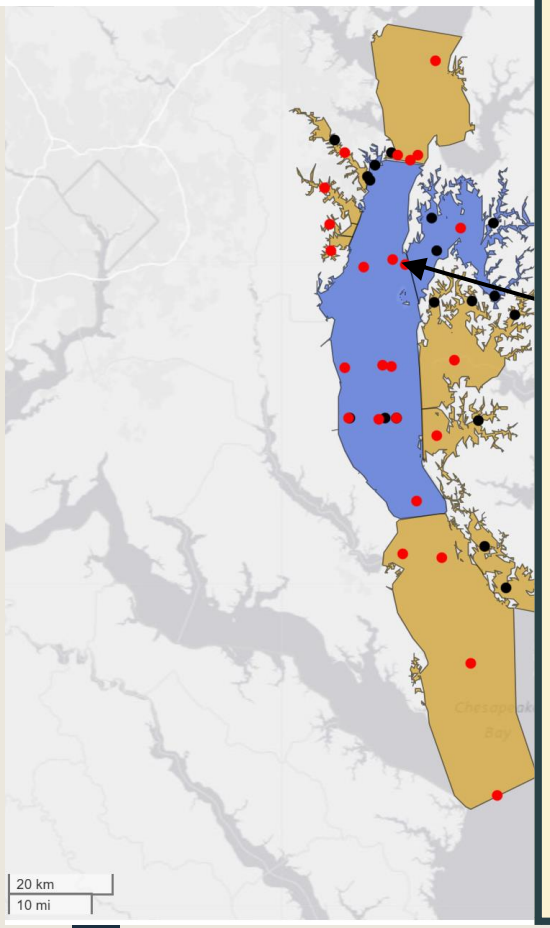
Graphical output of relationships fit to the data



Example region: mid-bay

Purple are target segments

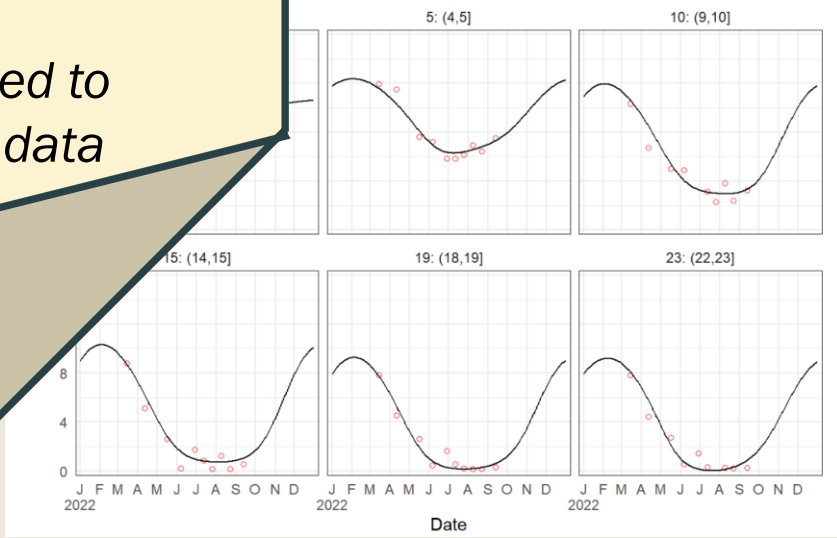
Tan are boundary segmen

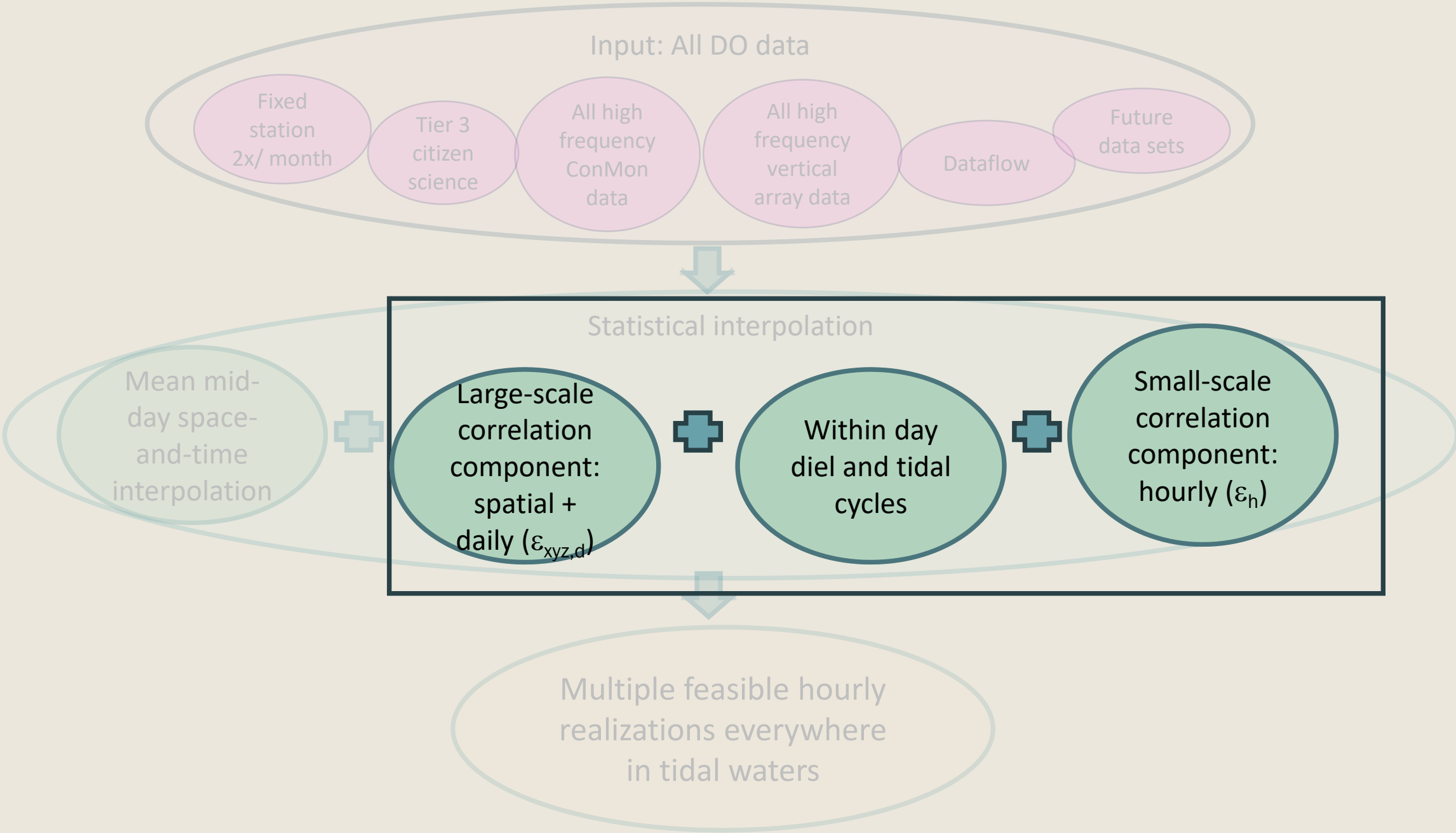


Which data?

