Overview of 4-D Interpolator Project

Elgin Perry
eperry@chesapeake.net
DIWG 2/28/2023

Statement of goal

Goal is to build a tool to be used in the improved assessment of short term criteria such as weekly mean and daily minimum.

We do NOT expect this tool to accurately predict DO in a specific hour on a specific date, but instead to assess the frequency of violations during an assessment period.

The Bay Oxygen Research Group (BORG) Team

CBP Coordination
Peter Tango, August Goldfischer, Breck Sullivan, Kaylyn Gootman, Lee McDonnell

Statistics and Programming
Rebecca Murphy, Elgin Perry, Jon Harcum (Tetra Tech), Erik Leppo (Tetra Tech)

Modeling Group
Gary Shenk, Isabella Bertani, Richard Tian

GIS support
Zhaoying (Angie) Wei

EPA Guidance
Jim Hagy

Conceptual diagram

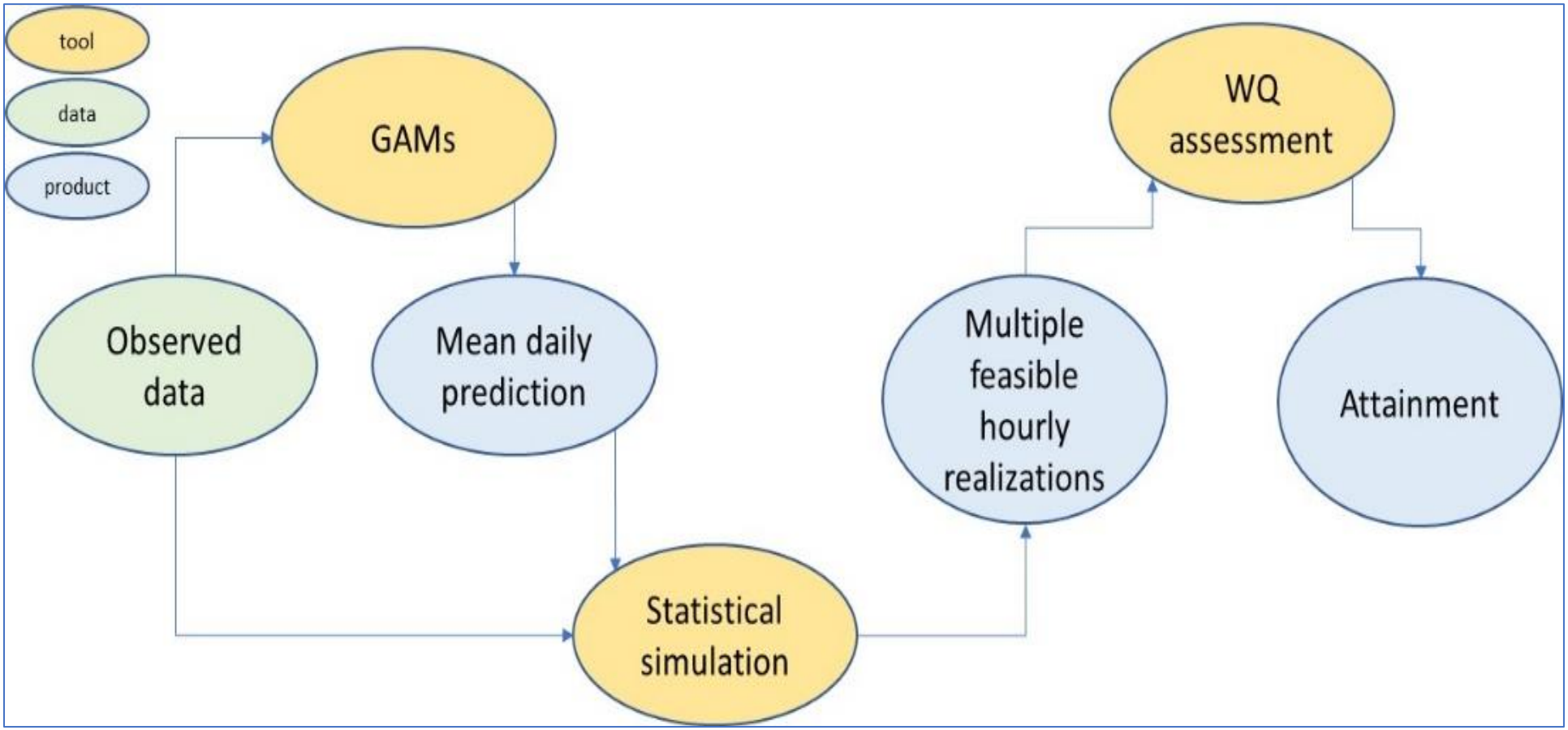
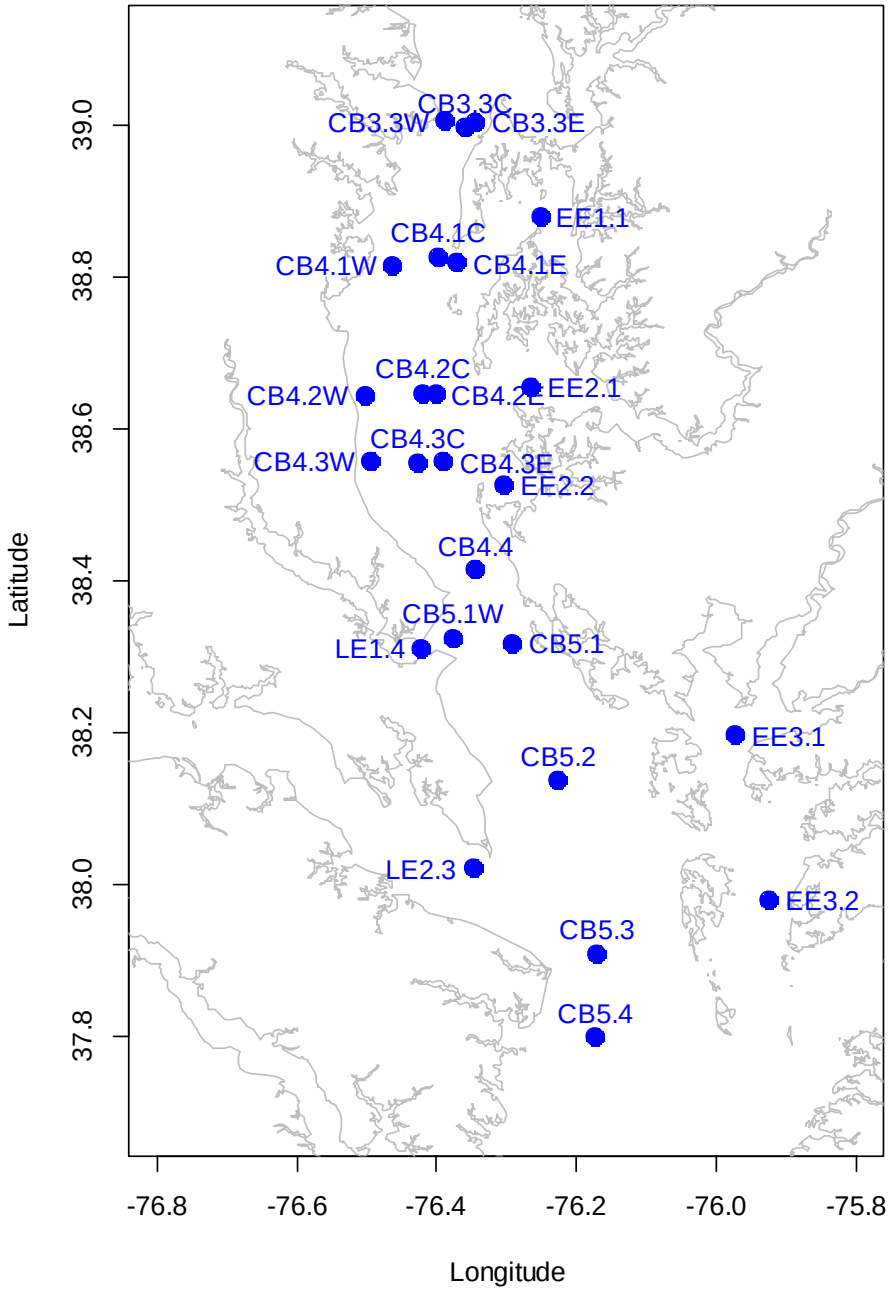


