A Trial Run of 4-D interpolation of DO Using GAMs

STAC Advanced Monitoring Workshop

11 May, 2022

Presented by

Elgin Perry

On behalf of the Bay Oxygen Research Group

Peter Tango, Gary Shenk, Rebecca Murphy, Isabella Bertani, Breck Sullivan

Outline:

The Study Area

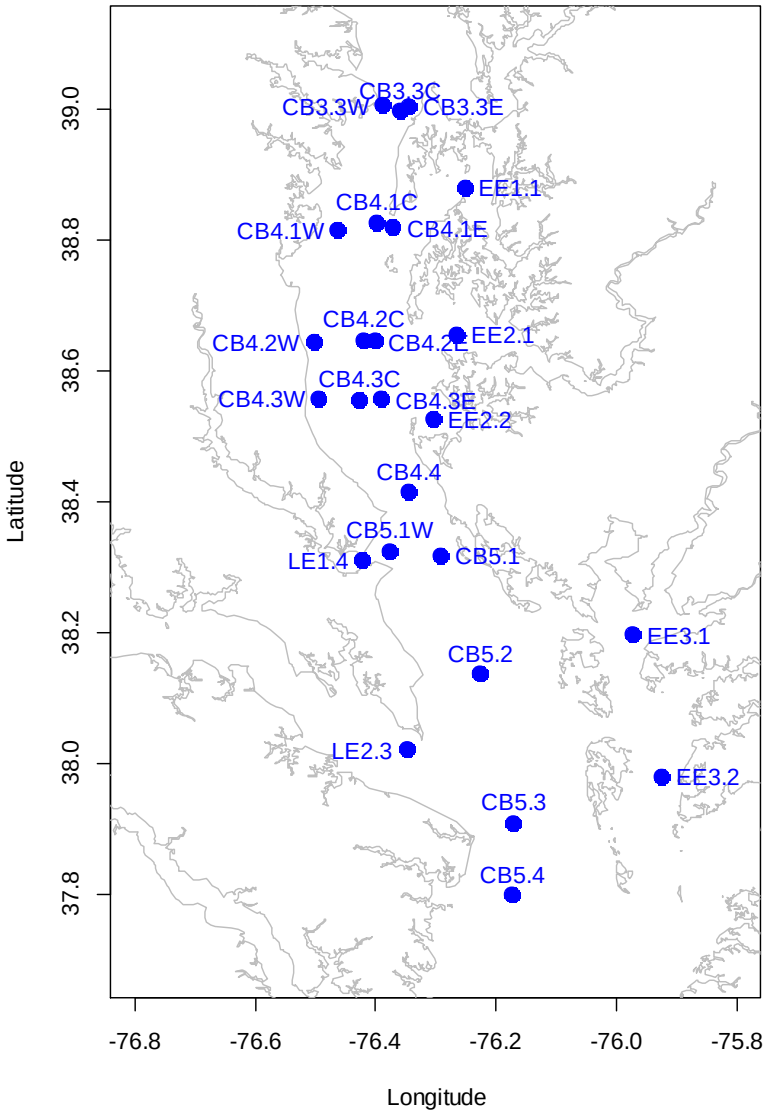
Predictor Variables

Trial Results

Adding lags of Flow

Animation of Predictions

Map of stations in the test region.



Predictor Variables:

Variable	Variable Name
Decimal Years from 1990-2010 centered on the year 2000	Centered Year
Day of Year	Day of Year
Depth of water at which sample was taken	Water Depth
Distance from fall line along estuary thalweg	Estuary Longitude
Nearest Distance to Thalweg	Estuary Latitude
Total water depth at a location	Bottom Depth

Predictor variable must be known in 4-d to be useful in 4-d.