- *All of the data (entire table on slide 7) collected in the shaded segments in one year is used to build these relationships. If frequency is <1 hour, the station is sub-sampled to an hour.*
- *The relationships are used to interpolate between the data points in space and time to get a DO estimate each day, everywhere for the year.*

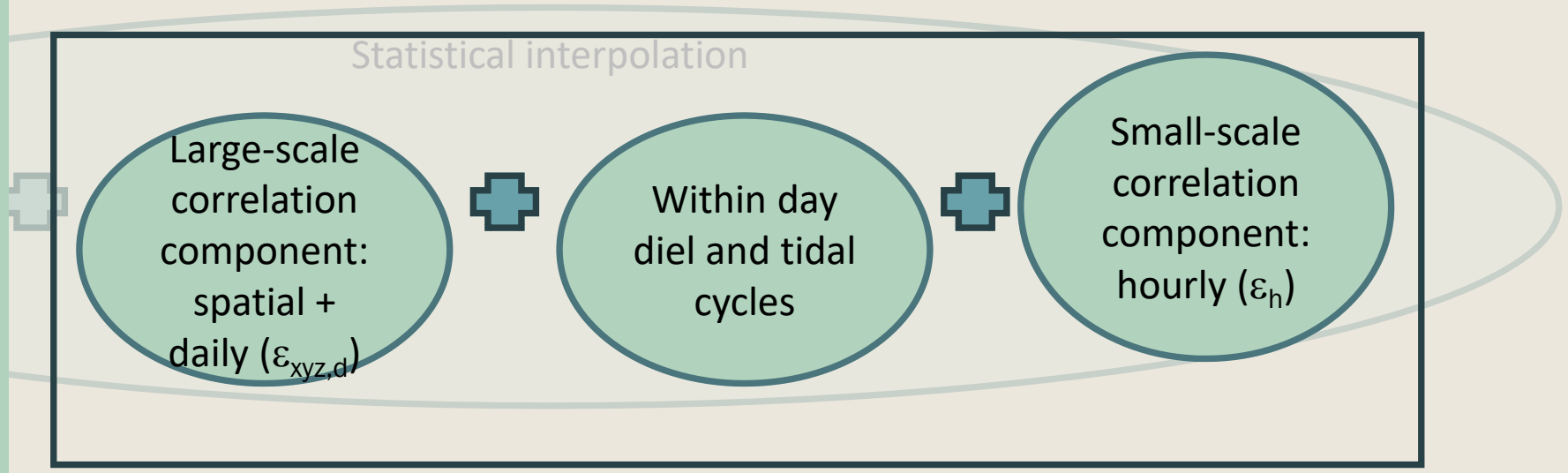
) data at a sample location




Input: All DO data
Fixed station 2x/ month
Tier 3 citizen science
All high frequency ConMon data
All high frequency vertical array data
Dataflow
Future data sets
Statistical interpolation
Mean mid-day space-and-time interpolation
Large-scale correlation component: spatial + daily ($\varepsilon_{xyz,d}$)
Within day diel and tidal cycles
Small-scale correlation component: hourly (ε_h)
Multiple feasible hourly realizations everywhere in tidal waters

Evaluate high-frequency data for within-day (Conmon) and location-to-location (Dataflow) variability.

Apply those patterns to generation hourly simulation results.

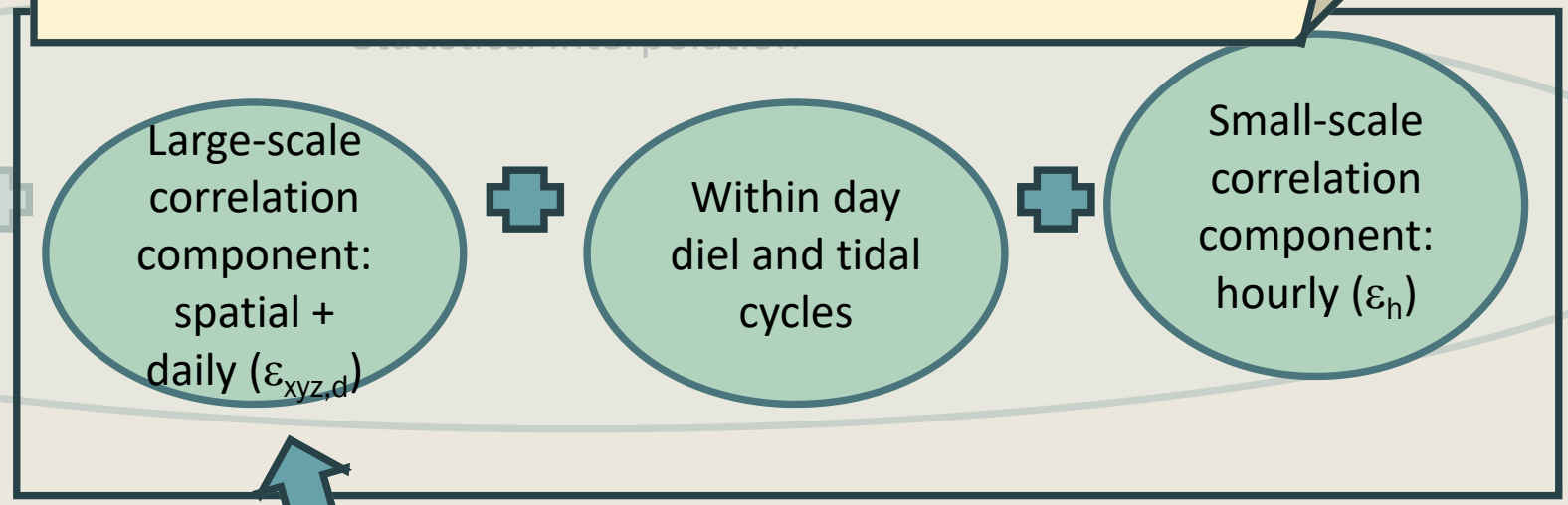


Evaluate high-frequency data for within-day (Conmon) and location-to-location (Dataflow) variability.

Apply those patterns to generation hourly simulation results.

Which data?

- *The data for these pieces is not available every year, in every segment.*
- *This led to the design where we: parameterize based on all the high frequency data we do have (back to 2003) and use it to generate* ***multiple feasible simulation*** *results.*
- *We will compare the simulation results to the co-located high frequency data to be sure they represent the key criteria-related features in the same way.*



example

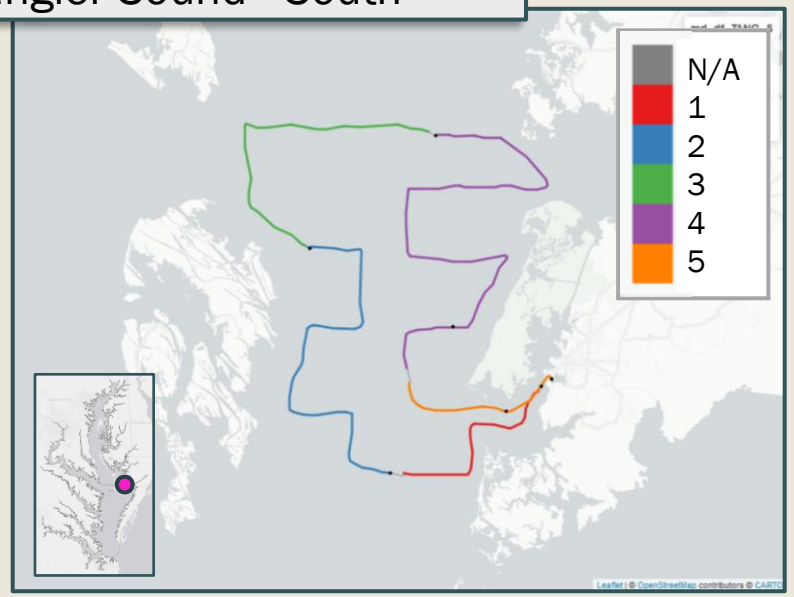
Example High freq data use

Correlation: Horizontal (along- and cross-bay)

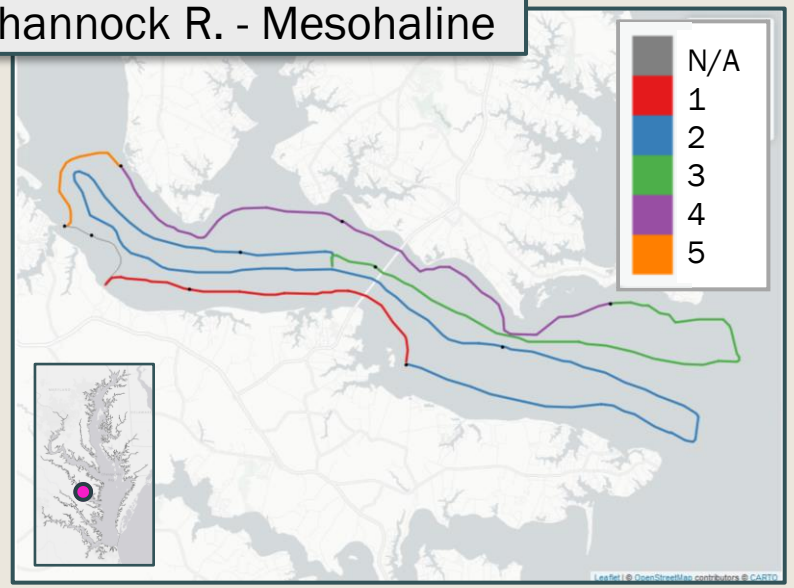
- DataFlow—high frequency underway surveys (~0.5 m intake)
  - **MD DNR Eyes on the Bay (EOTB)**: ~63 “stations”, 2001-2023
  - **Virginia Estuarine and Coastal Observing System (VECOS)**: ~36 “stations” 2003-2024
- Map lines show survey legs (1–5)

