

# 4D Interpolator – Day-to-Day Correlation

Bay Oxygen Research Group  
July 1, 2026

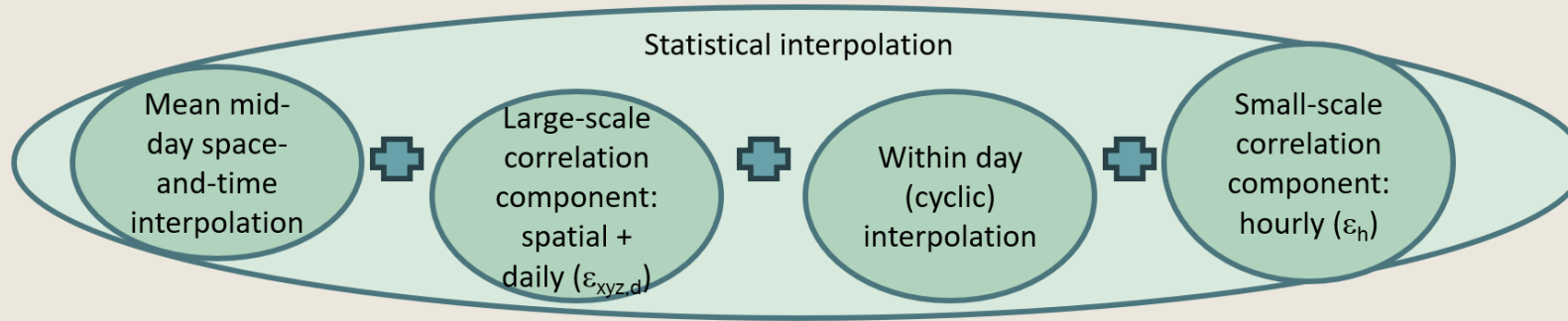
Jon Harcum<sup>1</sup> for development team:  
Elgin Perry<sup>2</sup>, Rebecca Murphy<sup>3</sup>, Breck Sullivan<sup>4</sup>, Peter Tango<sup>4</sup>

<sup>1</sup>Tetra Tech, <sup>2</sup>Statistics Consultant, <sup>3</sup>UMCES at CBP, <sup>4</sup>USGS at the CBP

# Outline

- Where day-to-day correlation fits within the 4-D interpolator?
- What “daily memory” means and how it affects simulated time series?
- How we estimate daily memory from observed monitoring data?
- Historical patterns, candidate predictors, and selected parameterization.

# Building a Realistic DO Time Series in Components

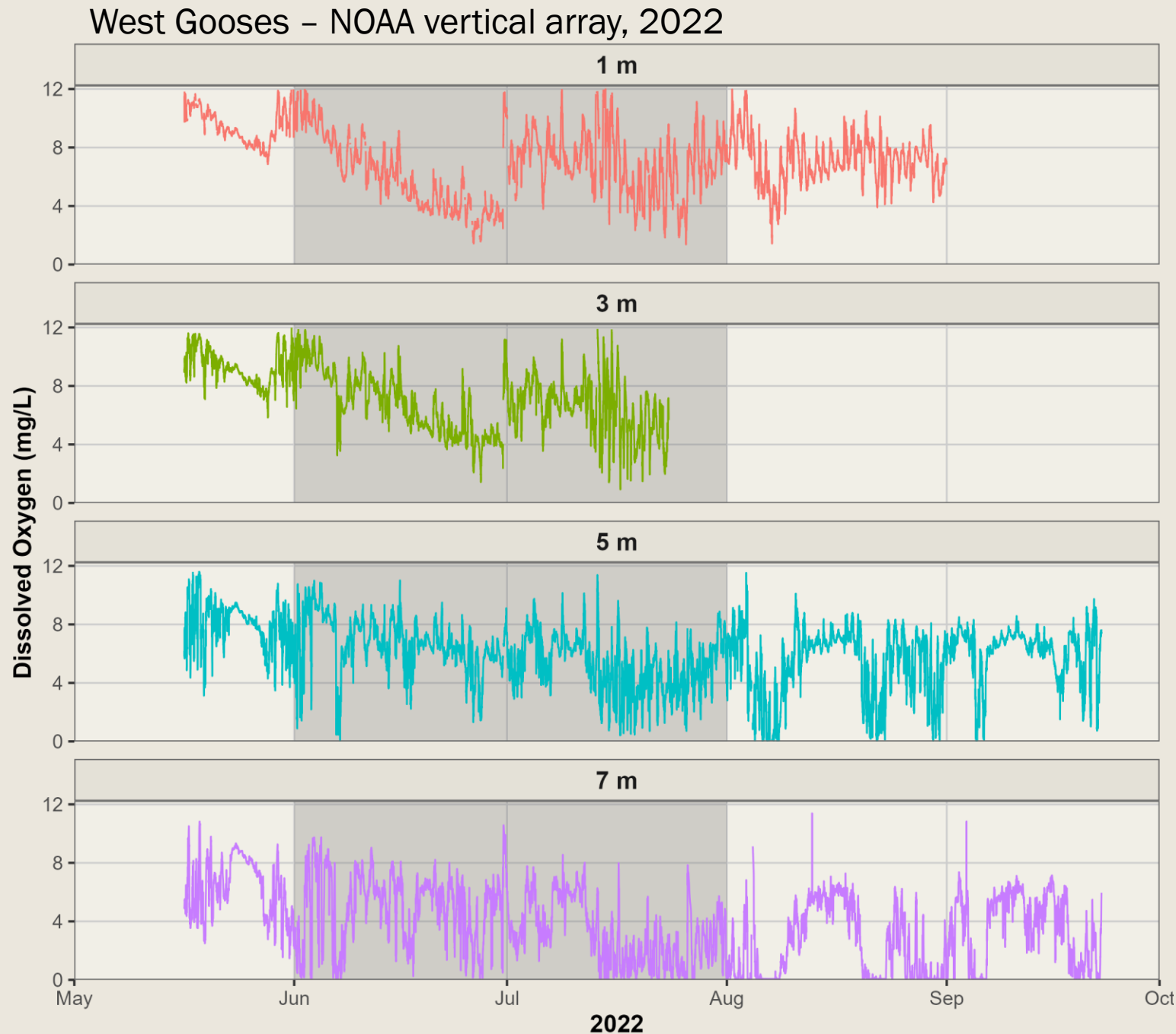


- Begin with the **expected mid-day DO pattern** across season, location, and depth.
- Add **large-scale correlated departures** from that pattern:
  - *spatial correlation*
  - *day-to-day correlation*
- Then add the **within-day cycle** and **hourly variation**.

**Today's focus:** How day-to-day correlation gives simulated conditions memory.

## The Data Already Show Why Correlation Matters

- Nearby days carry useful information, but conditions can change quickly.
- Depths move together, but deeper waters often behave differently.
- These patterns motivate separate day-to-day and depth correlation terms.
- **TODAY:** we focus on day-to-day correlation.

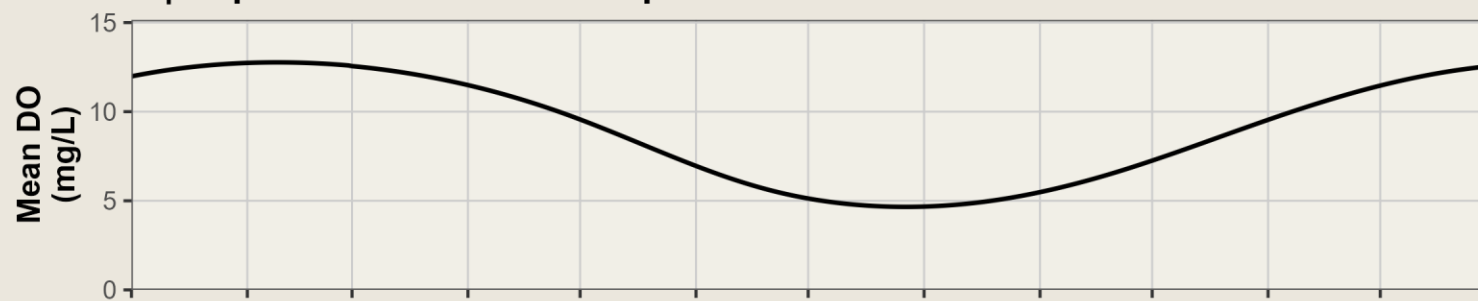


# Building One **Simulated** Mid-Day DO Series

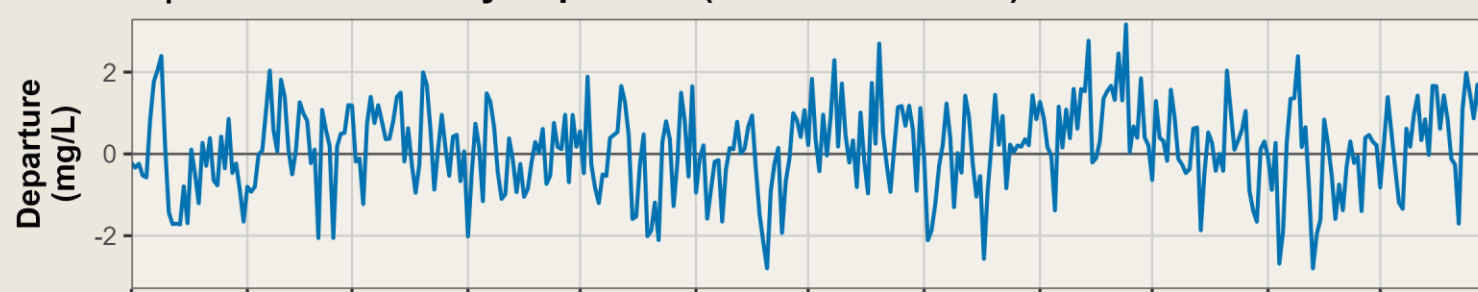
- Simulated mid-day DO =  
expected GAM value +  
correlated daily departure
- The GAM determines the expected level; the residual component determines the daily departure from it.

Decomposing a simulated dissolved-oxygen series  
Black line = GAM mean; colored line = departure or resulting simulation

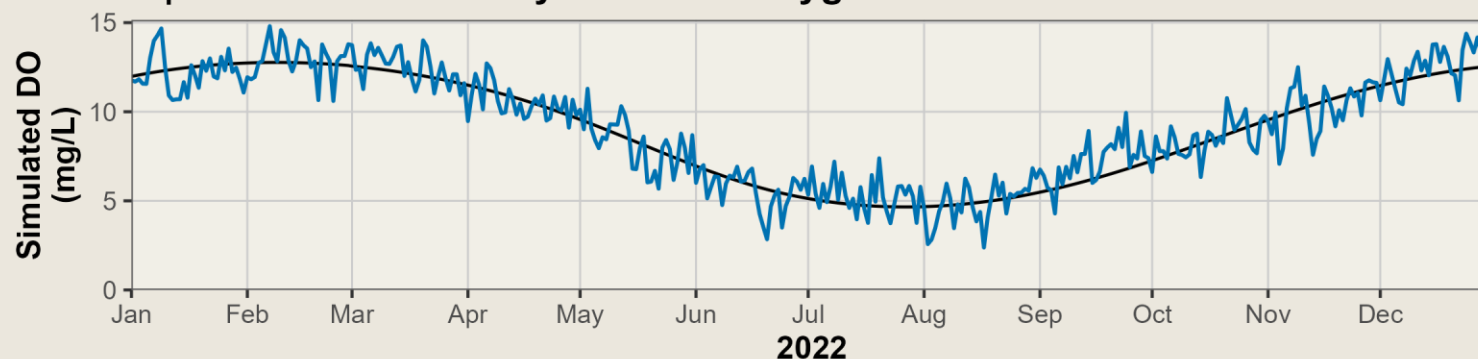
## 1 | Expected seasonal GAM pattern



## 2 | + correlated daily departure (correlation = 0.4)



## 3 | = simulated mid-day dissolved oxygen

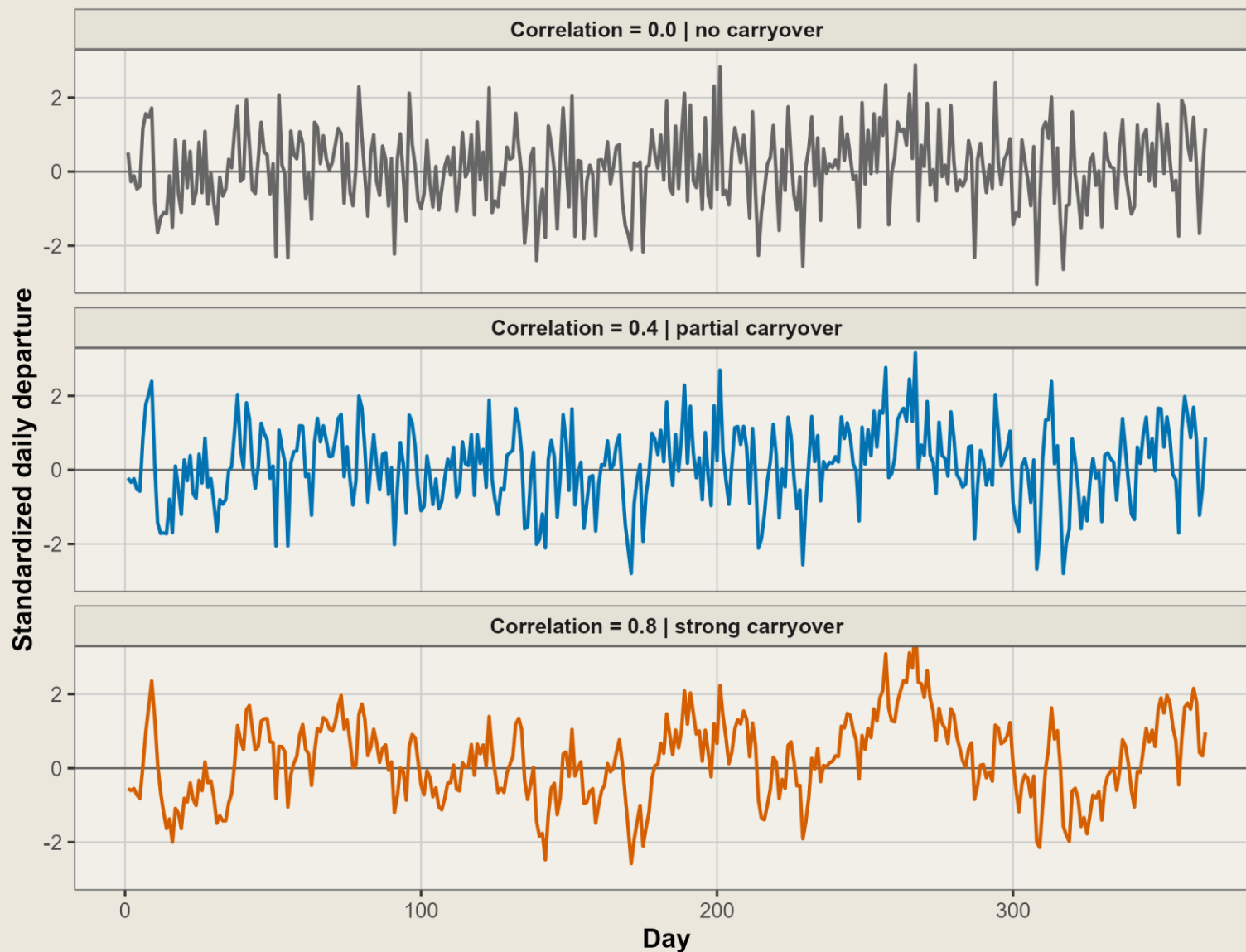


# Day-to-Day Correlation Represents Memory

- Higher correlation produces longer runs on the same side of the expected value.

## The missing ingredient is day-to-day memory

Each process has the same theoretical long-run  $N(0,1)$  distribution

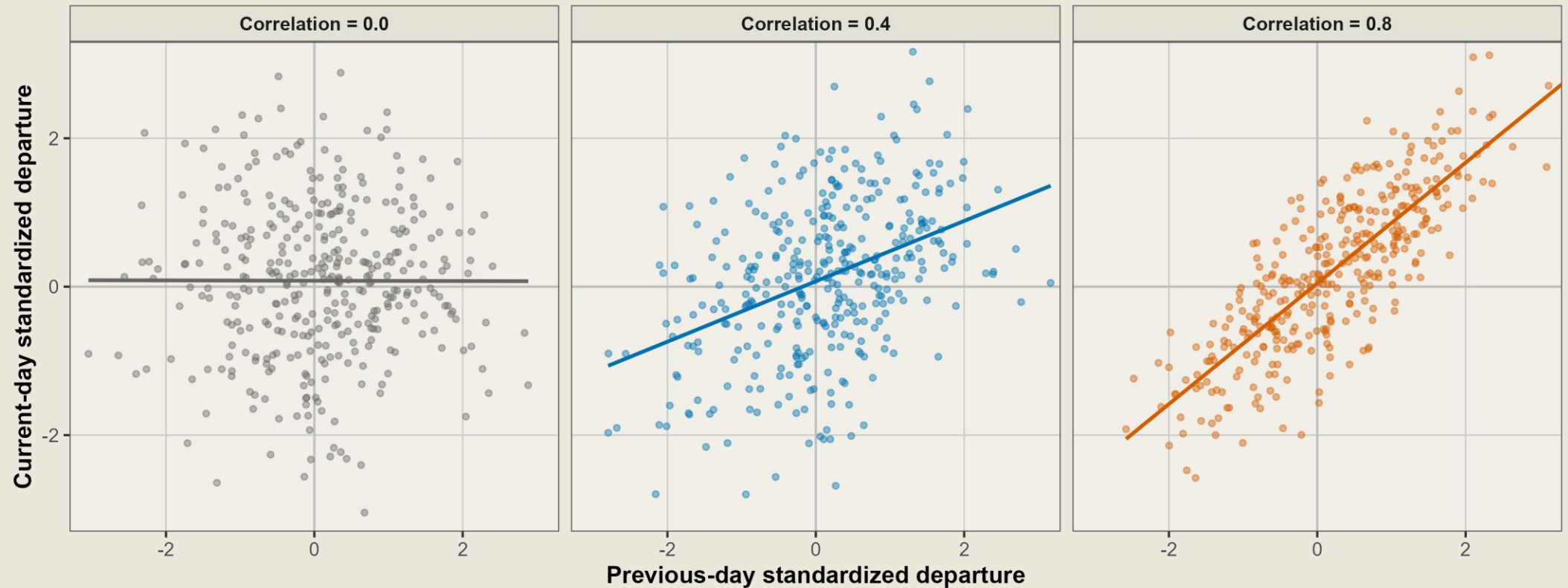


# Measuring Memory from One Day to the Next

- Each point pairs **yesterday's departure** with **today's departure**.
- Greater diagonal alignment indicates stronger lag-1 correlation.

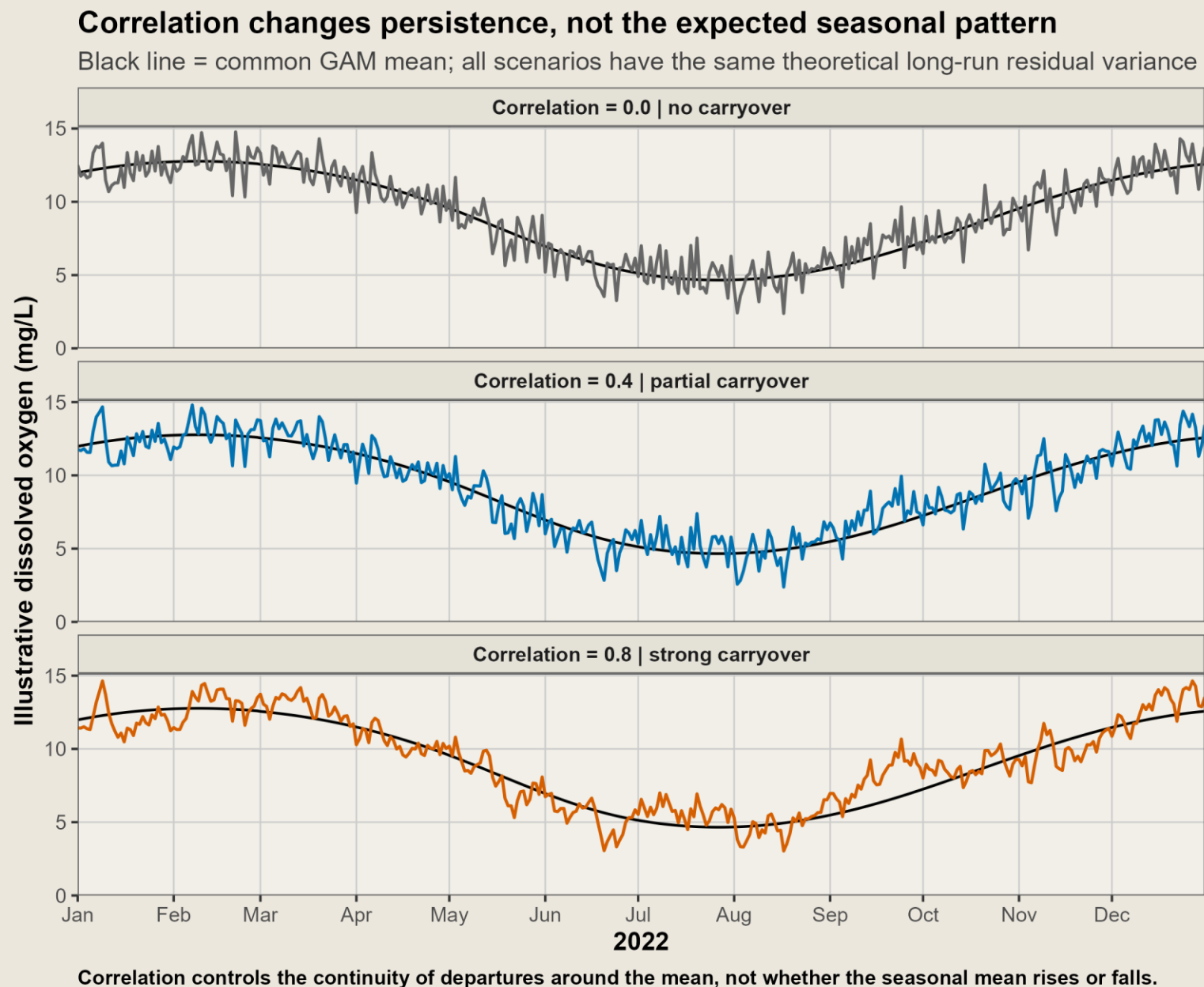
## Day-to-day correlation measures how strongly yesterday carries into today

Greater alignment around the diagonal indicates stronger lag-1 correlation

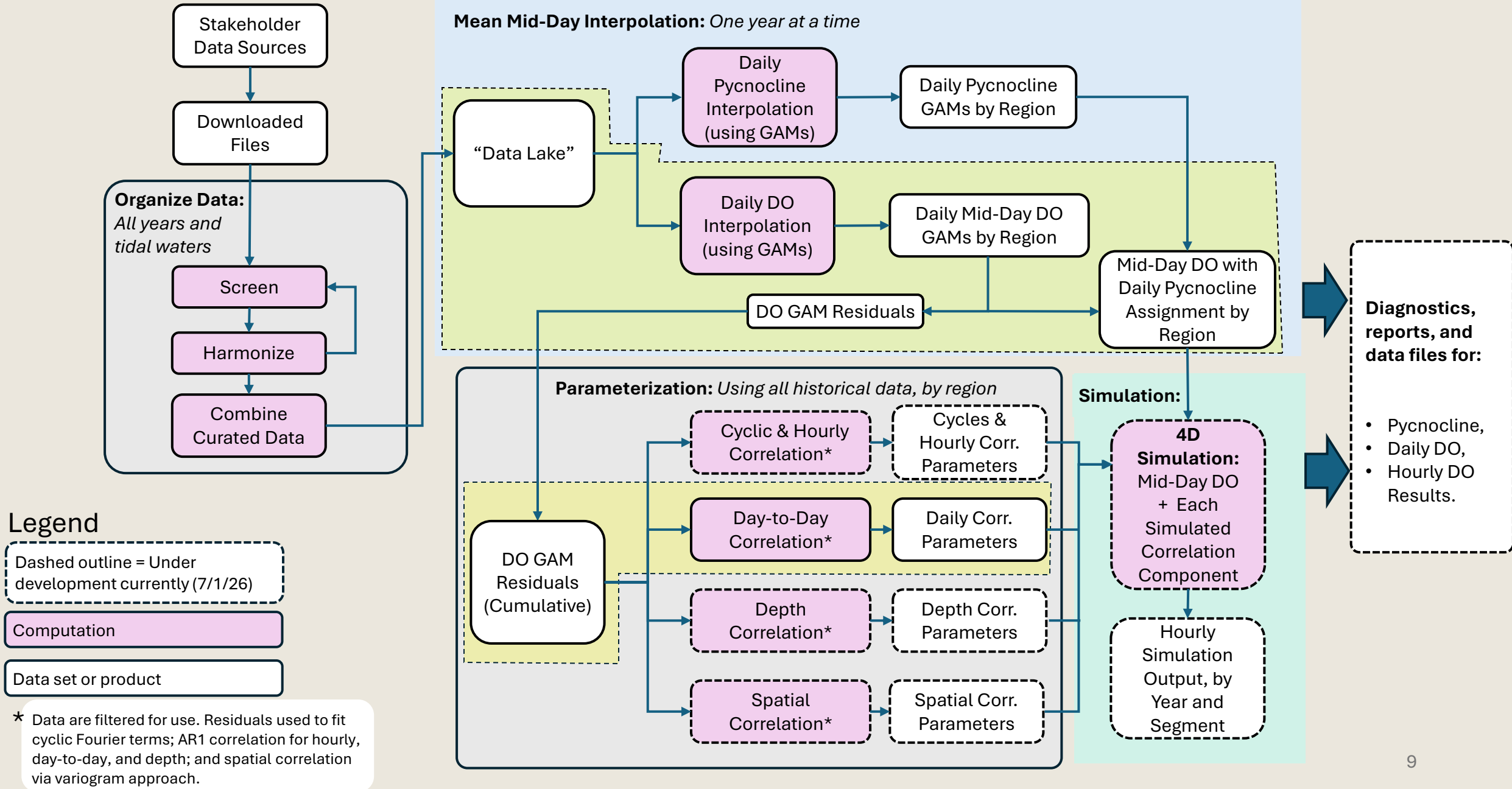


## Why Parameterize Day-to-Day Correlation?

- The black GAM mean is identical in every simulation.
- Correlation controls the **continuity of departures** around that mean.
- Realistic correlation is needed to represent realistic multi-day conditions.



# 4-D Interpolator Workflow



"Data Lake"

Daily DO  
Interpolation  
(using GAMs)

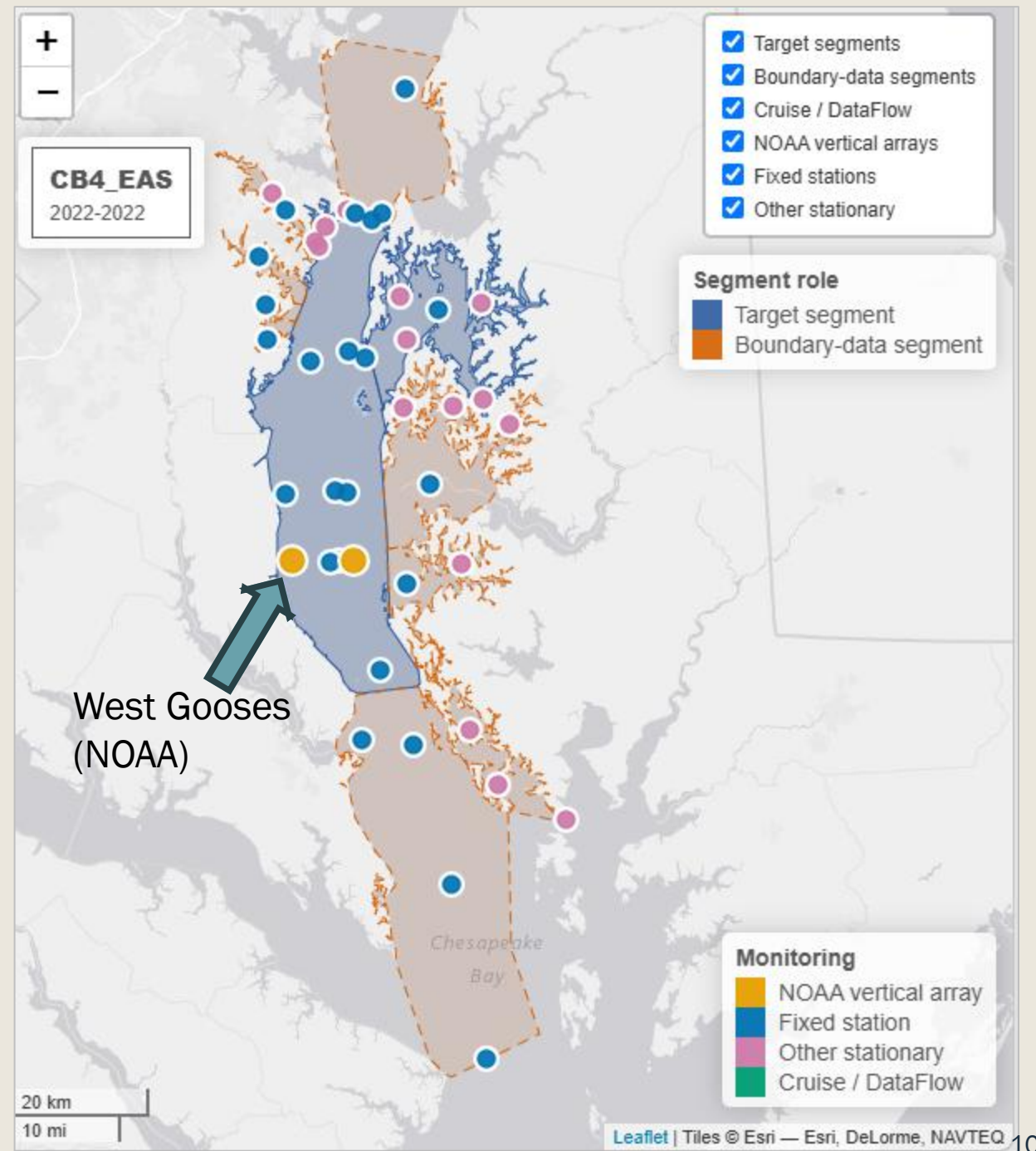
Daily Mid-Day DO  
GAMs by Region

DO GAM Residuals

Mid-Day DO with  
Daily Pycnocline  
Assignment by  
Region

## Mid-Day GAMs Are Fit by Interpolator Region

- 27 overlapping segment interpolator regions.
- Target segments define the primary interpolation area.
- Boundary-data segments provide supporting observations near regional edges.
- West Gooses is shown within the **CB4\_EAS** example region.



"Data Lake"

Daily DO  
Interpolation  
(using GAMs)

Daily Mid-Day DO  
GAMs by Region

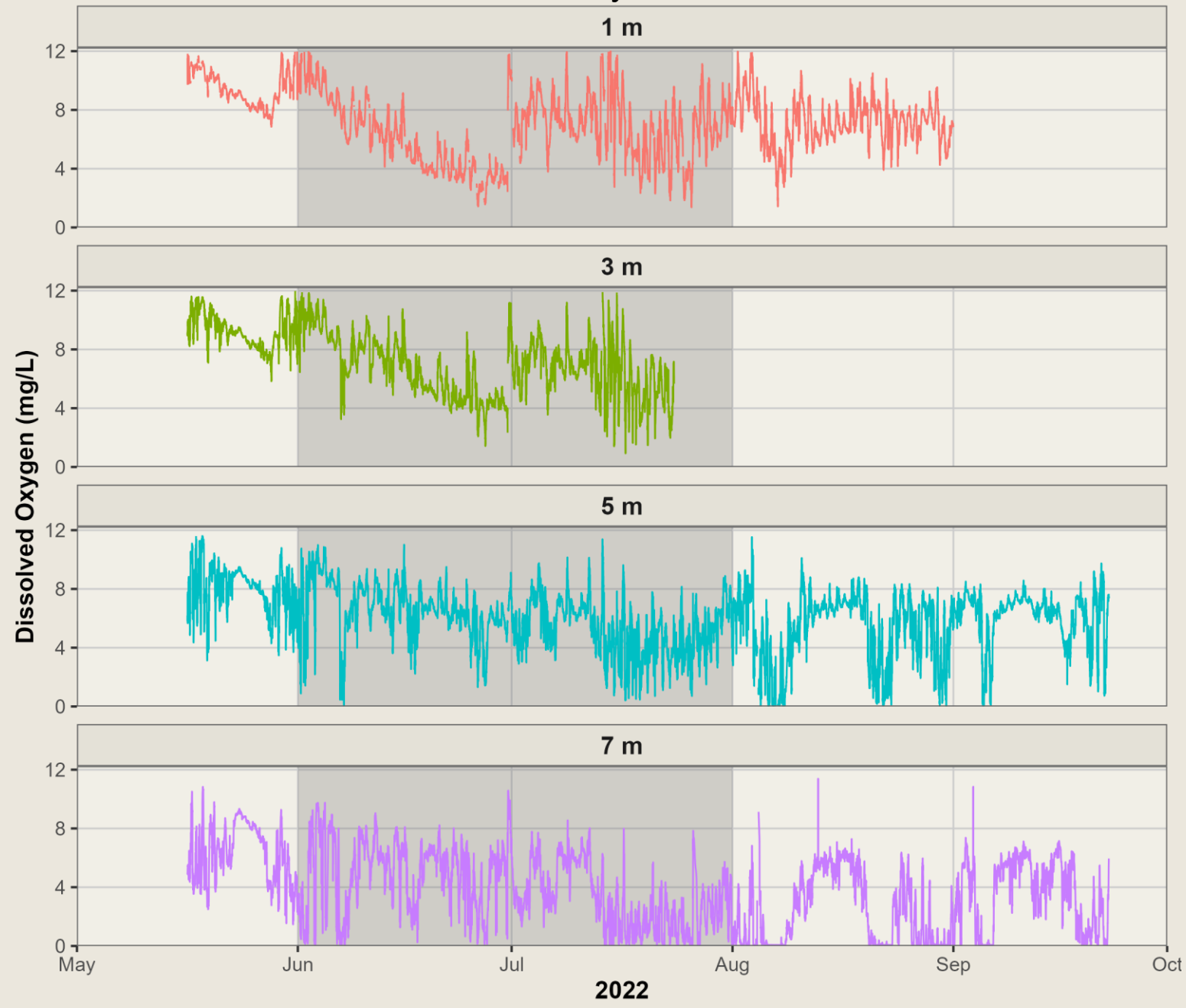
DO GAM Residuals

Mid-Day DO with  
Daily Pycnocline  
Assignment by  
Region

## Begin with the Available DO Record

- Includes CBP DataHub, Tier 3 CMC, EOTB, VECOS, and NOAA.
- High frequency data are subsampled to Observation Units.
  - 10- and 15-minute data subsampled to hourly observations.
  - DataFlow subsampled to 500m spacing.
- Preserves major information in the monitoring record while preventing unusually dense data sources from dominating the regional GAM fit.

West Gooses – NOAA vertical array, 2022



"Data Lake"

Daily DO  
Interpolation  
(using GAMs)

Daily Mid-Day DO  
GAMs by Region

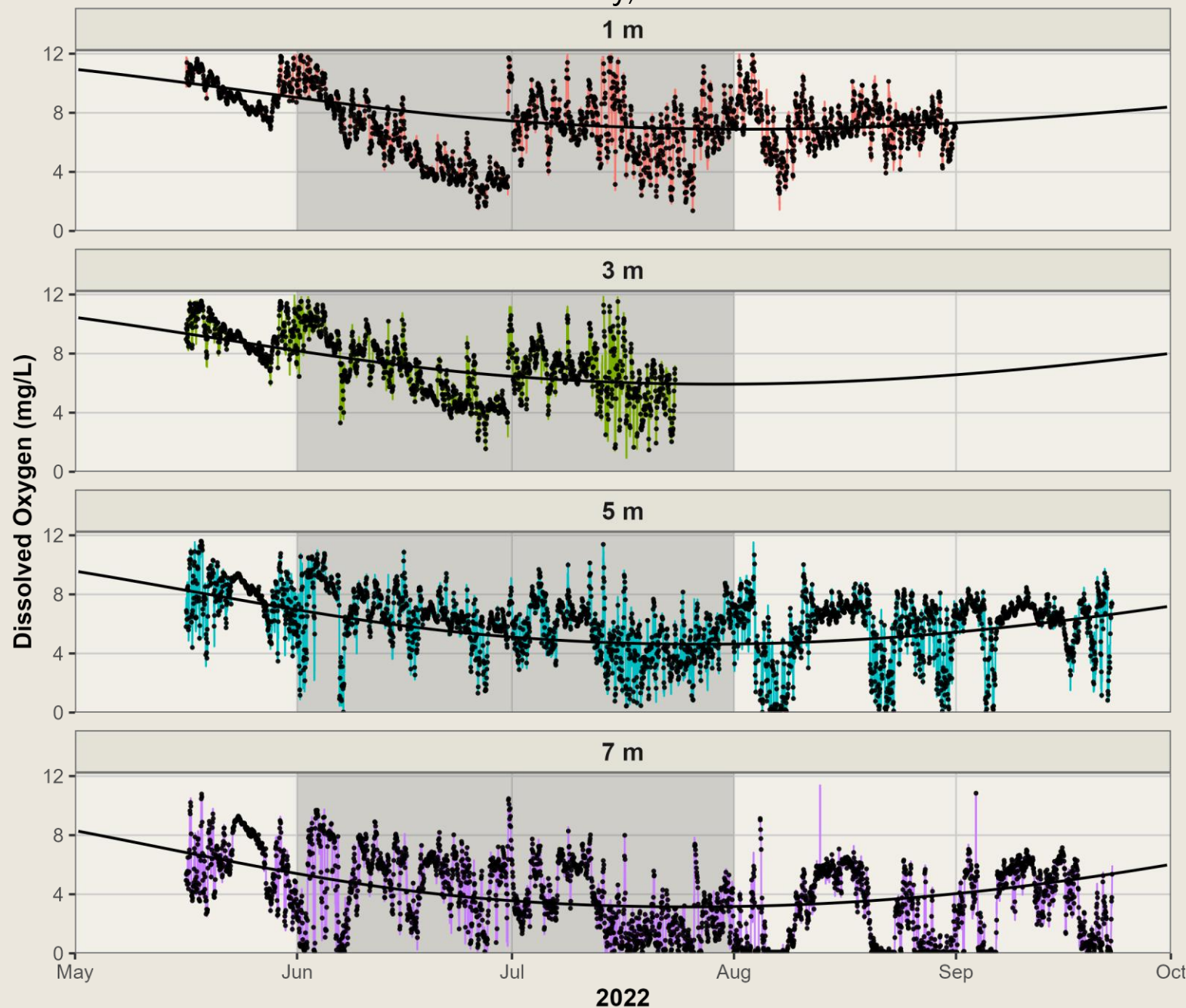
DO GAM Residuals

Mid-Day DO with  
Daily Pycnocline  
Assignment by  
Region

## Fit the Regional Mid-Day GAM to the Subsampled Data

- **Black points:** subsampled GAM-fitting observations.
- **Black line:** expected mid-day DO from the regional GAM.
- The GAM captures broad seasonal, spatial, and depth-related patterns.
- Short-term departures remain in the data.

## West Gooses – NOAA vertical array, 2022



*For clarity, plots show observed units (mg/L); analyses were performed with beta-logit-transformed DO, so results may differ slightly though patterns are unchanged.*

"Data Lake"

Daily DO  
Interpolation  
(using GAMs)

Daily Mid-Day DO  
GAMs by Region

DO GAM Residuals

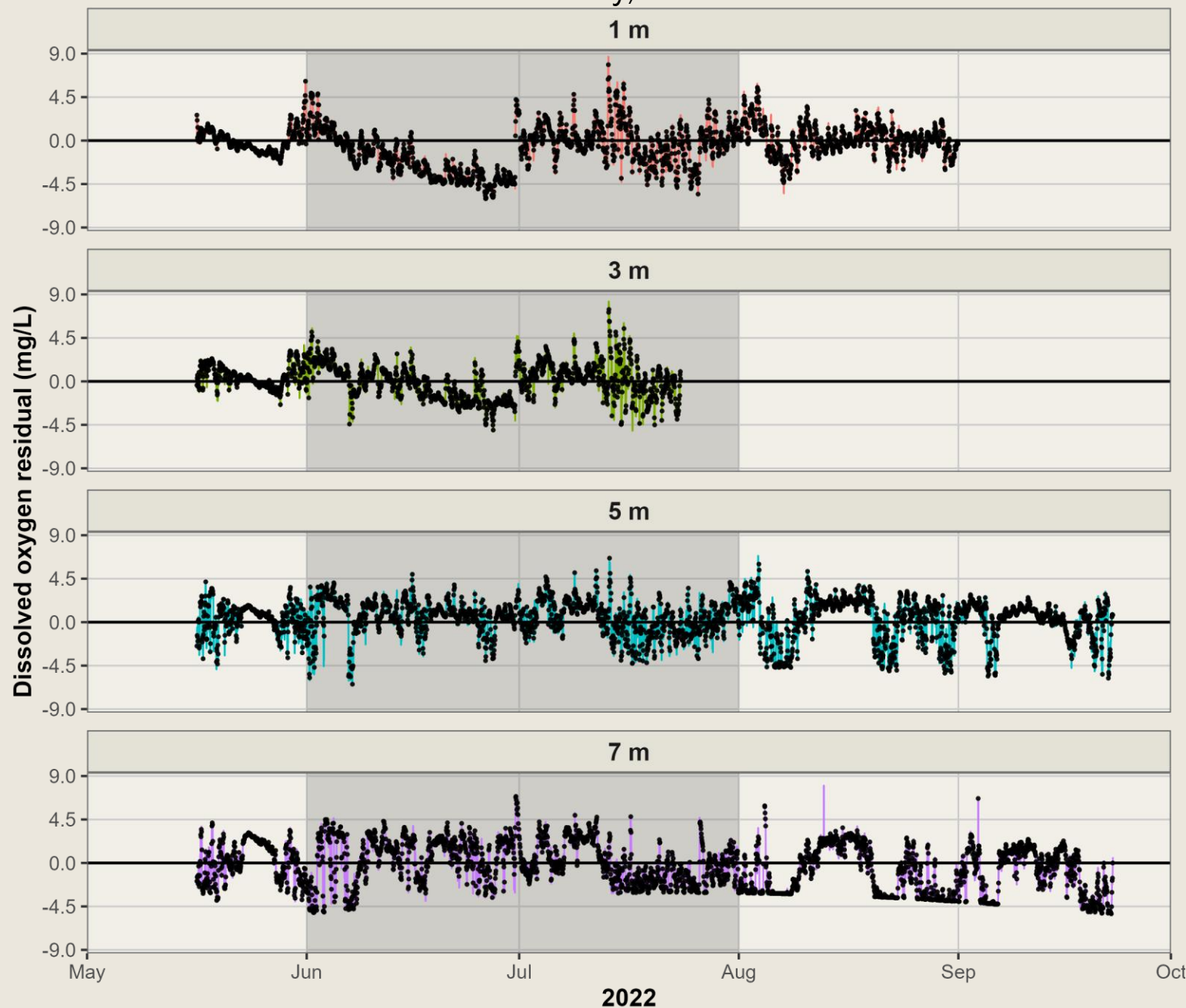
Mid-Day DO with  
Daily Pycnocline  
Assignment by  
Region

## GAM Residuals Retain the Unexplained Temporal Structure

Residual =  
observation –  
GAM prediction

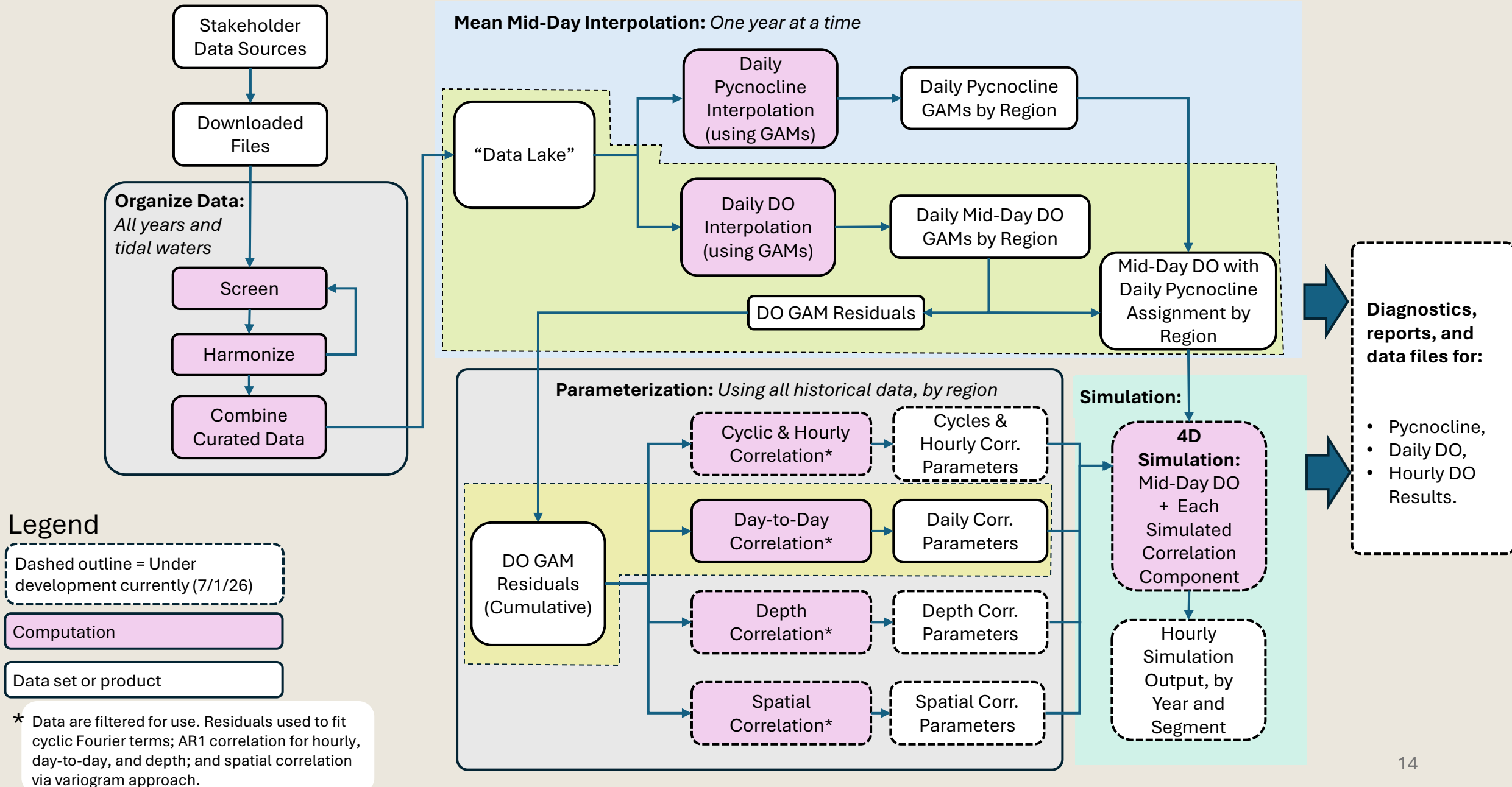
- Residuals are centered around the expected mid-day pattern.
- Positive residuals indicate conditions above the GAM expectation.
- Negative residuals indicate conditions below it.
- Their temporal structure is used to estimate correlation.

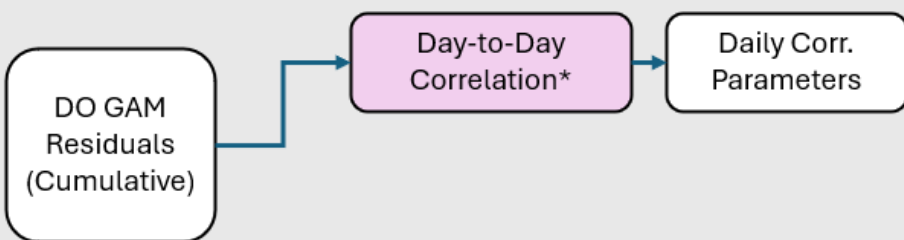
### West Gooses – NOAA vertical array, 2022



*For clarity, plots show observed units (mg/L); analyses were performed with beta-logit-transformed DO, so results may differ slightly though patterns are unchanged.*

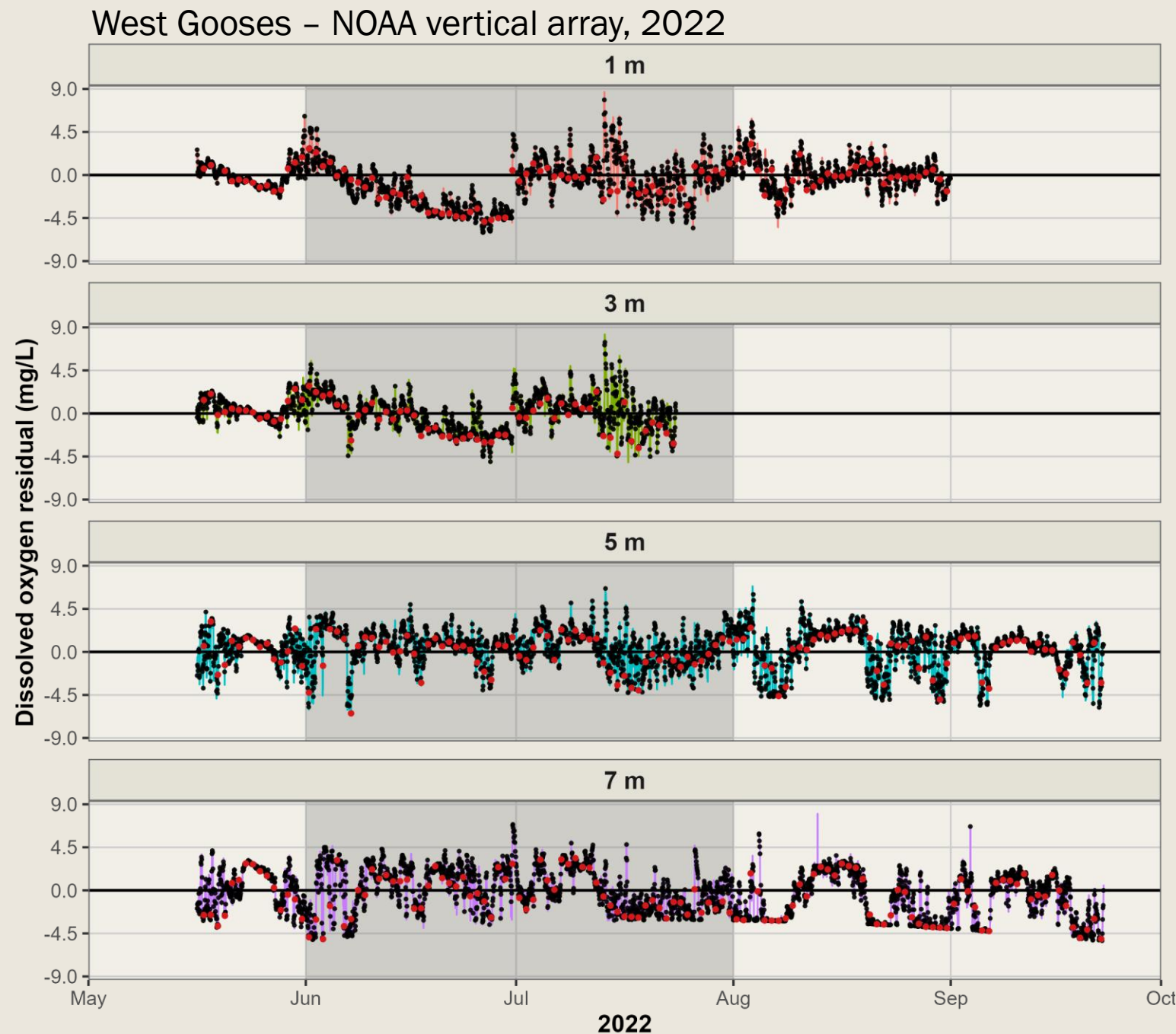
# 4-D Interpolator Workflow



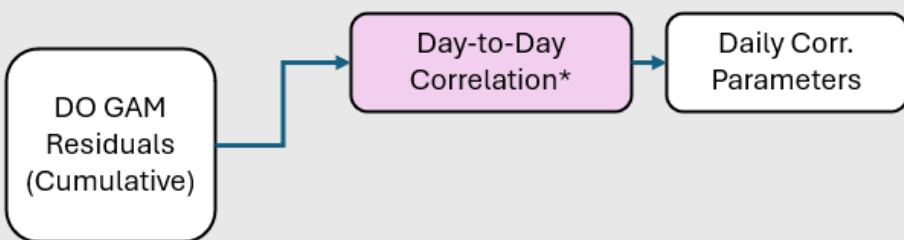


## Select One Residual per Day Near 11:00

- Begin with hourly residuals grouped by station and layer.
- Select the observation nearest **11:00**, within a  $\pm 2$ -hour window.
- Retain no more than one residual per calendar day.
- **Red points:** selected daily residuals.



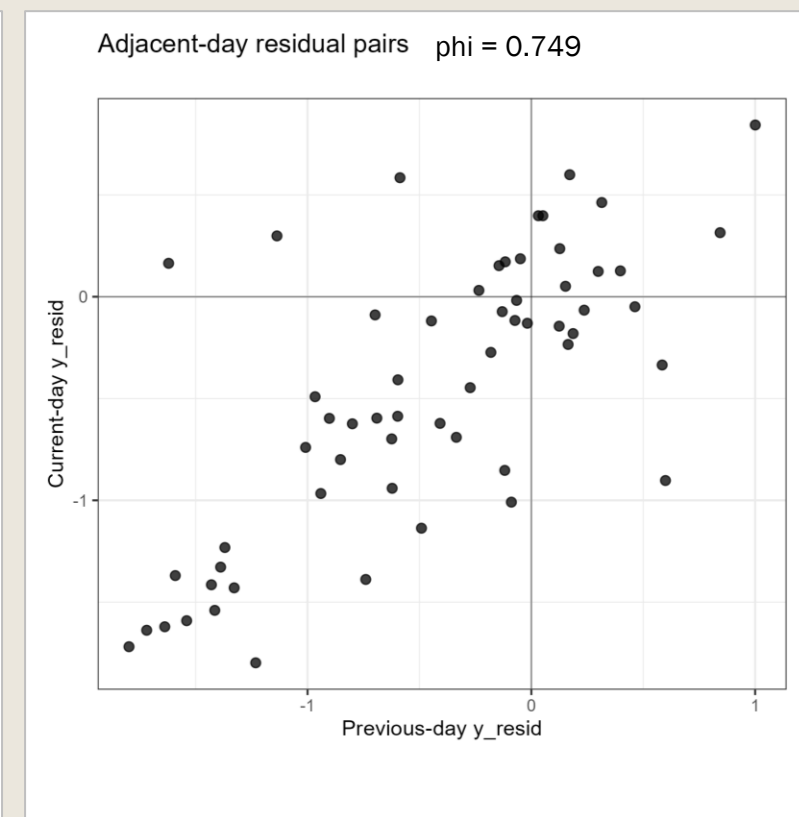
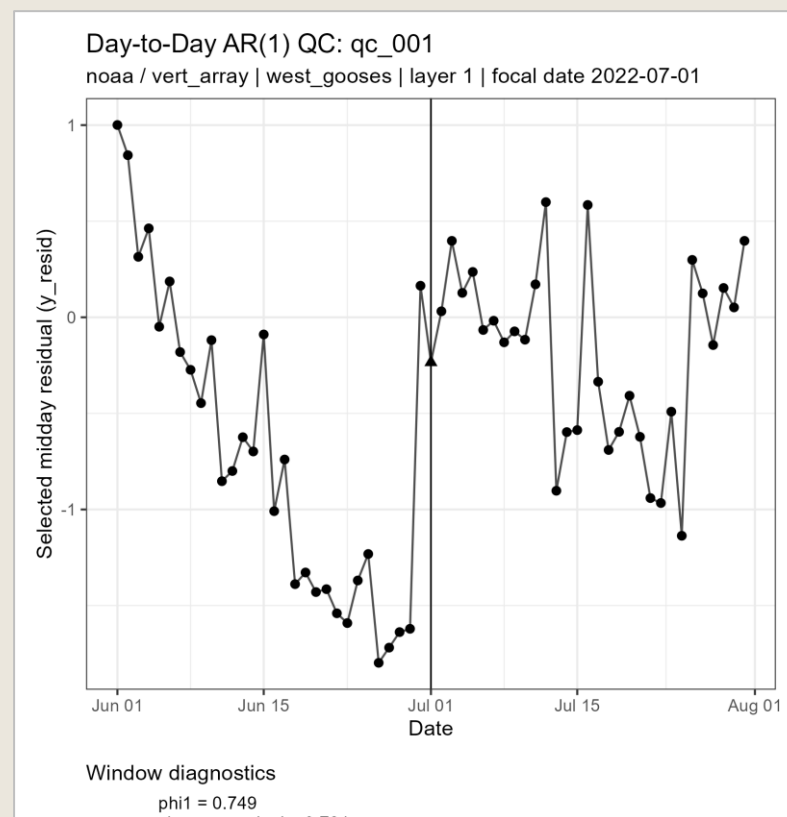
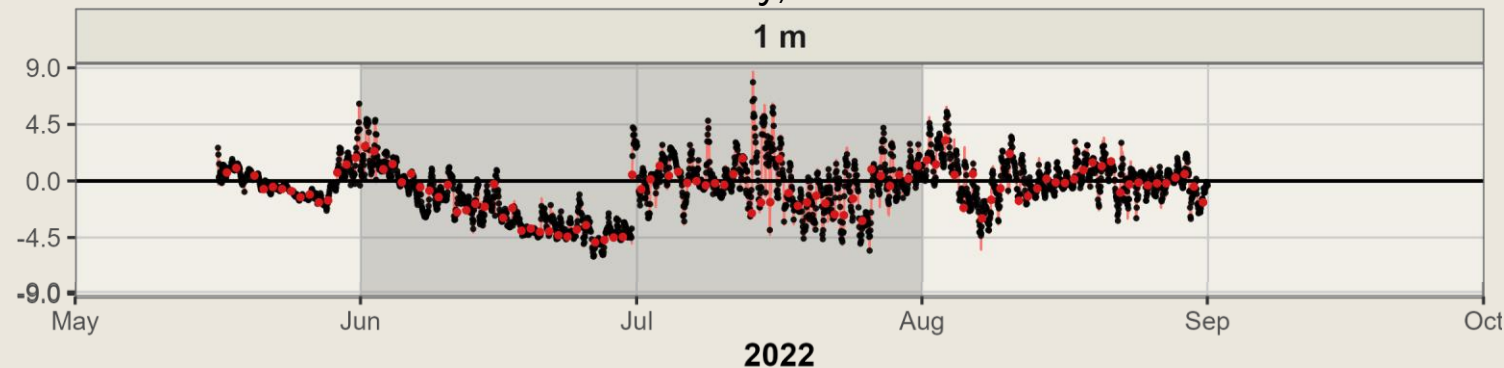
*For clarity, plots show observed units (mg/L); analyses were performed with beta-logit-transformed DO, so results may differ slightly though patterns are unchanged.*



## Estimate Lag-1 Correlation within a Local Moving Window

- Center a  $\pm 30$ -day window on each focal date.
- Require adequate coverage before and after the focal date.
- Pair each selected residual with the preceding day's residual.
- Estimate the local lag-1 correlation,  $\phi_1$ .
- **Example:** West Gooses, 1 m, July 1, 2022:  $\phi_1 = 0.749$

### West Gooses – NOAA vertical array, 2022



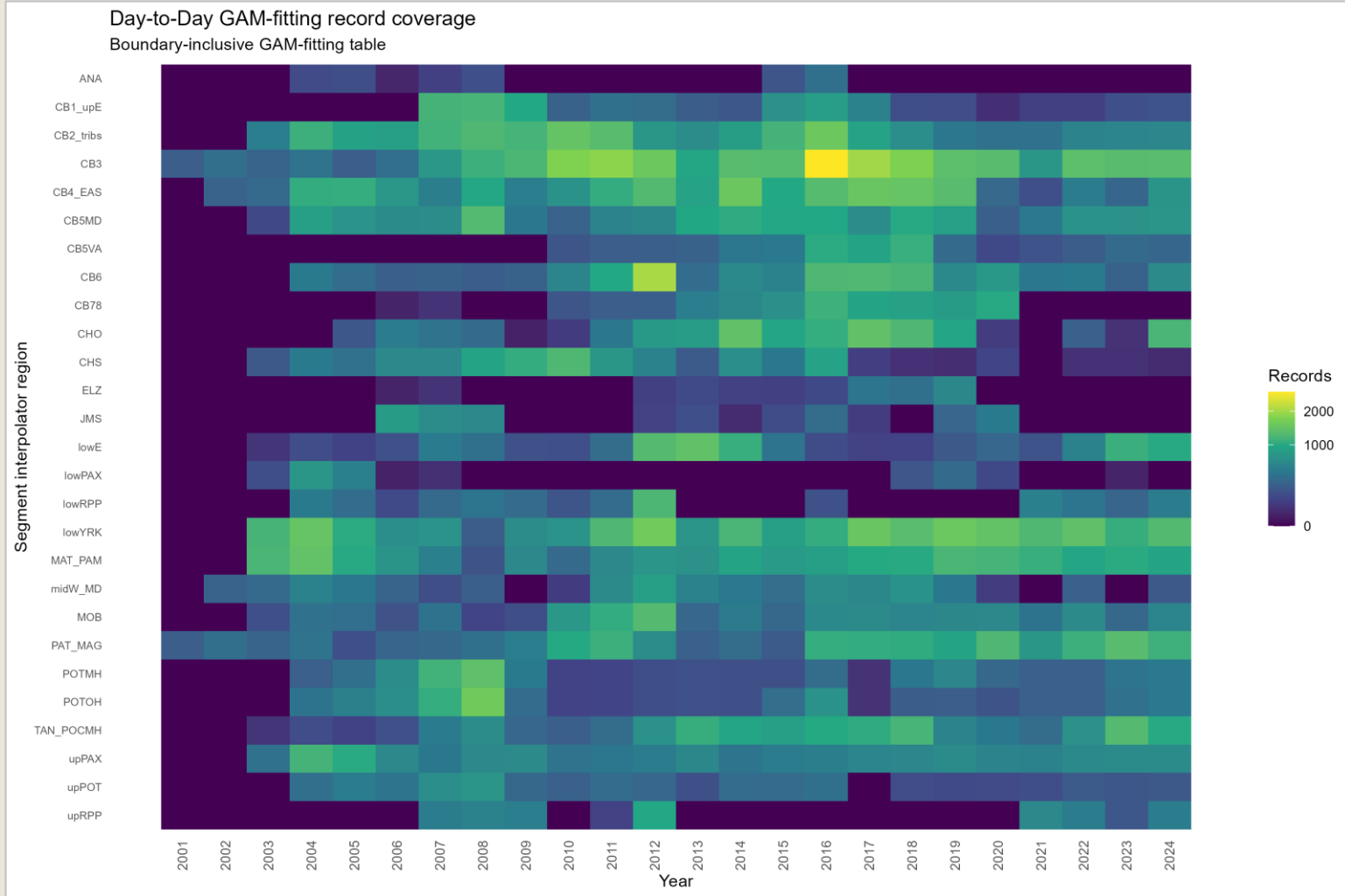
DO GAM  
Residuals  
(Cumulative)

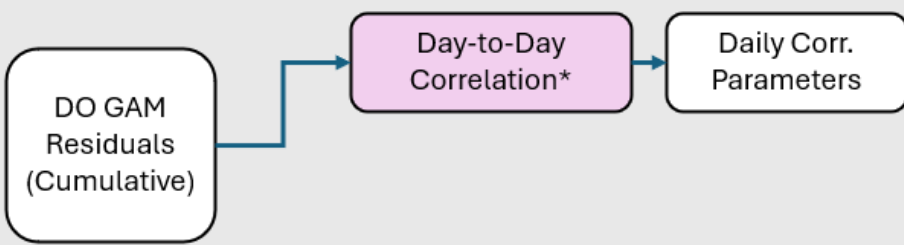
Day-to-Day  
Correlation\*

Daily Corr.  
Parameters

# Historical Data Provide Broad Regional Support

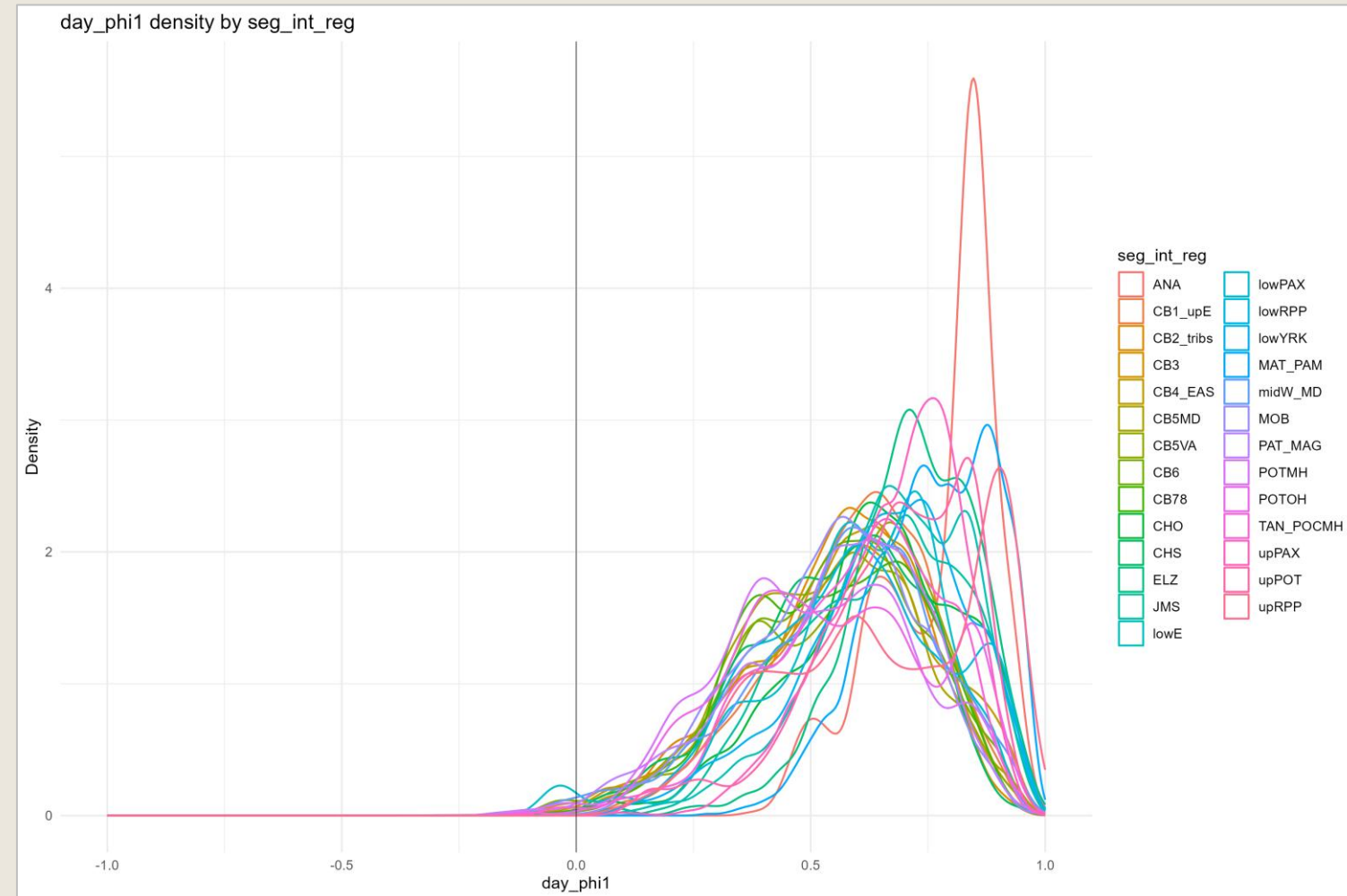
- 139,626 eligible target-region records.
- 24 years of monitoring data across 27 target regions.
- Boundary data expand the regional fitting table to:
  - 294,346 records
  - 27 interpolator regions
- Coverage varies by region, year, source, and monitoring type.

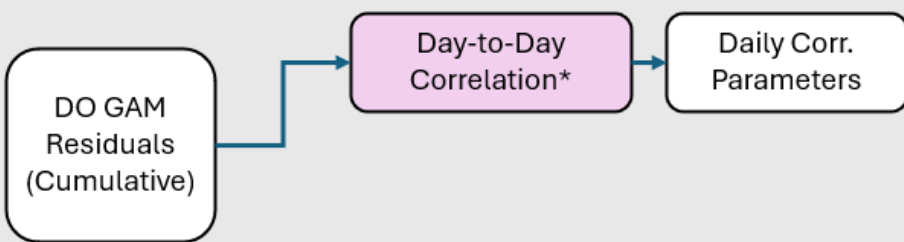




## Day-to-Day Residual Correlation Is Generally Positive

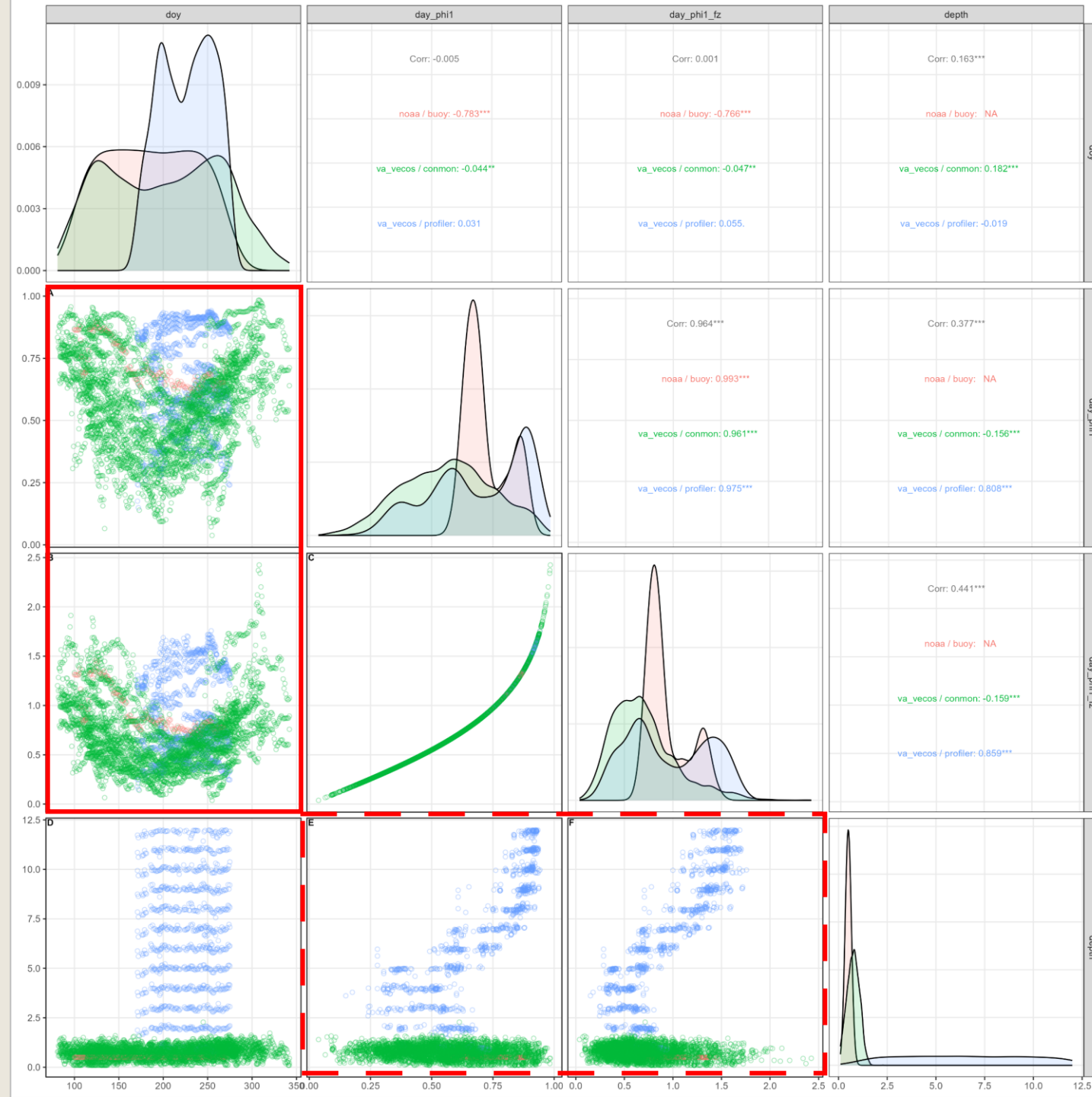
- Most local estimates indicate **moderate to strong positive daily memory**.
- Correlation varies:
  - seasonally within regions
  - among interpolator regions
- This supports **regional parameterization** rather than one Bay-wide value.
- Narrow or unusually high distributions must be interpreted in light of data coverage.

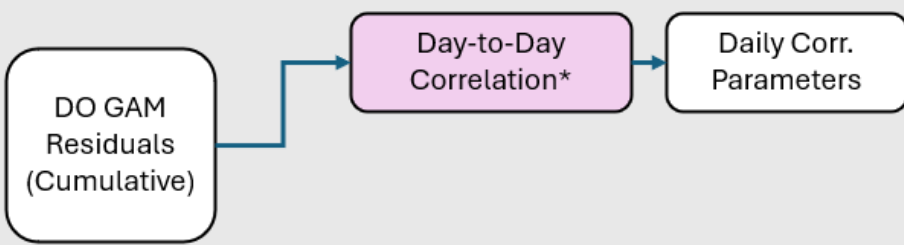




## Depth Structure Can Be Strong but Localized

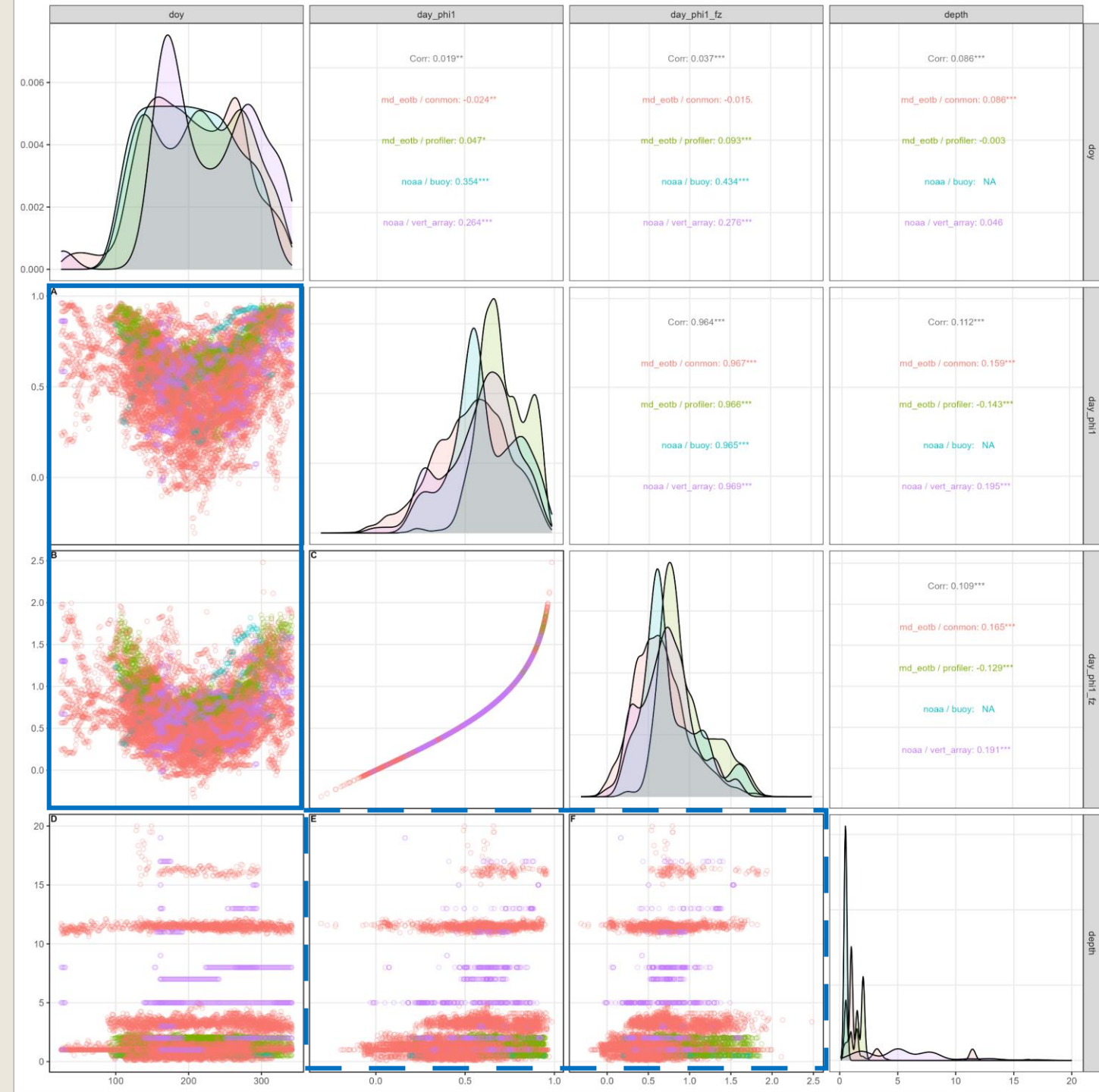
- Local correlation estimates are not evenly representative of season and potential other factors.
- GAM predictors can help adjust for this structure.
- In lowRPP, depth shows a visible relationship with daily correlation.
- But the depth signal is derived from a single profiler station.
- Strong local structure can inform model development without automatically becoming a common production term.





## CB4\_EAS Shows Seasonal Structure but Limited Sample Depth Support

- **More common example** – CB4\_EAS is the region used for the West Gooses examples.
- Day-to-day correlation changes through the year.
- Depth relationships are comparatively weak or inconsistent among data types.
- The broader evidence favors a seasonal term over a common depth term.



# Selected Day-to-Day Correlation Parameterization

- Regional day-to-day lag-1 correlation models are fitted on the Fisher-Z scale.

- **Selected common model form:**

$$\text{day\_phi1\_fz} \sim \text{s}(\text{doy}, \text{bs} = \text{"cc"}, k = 6)$$

- The restrained cyclic smooth captures broad seasonal structure.
- Other independent variables (e.g., localized depth and bottom-depth) were not included.
- Predictions are back-transformed to the correlation scale for simulation use.
- The current 4D interpolator incorporates regional variation but uses one constant day-to-day lag-1 correlation value within each region; future versions may use the fitted day-of-year predictions to incorporate seasonal variation.