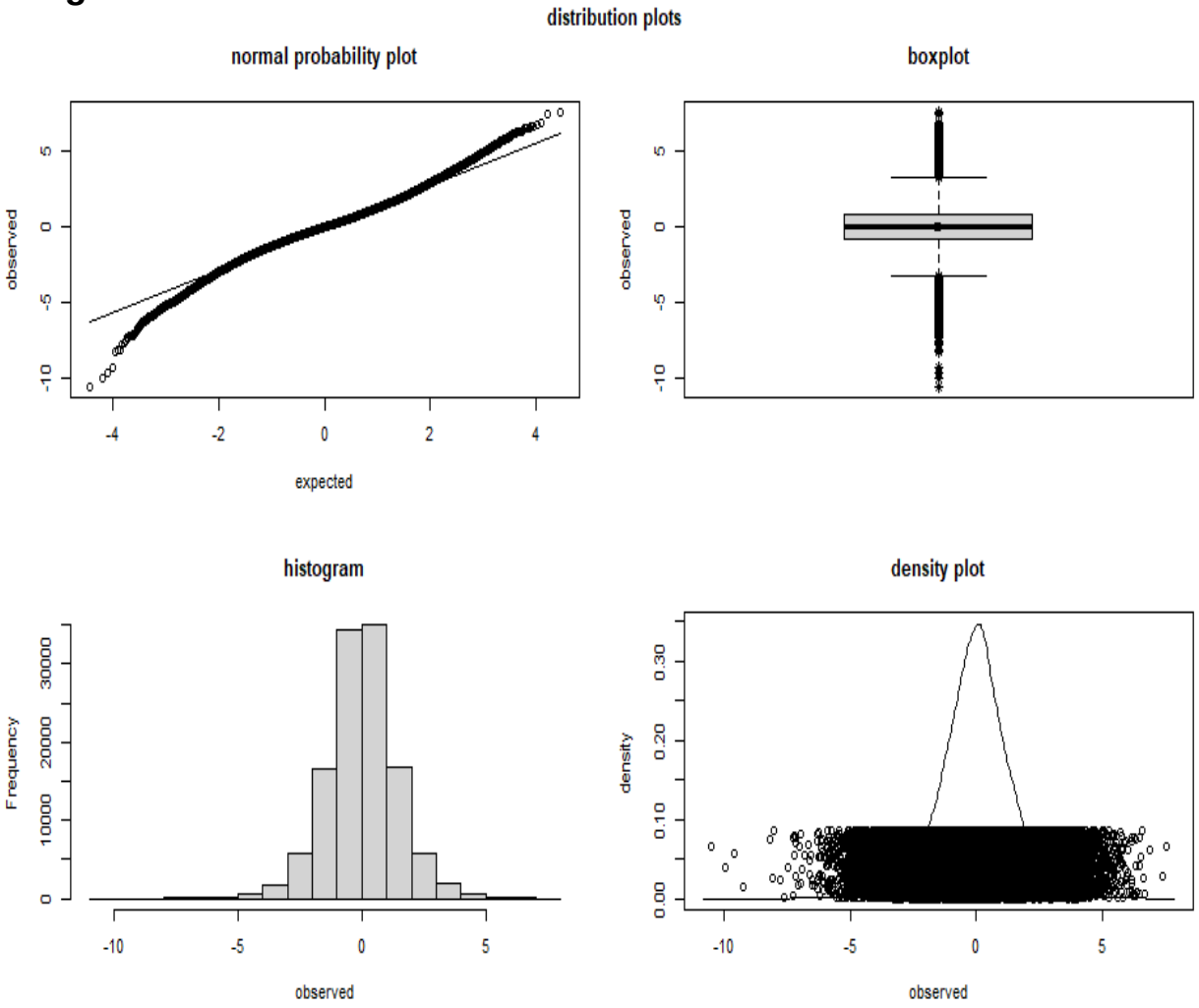
Using Variable Selection methods, a prototype model (gs6a) with the terms shown below was fitted to the test data.