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

Tangier Sound - South



Rappahannock R. - Mesohaline

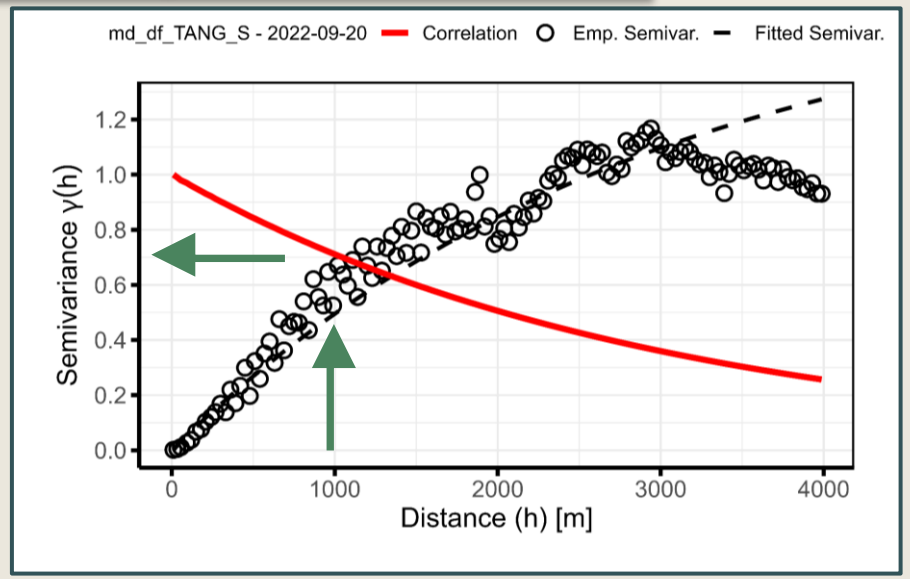


Example High freq data use

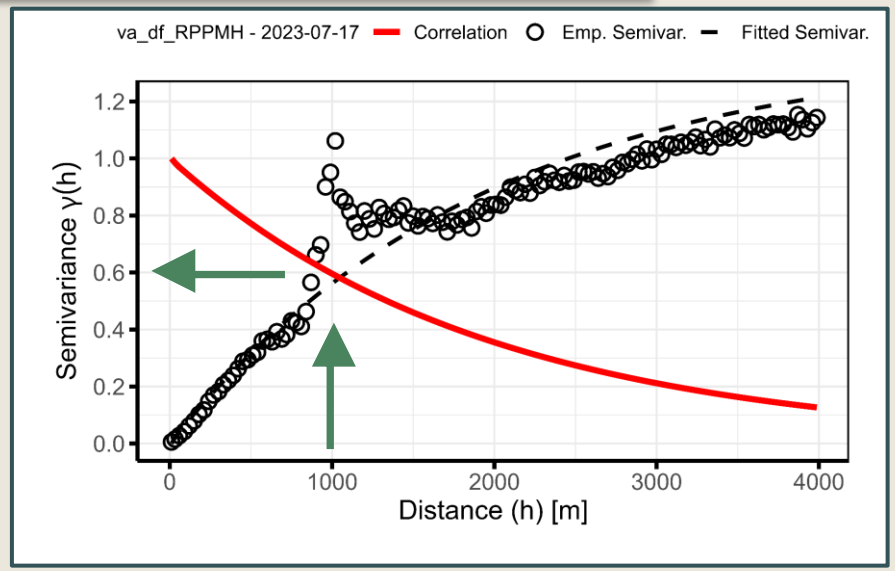
Correlation: Horizontal (along- and cross-bay)

- For each cruise, compute a variogram cloud by chunk (all point pairs within a chunk).
- Bin the clouds (30 m bins) and compute the empirical semivariogram.
- Fit exponential variogram.
- Use pairs out to 4,000 m maximum distance.
- Plot guide:
  - open circles = empirical semivariance.
  - dashed = fitted semivariogram.
  - red line = correlation decay vs. distance.