Figure 1: Interpolation and attainment assessment system

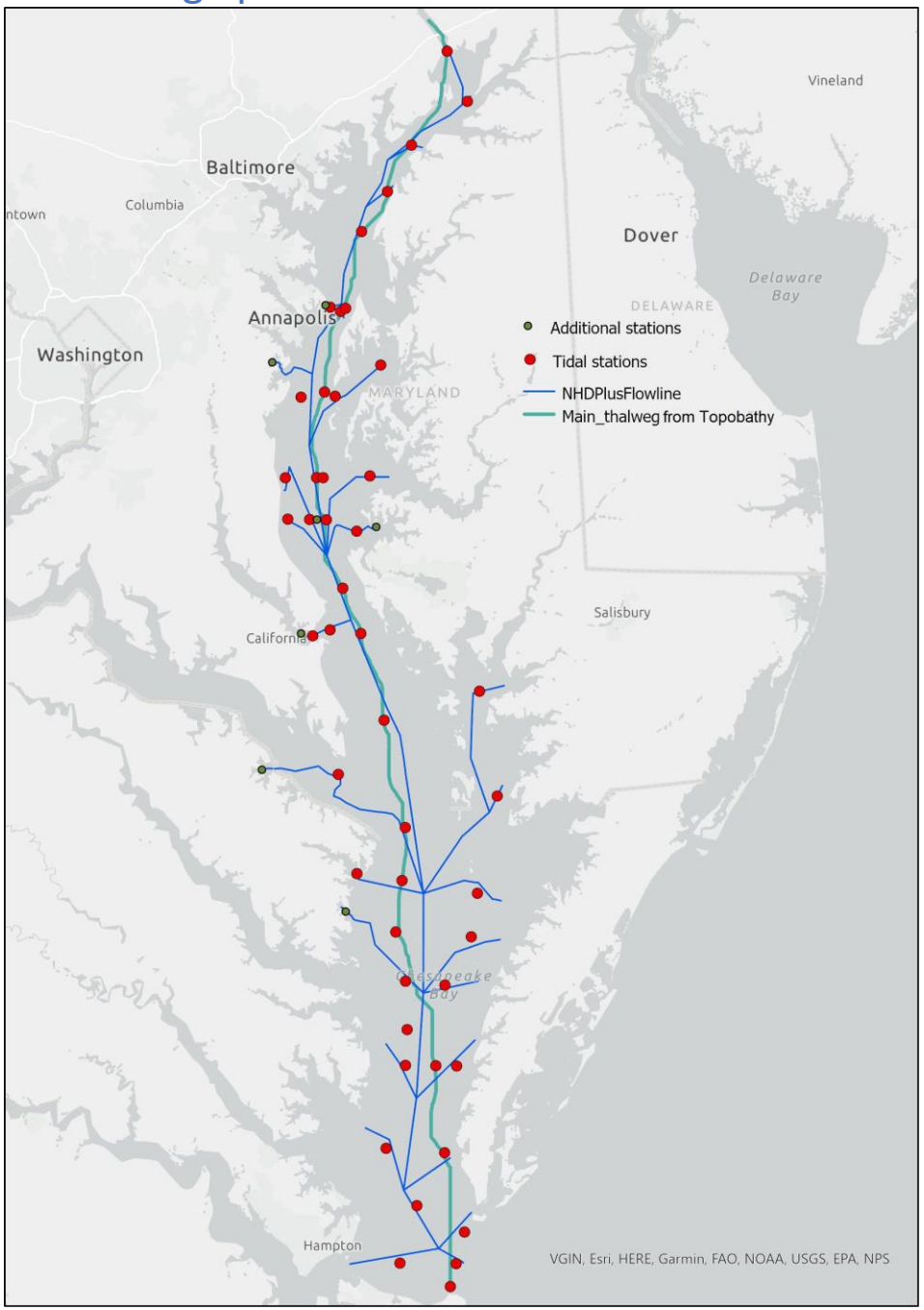
First: Consider what happens in the GAMs tool

Second: Consider what happens in the Statistical Simulation tool.

Map of stations in the test region. Test Period: 1990-2010, monthly and twice monthly observations.



Measuring Space



Current Test Model

Independent variables must be known in the 4-D space.

Rebecca Proposed Model (gs7):

```
Approximate significance of smooth terms:
                                F p-value
s(decimal year)           121.920  <0.0001 ***
s(day of year)          48806.475  <0.0001 ***
s(water Depth)           1696.372  <0.0001 ***
s(Estuarine Longitude)    568.577  <0.0001 ***
s(Estuarine Latitude)       2.263   0.132
s(bottom Depth)            16.625  <0.0001 ***
ti(LatKm,wDepth,LonKm)    239.602  <0.0001 ***
ti(wDepth,bDepth)          41.268  <0.0001 ***
ti(wDepth,ddate)           14.721  <0.0001 ***
ti(LonKm,ddate)            13.521  <0.0001 ***
ti(LatKm,ddate)            10.724  <0.0001 ***
ti(bDepth,ddate)            6.001  <0.0001 ***
ti(wDepth,doy)           2432.797  <0.0001 ***
ti(LonKm,doy)             176.734  <0.0001 ***
ti(LatKm,doy)             130.902  <0.0001 ***
ti(bDepth,doy)             25.453  <0.0001 ***
ti(ddate,doy)             162.919  <0.0001 ***
---
R-sq.(adj) =  0.863
```

Other potential independent Variables:

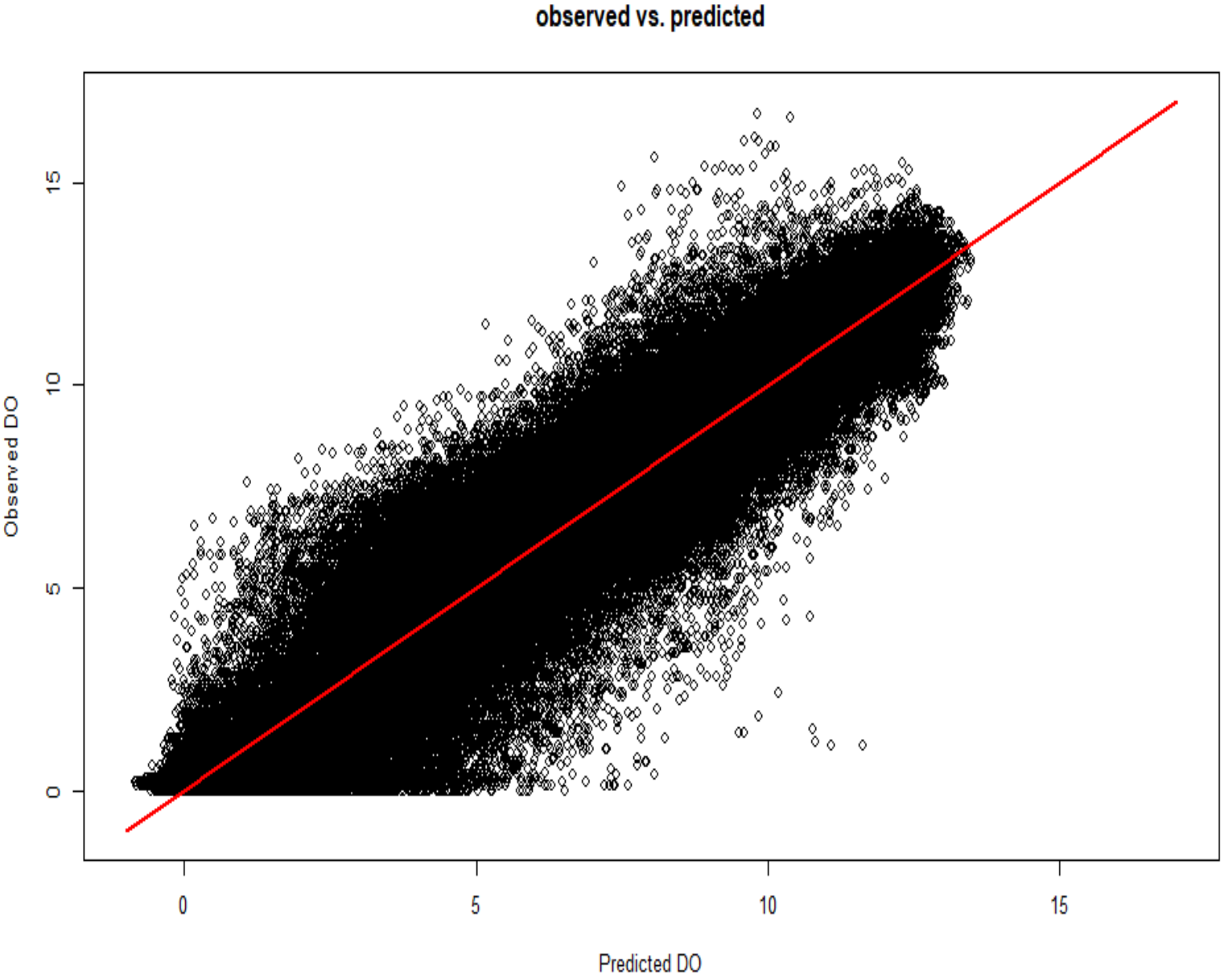
Flow (near continuous in time)

Have worked on adding flow with little success.