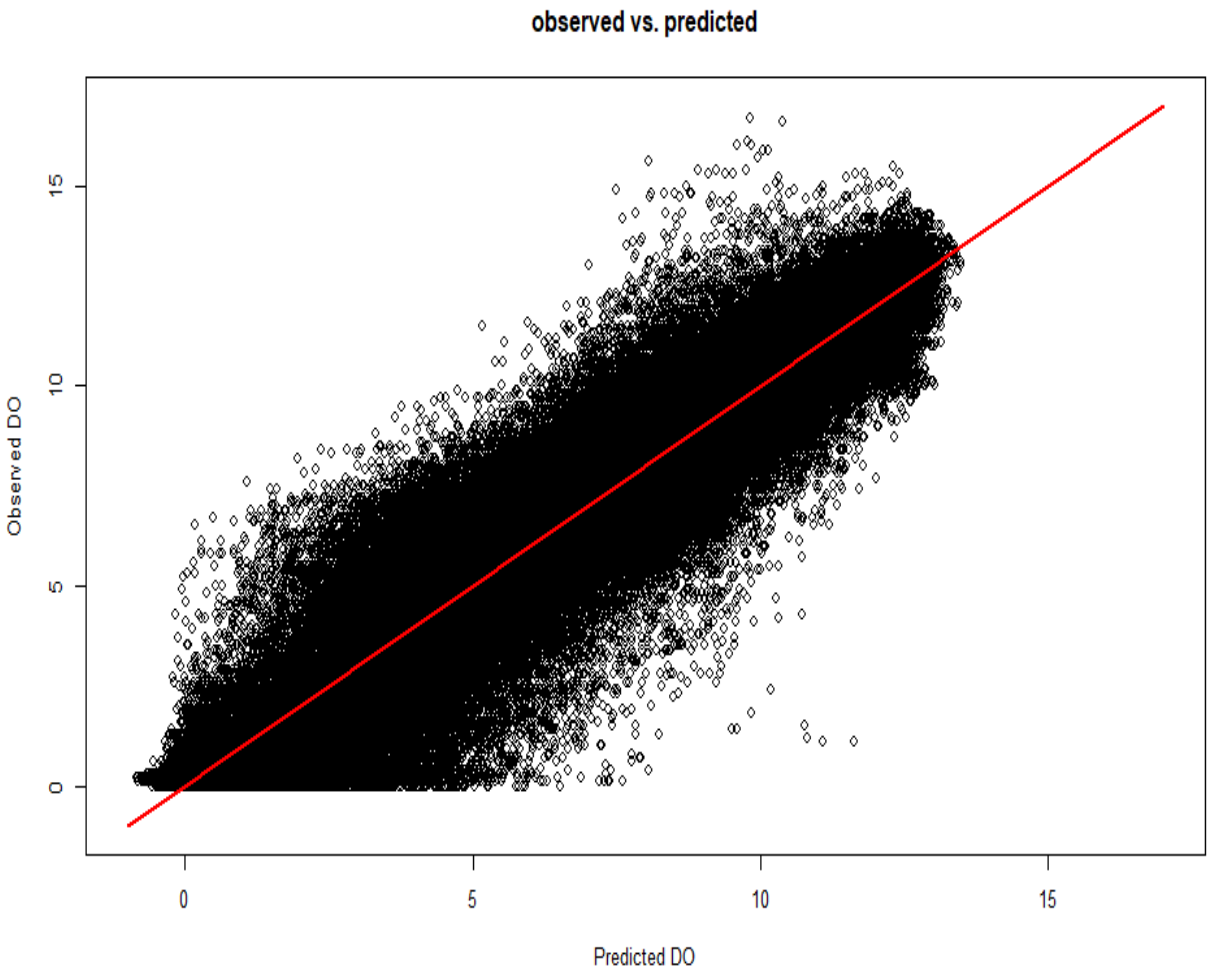
```
Parametric coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   8.3026     0.1526   54.42   <0.0001 ***

Approximate significance of smooth terms:
                                                    edf Ref.df         F  p-value
-------------------------------------------------------------------------------
s(Centered Year)                                 18.948 19.000   120.249  <0.0001
s(Day of Year)                                    7.961  8.000 49100.670  <0.0001
s(Water Depth)                                    9.000  9.000  8736.777  <0.0001
s(Estuary Longitude)                              9.000  9.000   400.597  <0.0001
s(Bottom Depth)                                   8.063  8.081    27.125  <0.0001
s(Estuary Latitude)                               6.442  6.743     9.965  <0.0001
ti(Water Depth,Day of Year)                      11.866 12.000  2628.248  <0.0001
ti(Estuary Longitude,Water Depth)                15.998 16.000   670.594  <0.0001
ti(Estuary Longitude,Day of Year)                11.901 12.000   197.701  <0.0001
ti(Centered Year,Day of Year)                    11.918 12.000   159.591  <0.0001
ti(Water Depth,Bottom Depth)                     15.965 15.999    42.566  <0.0001
ti(Centered Year,Water Depth)                     7.939  9.650    15.927  <0.0001
ti(Estuary Longitude,Centered Year)              12.309 14.004    17.944  <0.0001
ti(Water Depth,Day of Year,Estuary Longitude)    47.752 48.000    23.956  <0.0001
ti(Water Depth,Day of Year,Centered Year)        36.461 48.000    17.192  <0.0001
-------------------------------------------------------------------------------

R-sq.(adj) =  0.859   Deviance explained = 85.9%
GCV = 1.9399  Scale est. = 1.9362    n = 119283
```