Tangier Sound - South



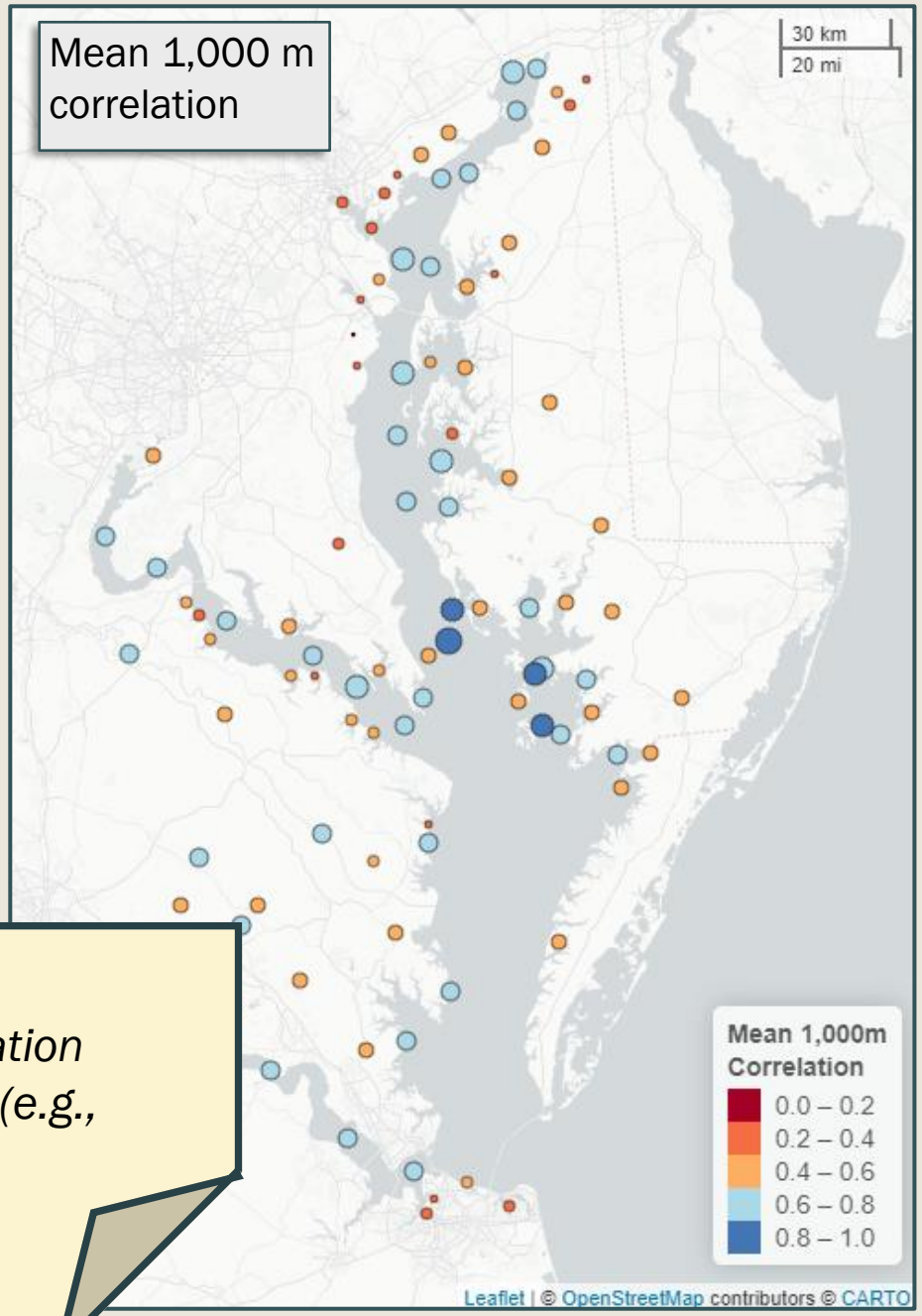
Rappahannock R. - Mesohaline



Example High freq data use

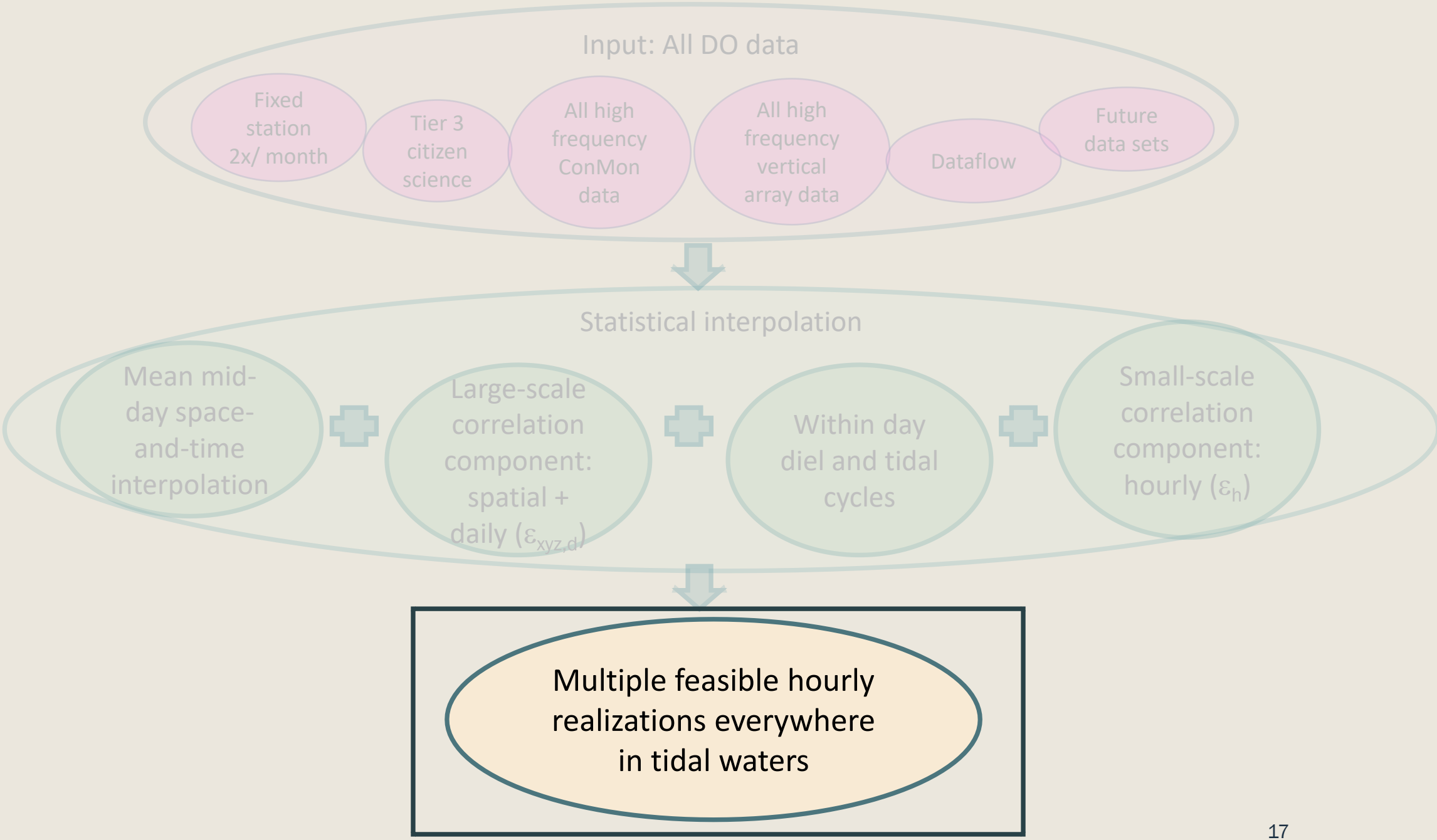
Correlation: Horizontal (along- and cross-bay)

- **Highest (0.8–1.0):** mid-Bay near CB5MH_MD and Tangier Sound.
- **Moderately high (0.6–0.8):** widespread in the upper mainstem and larger tributaries.
- **Lower (0.2–0.4):** common in upper tributaries, small embayments, and nearshore.
- Across ~100 stations, most means fall 0.4–0.7.
- Spatial differences dominate variability more than temporal effects.



How is this processed data used in the tool?

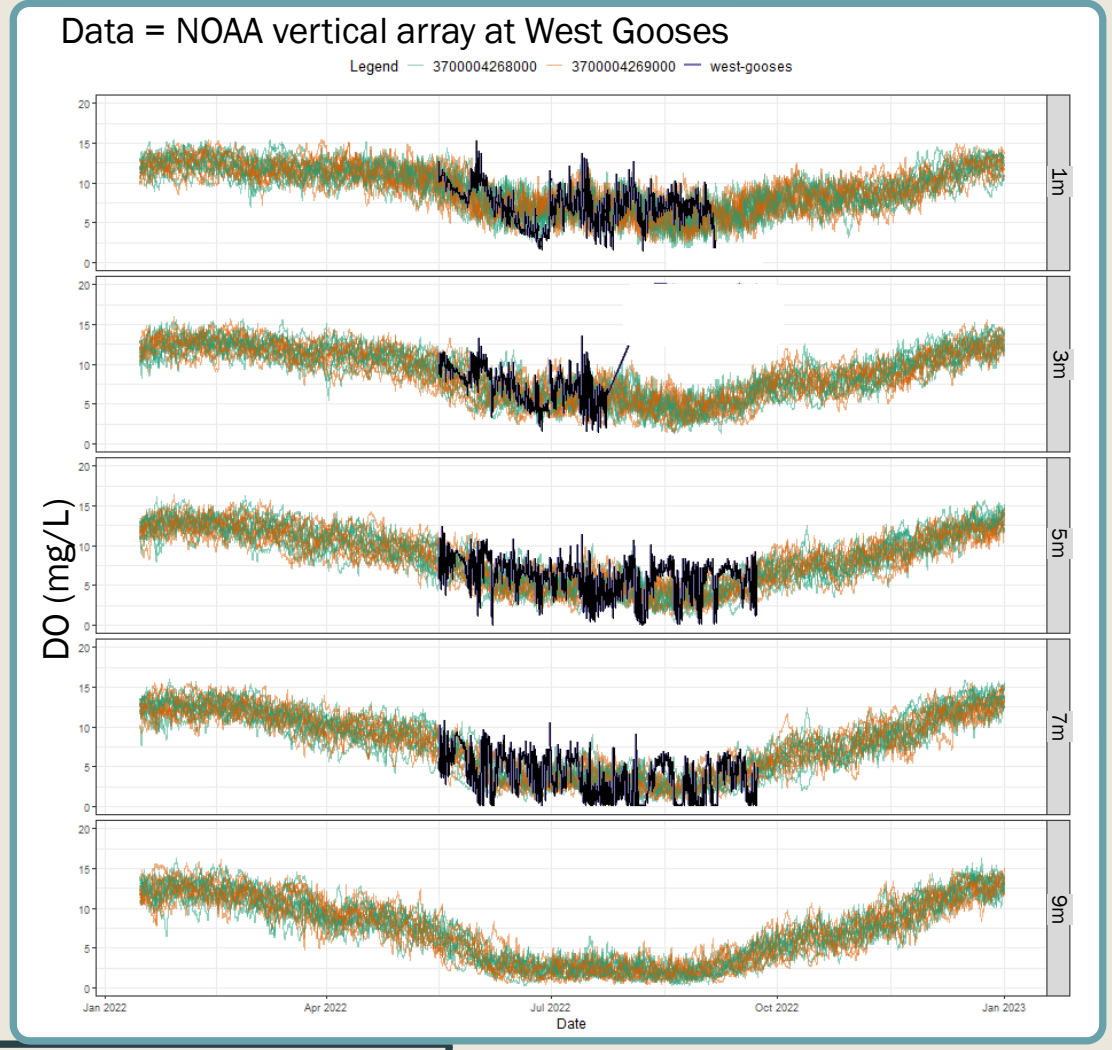
- *We will fit a smooth spatial function (GAM) to these correlation values so that we can represent how they vary by location (e.g., correlation is higher in deeper waters, lower in shallower).*
- *For any spot we are interpolating DO, we will pull out the appropriate correlation parameter for that location to use.*


Input: All DO data
Fixed station 2x/ month
Tier 3 citizen science
All high frequency ConMon data
All high frequency vertical array data
Dataflow
Future data sets
Statistical interpolation
Mean mid-day space-and-time interpolation
Large-scale correlation component: spatial + daily ($\varepsilon_{xyz,d}$)
Within day diel and tidal cycles
Small-scale correlation component: hourly (ε_h)
Multiple feasible hourly realizations everywhere in tidal waters

100 realizations of hourly estimates everywhere for a year.