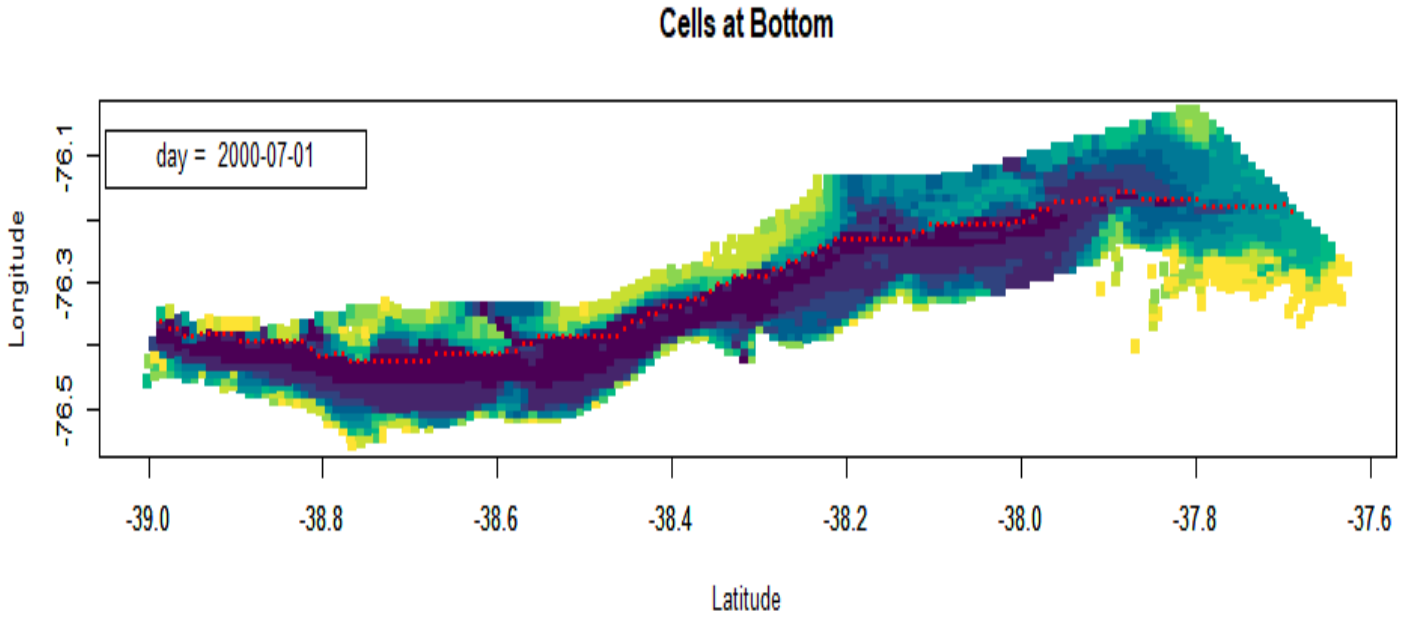
Solar Radiation (PAR)

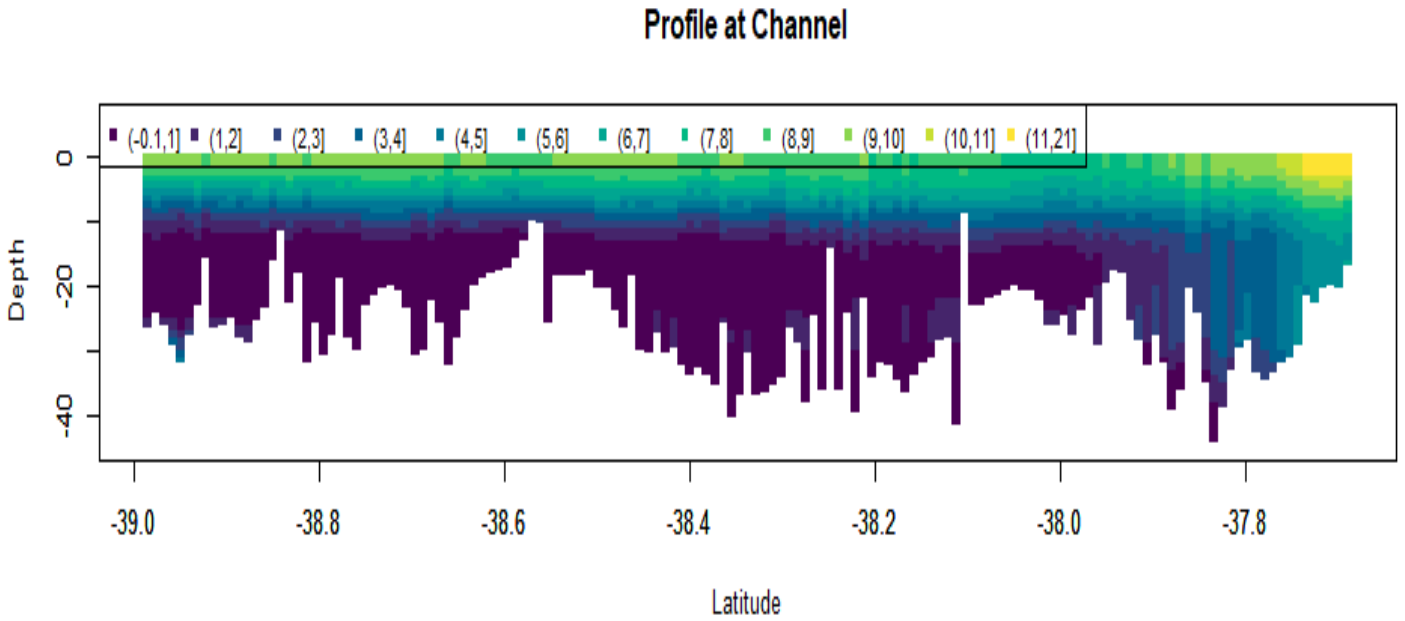
Wind

Temperature


observed vs. predicted
Observed DO
Predicted DO
0
5
10
15

Plane and Profile views of the Model Prediction Space.





Show Animation of Plan and Profile Predictions.
MidBayDO_profile_plane_200009Mar2022.mp4

Creating Cross-Validation (CV) data for time

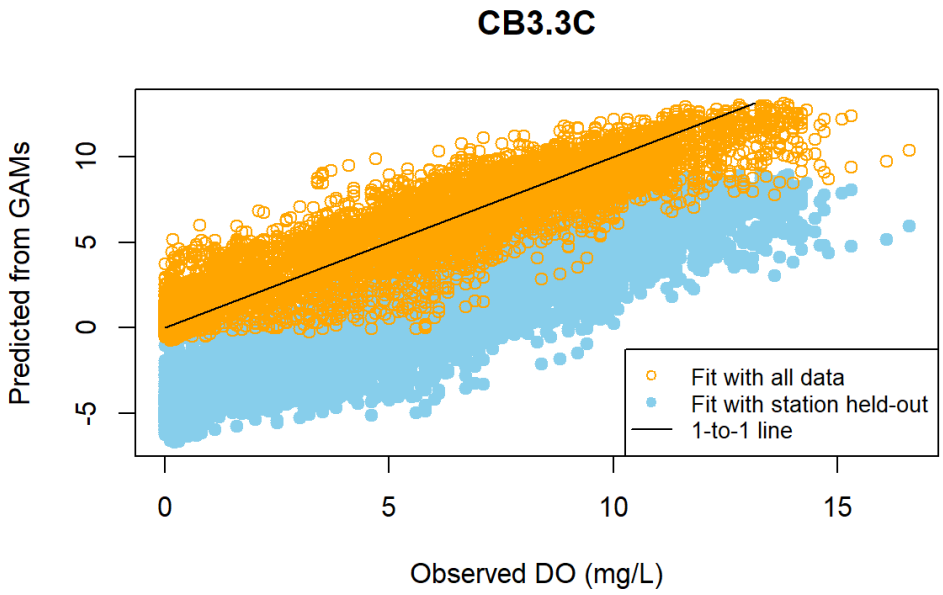
- **Trial uses mid-bay data 1990-2010 as in previous trials**
- **Treat each station-date as an event**
- **Identify months with two or more events**
- **In these months, randomly choose one event for CV**
- **Retain others for Training (TR).**