Diagnostic Residual Plots

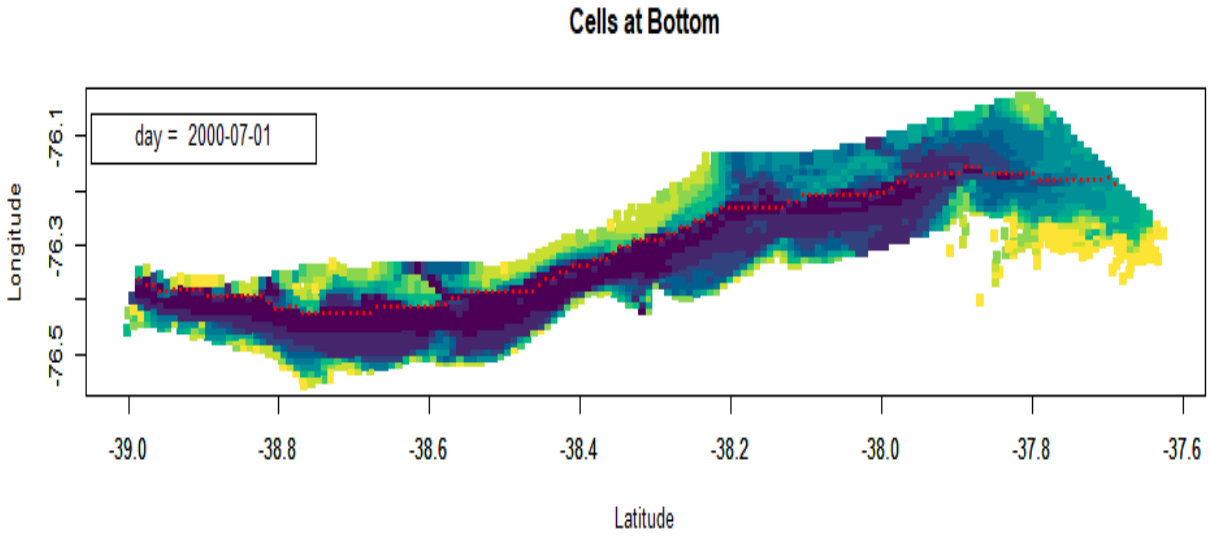


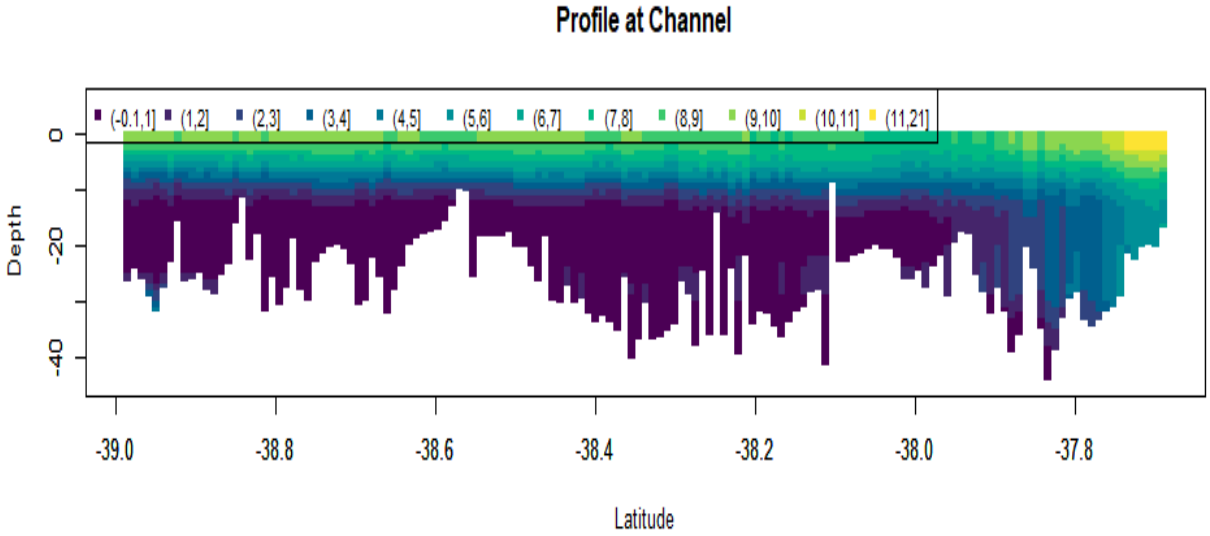

observed vs. predicted
Observed DO
Predicted DO
0
5
10
15

Summary of DO model runs in model development.

model	Term Added	rSquare	AIC	rmse	maxPres	maxNres	FitTime sec
gs1	Centered year, Day of Year	0.497	568892.9	2.63	12.2	10.2	0.98
gs2	Water Depth	0.77	475598.4	1.78	9.5	8.8	1.34
gs3	Estuary Longitude	0.792	463792.4	1.69	10	8.5	1.72
gs4a	Bottom Depth	0.794	462690.1	1.68	9.9	8.5	2.26
gs4b	Estuary Latitude	0.793	462895	1.68	9.6	8.3	2.21
gs4c	Water Depth & Estuary Latitude	0.795	462121.3	1.68	9.5	8.4	2.67
gs5	Water Depth*Day of Year	0.835	435780	1.5	8.3	9.1	3.49
gs5a	all two*way interactions	0.857	419385.1	1.4	7.5	10.5	18.33
gs6	Water Depth*Day of Year*Centered Year	0.858	418346.3	1.4	7.5	10.4	55.46
gs6a	Water Depth*Day of Year*Estuary Longitude	0.859	417555.3	1.39	7.5	10.5	60.19

Plane and Profile views of Model Predictions.





Show Animation of Plan and Profile Predictions.

List of variables yet to consider.

Flow (with appropriate lags)
Pycnocline
Wind
Temperature
Tide

Other ideas:

Outlier scrutiny
Kalman Filter
Cross Validation testing
Spatial correlation over depth

Conceptual model: Flow has multiple effects in the Estuary

1. Hydraulic Effect	**Hours to days**
2. Delivery Effect	**Days to Week**
3. Food Web Effect	**Months**

Access these concepts using variable selection on Moving Windows of Lags

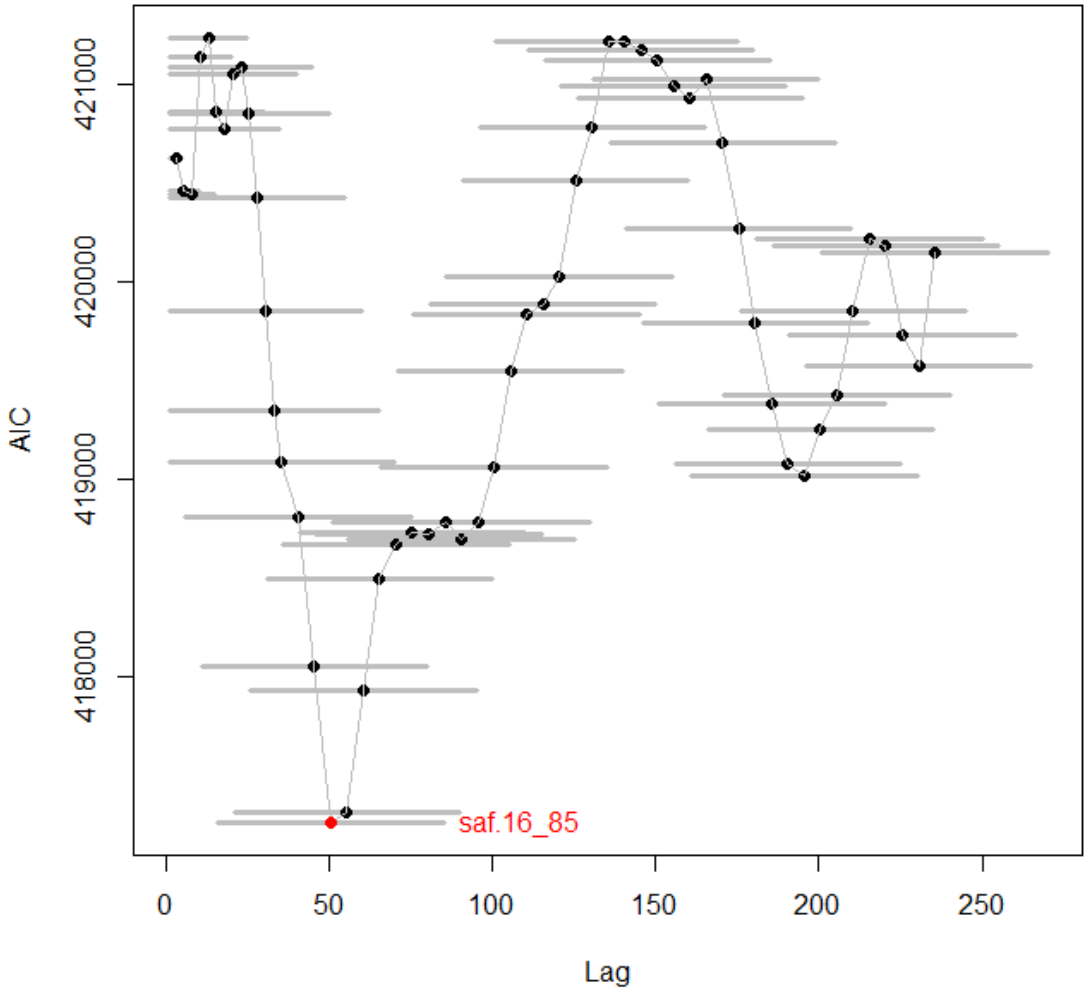
Flow values within a lag window are averaged to create a predictor.

Defined by a window size and a window increment

Example: Window = 70, increment = 5

1-1, 1-5, 1-10, . . . 1-70, 6-75, 11-80, . . . 201-270

Fit model with each lag of flow, select best fitting model based on AIC.



Approximate significance of smooth terms:

	edf	Ref.df	F	p-value
s(Day of Year)	7.982	8.000	14031.661	<0.0001
s(Water Depth)	8.322	8.816	24021.306	<0.0001
s(Estuary Longitude)	9.000	9.000	1049.892	<0.0001
s(Estuary Latitude)	8.611	8.870	162.550	<0.0001
ti(Water Depth,Day of Year)	11.809	12.000	2462.761	<0.0001
ti(Estuary Longitude,Water Depth)	15.980	16.000	786.237	<0.0001
ti(Estuary Longitude,Day of Year)	11.732	12.000	188.524	<0.0001
s(Seasonally Adjusted Flow.16_85)	9.000	9.000	154.785	<0.0001
ti(Day of Year,Seasonally Adjusted Flow.16_85)	11.717	12.000	153.805	<0.0001
ti(Water Depth,Seasonally Adjusted Flow.16_85)	13.551	14.840	60.255	<0.0001
ti(Estuary Longitude,Seasonally Adjusted Flow.16_85)	14.938	15.682	10.308	<0.0001
ti(Estuary Latitude,Seasonally Adjusted Flow.16_85)	14.943	15.797	9.027	<0.0001
s(Seasonally Adjusted Flow.196_265)	8.996	9.000	203.803	<0.0001
ti(Day of Year,Seasonally Adjusted Flow.196_265)	11.968	12.000	107.972	<0.0001
ti(Water Depth,Seasonally Adjusted Flow.196_265)	11.946	13.621	9.045	<0.0001
ti(Estuary Longitude,Seasonally Adjusted Flow.196_265)	12.143	13.487	15.032	<0.0001
ti(Estuary Latitude,Seasonally Adjusted Flow.196_265)	15.493	15.918	23.081	<0.0001
s(Seasonally Adjusted Flow.46_115)	9.000	9.000	96.849	<0.0001
ti(Day of Year,Seasonally Adjusted Flow.46_115)	11.773	12.000	116.353	<0.0001
ti(Water Depth,Seasonally Adjusted Flow.46_115)	14.654	15.498	30.317	<0.0001
ti(Estuary Longitude,Seasonally Adjusted Flow.46_115)	12.773	13.921	16.801	<0.0001
ti(Estuary Latitude,Seasonally Adjusted Flow.46_115)	15.455	15.929	16.494	<0.0001
ti(Water Depth,Day of Year,Estuary Longitude)	47.761	48.000	26.389	<0.0001

