~~Road Map~~ Exploration Tree of Stochastic Components of Daily Mean and Small Scale Variability.

Elgin Perry
eperry@chesapeake.net
BORG 5/20/2024

Conceptual diagram

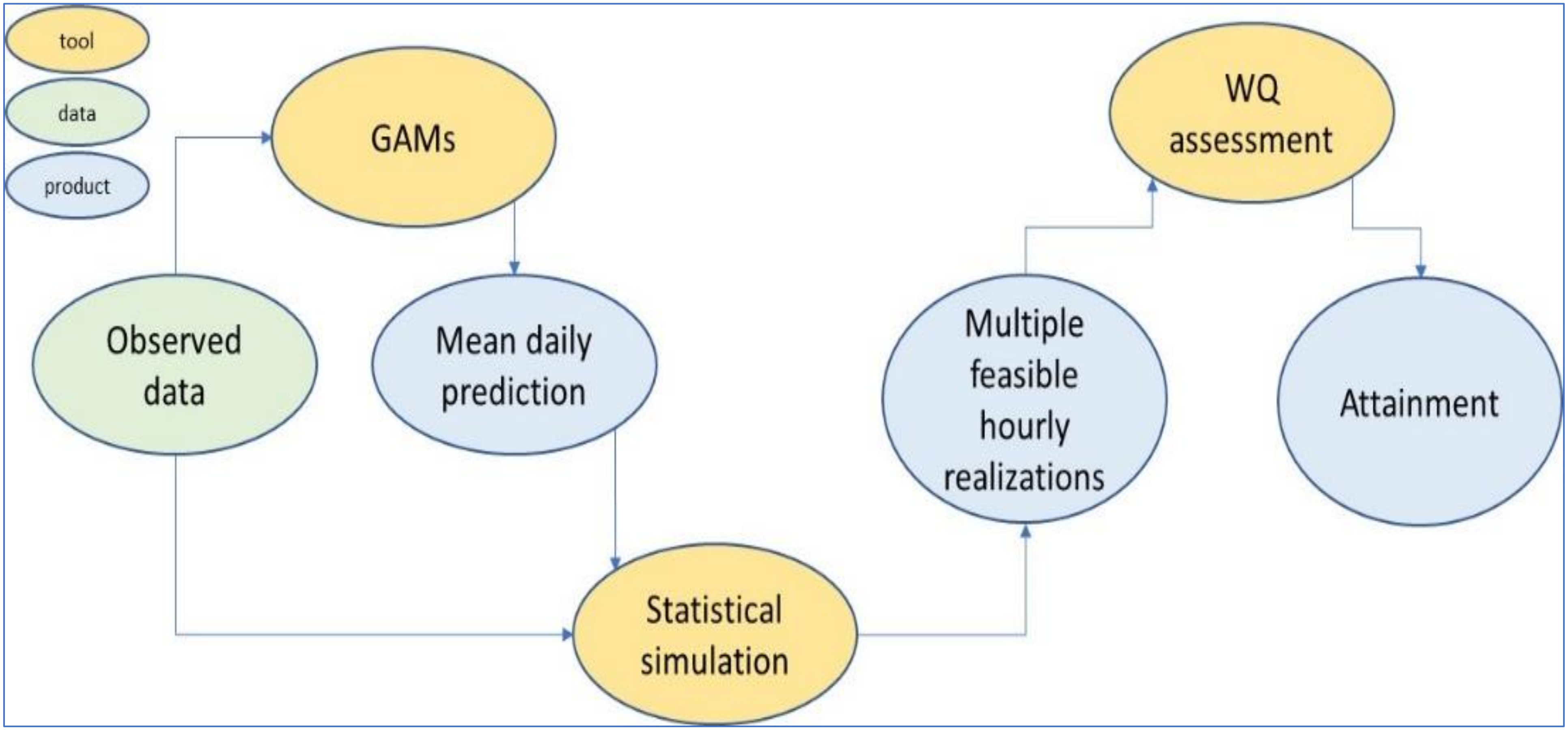


Figure 1: Interpolation and attainment assessment system