Statistic	Training	Validation
Route mean square error (rmse)	1.3354	1.4622
Median absolute deviation (mad)	0.7766	0.8755
R-square (rsq)	0.8656	0.838

Cross-Validation for space

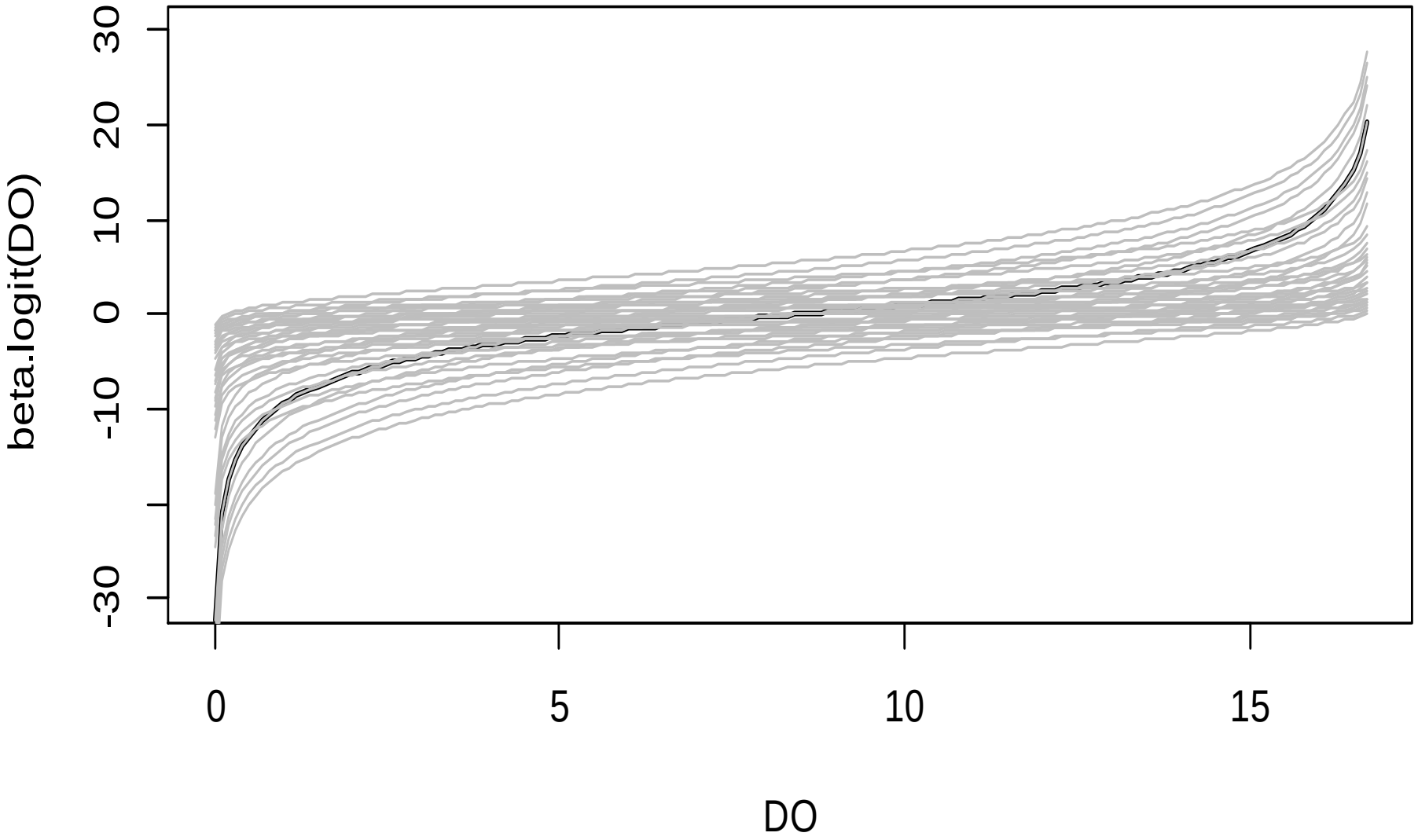
Jack Knife approach – leave out one station at a time (Rebecca's work))

Station	RMSE from full gs7 GAM fit	RMSE from gs7.cv model fit holding out this station's data
CB3.3C	1.50	5.06
CB3.3E	1.67	4.87
CB3.3W	1.77	4.17
CB4.1C	1.36	2.43
CB4.1E	1.42	2.50
CB4.1W	1.68	4.42
CB4.2C	1.33	1.35
CB4.2E	1.47	1.60
CB4.2W	1.85	3.00
CB4.3C	1.25	3.36
CB4.3E	1.55	1.73
CB4.3W	1.67	2.35
CB4.4	1.29	1.35
CB5.1	1.33	1.44
CB5.1W	1.10	12.55
CB5.2	1.25	1.72
CB5.3	1.10	1.40
CB5.4	1.22	2.90
EE1.1	1.57	1.71
EE2.1	1.05	1.72
EE2.2	1.79	3.07
EE3.1	1.05	12.99
EE3.2	1.13	6.15
LE1.4	1.23	27.07
LE2.3	1.43	26.59

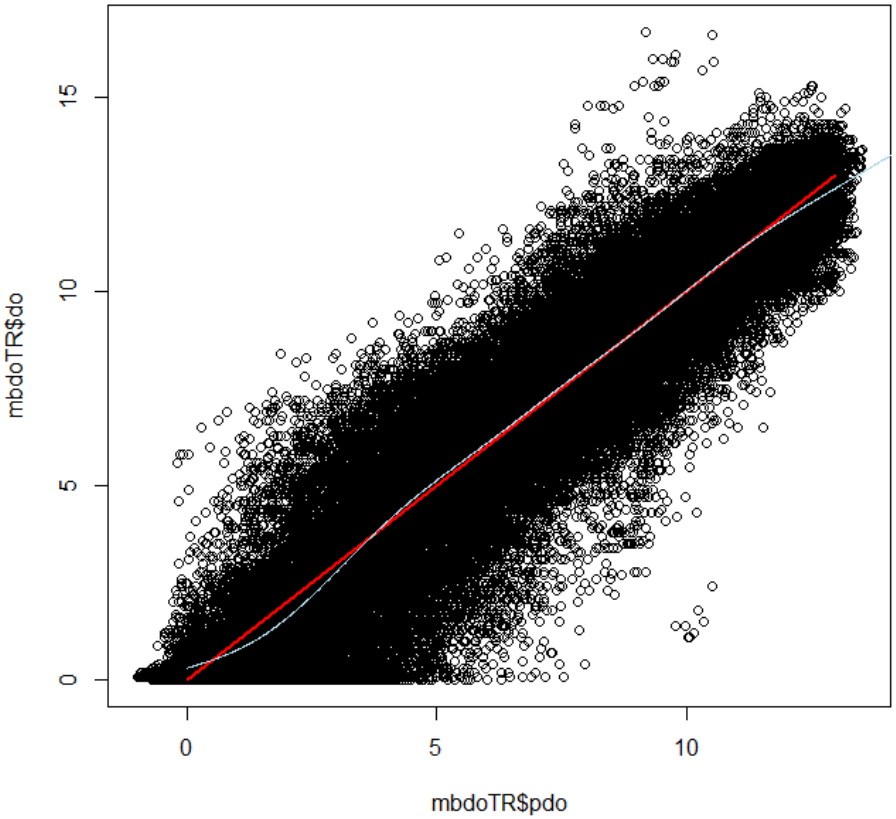


Lateral stations often have poor validation results. One "Fix" for this is to stiffen the gam smoothing functions.

