Dissolved Oxygen Interpolation Update:

Applying simulation approach to CB4MH 2022

January 15, 2025, represented February 10, 2025

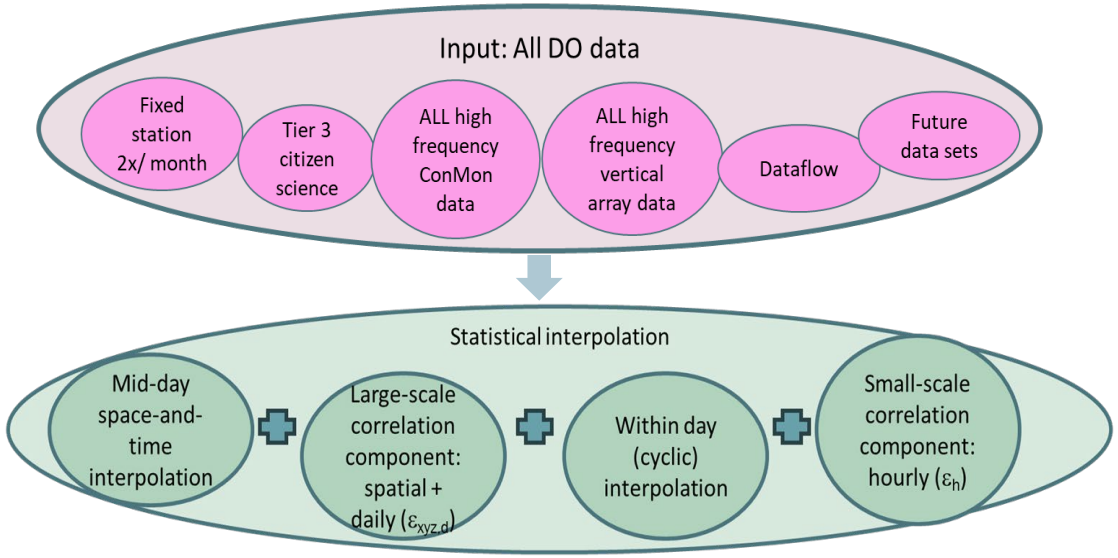
Rebecca Murphy (UMCES/CBP), Elgin Perry (consultant), Jon Harcum (Tetra Tech),
Breck Sullivan (USGS), & Peter Tango (USGS)
Contributions from Wes Slaughter (UMD)

Dissolved Oxygen Interpolation Update:

Applying simulation approach to CB4MH 2022

January 15, 2025

Rebecca Murphy (UMCES/CBP)
Elgin Perry (consultant)
Jon Harcum (Tetra Tech)
Contributions from Wes Slaughter (UMD)



CBP DataHub
- 1984-2022
- 835 stations
- 819k obs.

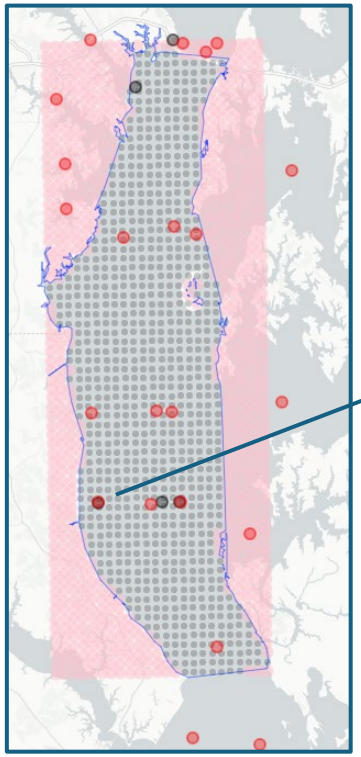
EOTB
- 2001-2022
- 126 stations
- 11,916k obs.

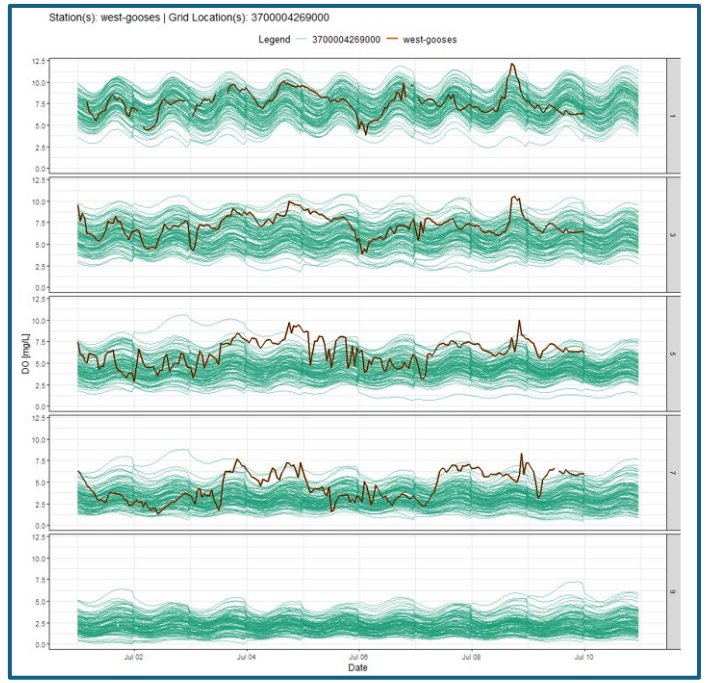
VECOS
- 2003-2022
- 54 stations
- 6,776k obs.

NOAA
- 2022-2023
- 5 stations
- 440k obs.

DATAFLOW*
- 2007-2008
- 576k obs.

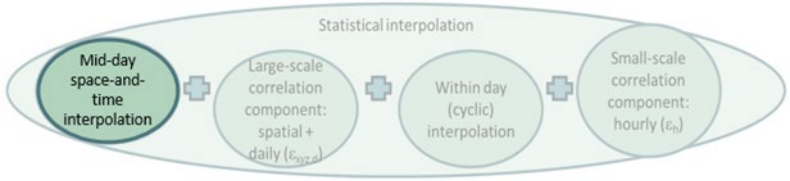
Tier 3†
- Citizen science





* *DATAFLOW: Potomac River Pilot*
† *Assessment would include Tier 3 citizen science data*

Mid-day space-and-time interpolation



Input ~ data from 2022

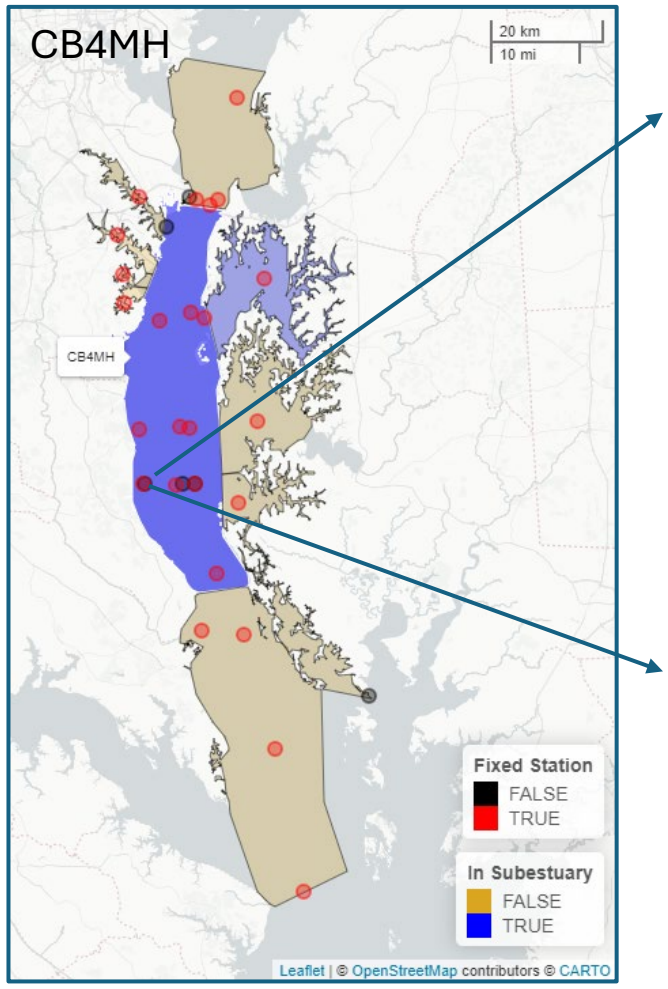
- CBP DataHub
- High frequency data subsampled to mimic CBP DataHub

GAM formula

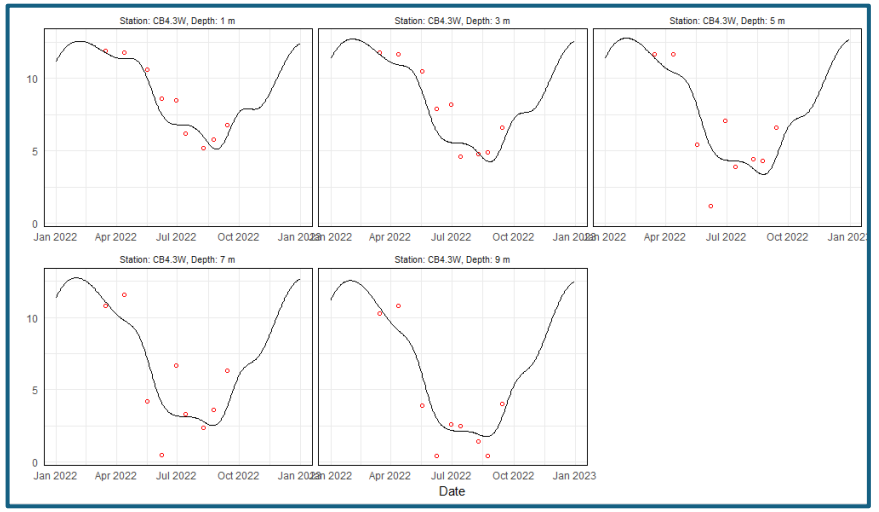
- Univariate smooth terms
  - hod
  - doy
  - depth
  - depth_b
  - wb_lon_km
  - wb_lat_km
- 2-Way Interactions
  - depth × depth_b
  - depth × doy
  - depth_b × doy
  - depth × hod
  - depth_b × hod
  - doy × hod
  - wb_lon_km × doy
  - wb_lat_km × doy
  - wb_lon_km × hod
  - wb_lat_km × hod
- 3-Way Interactions
  - wb_lat_km × depth × wb_lon_km

Output

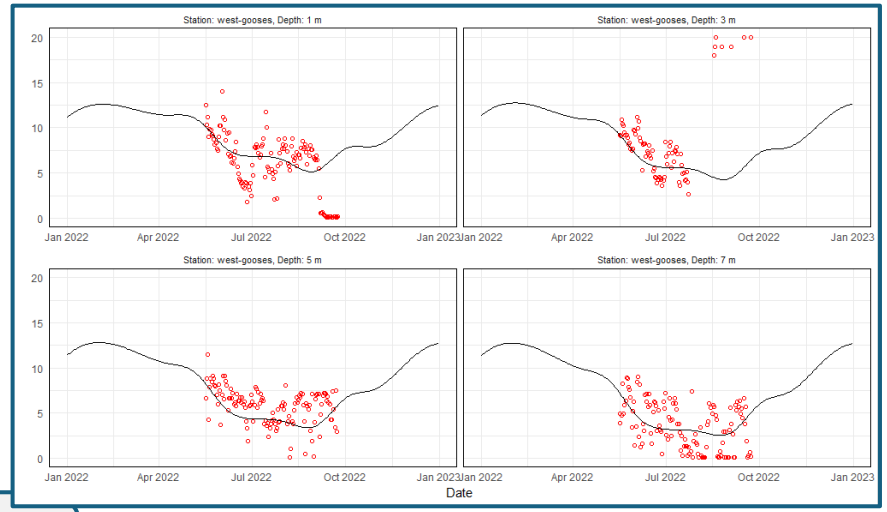
- Mid-day space and time interpolation GAM for 2022



CB4.3W observed DO with mean mid-day GAM



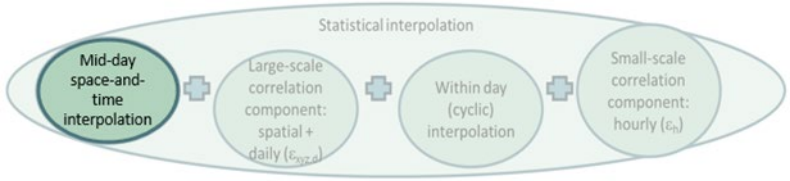
west-gooses observed DO with mean mid-day GAM



Terms

- *hod: time of day*
- *doy: day of year*
- *depth: sample depth*
- *depth_b: bottom depth*
- *wb_lon_km: distance along primary direction of flow*
- *wb_lat_km: distance in transverse direction of flow*

Mid-day space-and-time interpolation

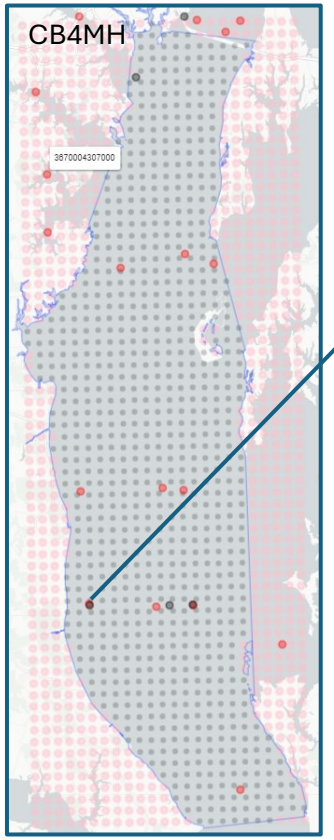


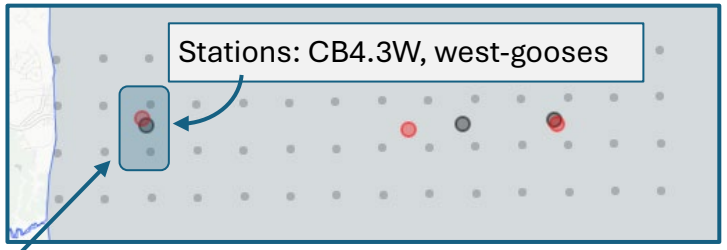
Input:

- Mid-day space and time interpolation GAM
  - Includes mean and variance-covariance matrix for GAM coefficients
- Water quality interpolator grid

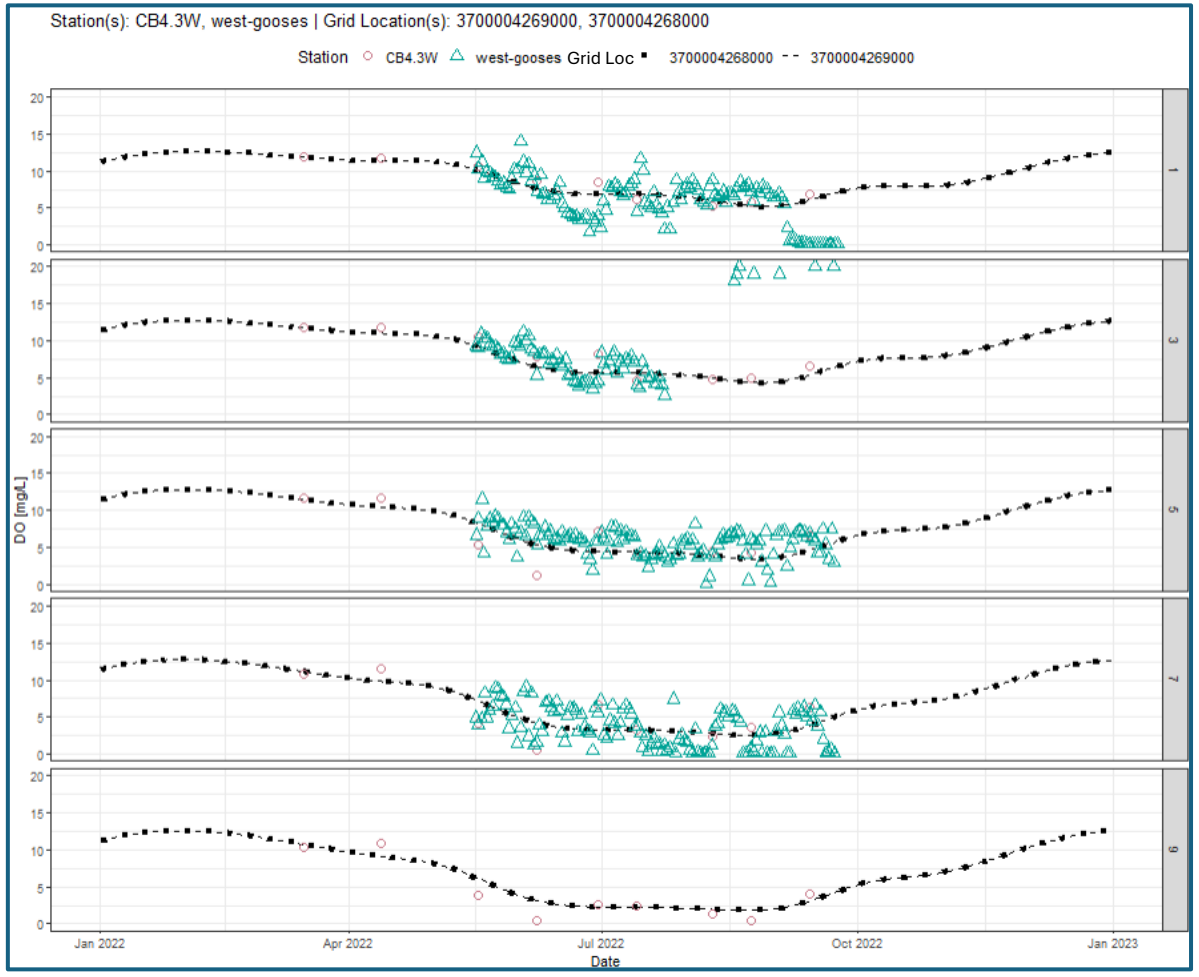
Procedure:

- Construct 3D spatial grid (cuboid)
  - Expand with daily time series to create 4D spatial-temporal grid
- Compute multiple `realizations` of mid-day GAM at grid points
  - Generate a vector of coefficients from a multivariate normal distribution based on the coefficient means and corresponding coefficient variance-covariance matrix.

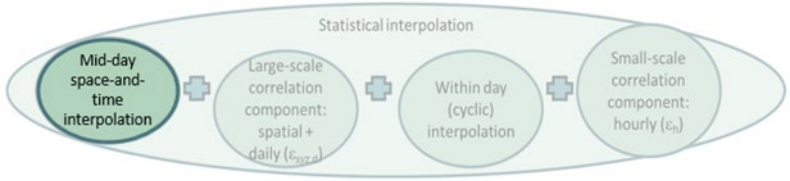




Mean mid-day GAM



Mid-day space-and-time interpolation

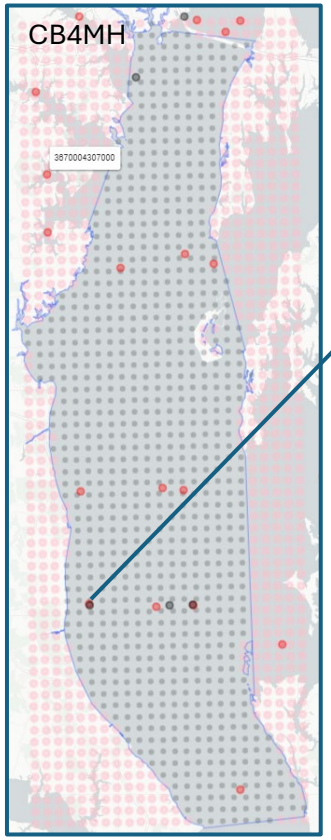


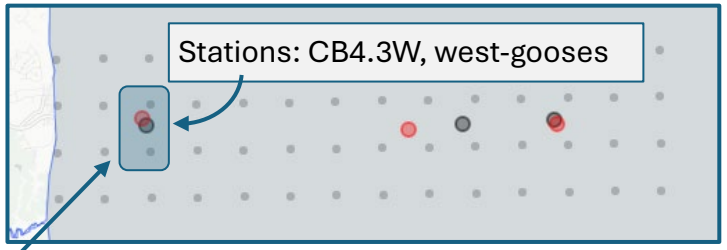
Input:

- Mid-day space and time interpolation GAM
  - Includes mean and variance-covariance matrix for GAM coefficients
- Water quality interpolator grid

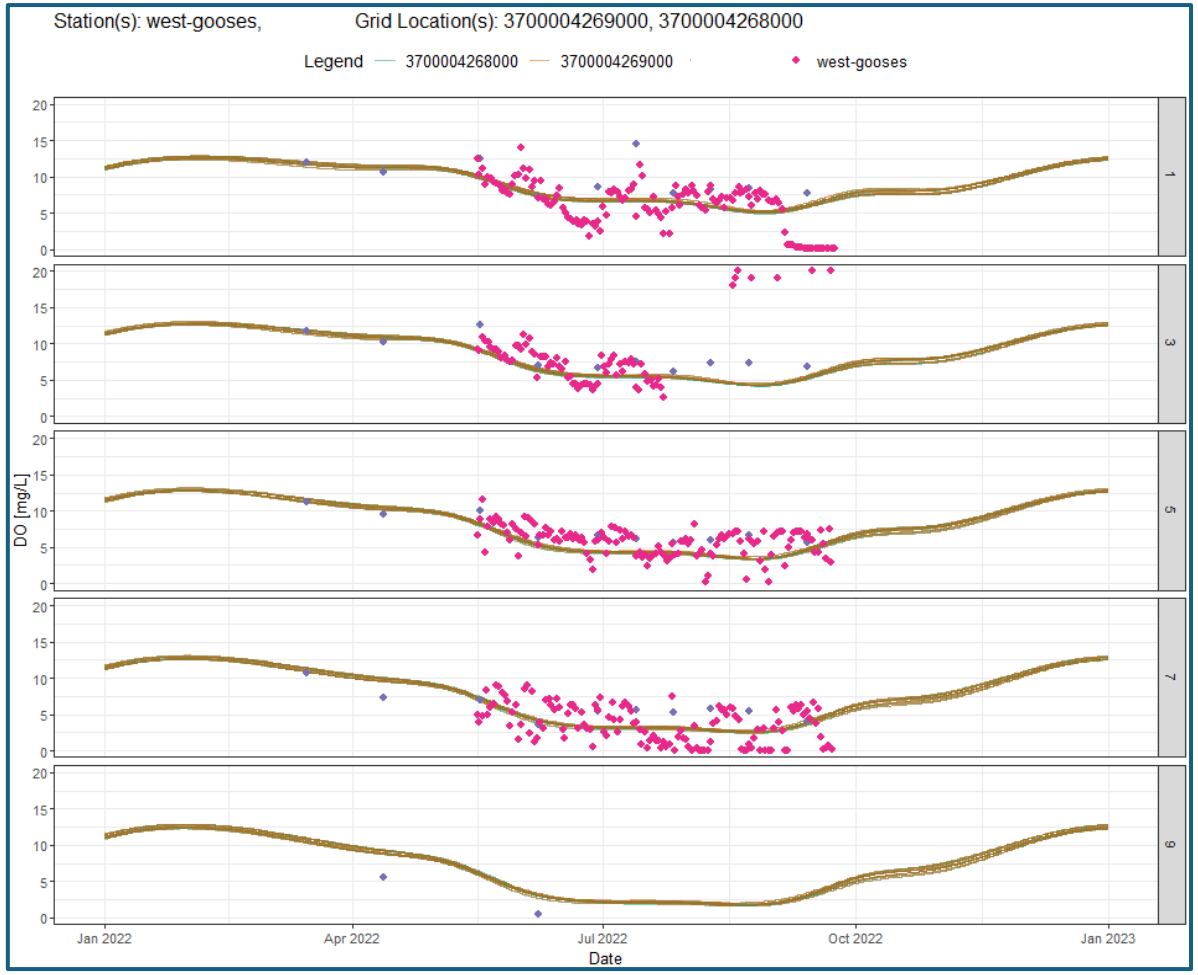
Procedure:

- Construct 3D spatial grid (cuboid)
  - Expand with daily time series to create 4D spatial-temporal grid
- Compute multiple `realizations` of mid-day GAM at grid points
  - Generate a vector of coefficients from a multivariate normal distribution based on the coefficient means and corresponding coefficient variance-covariance matrix.

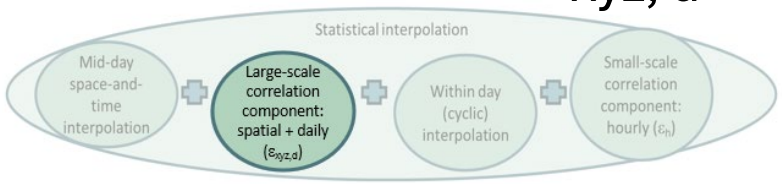




10 Mid-day Realizations at 2 grid locations



Large-scale correlation: spatial + daily ($\varepsilon_{xyz,\,d}$)

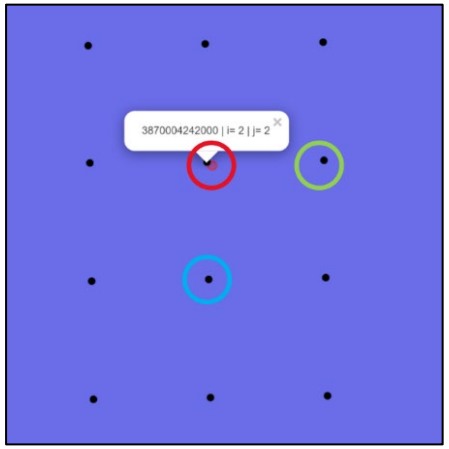


Input:

- Mid-day space and time interpolation GAM
  - Includes an estimate of interpolation uncertainty due to day-to-day noise
- 4D spatial-temporal grid
- Correlation
  1. Day-to-day: high frequency data
  2. Spatial: DATAFLOW data
  3. Depth: Vertical array data

Output:

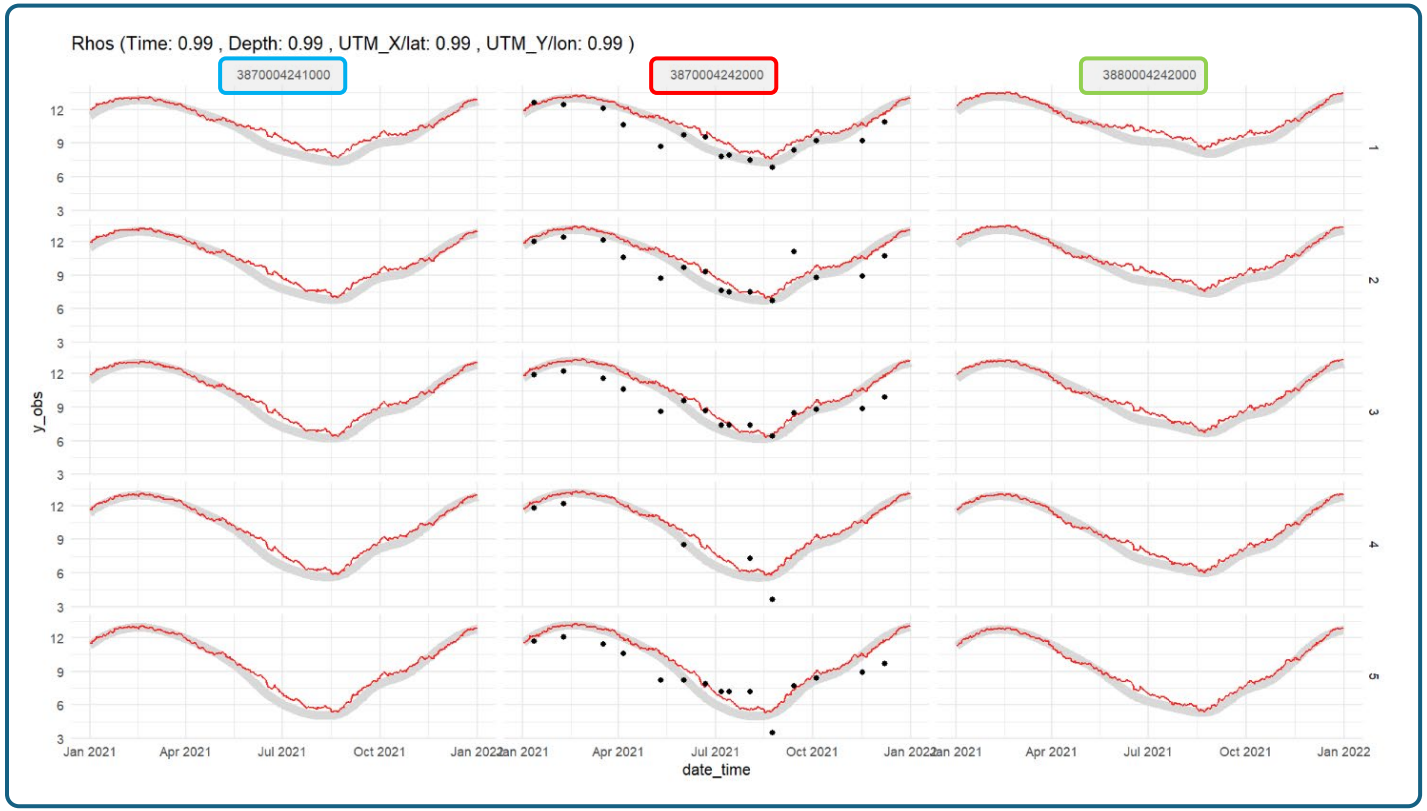
- Large scale correlation term



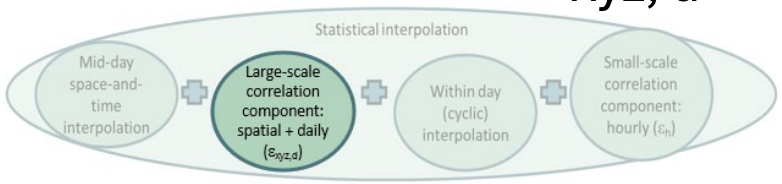
Testing performance and impact of correlation in all directions and time.

- Red: "3870004242000"
- Blue: "3870004241000" ~ 1 grid point south
- Green: "3880004242000" ~ 1 grid point east

➢ All dimensions highly correlated



Large-scale correlation: spatial + daily ($\varepsilon_{xyz,\,d}$)

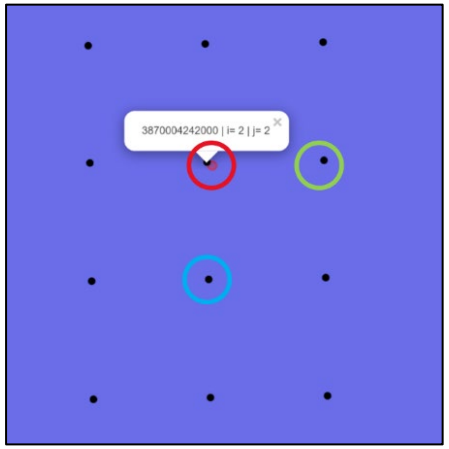


Input:

- Mid-day space and time interpolation GAM
  - Includes an estimate of interpolation uncertainty due to day-to-day noise
- 4D spatial-temporal grid
- Correlation
  1. Day-to-day: high frequency data
  2. Spatial: DATAFLOW data
  3. Depth: Vertical array data

Output:

- Large scale correlation term

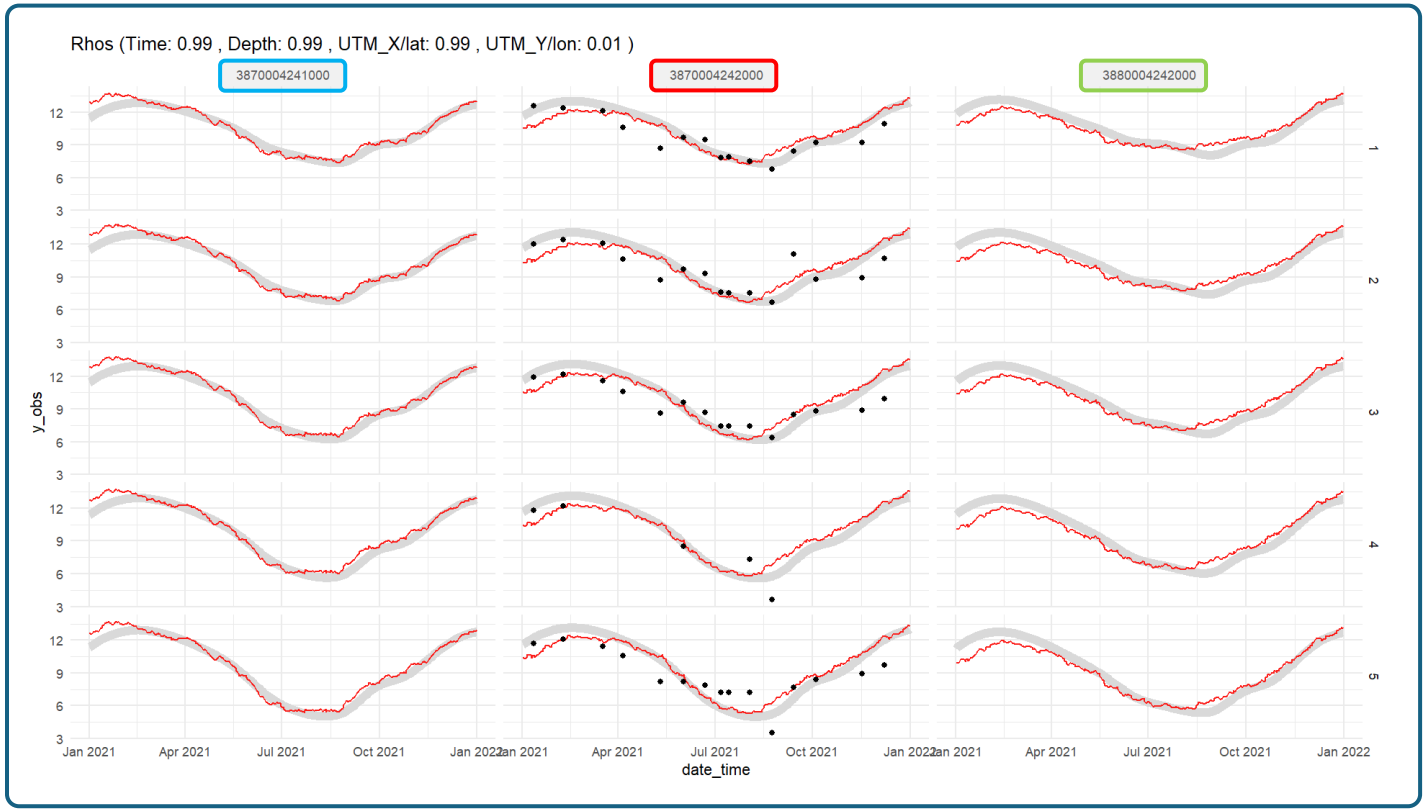


Testing performance and impact of correlation in all directions and time.

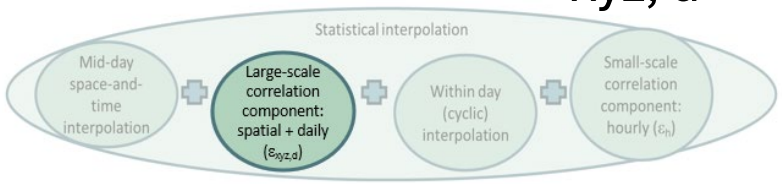
- Red: "3870004242000"
- Blue: "3870004241000" ~ 1 grid point south
- Green: "3880004242000" ~ 1 grid point east

➢ Low North/South correlation

➢ 1st/2nd columns have lower similarity



Large-scale correlation: spatial + daily ($\varepsilon_{xyz,\,d}$)

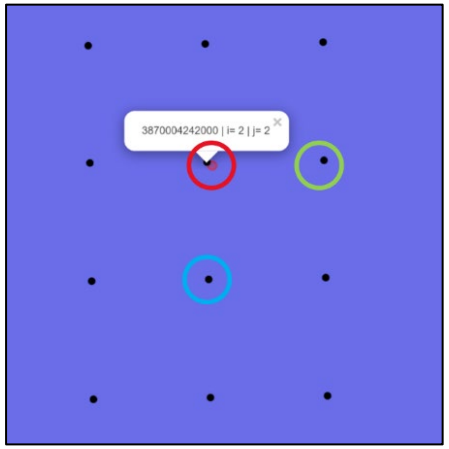


Input:

- Mid-day space and time interpolation GAM
  - Includes an estimate of interpolation uncertainty due to day-to-day noise
- 4D spatial-temporal grid
- Correlation
  1. Day-to-day: high frequency data
  2. Spatial: DATAFLOW data
  3. Depth: Vertical array data

Output:

- Large scale correlation term

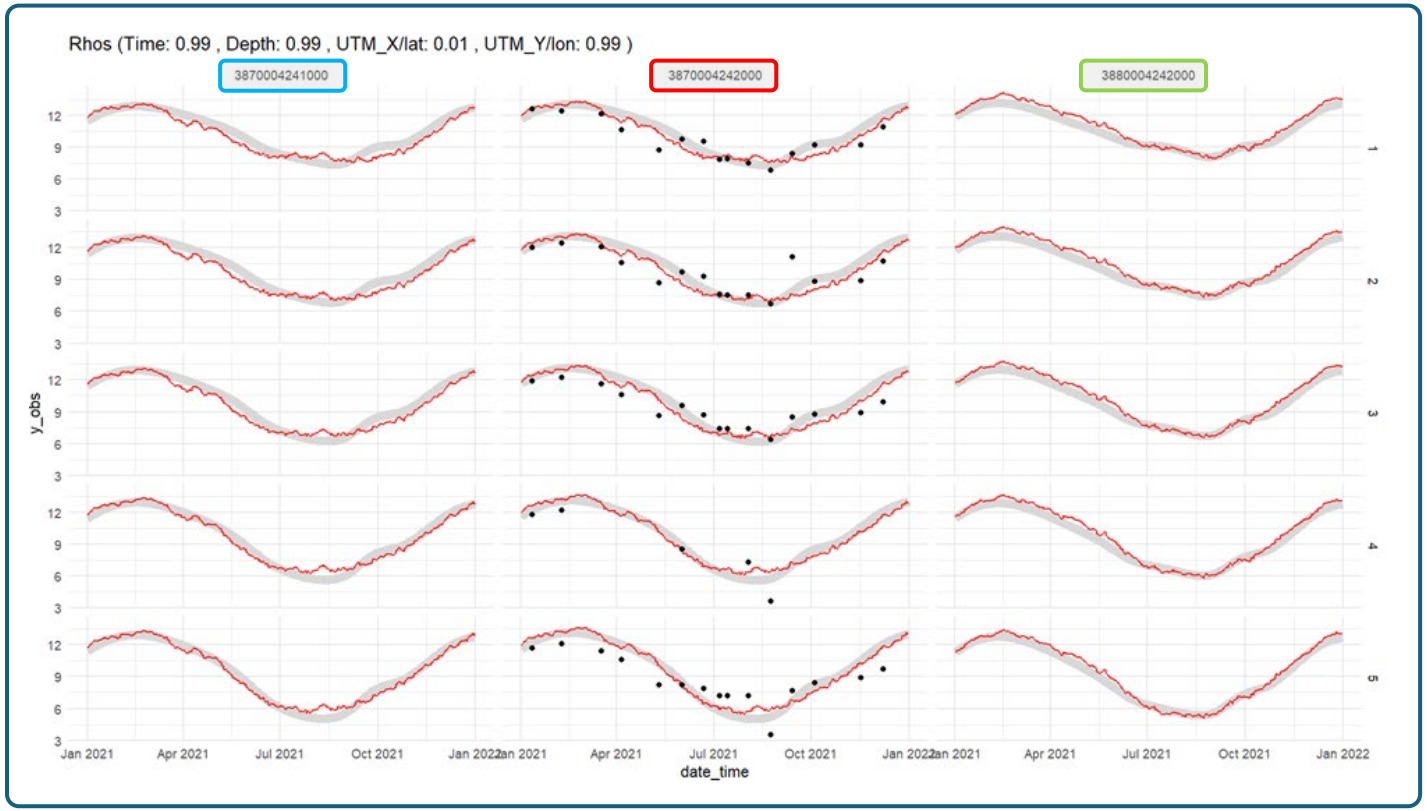


Testing performance and impact of correlation in all directions and time.