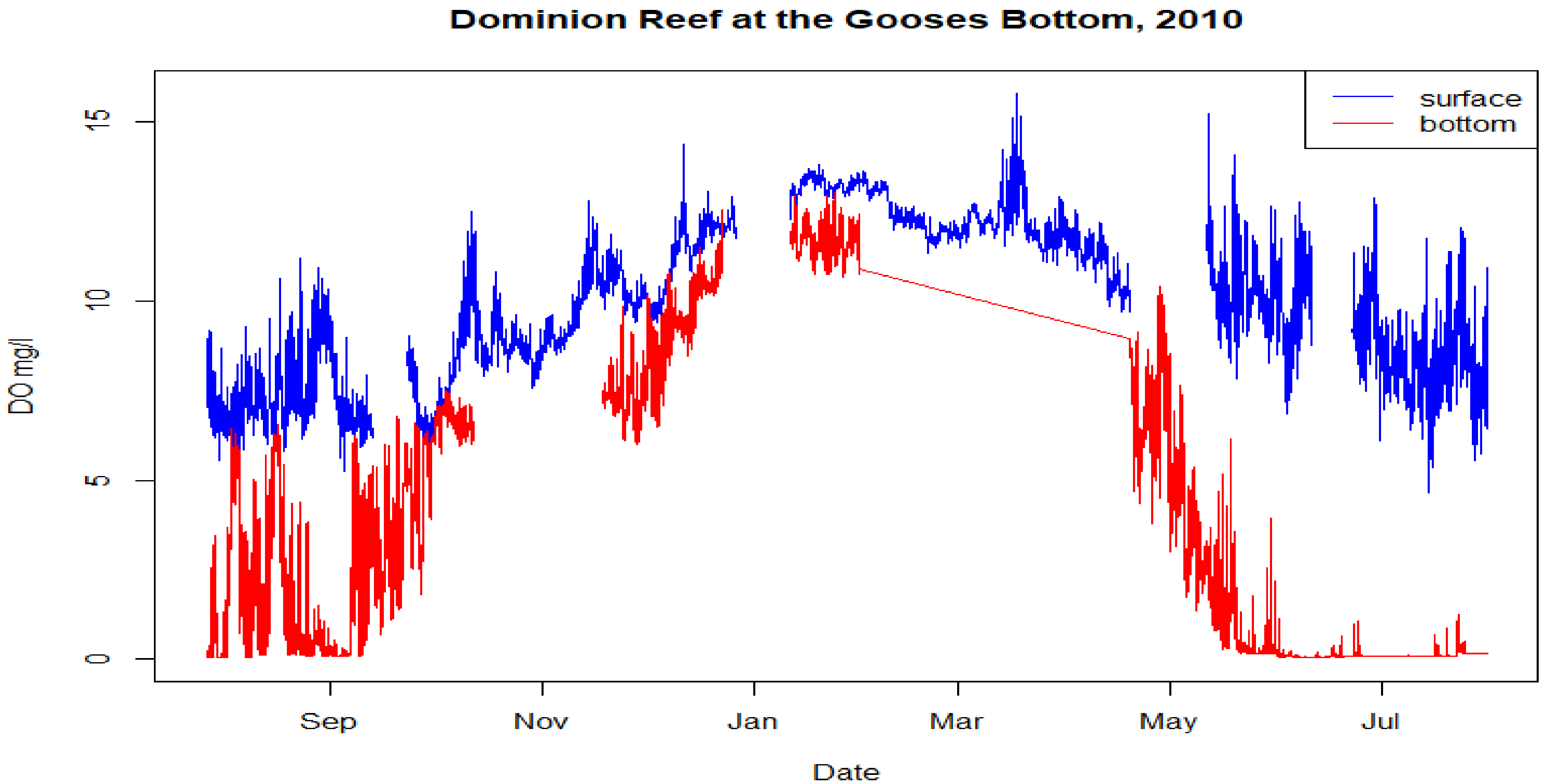
First: Consider stochastic components in the **Statistical Simulation** of small scale variability.

Second: Consider stochastic components in the **GAMs** tool for Daily Prediction.

Three Approaches to Stochasticity for Small Scale Variability.

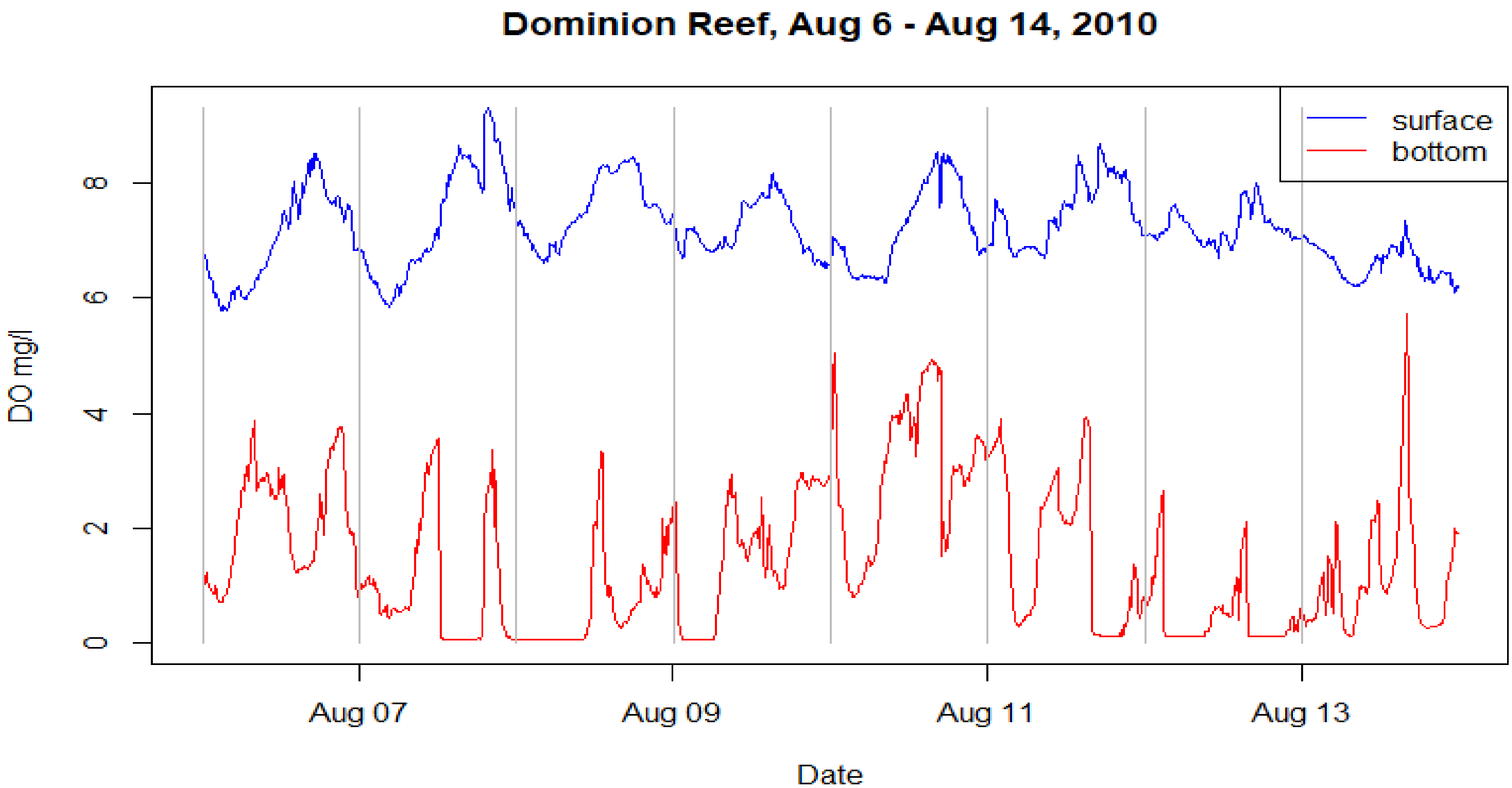
1. **Resample mean adjusted residuals with Day as Experimental Unit.**
2. **Model cycles with Fourier terms and resample cycle detrended residuals.**
3. **Model cycles with Fourier terms and Simulate cycle detrended residuals.**

ConMon data.



Dominion Reef at the Gooses Bottom, 2010
surface
bottom
DO mg/l
15
10
5
0
Sep
Nov
Jan
Mar
May
Jul
Date

Taking a closer look at a narrow window.



This week of data shows a clear diel signal in the surface water and evidence of a tidal signal in the bottom water.

1. Resample mean adjusted residuals.

2. Model cycles with Fourier terms and resample cycle detrended residuals.

3. Model cycles with Fourier terms and Simulate cycle detrended residuals.

Partition ConMon into daily units.

Stratify units by Season and Space.

Compute a mean adjusted vector for each Unit

$\boldsymbol{r_i = (DO_i - \overline{DO})}$ i = 1,2, . . . 24 where: $\boldsymbol{\overline{DO} = \frac{1}{24}\sum_1^{24} DO_i}$

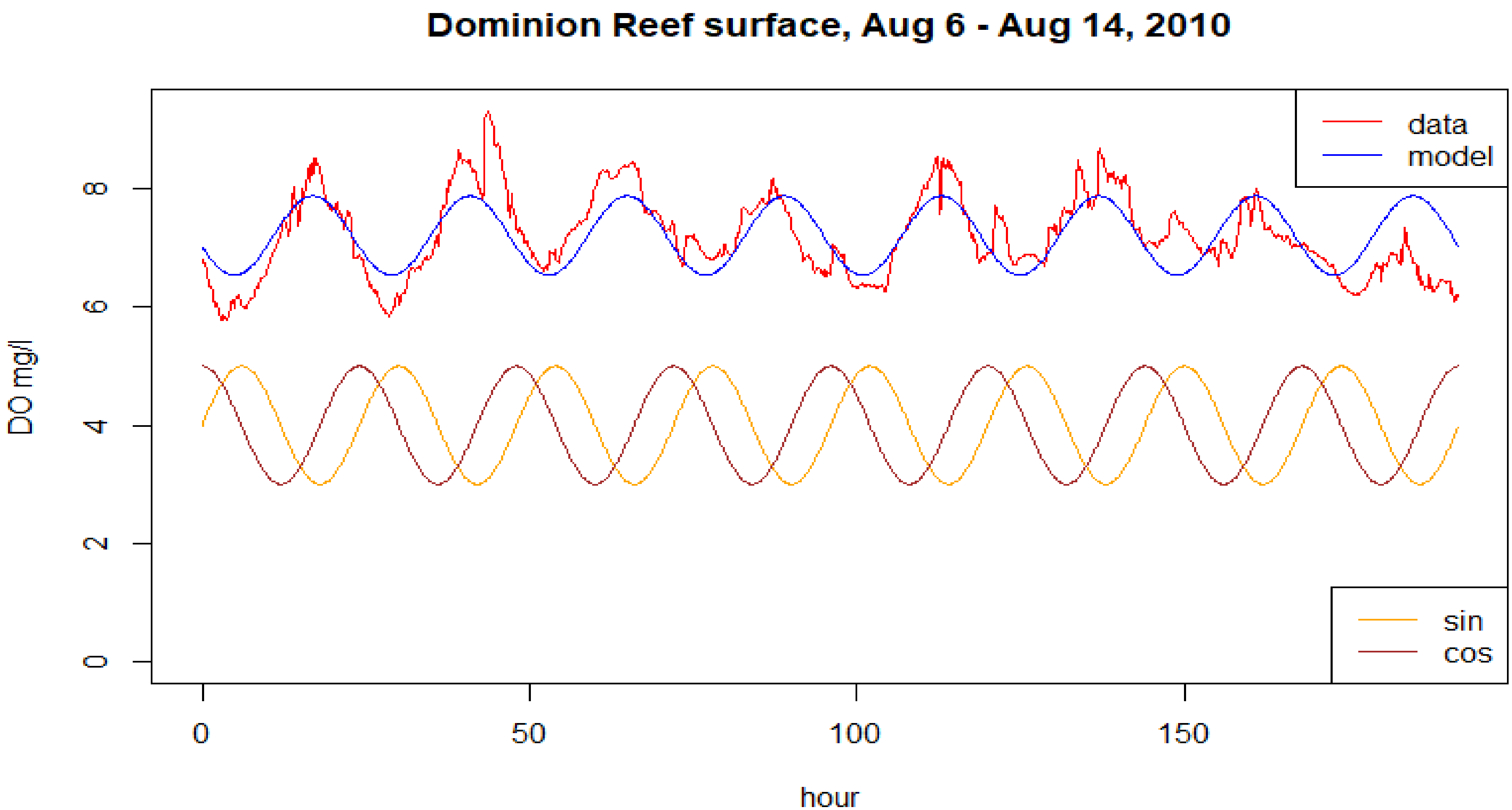
Simulation for 1 day = 24 hours

$\boldsymbol{\left(\widehat{DO_g} + r_i\right)}$ i = 1,2, . . . 24

Where: $\boldsymbol{\widehat{DO_g}}$ is a daily prediction from GAM.

This takes care of deterministic cycles and autocorrelated error all at once.

1. **Resample mean adjusted residuals.**
2. **Model cycles with Fourier terms and resample cycle detrended residuals.**
3. **Model cycles with Fourier terms and Simulate cycle detrended residuals.**



$$DO_h = lc * h + Har(24) + Har(12.42) + v_h \quad h = 1:24$$

$$Har(p,h) = sc^p * sin\left(\frac{2\pi * h}{p}\right) + cc^p * cos\left(\frac{2\pi * h}{p}\right)$$

Partition ConMon into daily units.

Fit a Fourier model F(h) to obtain Harmonic Coefficients (HC) and residuals.

Use GAM to model *HC* (GAM_{HC}) as a function of space, season, and trend.

Simulation for 1 day = 24 hours

$\left(\widehat{\boldsymbol{DO}_g} + \boldsymbol{F}_h\widehat{(\boldsymbol{HC})} + \boldsymbol{r}_h\right)$ h = 1,2, . . . 24

Where: $\widehat{\boldsymbol{DO}_g}$ is a daily prediction from GAM.

$\boldsymbol{F}(\boldsymbol{HC})$ **Is a hourly prediction from the Fourier model**

HC **could be a prediction from the GAM_{HC} for *HC***

or a random draw from a population of *HC* values.

$\boldsymbol{r}_h$ h = 1,2, . . . 24, **is a one day set of hourly residuals** $\boldsymbol{r}_h = \left(\boldsymbol{DO}_h - \boldsymbol{F}_h\widehat{(\boldsymbol{HC})}\right)$

Randomly drawn from a suitable stratum.

1. Resample mean adjusted residuals.
2. Model cycles with Fourier terms and resample cycle detrended residuals.
3. Model cycles with Fourier terms and Simulate cycle detrended residuals.

This 3rd method follows method 2 up to the simulated residuals. Instead of resampling, simulated error estimates are obtained through the recursion

$$\tau_i = \phi_h \tau_{i-1} + \epsilon_i \quad \epsilon \sim N(0, \sigma_h^2)$$

Or equivalently

$$\boldsymbol{\tau = chol(\Sigma_h)\ x\ \epsilon}$$

Where:

ϵ is a vector with distribution $N\left(0, \sigma^2 I\right)$,

τ is a vector with distribution $N\left(0, \sigma^2 \Sigma_h\right)$, and

$chol(\Sigma_h)$ is the Choleski decomposition of Σ_h

AR(1) correlation matrix

$$\Sigma_H = \begin{bmatrix} 1 & \varphi & \cdots & \varphi^{24} \\ \varphi & 1 & \varphi & \varphi^{23} \\ \vdots & \varphi & \ddots & \vdots \\ \varphi^{24} & \varphi^{23} & \cdots & 1 \end{bmatrix}_{24x24} \quad \sigma^2$$

It turns out

$$chol(\Sigma_h) = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ \varphi & c & 0 & 0 & 0 \\ \varphi^2 & \varphi^1 c & c & 0 & 0 \\ \varphi^3 & \varphi^2 c & \varphi^1 c & c & 0 \\ \varphi^4 & \varphi^3 c & \varphi^2 c & \varphi^1 c & c \end{bmatrix} \text{ where } c = \sqrt{1 - \varphi^2}$$

The GAM tool for Daily Means (GAM_{dm})

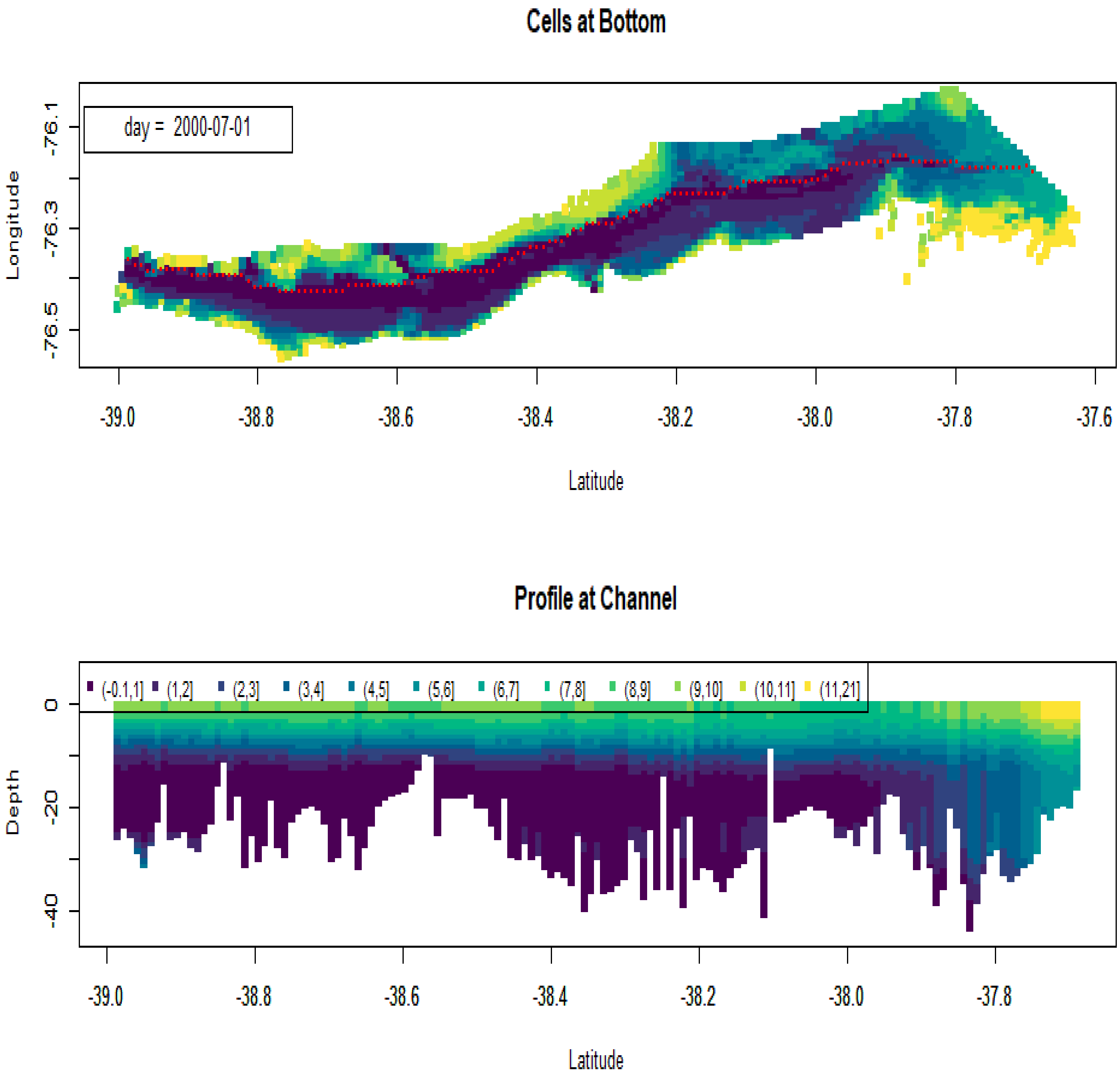
Captures large scale variation in DO as a function of

Estuarine Longitude
Estuarine Latitude
Sample Depth
Bottom Depth
Long term trend
Seasonal Trend.

Produces a predictions in a time x depth x longitude x latitude lattice at a resolution of:

Time:	**one per day**
Depth:	**one per meter**
Longitude:	**1 per kilometer**
Latitude:	**1 per kilometer**

Plane and Profile views of the Model Prediction Space.



Two Stochastic Elements in Daily Predictions:

Model uncertainties in GAM_{dm}

Space-Time correlated Errors.

Model uncertainty:

The daily prediction GAM_{dm} fit yields:

A parameter vector, $\widehat{\boldsymbol{\beta}}$, and

an associated variance-covariance matrix $\widehat{\boldsymbol{\Sigma_\beta}}$

To account for model uncertainty, simulate a prediction parameter vector $\boldsymbol{\beta}^*$ using a multivariate normal distribution,

$$\boldsymbol{\beta}^* \sim \boldsymbol{MVN}\left(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\Sigma_\beta}}\right)$$

$$\boldsymbol{\beta}^* = \widehat{\boldsymbol{\beta}} + \boldsymbol{chol}(\widehat{\boldsymbol{\Sigma_\beta}})\boldsymbol{\epsilon} \quad \textbf{where} \quad \boldsymbol{\epsilon} \sim \boldsymbol{MVN}(\mathbf{0}, \mathbf{I})$$

Length($\widehat{\boldsymbol{\beta}}$) < 100, chol() requires order n^3 operation

Daily Mean predictions are computed as

$$\boldsymbol{d}^* = \boldsymbol{X}\boldsymbol{\beta}^* + \boldsymbol{\varepsilon}$$

We would like $\boldsymbol{\varepsilon}$ to have realistic space-time dependence.

Space-Time correlated Errors

Consider building a small example of a Space Time Variance Covariance matrix:

Assume the predictions are sorted by Longitude, Latitude, Depth and Day.

For simplicity, assume five levels of each dimension.

The variance-covariance matrix = Σ_{st} has dimensions 625 x 625.

5 days at a point in space would have an AR(1) correlation matrix

$$\Sigma_T = \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 & \rho^4 \\ \rho & 1 & \rho & \rho^2 & \rho^3 \\ \rho^2 & \rho & 1 & \rho & \rho^2 \\ \rho^3 & \rho^2 & \rho & 1 & \rho \\ \rho^4 & \rho^3 & \rho^2 & \rho & 1 \end{bmatrix} \sigma^2$$

Assume that observations across space are independent. Then the full VC matrix has the form.

$$\Sigma_{ST} = \sigma^2 \begin{bmatrix} \Sigma_T & 0 & \cdots & \cdots & 0 \\ 0 & \Sigma_T & & & \vdots \\ \vdots & & \ddots & & \vdots \\ \vdots & & & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & \Sigma_T \end{bmatrix}_{625x625}$$

Assume that adjacent layers in depth are dependent.

$$\Sigma_{TS10} = \begin{bmatrix} \square & 1111 & 1112 & 1113 & 1114 & 1115 & 1121 & 1122 & 1123 & 1124 & 1125 \\ 1111 & 1 & \rho & \rho^2 & \rho^3 & \rho^4 & \tau & 0 & 0 & 0 & 0 \\ 1112 & \rho & 1 & \rho & \rho^2 & \rho^3 & 0 & \tau & 0 & 0 & 0 \\ 1113 & \rho^2 & \rho & 1 & \rho & \rho^2 & 0 & 0 & \tau & 0 & 0 \\ 1114 & \rho^3 & \rho^2 & \rho & 1 & \rho & 0 & 0 & 0 & \tau & 0 \\ 1115 & \rho^4 & \rho^3 & \rho^2 & \rho & 1 & 0 & 0 & 0 & 0 & \tau \\ 1121 & \tau & 0 & 0 & 0 & 0 & 1 & \rho & \rho^2 & \rho^3 & \rho^4 \\ 1122 & 0 & \tau & 0 & 0 & 0 & \rho & 1 & \rho & \rho^2 & \rho^3 \\ 1123 & 0 & 0 & \tau & 0 & 0 & \rho^2 & \rho & 1 & \rho & \rho^2 \\ 1124 & 0 & 0 & 0 & \tau & 0 & \rho^3 & \rho^2 & \rho & 1 & \rho \\ 1125 & 0 & 0 & 0 & 0 & \tau & \rho^4 & \rho^3 & \rho^2 & \rho & 1 \end{bmatrix} \sigma^2$$

At this point, Cholesky decomposition becomes a problem for a matrix of this size.

- **Solving Cholesky iteratively did not yield a simple pattern like for AR1**
- **Possibly some expediency using block matrices.**
- **Possibly come at this by postulating chol(Σ_{TS}) that yields a reasonable Σ_{TS}.**
- **Possibly construct this as a nested random effects model.**
- **Possibly assume dependence in space behaves like dependence in time so we have a big AR1.**
- **Last resort, assume independence in space.**

Resampling is not an option because we don't have data on a 1km x 1km grid.