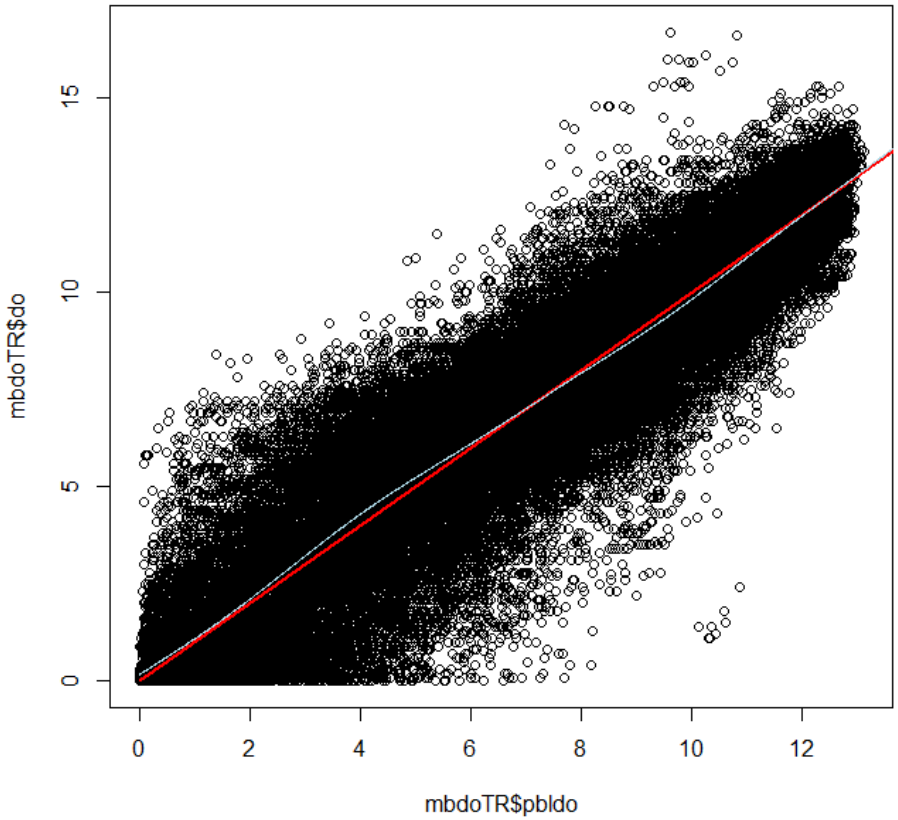
“Fix” negative predictions using Beta-Logit Transformation



Observed vs. Predicted Plots



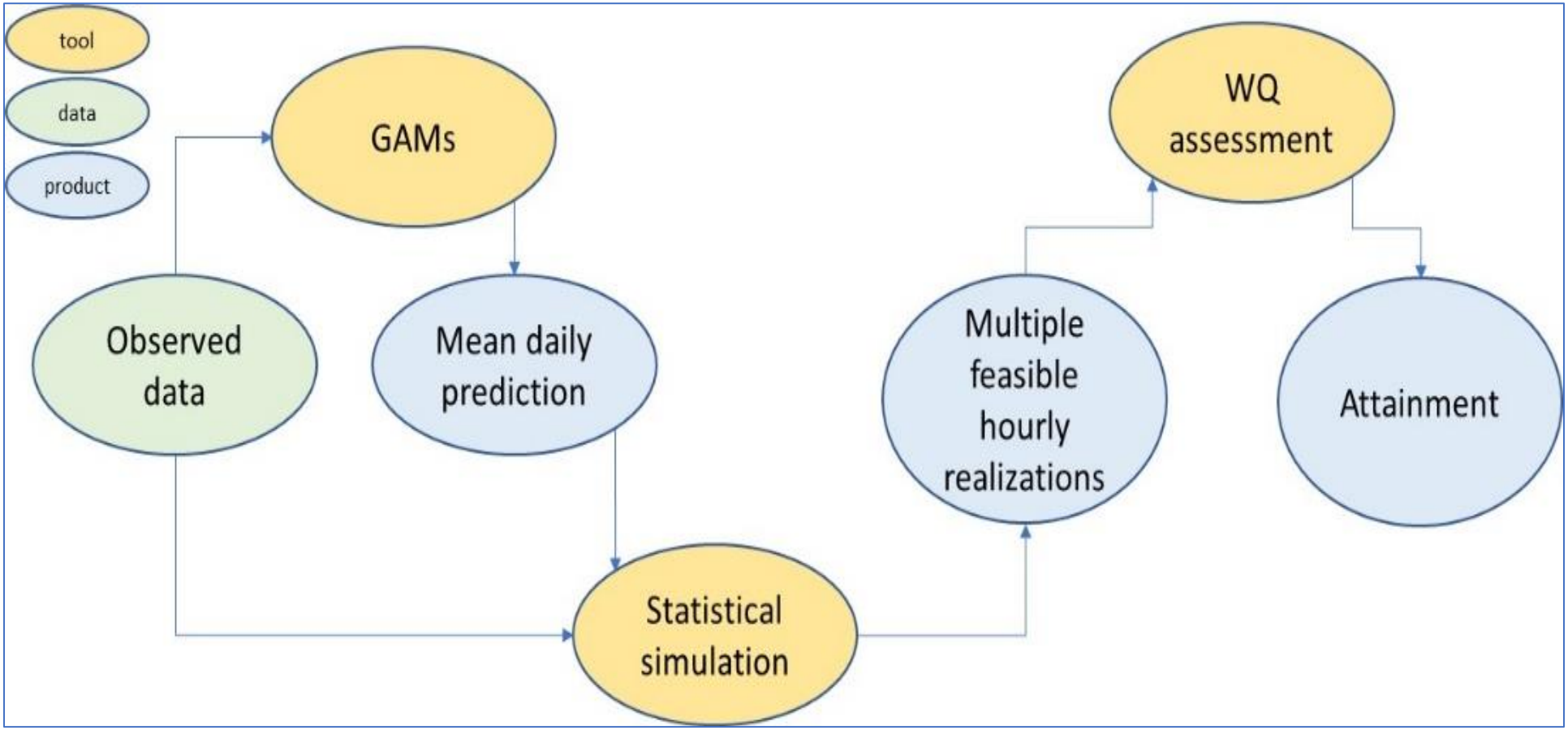
Original Scale



Beta-Logit Transform

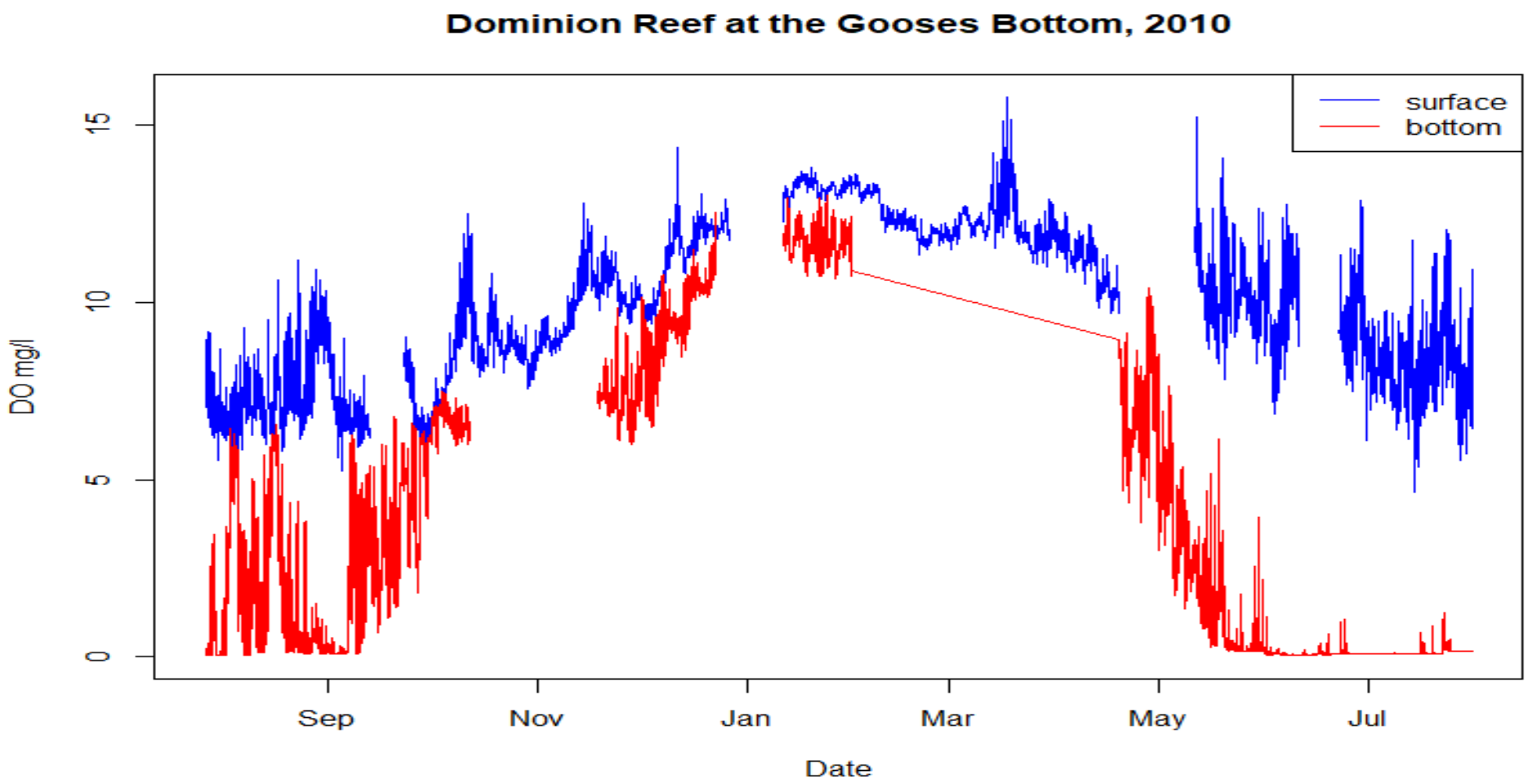
Training data:	R-square original scale	0.866
	R-square beta logit:	0.8666
Validation Data:	R-square original scale:	0.8381
	R-square beta logit:	0.8396

Next we look at High Frequency Temporal variation.