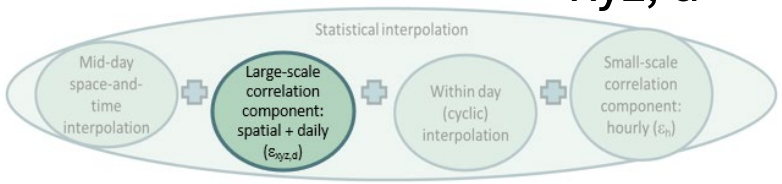
- Red: "3870004242000"
- Blue: "3870004241000" ~ 1 grid point south
- Green: "3880004242000" ~ 1 grid point east

➢ Low East/West correlation

➢ 2nd/3rd columns have lower similarity



Large-scale correlation: spatial + daily ($\varepsilon_{xyz,\,d}$)

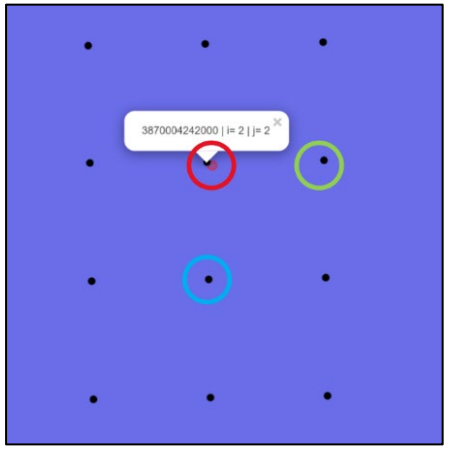


Input:

- Mid-day space and time interpolation GAM
  - Includes an estimate of interpolation uncertainty due to day-to-day noise
- 4D spatial-temporal grid
- Correlation
  1. Day-to-day: high frequency data
  2. Spatial: DATAFLOW data
  3. Depth: Vertical array data

Output:

- Large scale correlation term

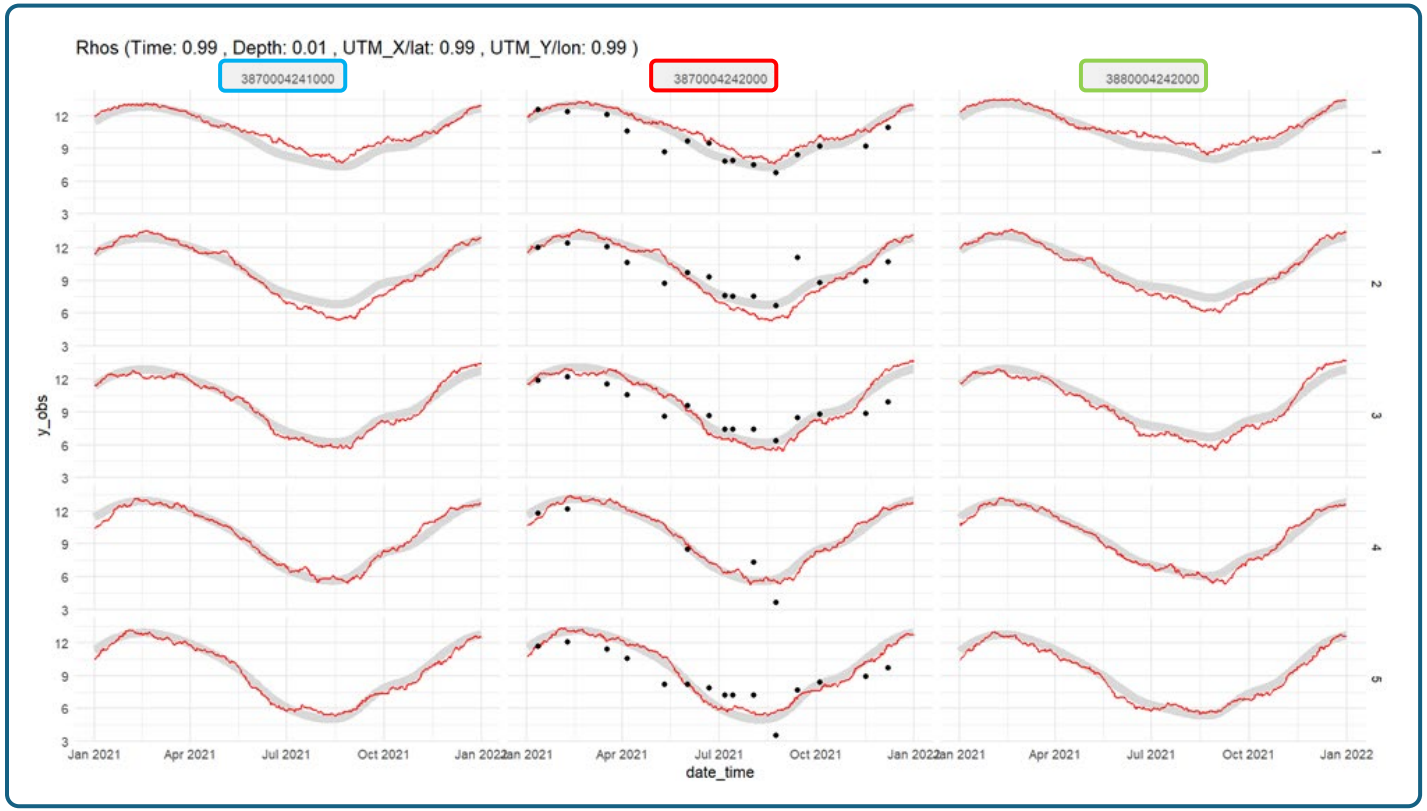


Testing performance and impact of correlation in all directions and time.

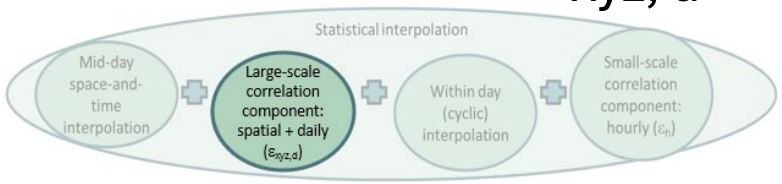
- Red: "3870004242000"
- Blue: "3870004241000" ~ 1 grid point south
- Green: "3880004242000" ~ 1 grid point east

➢ Low correlation in depth

➢ Rows have low similarity



Large-scale correlation: spatial + daily ($\varepsilon_{xyz,\,d}$)

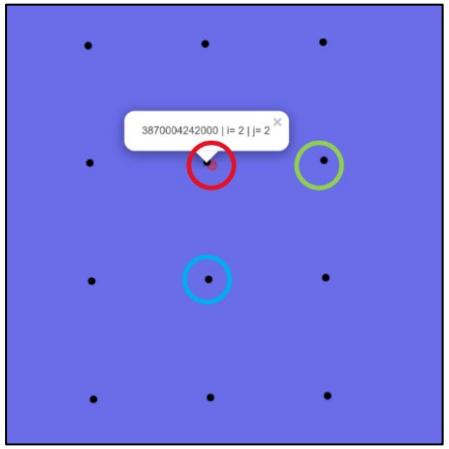


Input:

- Mid-day space and time interpolation GAM
  - Includes an estimate of interpolation uncertainty due to day-to-day noise
- 4D spatial-temporal grid
- Correlation
  1. Day-to-day: high frequency data
  2. Spatial: DATAFLOW data
  3. Depth: Vertical array data

Output:

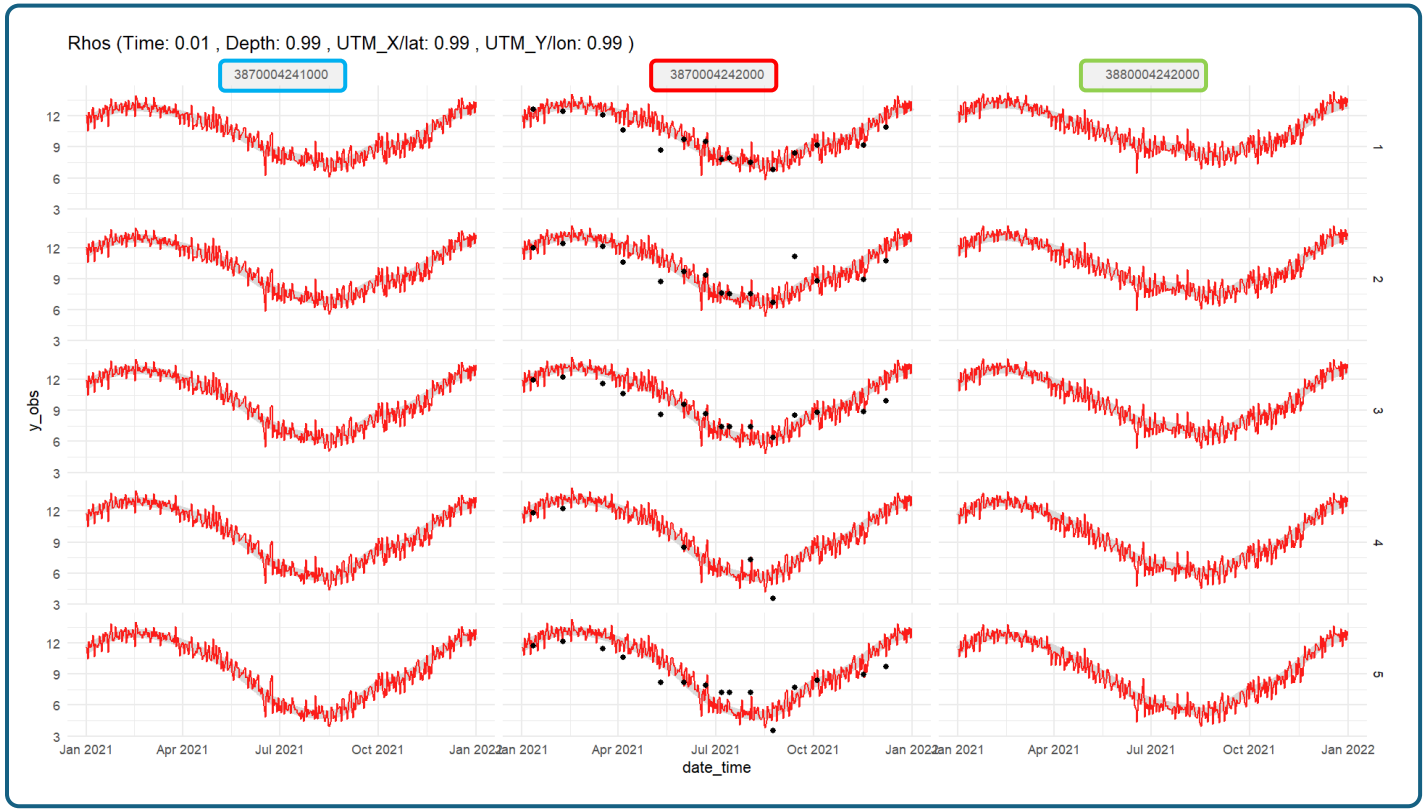
- Large scale correlation term



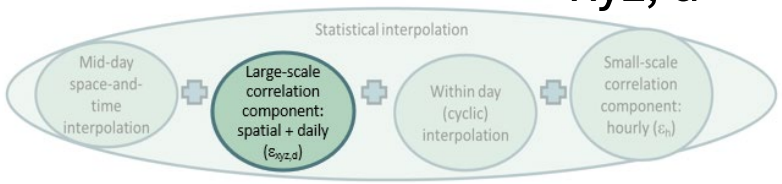
Testing performance and impact of correlation in all directions and time.

- Red: "3870004242000"
- Blue: "3870004241000" ~ 1 grid point south
- Green: "3880004242000" ~ 1 grid point east

➢ Low correlation in time



Large-scale correlation: spatial + daily ($\varepsilon_{xyz,\,d}$)



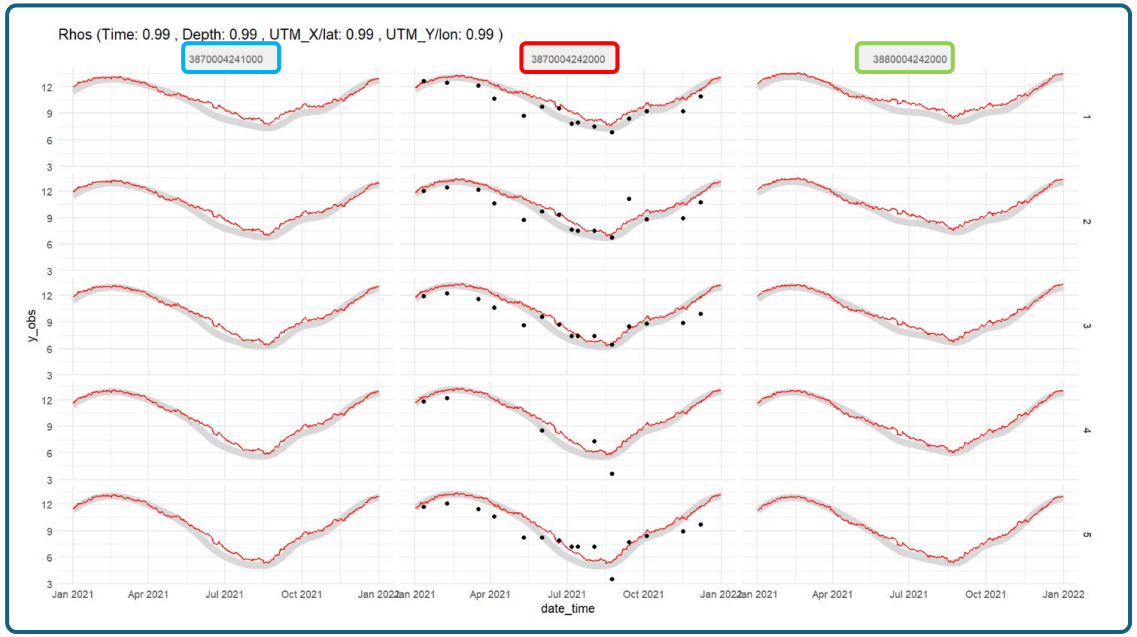
Input:

- Mid-day space and time interpolation GAM
  - Includes an estimate of interpolation uncertainty due to day-to-day noise
- 4D spatial-temporal grid
- Correlation
  1. Day-to-day: high frequency data
  2. Spatial: DATAFLOW data
  3. Depth: Vertical array data

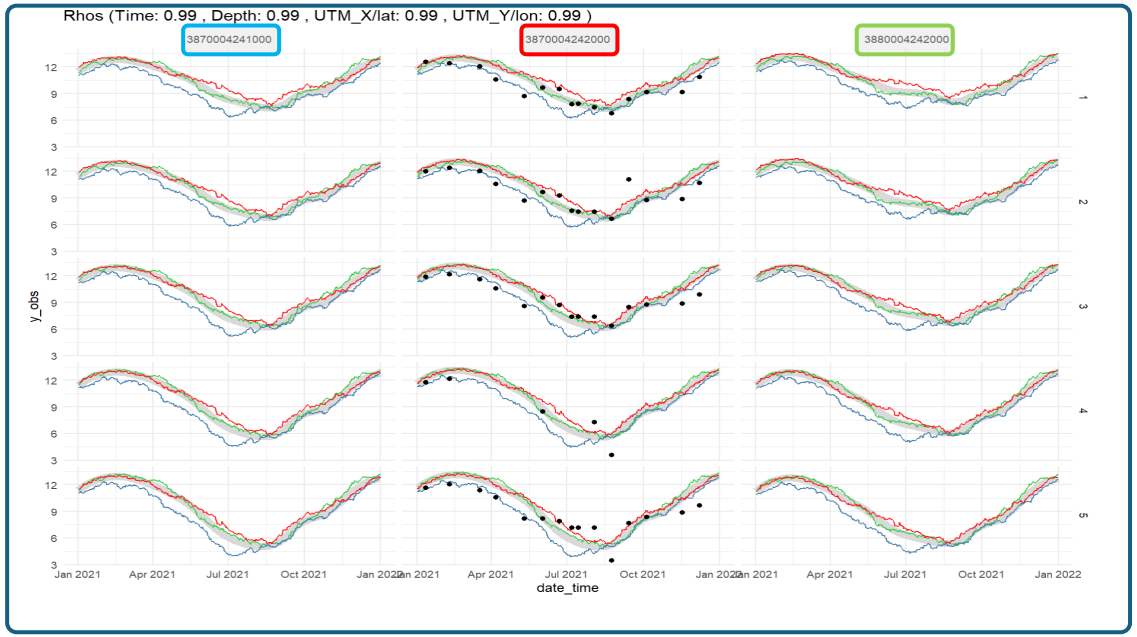
Output:

- Multiple realizations of large-scale correlation term

Single realization*

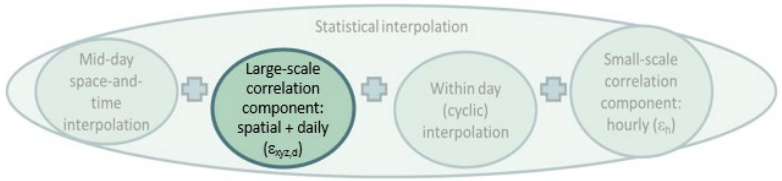


Multiple realizations*



* All dimensions with high correlation

1) Day-to-day correlation: high frequency data



Panel A: High frequency data (gray line) are subsampled to one value per day (symbols).*

** Data were subsampled to the observation closest to 11am.*

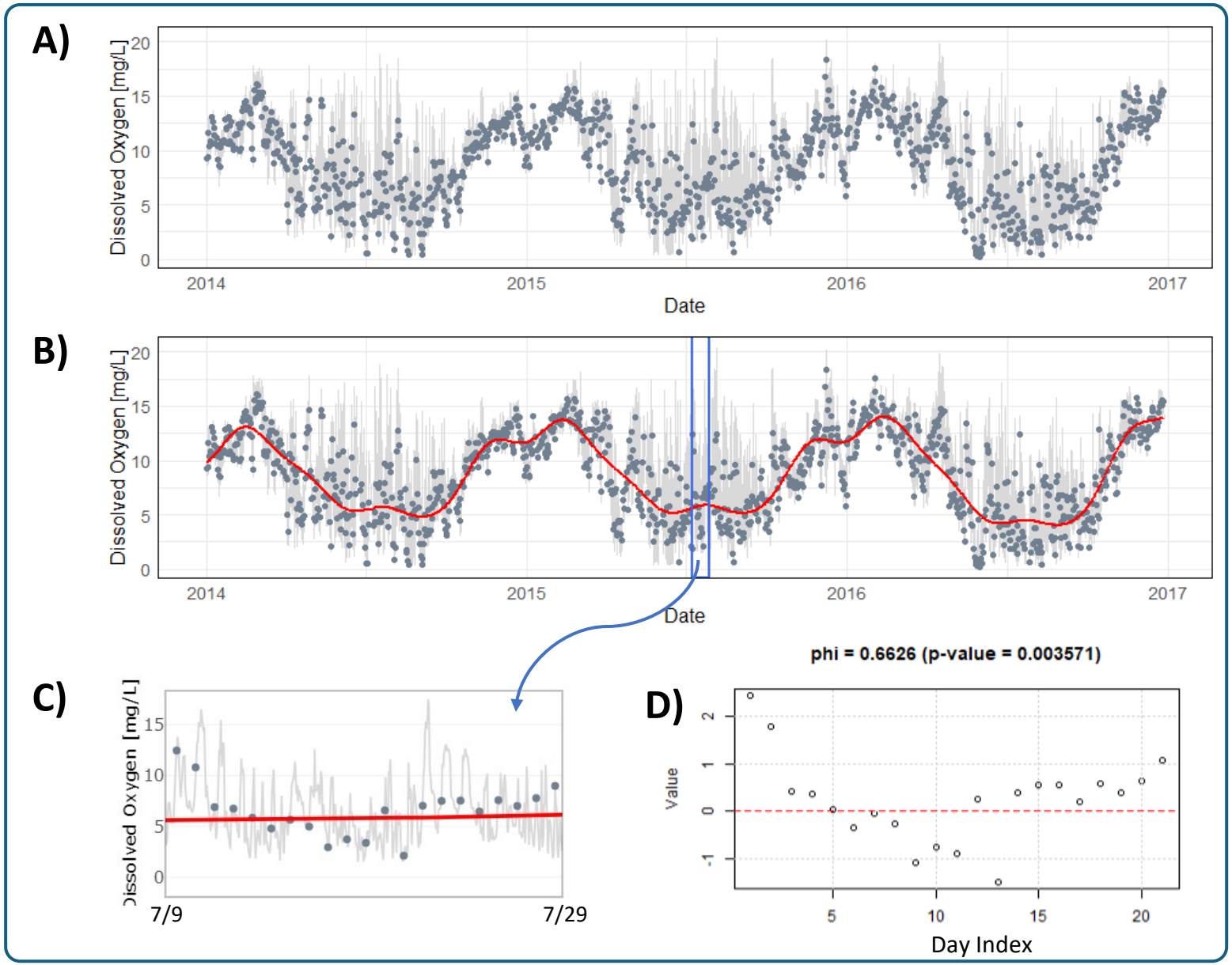
Panel B: Seasonal DO cycles are estimated (red line).†

† Red line is the mid-day interpolation.

Panel C: Sliding window of data selected.

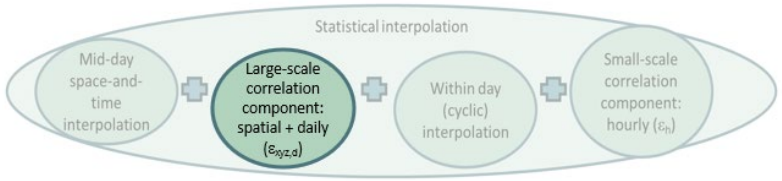
Panel D: Day-to-day correlation coefficient is computed using residuals from Panel C.

Parameterizing correlation components with data



Calculations performed with beta-logit transformed DO. Panels A, B, and C displayed in observed units (mg/L). Panel D displayed in transformed units.

2) Spatial Correlation: DATAFLOW data*



Panel A: Boat velocity used to chunk data into “segments”.

Panel B: Spatial depiction of “segments.”

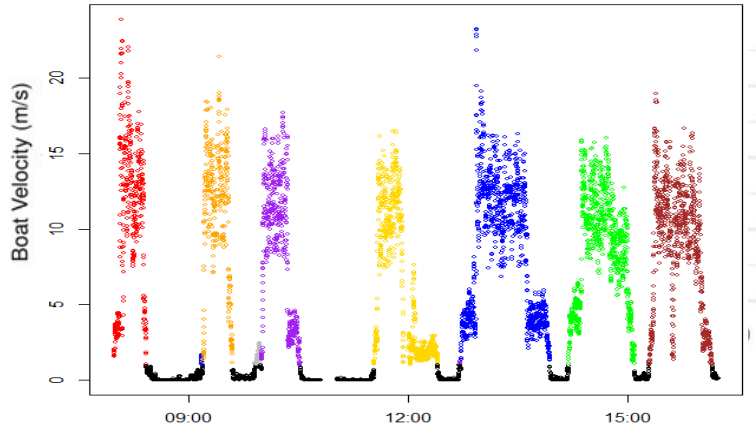
Panel C: Dissolved oxygen.

Panel D: Variogram.

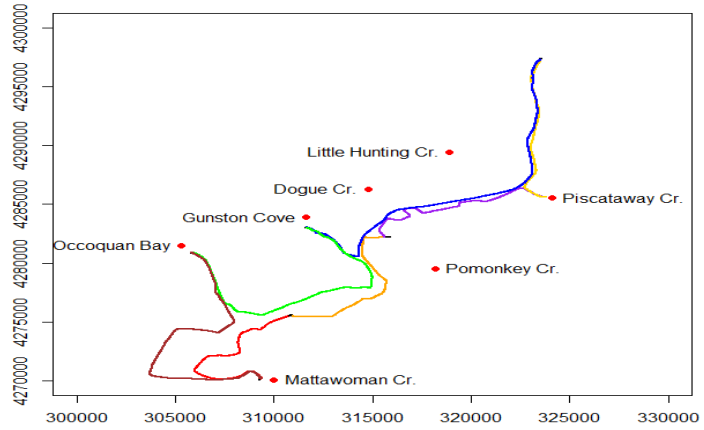
- Standardize data to N(0,1)
- Compute variogram cloud for each segment
- Combine clouds across segments
- Bin combined clouds
- Fit variogram (exponential)
- Compute correlation

Parameterizing correlation components with data

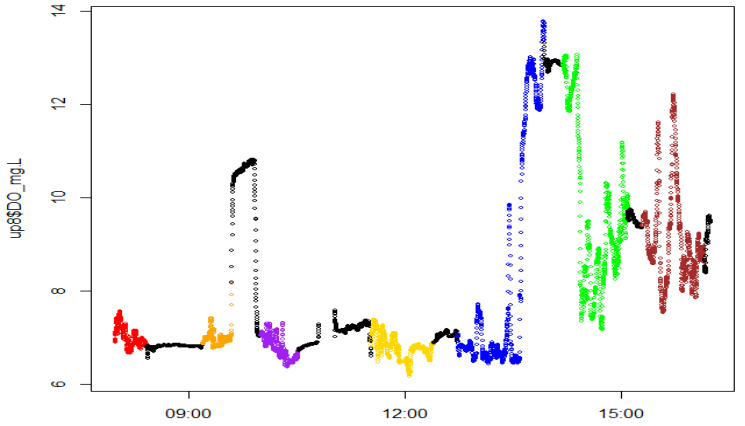
A) Velocity Over Survey (m/s)



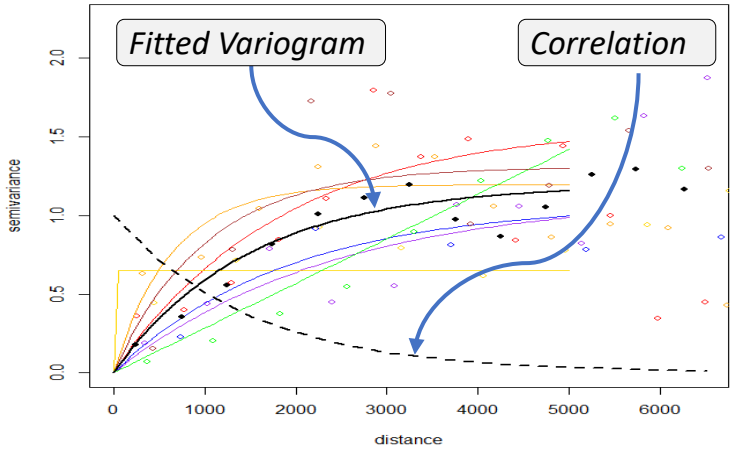
B) Survey "Segments"



C) DO over a cruise



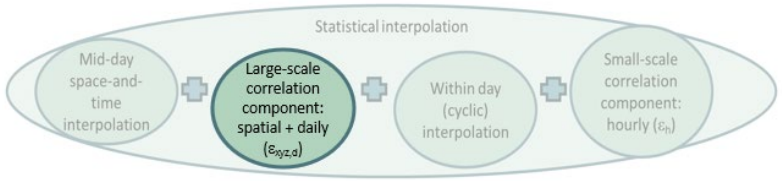
D) Variogram



** Work piloted with Potomac River DATAFLOW.*

3) Depth Correlation*

Parameterizing correlation components with data



Input Data:

- Vertical array data
- Fixed station vertical profiles

Correlations:

- Correlation at 2-m interval
  - Range = 0.37-0.85
  - Mean = 0.7
- 1-m correlation using an exponential variogram
  - $\sqrt{0.7} \approx 0.84$

Mean raw lag 2 depth correlation = 0.81
Correlations between depths across days for residual bl_do data.

	rbldo1	rbldo3	rbldo5	rbldo7	rbldo9	rbldo11	rbldo13	rbldo15	rbldo17	rbldo19
rbldo1	1.00	0.37	0.36	0.20	0.14	-0.01	0.03	-0.16	-0.17	-0.12
rbldo3	0.37	1.00	0.56	-0.07	0.20	0.11	0.16	-0.07	0.02	-0.04
rbldo5	0.36	0.56	1.00	0.48	0.53	0.28	0.19	0.04	0.14	0.21
rbldo7	0.20	-0.07	0.48	1.00	0.81	0.47	0.20	0.11	0.10	0.13
rbldo9	0.14	0.20	0.53	0.81	1.00	0.78	0.45	0.29	0.26	0.23
rbldo11	-0.01	0.11	0.28	0.47	0.78	1.00	0.77	0.57	0.42	0.24
rbldo13	0.03	0.16	0.19	0.20	0.45	0.77	1.00	0.80	0.56	0.26
rbldo15	-0.16	-0.07	0.04	0.11	0.29	0.57	0.80	1.00	0.85	0.42
rbldo17	-0.17	0.02	0.14	0.10	0.26	0.42	0.56	0.85	1.00	0.56
rbldo19	-0.12	-0.04	0.21	0.13	0.23	0.24	0.26	0.42	0.56	1.00

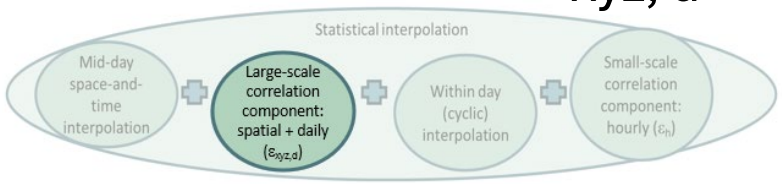
Mean residual lag 2 depth correlation = 0.7

Mean residual depth lag 4 correlation = 0.43
Square of depth lag 2 correlation = 0.49

Mean residual depth lag 6 correlation = 0.27
Cube of depth lag 2 correlation = 0.34

** Work piloted with "East Gooses" vertical array*

Large-scale correlation: spatial + daily ($\varepsilon_{xyz,\,d}$)



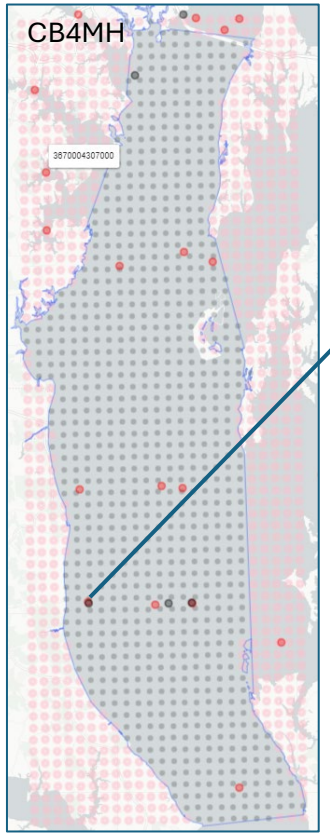
Input:

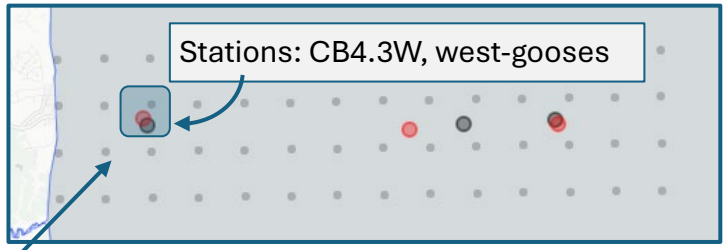
- Mid-day space and time interpolation GAM
  - Includes an estimate of interpolation uncertainty
- 4D spatial-temporal grid
- Correlation
  1. Day-to-day: high frequency data
  2. Spatial: DATAFLOW data
  3. Depth: Vertical array data

Output:

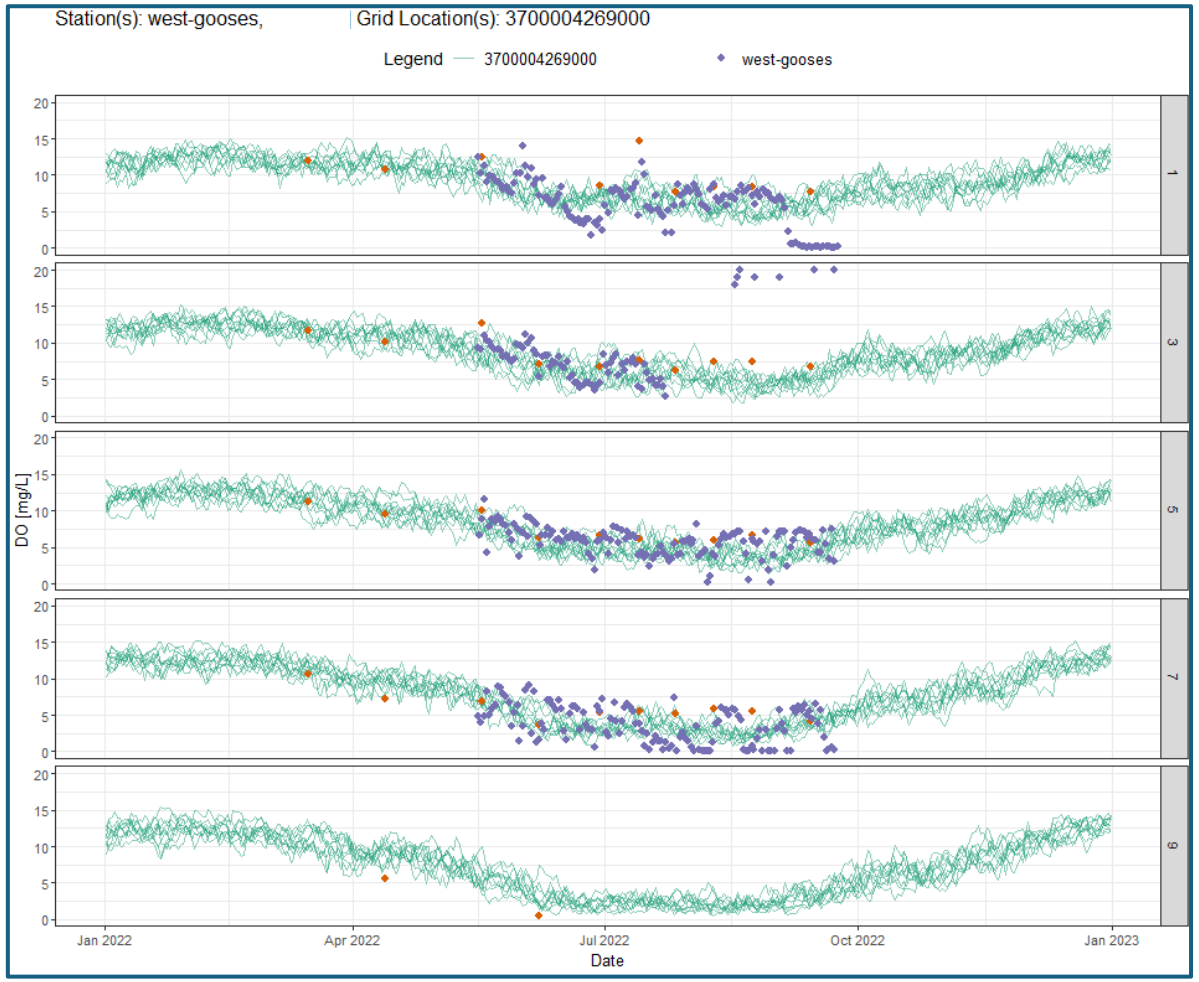
- Multiple realizations of large-scale correlation

Preliminary results with preliminary parameters compared to daily West Gooses observations



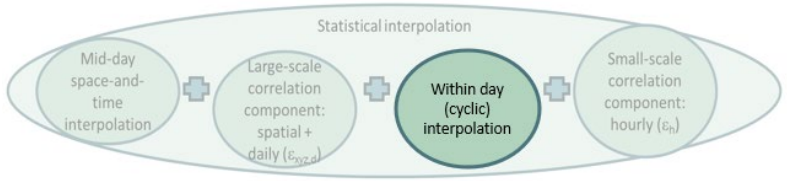


Multiple Realizations: Mid-day + Large-scale correlation term



** Correlations: day=0.8, depth=0.7, UTM_X=0.4, UTM_Y=0.4)*

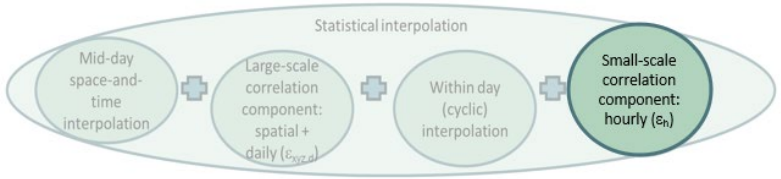
Within day (cyclic) interpolation



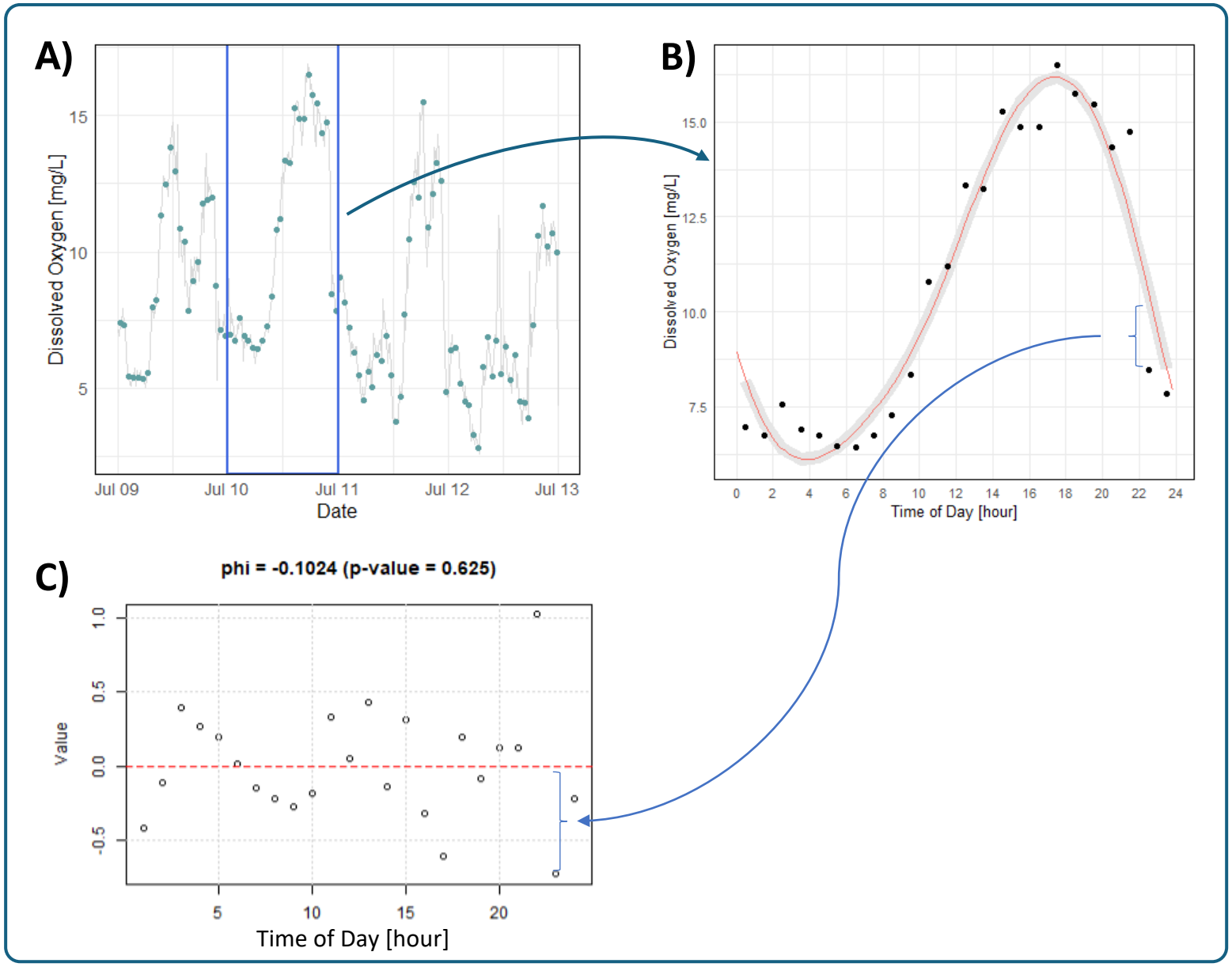
Panel A: High frequency data are used at an hourly frequency (symbols) for consistency across all data sources.

Panel B: Diel and tidal harmonic coefficients are computed for each `station/layer/day` of data.

† *Red line represents daily cyclic fit.*

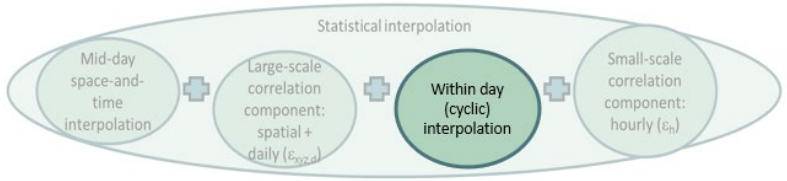


Panel C: Small-scale (hourly) correlation coefficient is computed using residuals from Panel B.

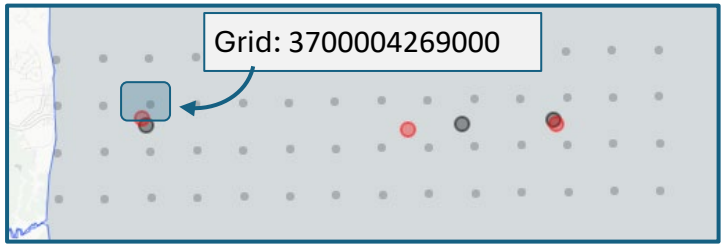


Calculations performed with beta-logit transformed DO. Panels A and B displayed in observed units (mg/L). Panel C displayed in transformed units.

Within day (cyclic) interpolation



Panel A: Daily cyclic interpolation for grid point 3700004269000 at 1 m depth

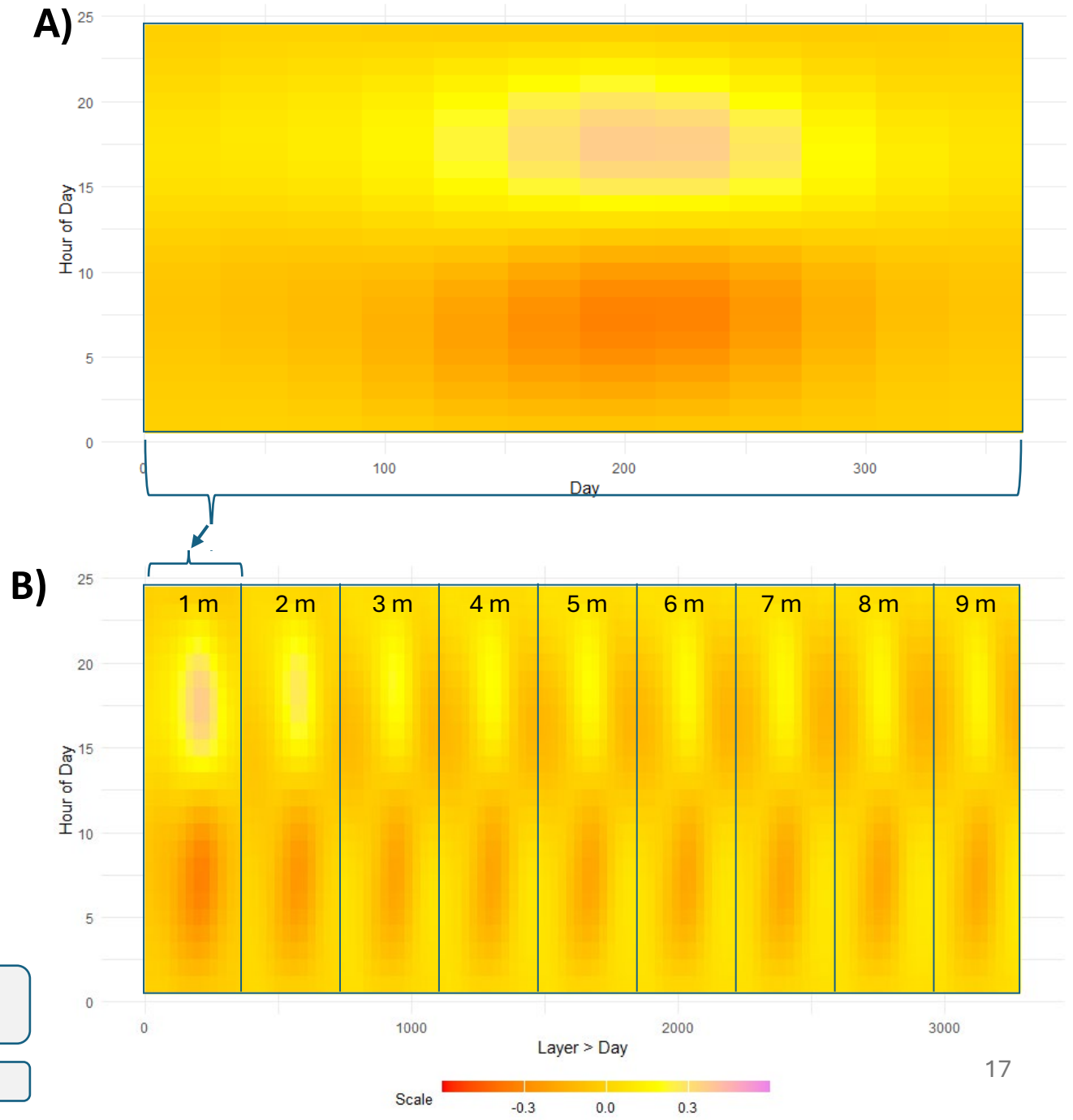


Panel B: Daily cyclic interpolation for grid point 3700004269000 for 1-9 m depth

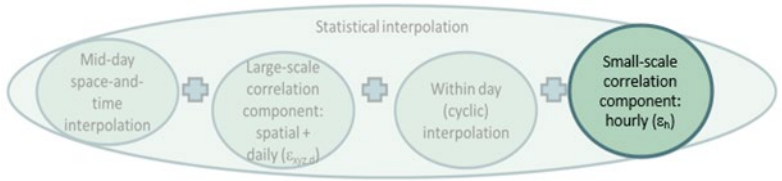
Calculations performed with beta-logit transformed DO. Panels A and B displayed in transformed units.

Single realization depicted in figures—multiple realizations run in practice.

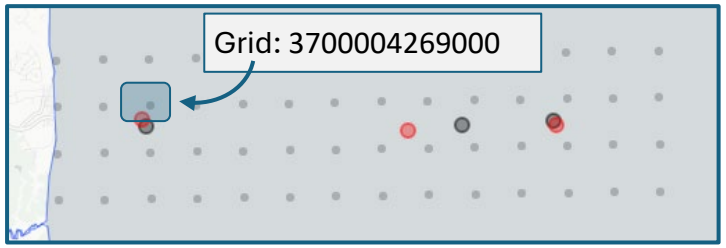
Single realization from cyclic DO interpolation in transformed units



Small-scale correlation component (ε_h)



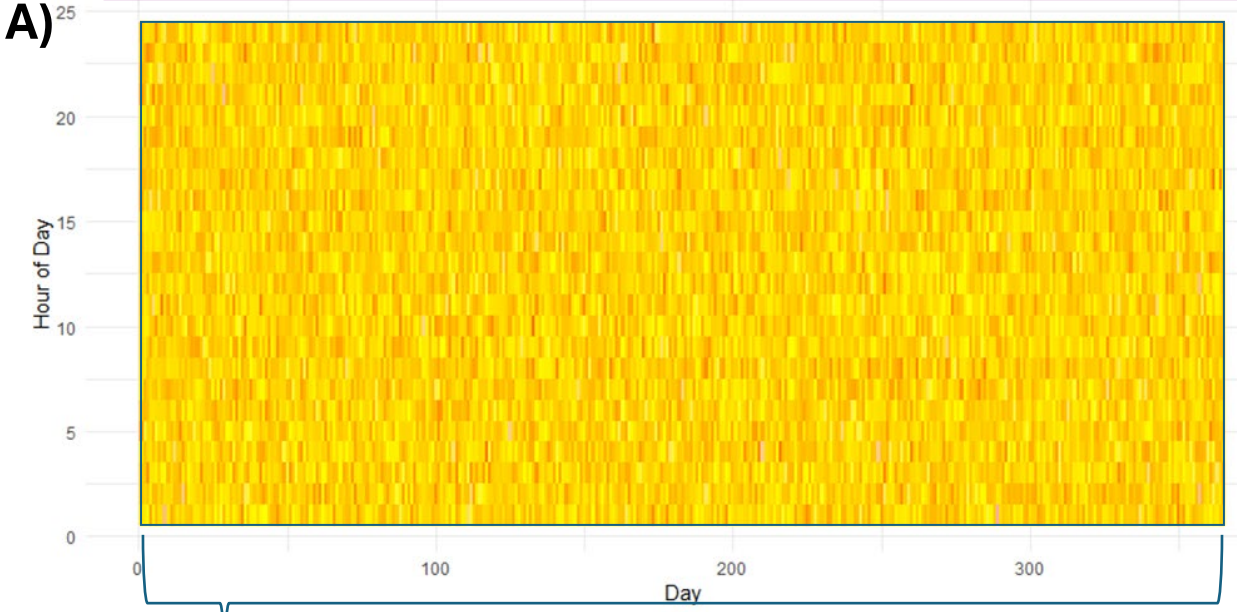
Panel A: Small-scale correlation component for grid point 3700004269000 at 1 m depth



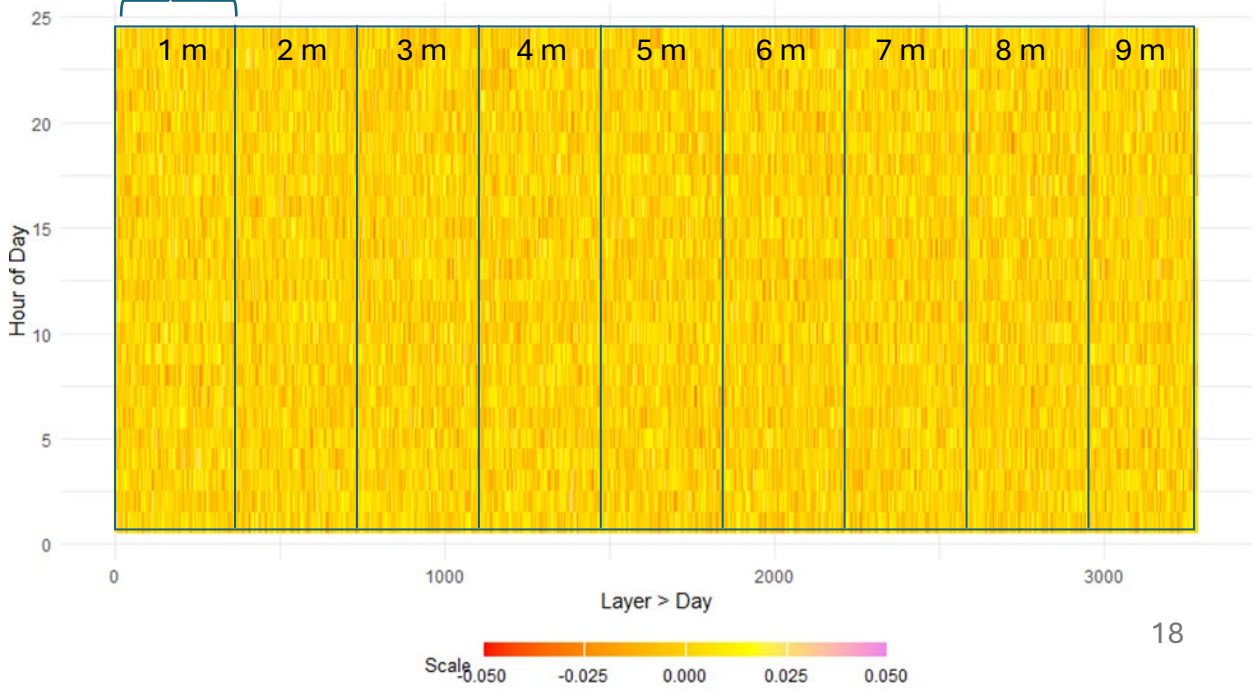
Panel B: Small-scale correlation component for grid point 3700004269000 for 1-9 m depth

Single realization from small-scale correlation component.

A)



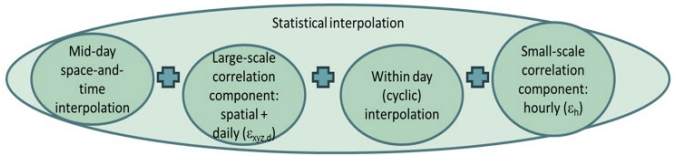
B)

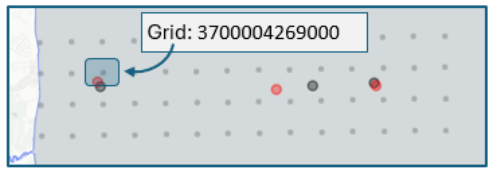


Calculations performed with beta-logit transformed DO. Panels A and B displayed in transformed units.

Single realization depicted in figures—multiple realizations run in practice.

Statistical Interpolation – the “pieces”

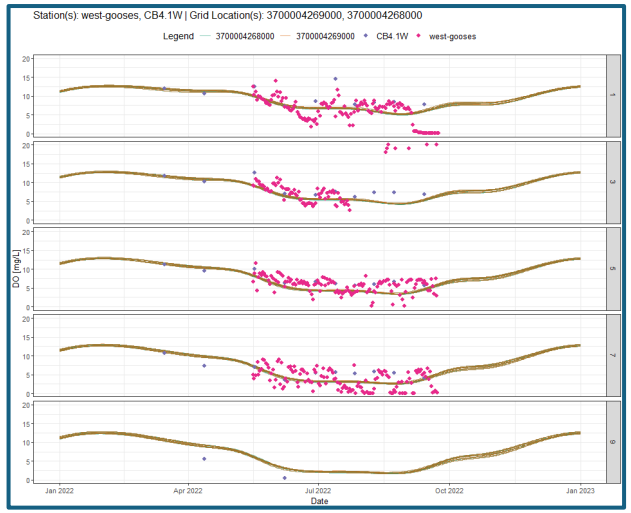




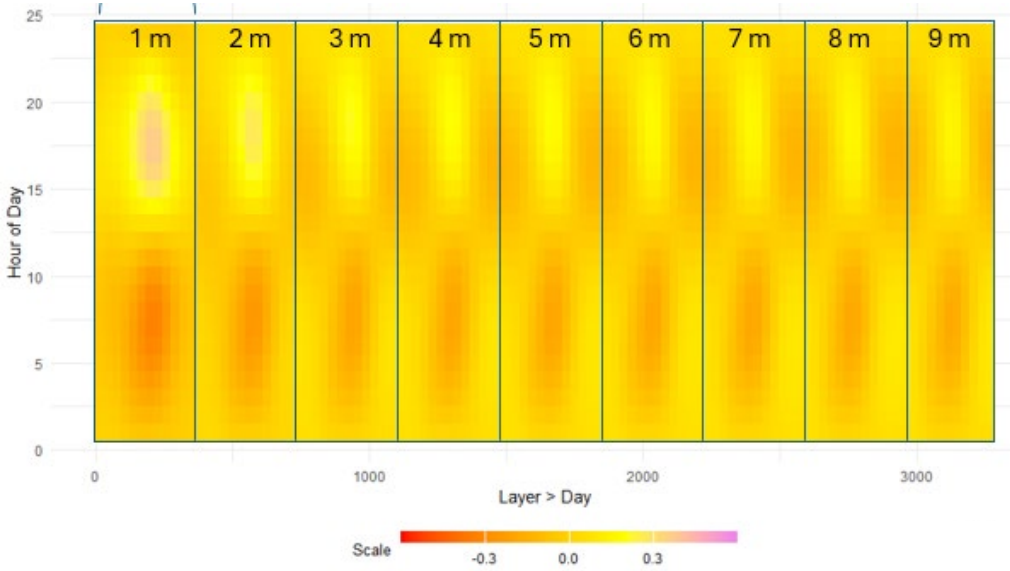
Calculations performed with beta-logit transformed DO. Panels A and B displayed in observed units (mg/L). Panels C and D displayed in transformed units.