This example was an early test at one location, compared to observed high frequency data (black lines).

Evaluation with recent 4-D updates is being done now with data like this and results will be shared with BORG.



Multiple feasible hourly realizations everywhere in tidal waters

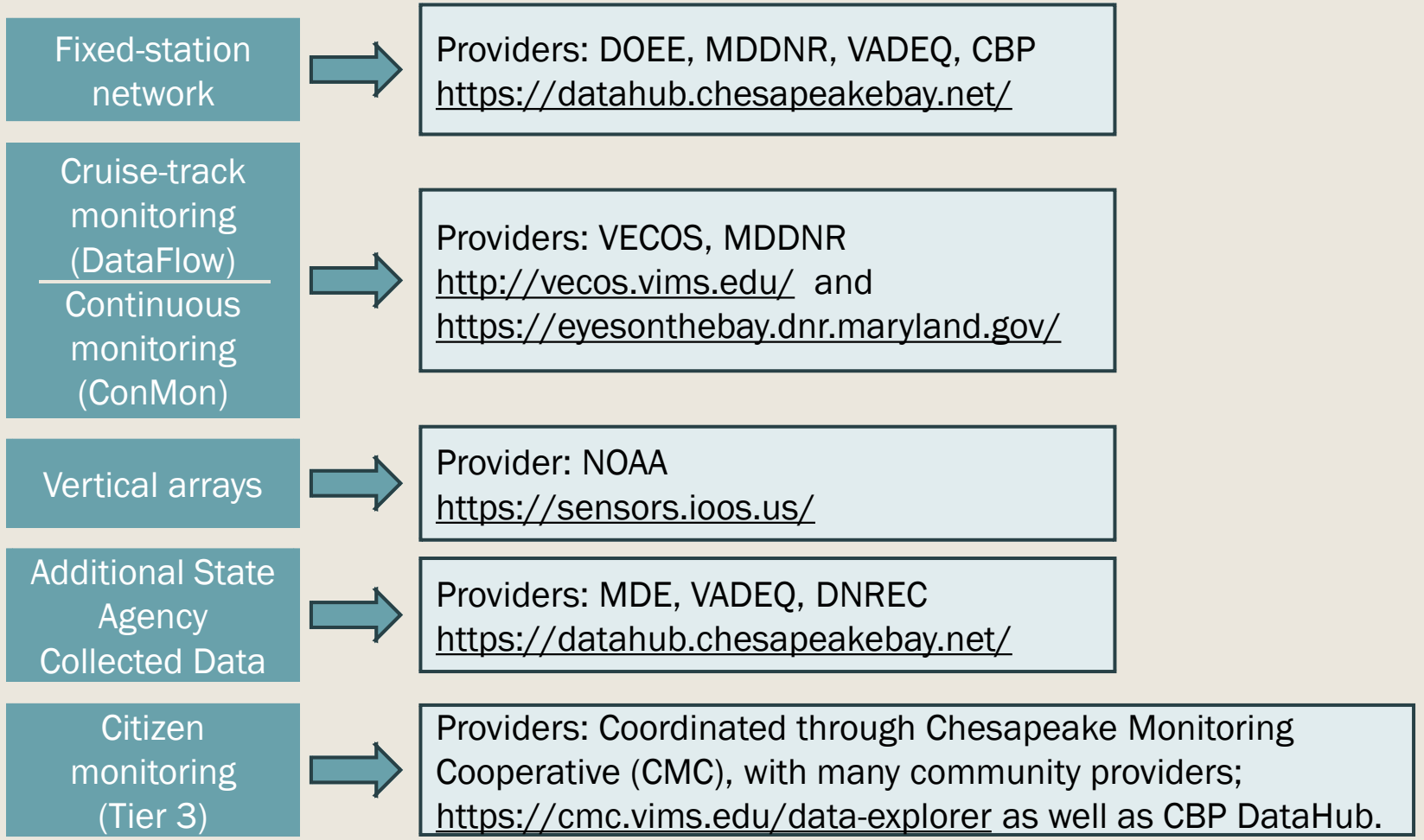
Data Landscape

Informs mean mid-day space–time interpolation

Fixed-station network	Cruise-track monitoring (**DataFlow***)	Continuous monitoring (**ConMon***)	Vertical arrays (**NOAA**)	Additional State Agency Collected Data
✓Fixed location; broad spatial coverage ✓Long-term consistency ✓Profiles (every 1-2 meter to bottom) ✓150+ sites ✓1-2x/month	✓Surface mapping continuous data (~0.5 m) ≈ 8-10 sites/yr* ≈ 4-7 cruises per year (Apr-Oct)	✓Fixed location ✓15-minute sampling ✓Fixed depth near surface or bottom ≈ 25-30 sites/yr*; typically, 6-9 months/yr; some year round	✓Fixed location; multi-depth ✓10-minute sampling ✓New since 2022 ≈ 2-3 sites/yr ≈ 5-9 month deployments	✓Expands monitoring breadth ✓1-2x/month **Citizen monitoring (Tier 3)** ✓Expands monitoring breadth ✓1-2x/month
Vertical correlation	Horizontal correlation	Daily correlation; Daily cycle; Hourly correlation	Vertical correlation; Daily correlation; Daily cycle; Hourly correlation	

*Sites commonly rotated about every 3 years to broaden coverage

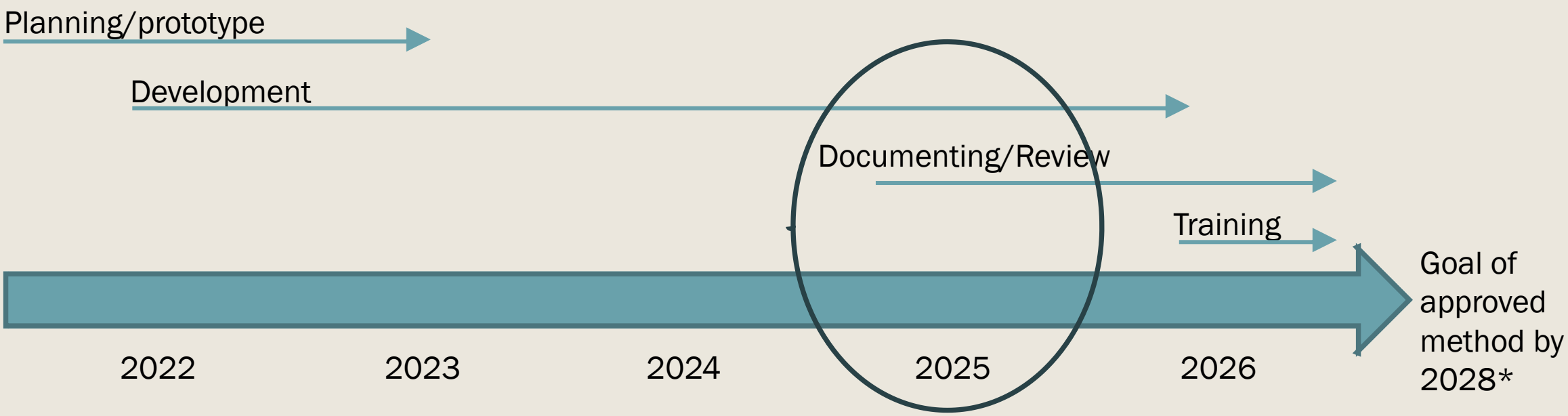
Data sources



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); Liz Chudoba (Alliance for CB).

extras

4-D interpolator development timeline



*with 2030 goal of reporting on all criteria

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.