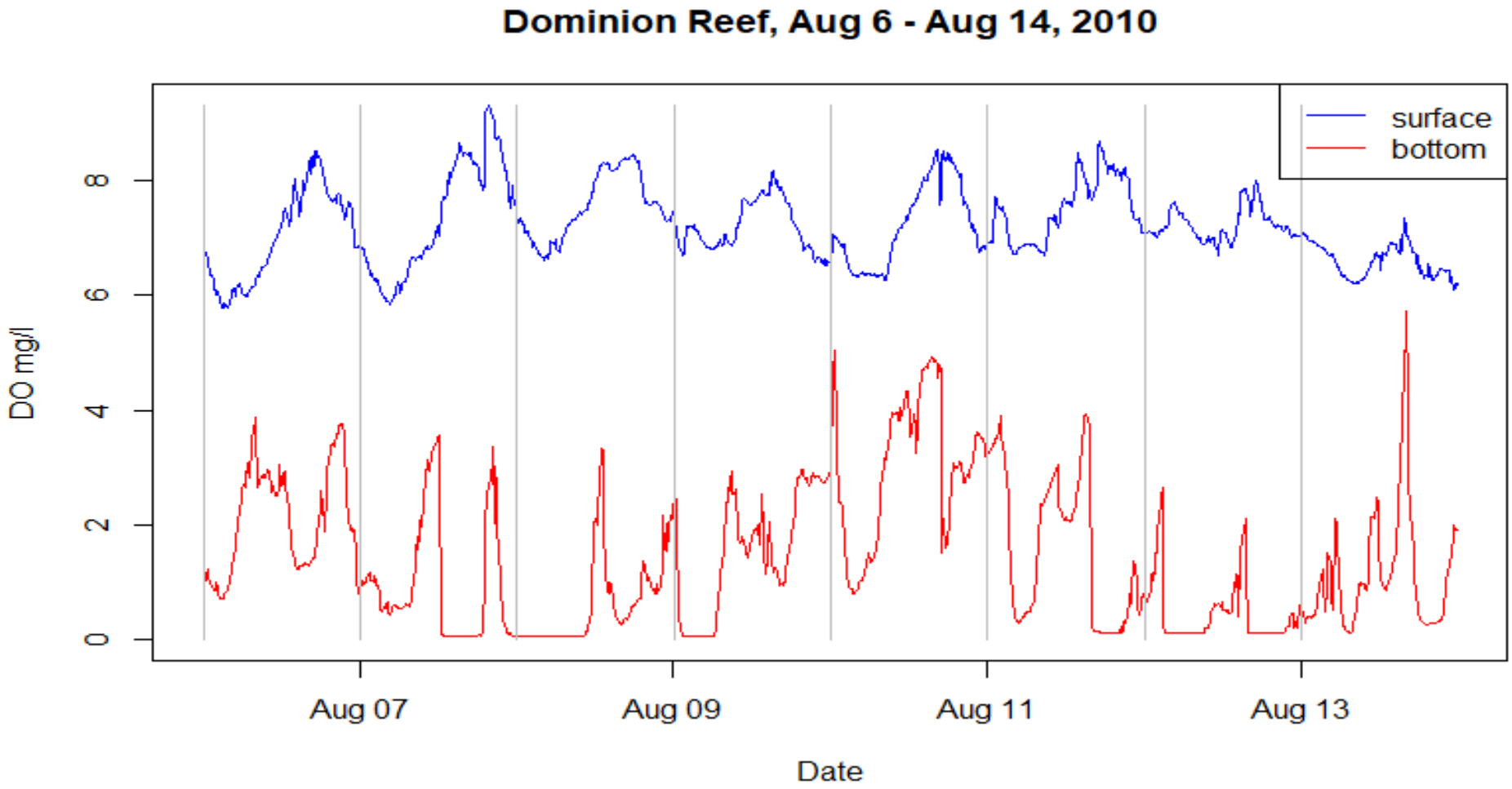


Second: Consider what happens in the Statistical Simulation tool.

Test data.

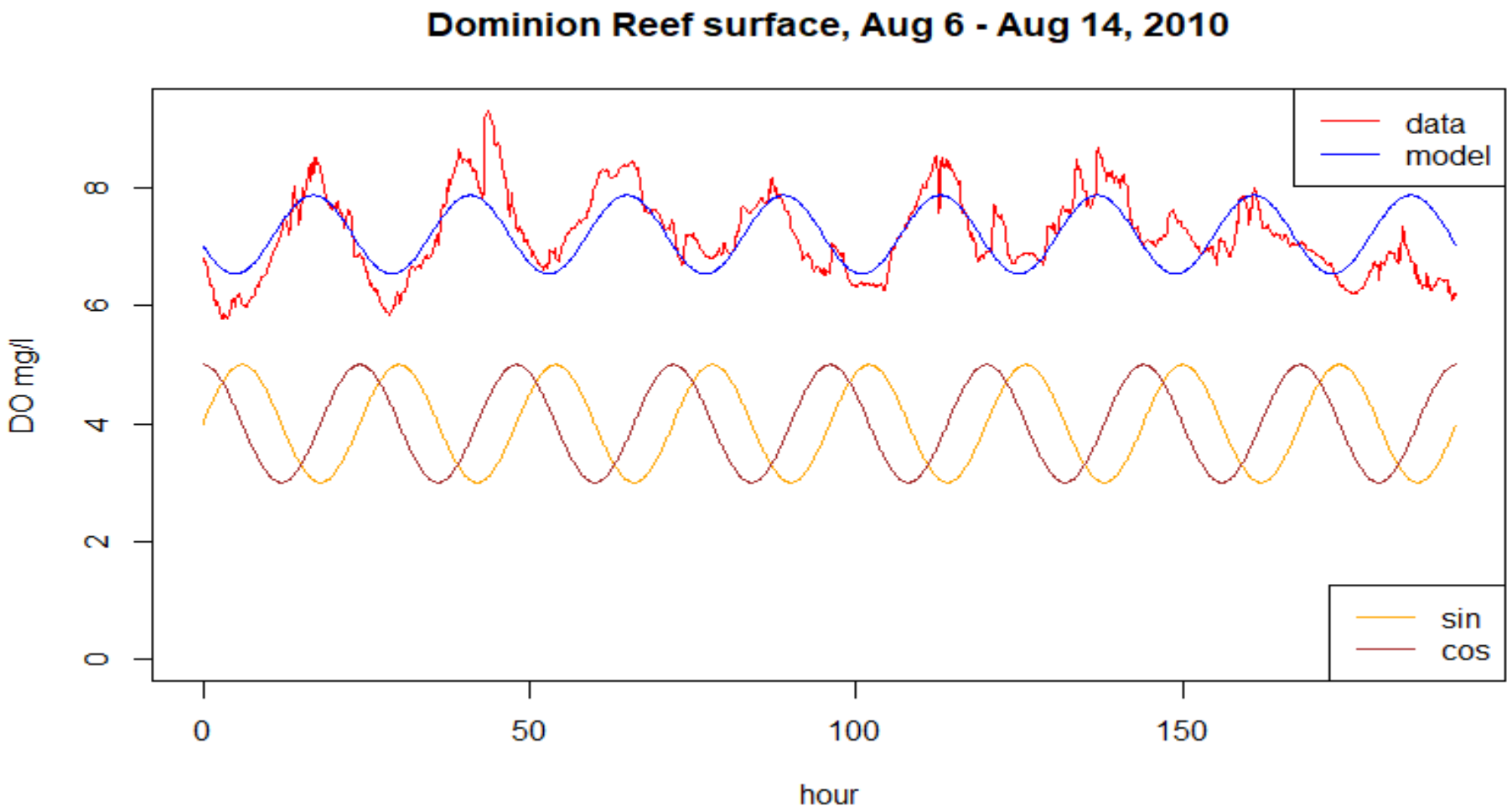

Dominion Reef at the Gooses Bottom, 2010
surface
bottom
DO mg/l
0
5
10
15
Sep
Nov
Jan
Mar
May
Jul
Date

Taking a closer look at a narrow window.