Deviance explained = 86.7%

As compared to Deviance explained = 85.2% for no-flow/no-time model

As compared to Deviance explained = 85.9% for time model

Conclusion: works better than I expected.

(End of Presentation)

Extra Slides:

Slides from here down may help with fielding questions

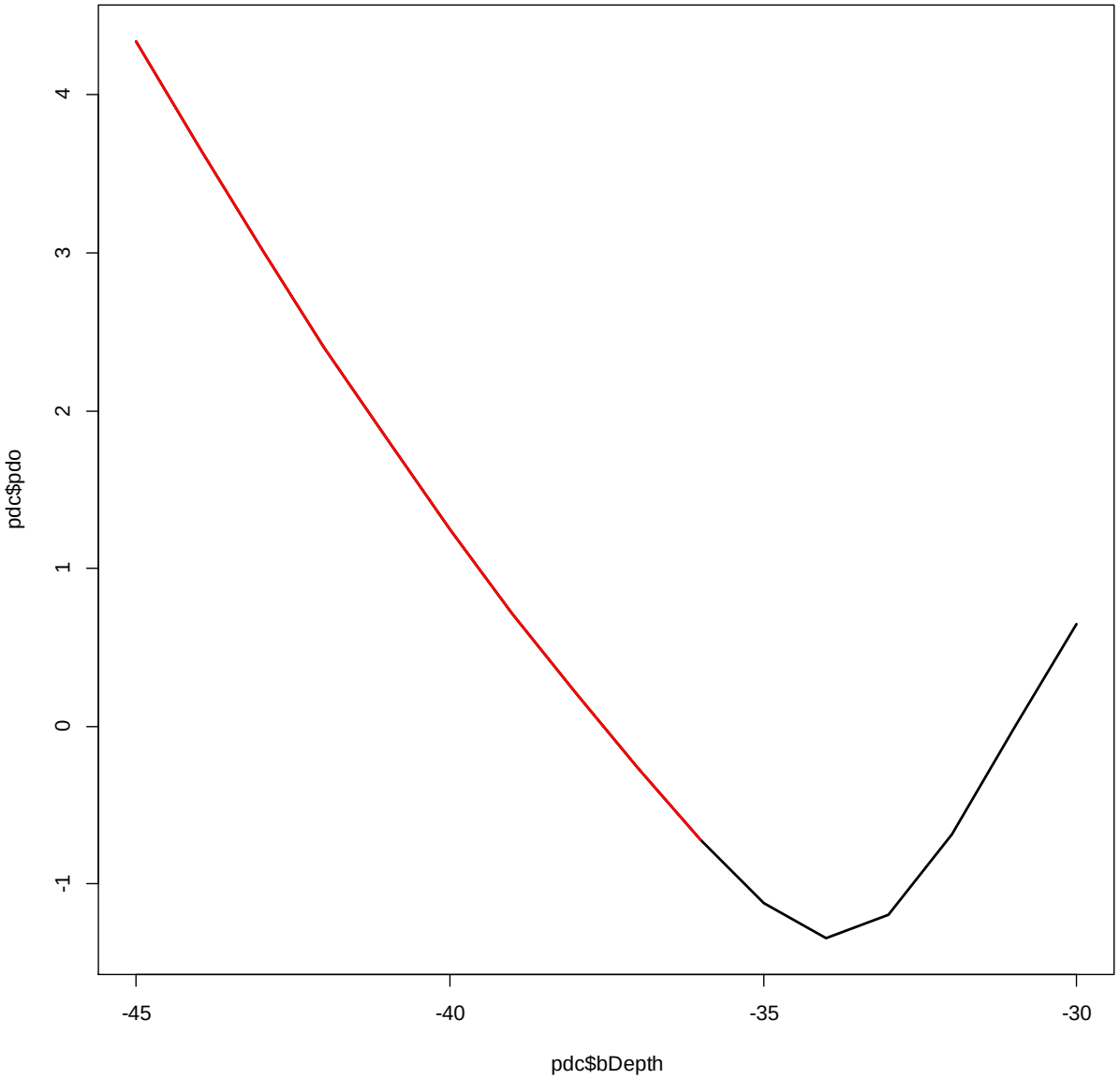
Here we look at extreme residuals in context.
Some look like bad data

	Station	date	Depth	TotalDepth	do	pdo	resdo
12164	CB4.1C	1990-07-17	0.5	31	7.8	8.182053678	-0.38205369
12165	CB4.1C	1990-07-17	0.5	31	7.8	8.182053678	-0.38205369
12166	CB4.1C	1990-07-17	1.0	31	4.8	7.886021384	-3.08602140
12167	CB4.1C	1990-07-17	2.0	31	4.8	7.285411499	-2.48541151
12168	CB4.1C	1990-07-17	3.0	31	4.7	6.661970034	-1.96197005
12169	CB4.1C	1990-07-17	5.0	31	4.4	5.305035118	-0.90503513
12170	CB4.1C	1990-07-17	7.0	31	3.9	3.790182615	0.10981737
12171	CB4.1C	1990-07-17	9.0	31	3.6	2.277665318	1.32233467
12172	CB4.1C	1990-07-17	11.0	31	2.7	1.057946164	1.64205382
12173	CB4.1C	1990-07-17	12.0	31	2.4	0.619820766	1.78017922
12174	CB4.1C	1990-07-17	13.0	31	2.3	0.300122710	1.99987728
12175	CB4.1C	1990-07-17	14.0	31	9.9	0.085534142	9.81446585
12176	CB4.1C	1990-07-17	15.0	31	0.5	-0.043401216	0.54340120
12177	CB4.1C	1990-07-17	16.0	31	0.3	-0.106899198	0.40689919
12178	CB4.1C	1990-07-17	17.0	31	0.0	-0.123412533	0.12341252
12179	CB4.1C	1990-07-17	18.0	31	0.0	-0.108698250	0.10869824
12180	CB4.1C	1990-07-17	19.0	31	0.0	-0.075225377	0.07522537
12181	CB4.1C	1990-07-17	21.0	31	0.0	0.014471288	-0.01447130
12182	CB4.1C	1990-07-17	23.0	31	0.0	0.100936030	-0.10093604
12183	CB4.1C	1990-07-17	25.0	31	0.0	0.155756484	-0.15575650
12184	CB4.1C	1990-07-17	27.0	31	0.0	0.135927286	-0.13592730
12185	CB4.1C	1990-07-17	29.0	31	0.1	-0.006995027	0.10699502
12186	CB4.1C	1990-07-17	30.0	31	0.1	-0.122497509	0.22249750

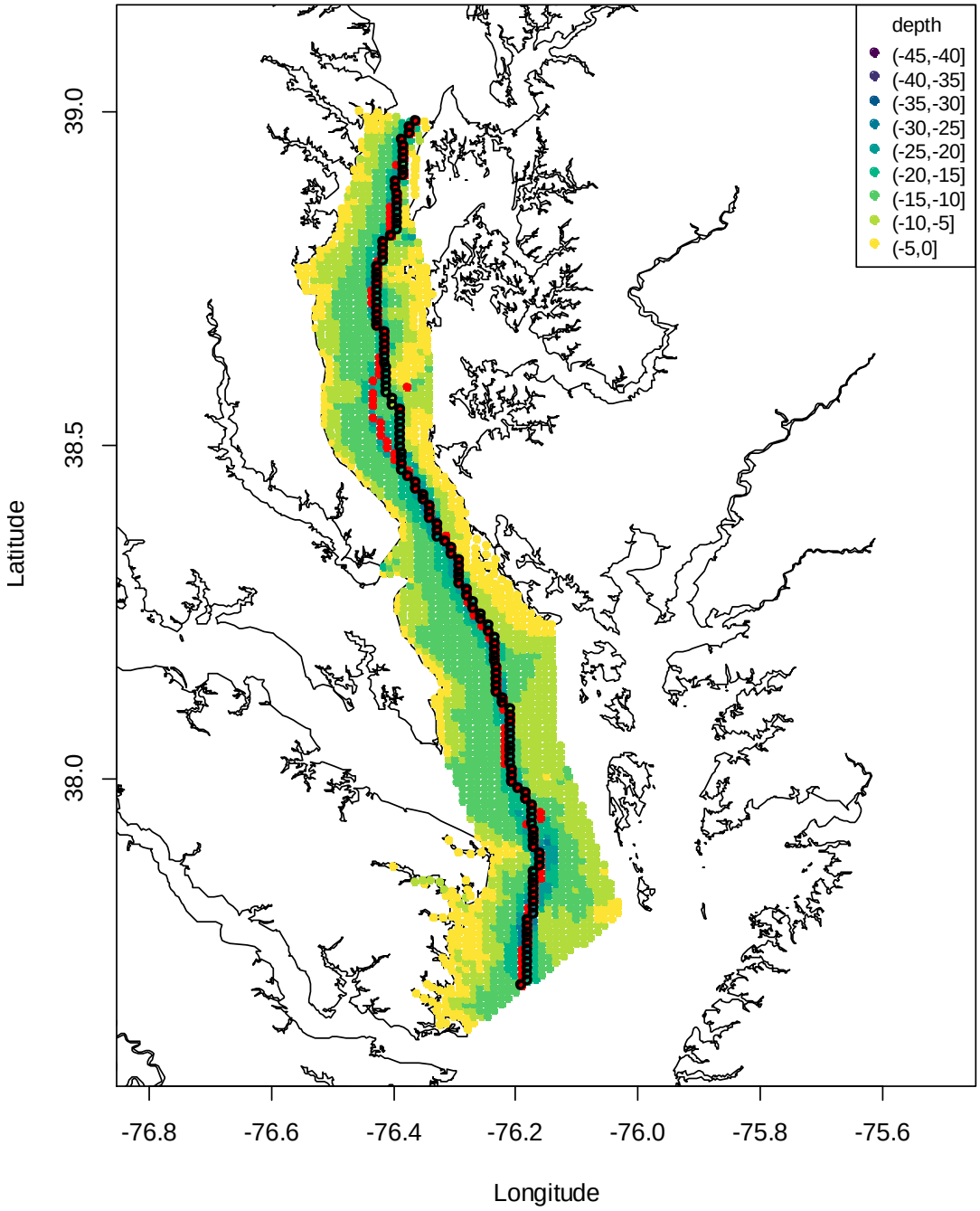
Some are clearly Low DO events that the model misses

	Station	date	Depth	TotalDepth	do	pdo	resdo
99102	EE2.2	1997-03-05	0.5	15	10.1	12.07835	-1.978347
99103	EE2.2	1997-03-05	1.0	15	10.1	12.15059	-2.050590
99104	EE2.2	1997-03-05	2.0	15	10.1	12.28256	-2.182563
99105	EE2.2	1997-03-05	3.0	15	10.0	12.37938	-2.379385
99106	EE2.2	1997-03-05	4.0	15	9.8	12.41955	-2.619548
99107	EE2.2	1997-03-05	5.0	15	9.7	12.38526	-2.685262
99108	EE2.2	1997-03-05	6.0	15	9.6	12.26469	-2.664692
99109	EE2.2	1997-03-05	7.0	15	9.3	12.05986	-2.759856
99110	EE2.2	1997-03-05	8.0	15	8.8	11.79461	-2.994606
99111	EE2.2	1997-03-05	9.0	15	8.1	11.50791	-3.407911
99112	EE2.2	1997-03-05	10.0	15	6.7	11.24075	-4.540752
99113	EE2.2	1997-03-05	11.0	15	1.5	11.02867	-9.528674
99114	EE2.2	1997-03-05	12.0	15	1.2	10.89722	-9.697217
99115	EE2.2	1997-03-05	13.0	15	1.1	10.85913	-9.759133
99116	EE2.2	1997-03-05	14.0	15	1.1	10.91685	-9.816854

glossing over the problem of depths that exceed depths with observed DO.



Glossing over thalweg location issue



By depthCat model.

Approximate significance of smooth terms:

	edf	Ref.df	F	p-value	
s(doy)	7.481	8.000	391.244	< 2e-16	***
s(wDepth):depthCat_surf	7.917	7.993	5181.324	< 2e-16	***
s(wDepth):depthCat_deep	6.094	6.688	1057.981	< 2e-16	***
s(LonKm)	9.000	9.000	49.173	< 2e-16	***
s(LatKm)	8.774	8.955	146.848	< 2e-16	***
ti(wDepth,doy):depthCat_surf	11.866	12.000	873.368	< 2e-16	***
ti(wDepth,doy):depthCat_deep	10.835	12.000	40.340	< 2e-16	***
ti(LonKm,wDepth):depthCat_surf	10.360	10.806	495.784	< 2e-16	***
ti(LonKm,wDepth):depthCat_deep	14.152	14.872	23.166	< 2e-16	***
ti(LonKm,doy)	11.639	12.000	220.795	< 2e-16	***
s(saf.1_90):depthCat_surf	8.879	8.994	61.413	< 2e-16	***
s(saf.1_90):depthCat_deep	8.965	8.999	92.630	< 2e-16	***
ti(doy,saf.1_90):depthCat_surf	11.910	12.000	46.324	< 2e-16	***
ti(doy,saf.1_90):depthCat_deep	11.923	12.000	116.957	< 2e-16	***
ti(wDepth,saf.1_90):depthCat_surf	3.844	4.082	117.884	< 2e-16	***
ti(wDepth,saf.1_90):depthCat_deep	3.615	4.072	8.139	1.61e-06	***
ti(LonKm,saf.1_90):depthCat_surf	15.145	15.866	7.832	< 2e-16	***
ti(LonKm,saf.1_90):depthCat_deep	12.554	14.056	17.469	< 2e-16	***
ti(LatKm,saf.1_90):depthCat_surf	12.197	13.615	9.619	< 2e-16	***
ti(LatKm,saf.1_90):depthCat_deep	10.885	12.294	6.011	< 2e-16	***
ti(wDepth,doy,LonKm)	42.739	48.000	24.163	< 2e-16	***

AIC of model with by depthCat 416821.98

Trends of DO with lags of flow and depth