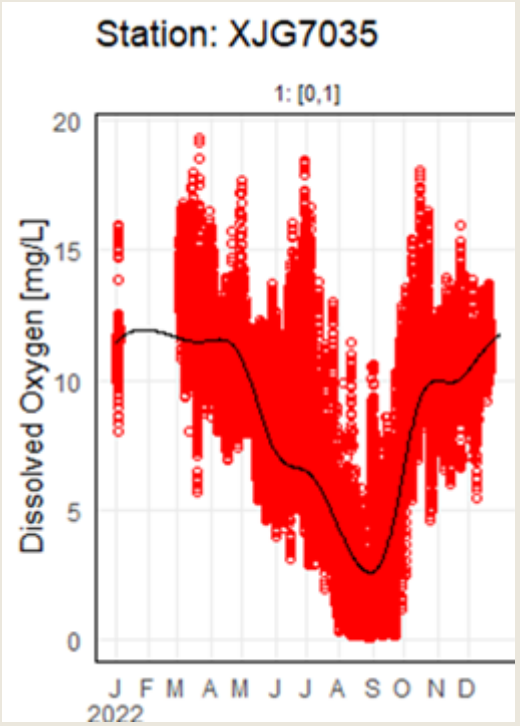
*Note a 30-day mean 6 mg/L MSN value is evaluated for purpose of the WQ indicator.

From EPA 2003 Ambient Water Quality Criteria

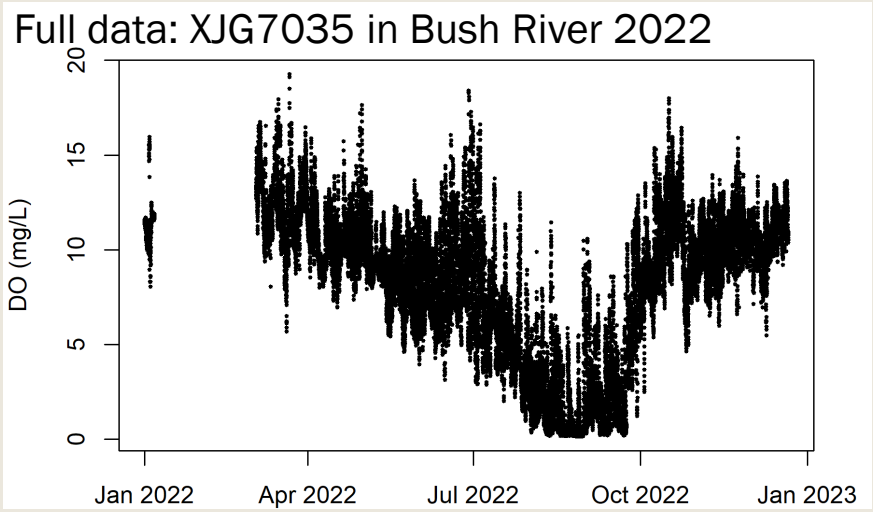
For this part only

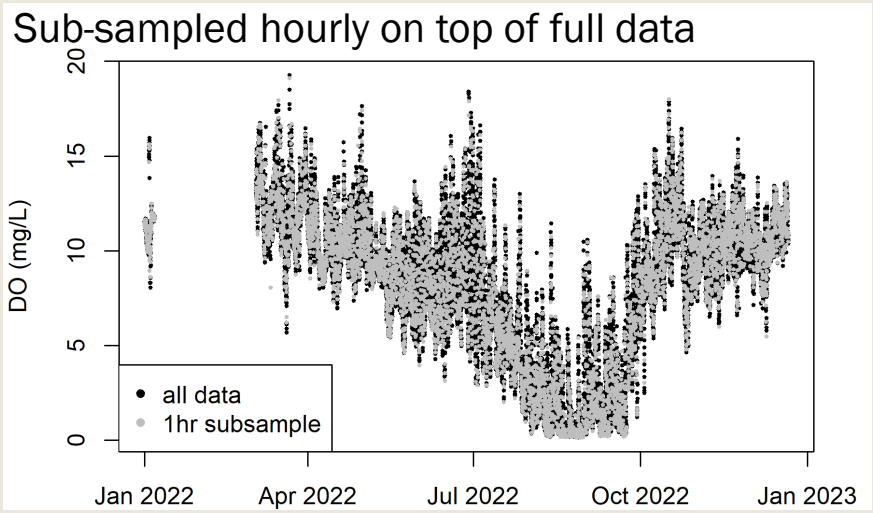
Mean mid-day space-and-time interpolation

What if ConMon is sub-sampled to hourly?

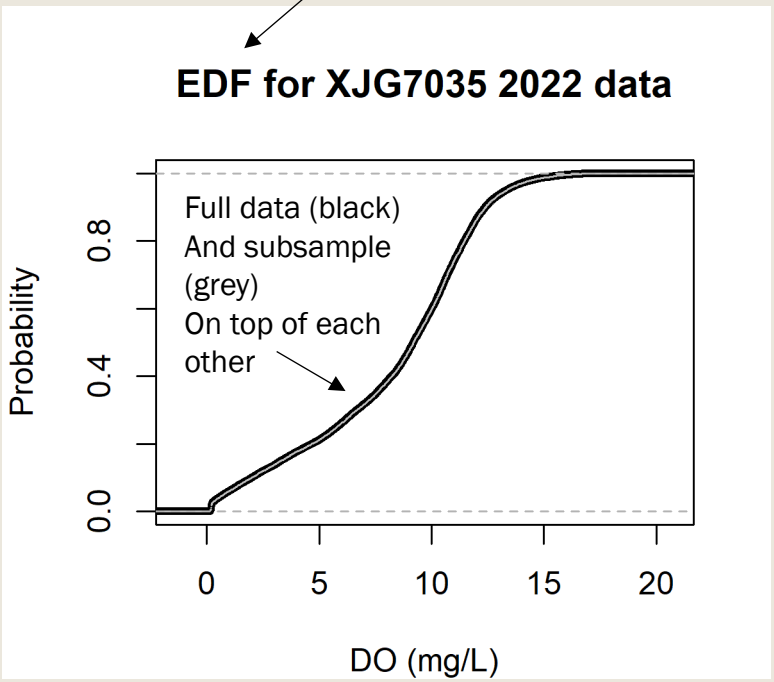


=





EDF = Empirical density function



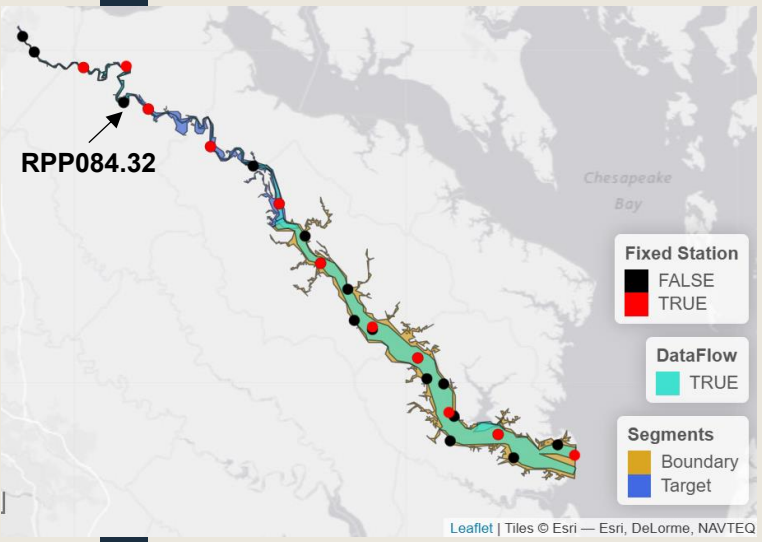
Data set	count	10th percentile	Fraction <3.2 mg/L	Fraction < 5mg/L
All	28,127	1.97	0.147	0.212
Sub-sample	7,071	1.98	0.146	0.211

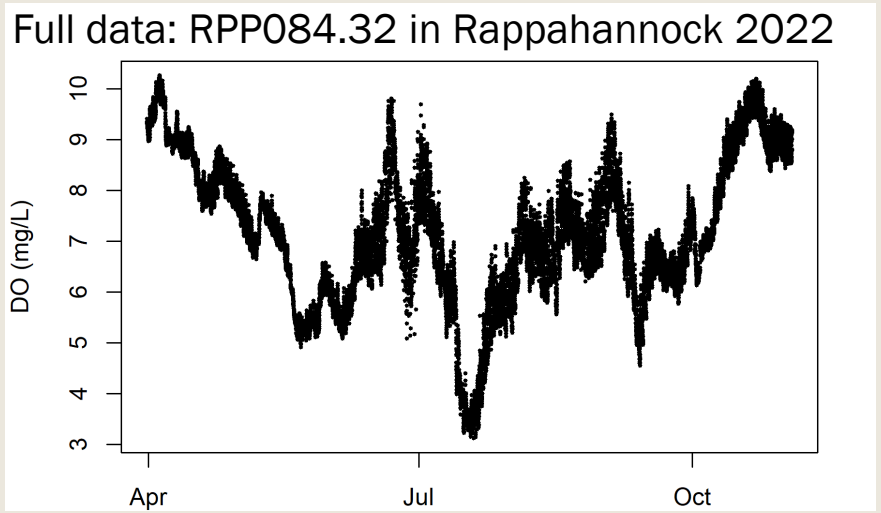
These summaries suggest we are not changing the important features of this dataset by sub-sampling this 15 min data to 1 hour.

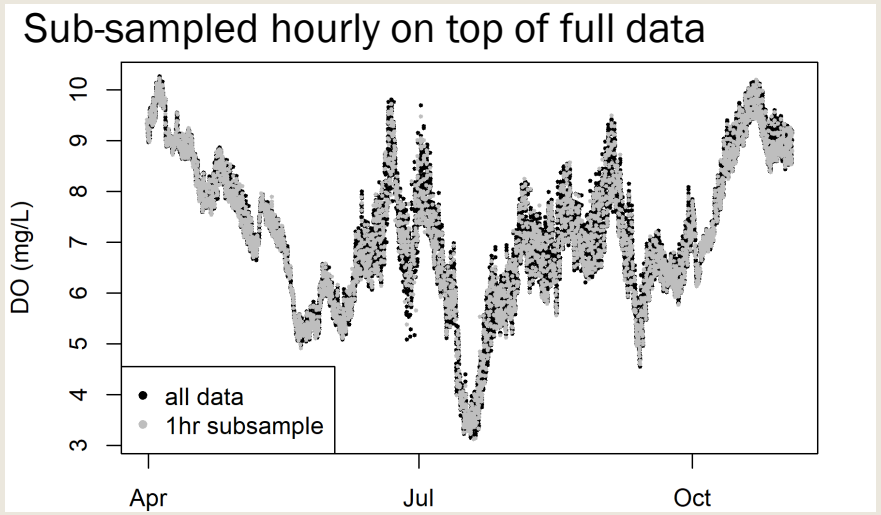
For this part only

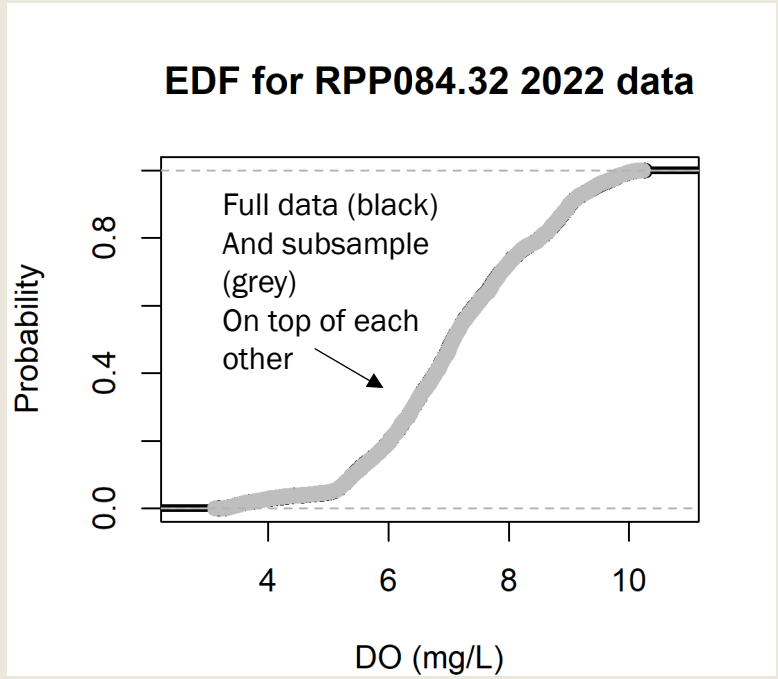
Mean mid-day space-and-time interpolation

Another example









Data set	count	10th percentile	Fraction <3.2 mg/L	Fraction < 5mg/L
All	20,927	5.42	0.00029	0.0467
Sub-sample	5,316	5.42	0.00038	0.0478

These summaries suggest we are not changing the important features of this dataset by sub-sampling this 15 min data to 1 hour.

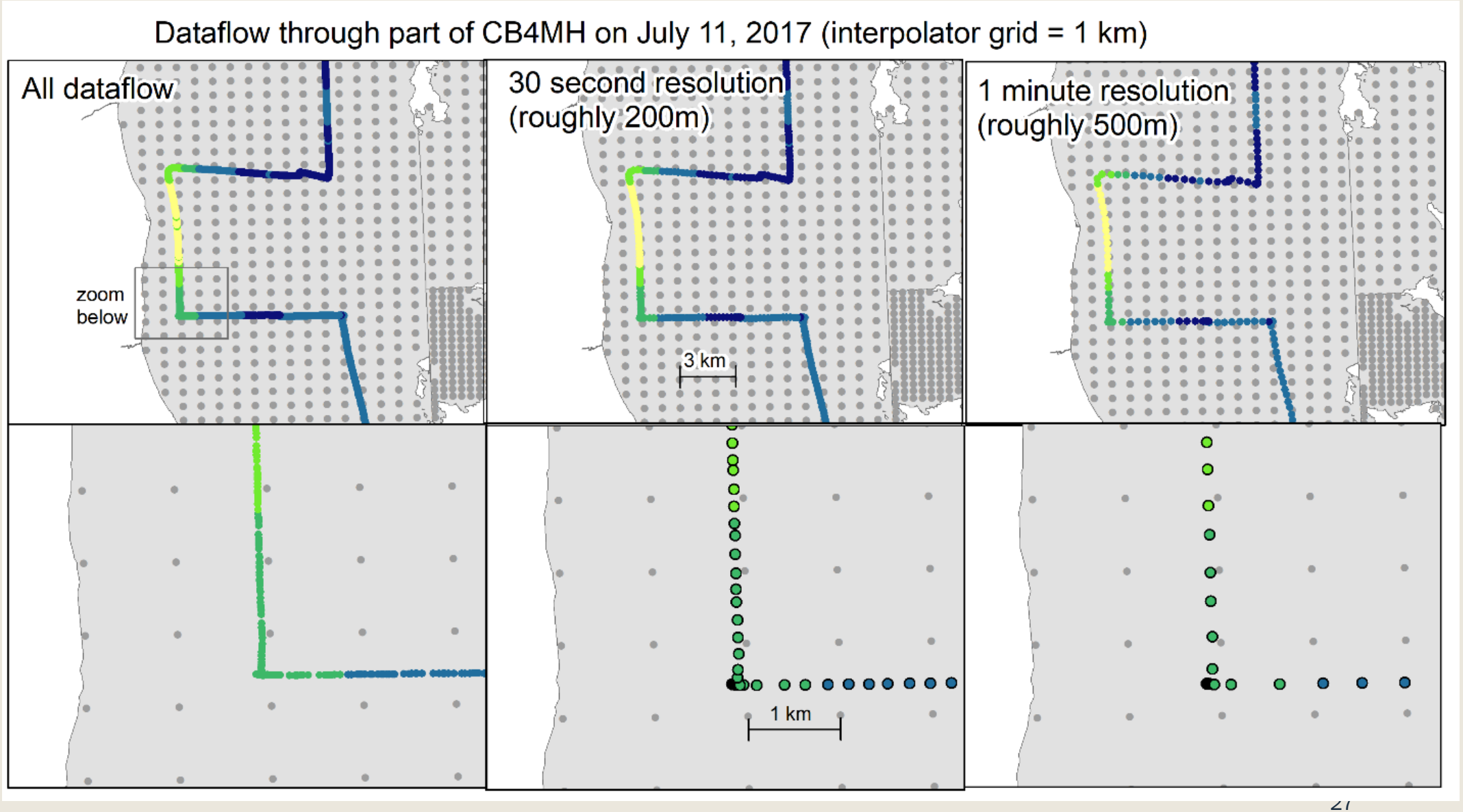
For this part only

Mean mid-day space-and-time interpolation

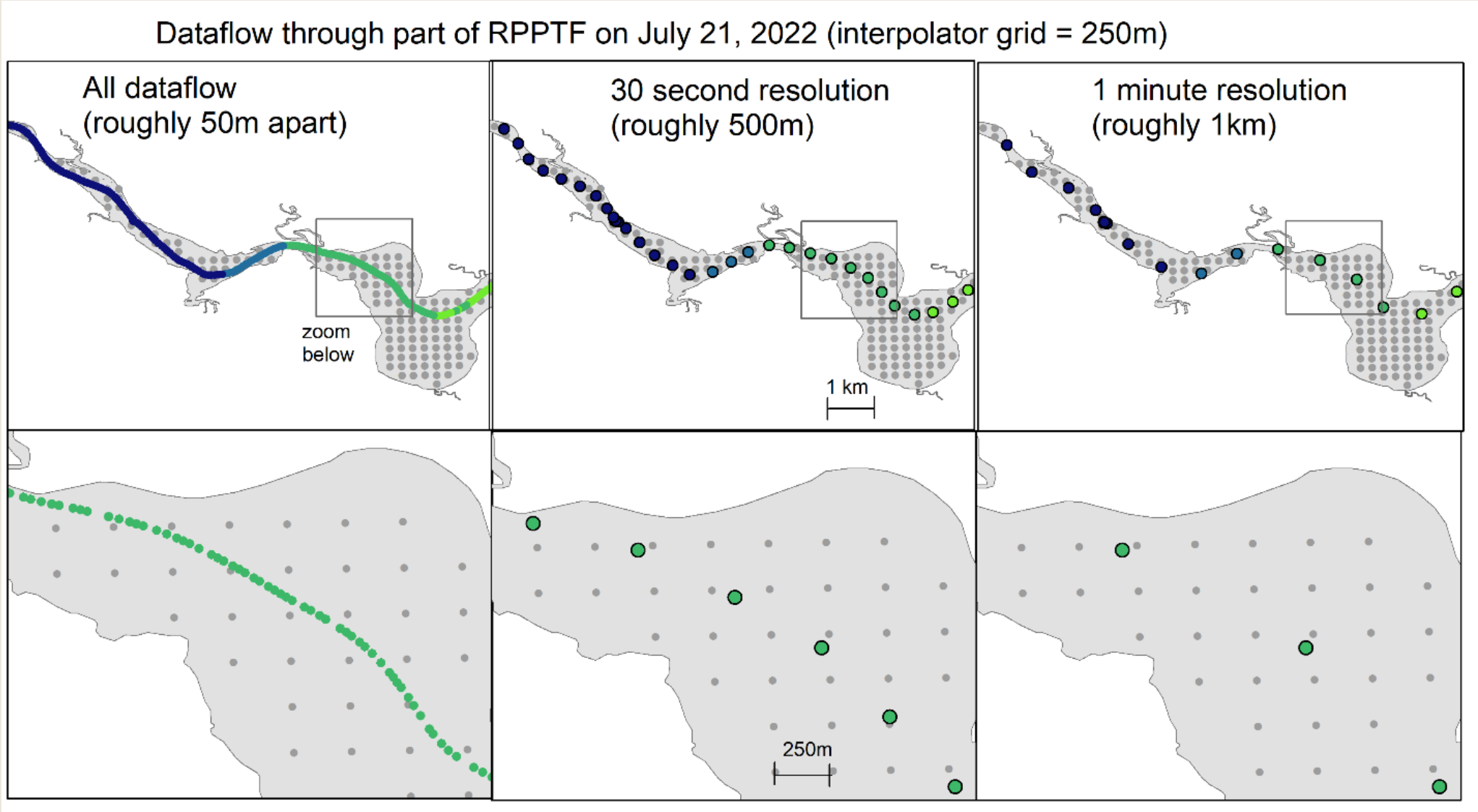
Dataflow subsampling

- Dataflow is different from ConMon in that the location is changing as high frequency data is collected.
- To subsample the Dataflow, we are suggesting to retain a sample every 500m or 1km. This for consistency with the interpolator grid throughout most of the bay.
- Testing is needed on the exact distance to use.
- Currently we tested sub-sampling in time, as a proxy for distance.

Dataflow: very dense spatial resolution



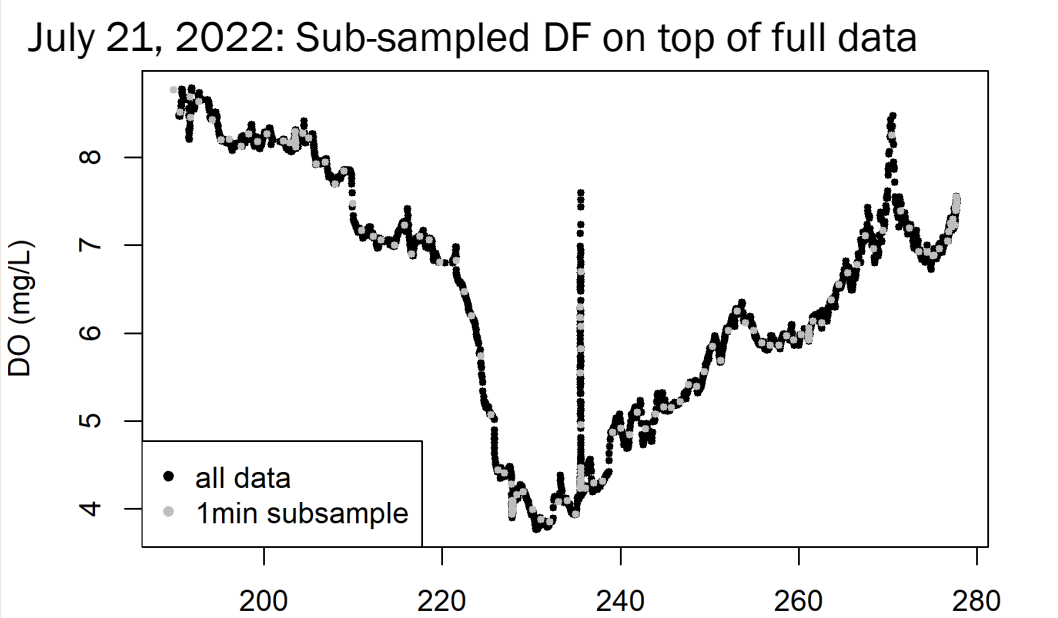
Dataflow: very dense spatial resolution

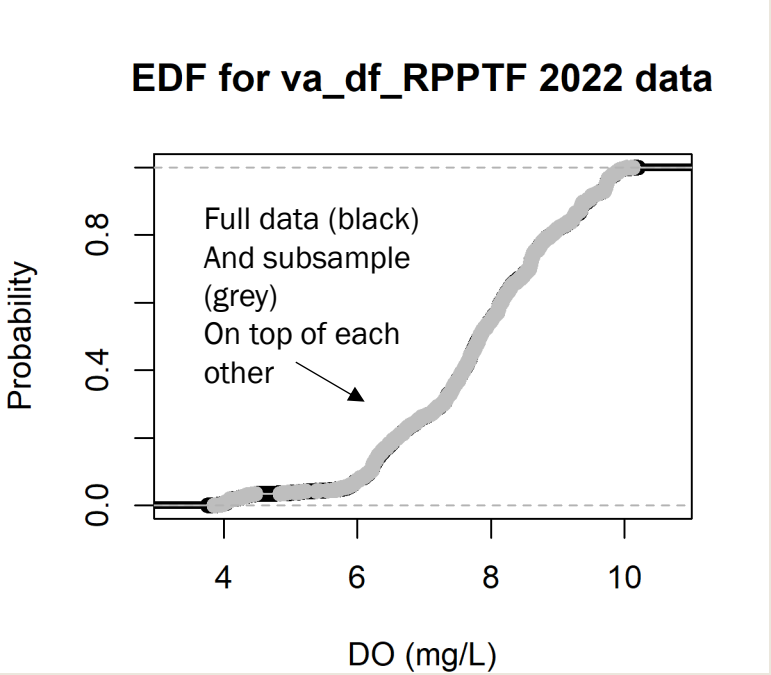


For this part only

Mean mid-day space-and-time interpolation

Subsampling impact: RPPTF dataflow in 2022





Data set	count	10th percentile	Fraction <3.2 mg/L	Fraction < 5mg/L
All	29,295	6.20	0	0.0374
Sub-sample	1,423	6.19	0	0.0379

Similar conclusion for sub-sampling dataflow. However, it is likely we will subsample based on distance instead of time in the final application.