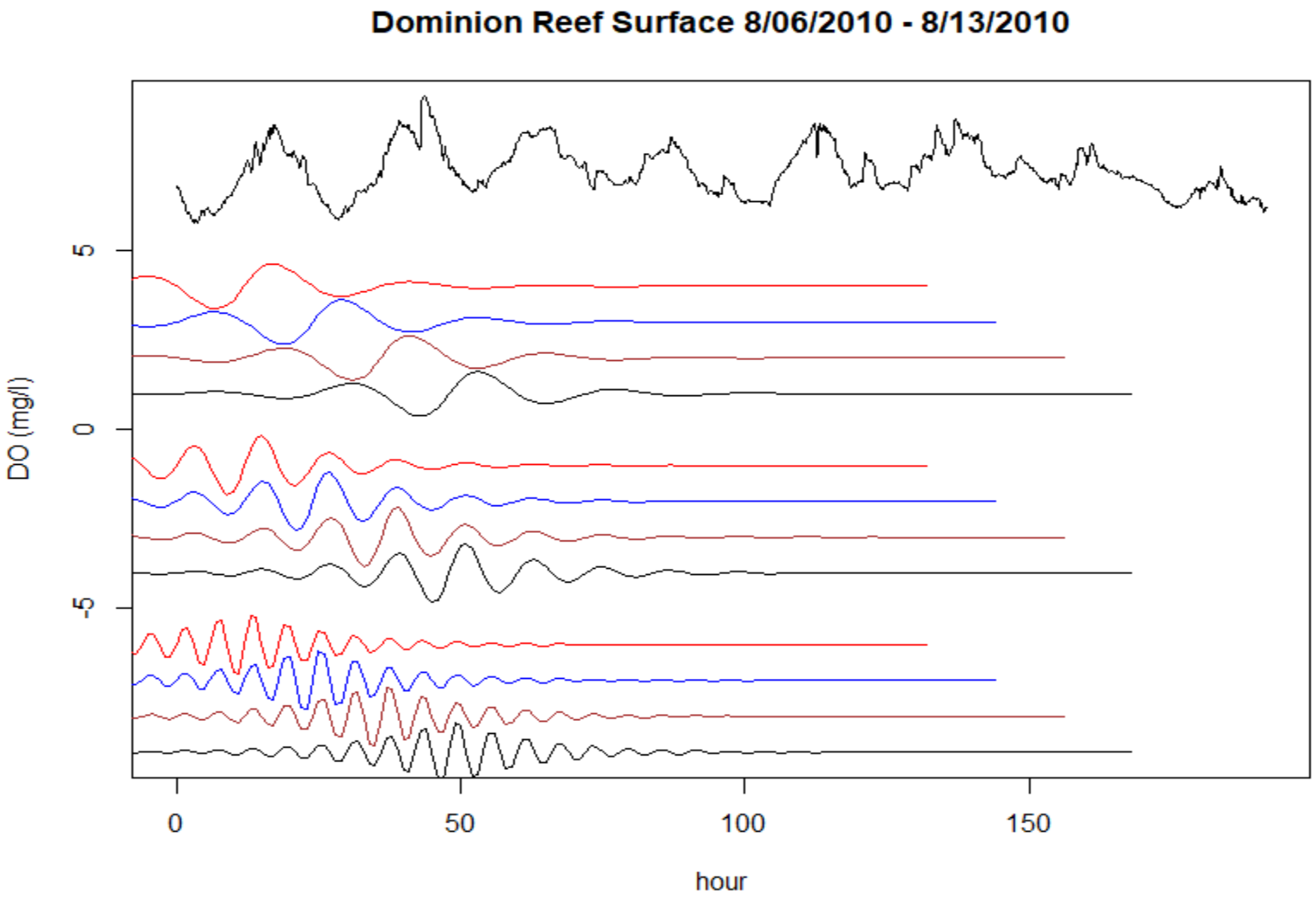
This week of data shows a clear diel signal in the surface water and evidence of a tidal signal in the bottom water (Jim Hagy suggested looking for this).