Single realization depicted in Panels C and D—multiple realizations run in practice.

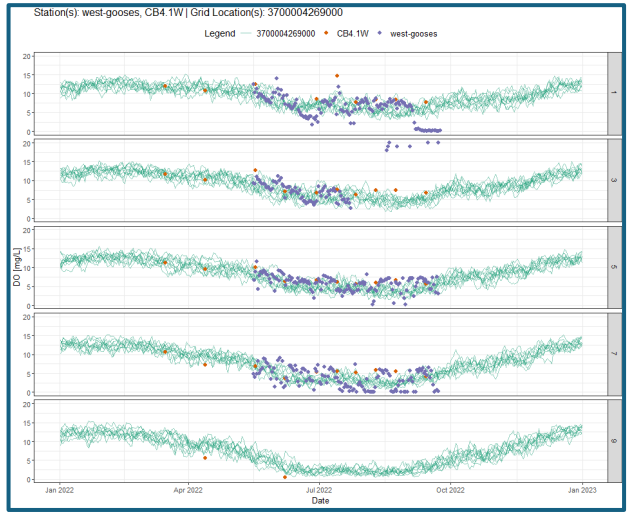
A) Mid-day space-and-time interpolation



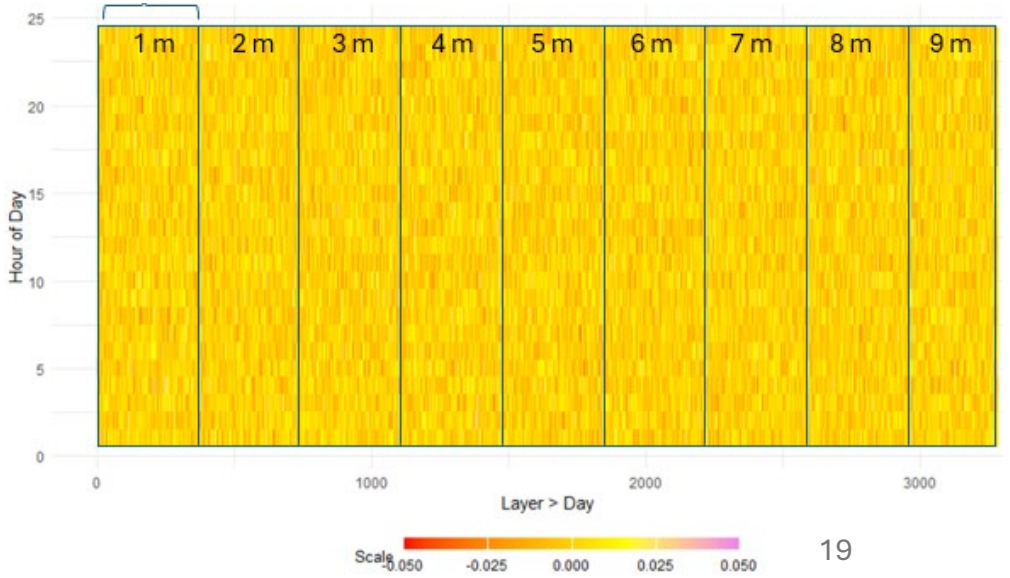
C) Within day (cyclic) interpolation



B) Mid-day + Large-scale correlation: spatial + daily ($\varepsilon_{xyz,\,d}$)

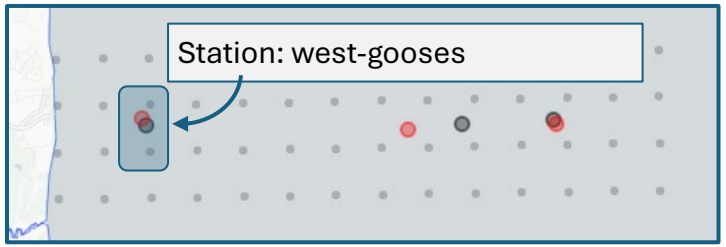


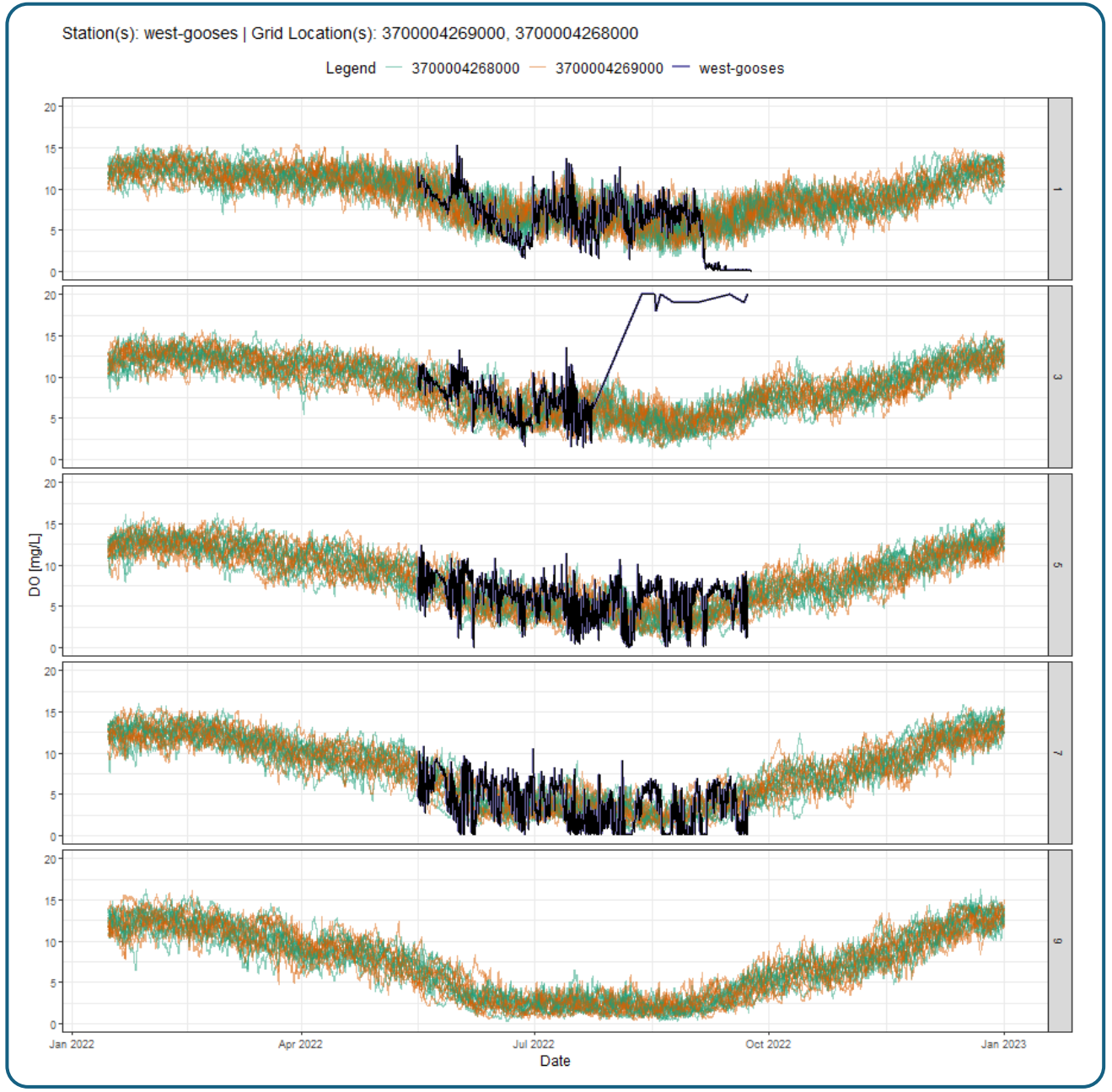
D) Small-scale correlation component (ε_h)



Statistical Interpolation – comparison with high frequency data

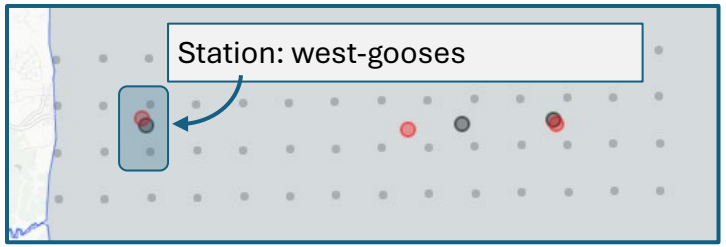
- 10 realizations from grid cells near “west-gooses” vertical array
  - 2022

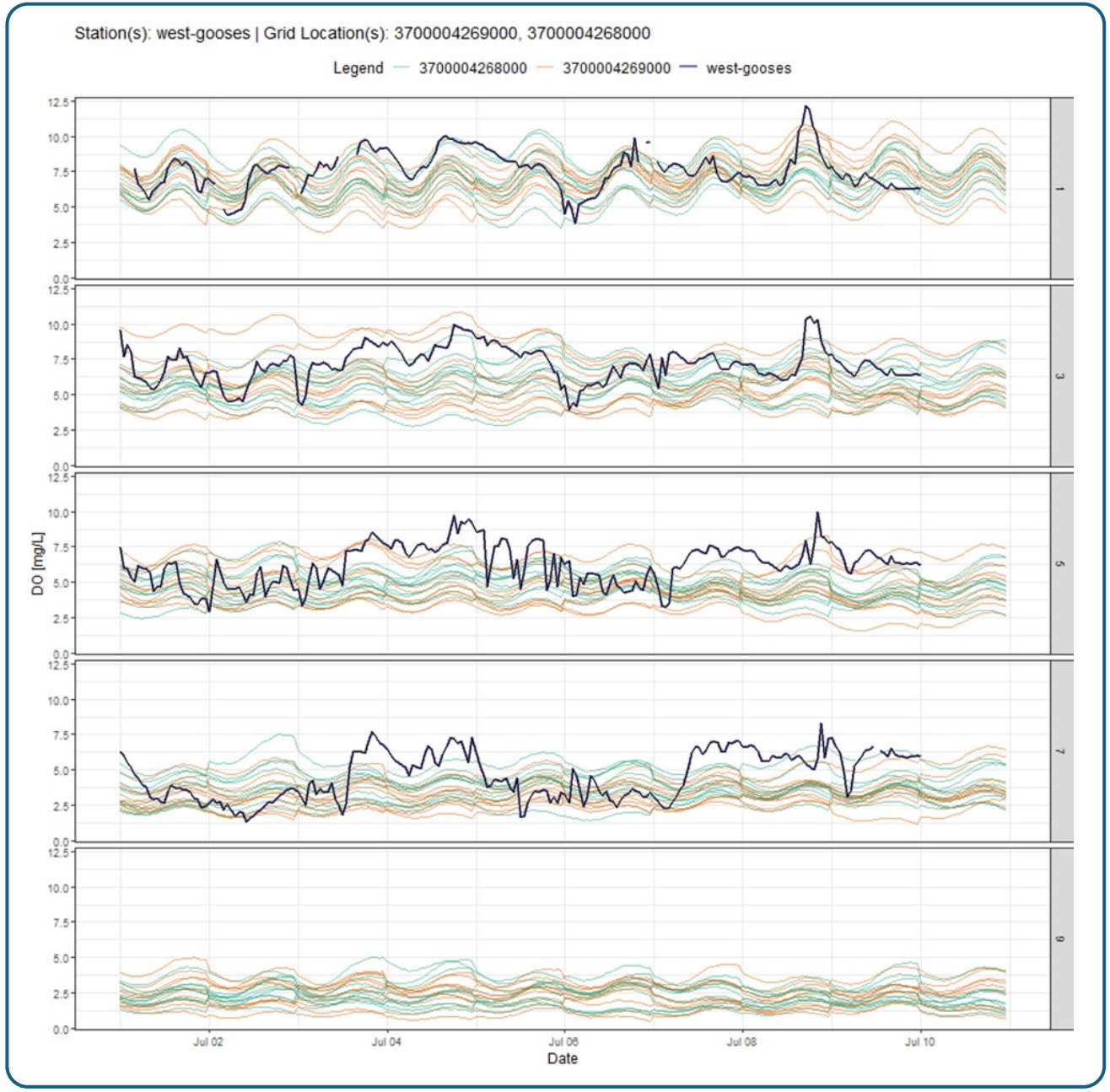




Statistical Interpolation – comparison with high frequency data

- 10 realizations from grid cells near “west-gooses” vertical array
  - July 1-10, 2022





Current and next development steps

- **Criteria assessment link:** Work with CAP team on how results would be used in criteria assessment (starting now)
  - *Settle on method for interpolating pycnocline (and build in code).*
  - *Assess if any changes are needed to prediction grid.*
  - *Brainstorm options for using 4D results in CFD or other approaches.*
  - *Conduct test cases of various options.*
- **Method development:** Continue necessary parametrization of all 4 parts of the 4-D tool (continuing through entire 2025)
  - *Continue work on cyclic interpolation.*
  - *Parameterize vertical correlation.*
  - *Expand Dataflow analysis beyond Potomac for horizontal correlation.*
  - *Scale up simulation to bay-wide, examine results, and validate with additional or hold-out data.*
  - *Consider non-stationarity in correlation parameters.*
- **Software development** (continuing through entire 2025 into 2026)
  - *Continue data compilation and use of all data sets.*
  - *Determine file management and storage options.*
  - *Work with future users on features.*