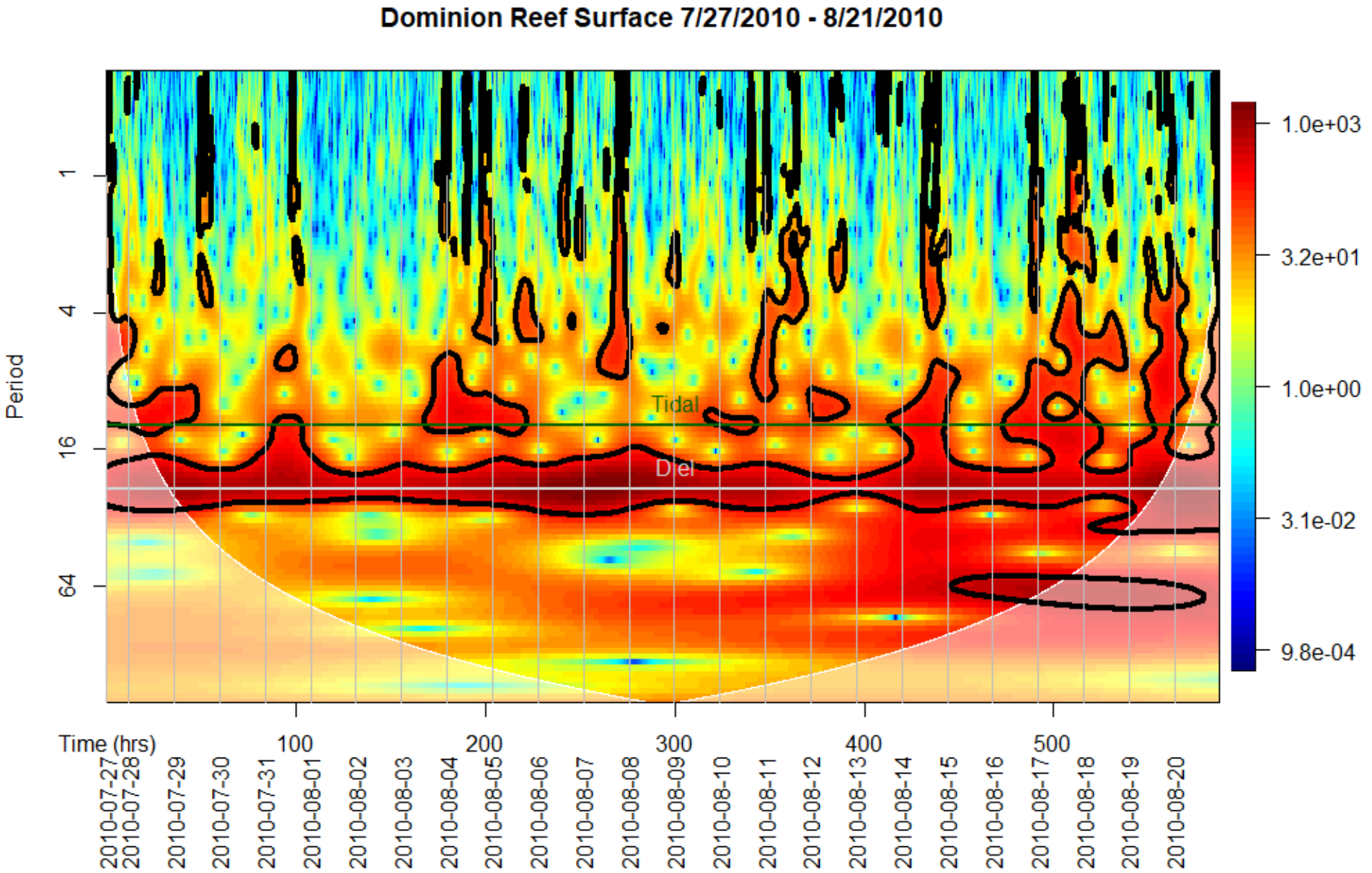
Basic Fourier Analysis

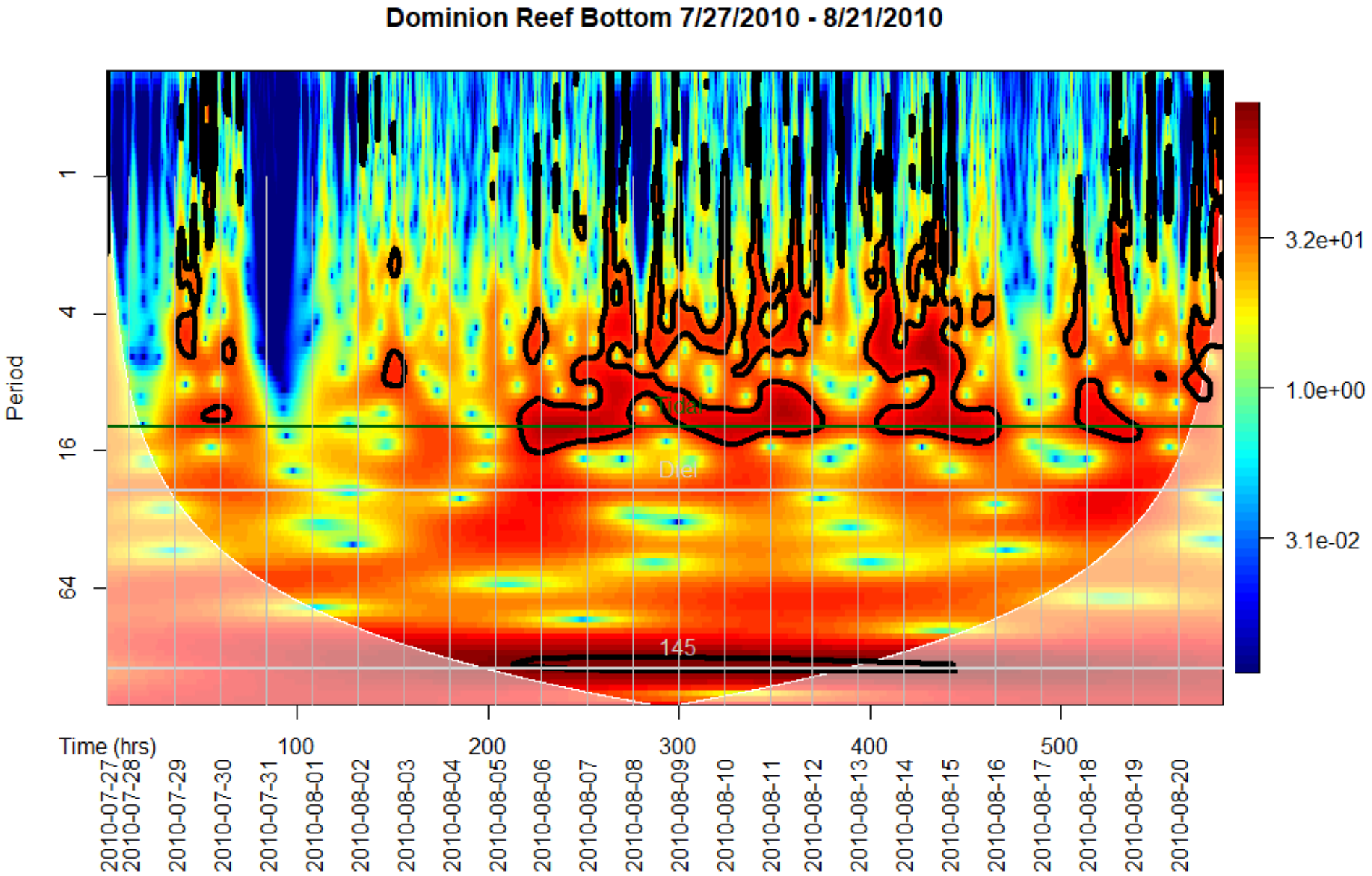


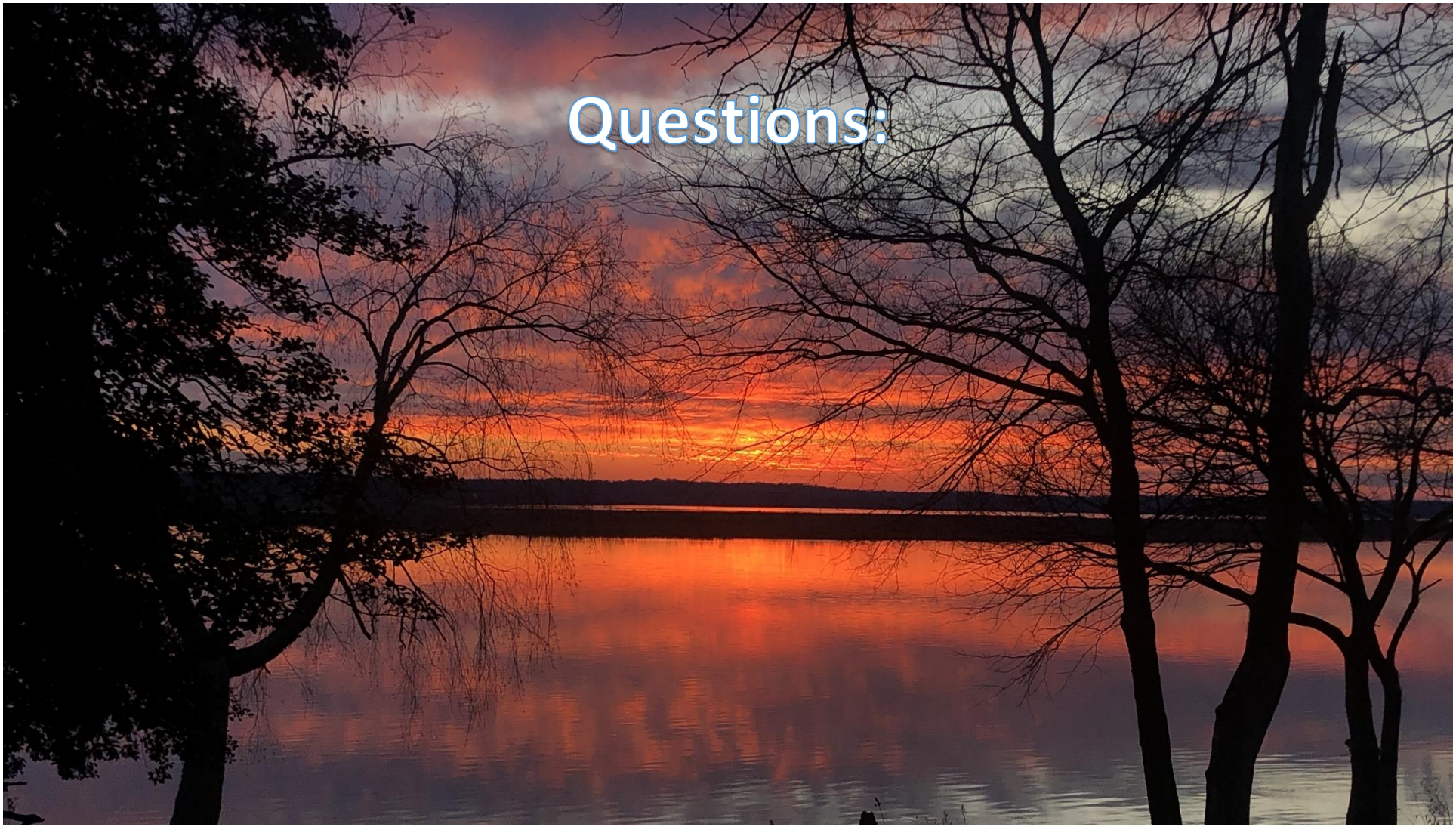
+ targets specific known periodicity
+ does not require evenly spaced data
+ yields amplitude information of signal
- targets specific known periodicity
- assumes stationarity (constant periodic cycle)
- computationally burdensome

Wavelet Analysis – The Concept




Dominion Reef Surface 7/27/2010 - 8/21/2010
Period
1
4
16
64
Tidal
Diel
1.0e+03
3.2e+01
1.0e+00
3.1e-02
9.8e-04
Time (hrs)
100
200
300
400
500
2010-07-27
2010-07-28
2010-07-29
2010-07-30
2010-07-31
2010-08-01
2010-08-02
2010-08-03
2010-08-04
2010-08-05
2010-08-06
2010-08-07
2010-08-08
2010-08-09
2010-08-10
2010-08-11
2010-08-12
2010-08-13
2010-08-14
2010-08-15
2010-08-16
2010-08-17
2010-08-18
2010-08-19
2010-08-20


Dominion Reef Bottom 7/27/2010 - 8/21/2010
Period
1
4
16
64
Tidal
Diel
145
3.2e+01
1.0e+00
3.1e-02
Time (hrs)
100
200
300
400
500
2010-07-27
2010-07-28
2010-07-29
2010-07-30
2010-07-31
2010-08-01
2010-08-02
2010-08-03
2010-08-04
2010-08-05
2010-08-06
2010-08-07
2010-08-08
2010-08-09
2010-08-10
2010-08-11
2010-08-12
2010-08-13
2010-08-14
2010-08-15
2010-08-16
2010-08-17
2010-08-18
2010-08-19
2010-08-20


Questions: