Overview of BayTrends, a CBP software package for Water Quality Trend Detection.

Jeni Keisman and Rebecca Murphy, Bay Program
Elgin Perry, statistics consultant
Jon Harcum, Tetra Tech

6/21/2018

Talk outline

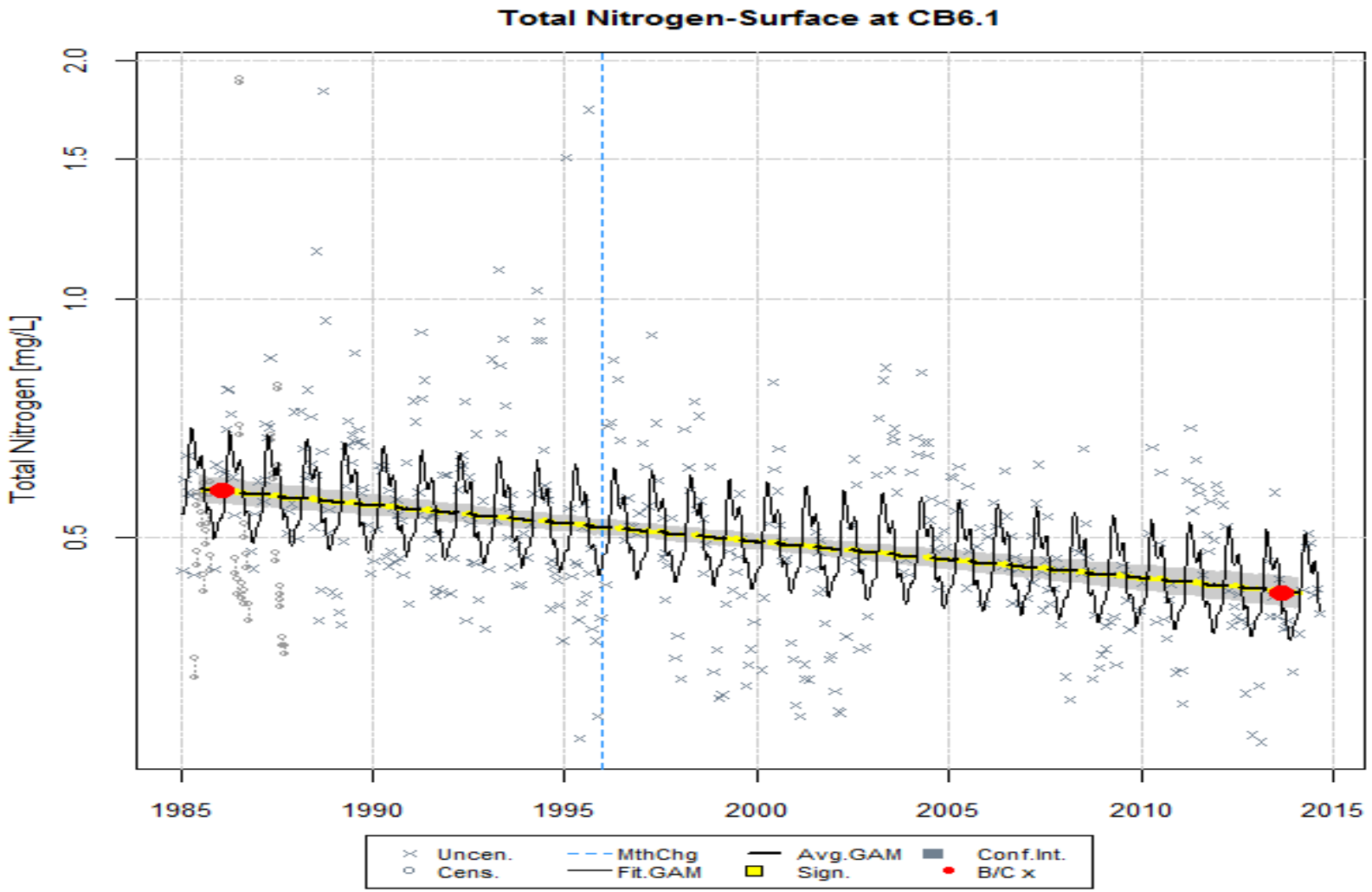
1. Overview of BayTrends
2. Handling of censored data
3. Modeling of Laboratory methods changes

Key Features of BayTrends:

- Written in R, an open source statistical programming environment
- Makes extensive use of Generalized Additive Models (gams)
- Estimates non-monotonic trends
- Does Flow adjustment

Total Nitrogen -- Surface

CB6.1 - S - tn



gam number = 0

title = Linear Trend with Seasonality

model = ~ cyear + s(doy,bs='cc')

Table: GAM Analysis of Variance - CB6.1 - S - tn

type	source	df	F	p.value
parametric terms	cyear	1	64.944	<0.0001
smoothed terms	s(doy)	7.07	9.1123	<0.0001

Table: GAM Parameter Coefficients. - CB6.1 - S - tn

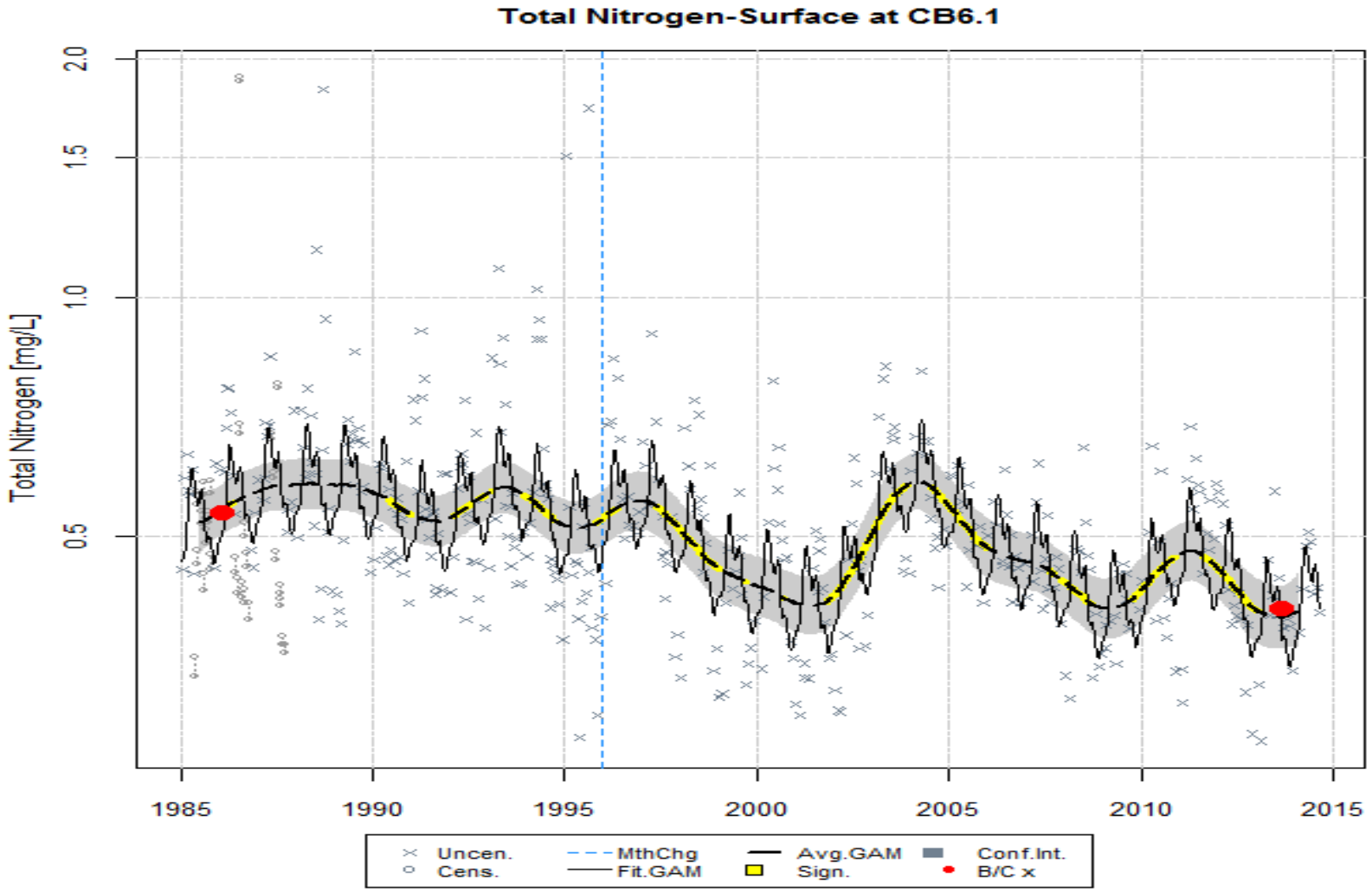
source	estimate	std.error	t.value	p.value
(Intercept)	-0.696489	0.011518	-60.4708	<0.0001
cyear	-0.010527	0.001306	-8.0588	<0.0001

Table: GAM Diagnostics. - CB6.1 - S - tn

AIC	RMSE	AdjRsquare
-7.53	0.237	0.235

Table: Estimates of Change from 1986-2016. - CB6.1 - S - tn

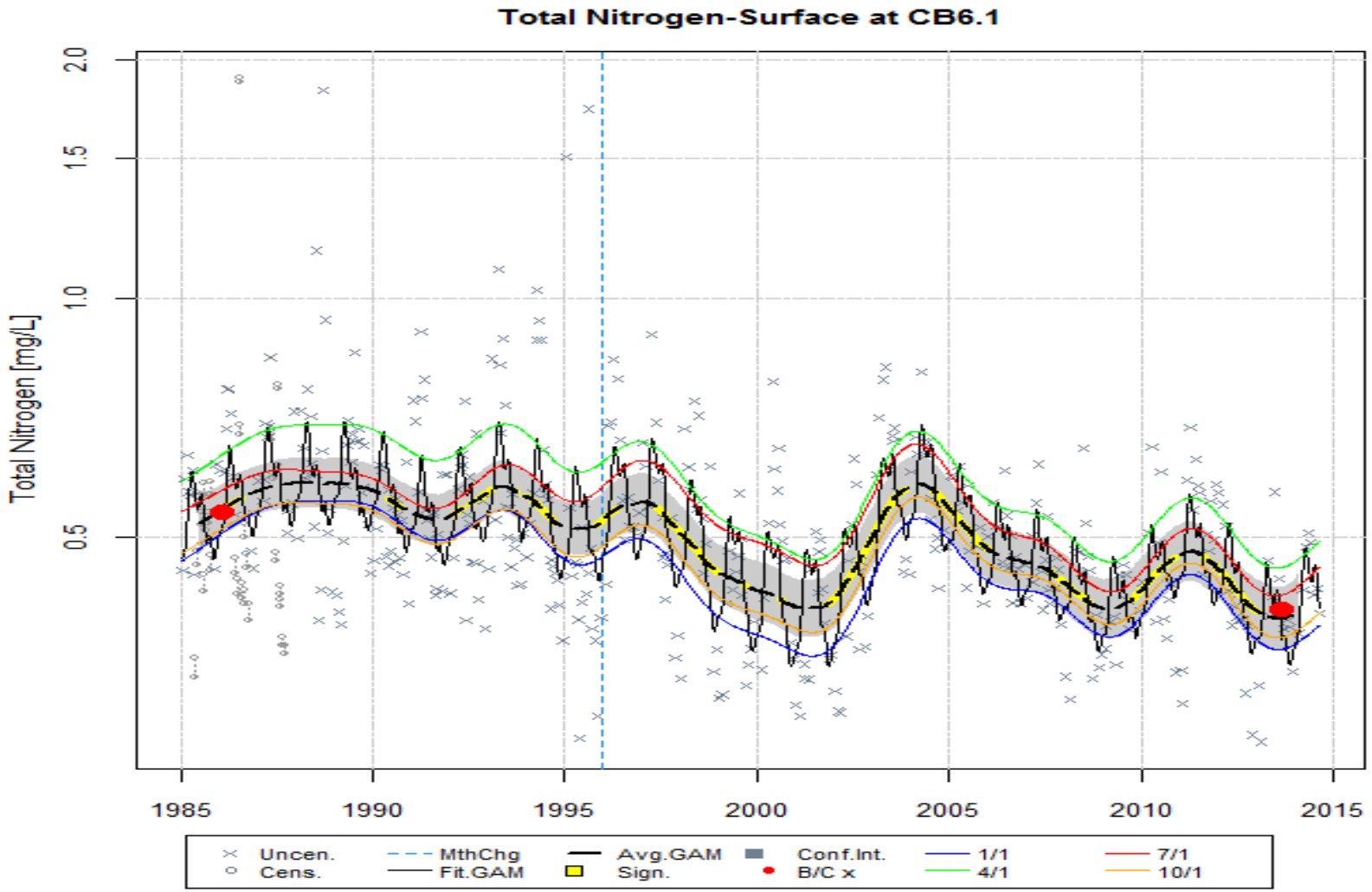
Calculation	Estimate
Baseline log mean (geometric mean)	-0.5571 (0.5729)
Current log mean (geometric mean)	-0.8518 (0.4266)
Estimated log difference	-0.2948
Std. Err. log difference	0.0366
95% Confidence interval for log difference	(-0.3665 , -0.2231)
Difference p-value	<0.0001
Period of Record Percent Change Estimate (%)	-25.53%
Period of Record	1985 - 2014



gam number = 1
title = Non-linear Trend with Seasonality
model = ~ cyear + s(cyear, k=gamK1) + s(doy,bs='cc')

Table: Estimates of Change from 1986-2016. - CB6.1 - S - tn

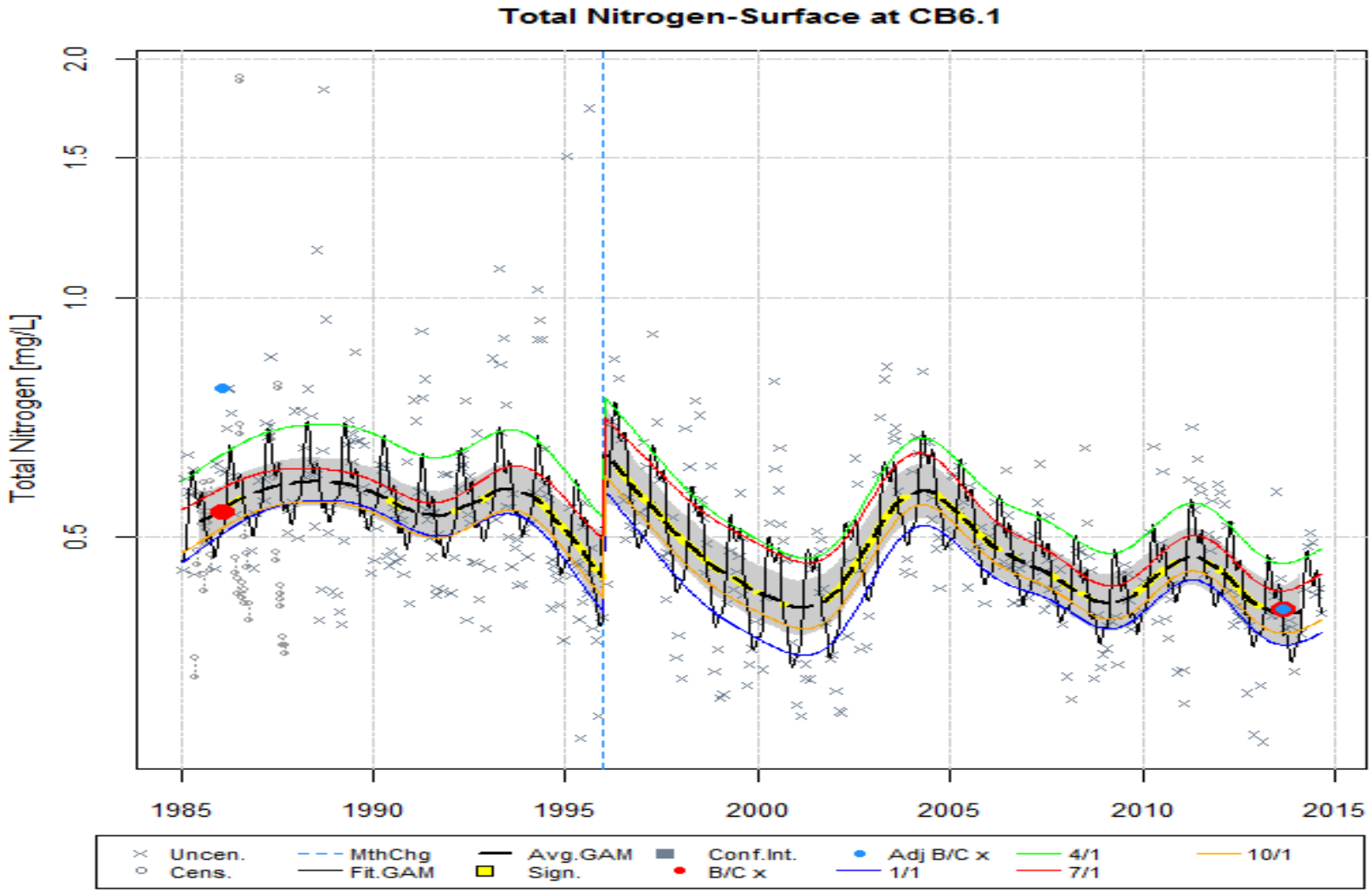
Calculation	Estimate
Baseline log mean (geometric mean)	-0.623 (0.5363)
Current log mean (geometric mean)	-0.9008 (0.4062)
Estimated log difference	-0.2778
Std. Err. log difference	0.0611
95% Confidence interval for log difference	(-0.3976 , -0.1581)
Difference p-value	<0.0001
Period of Record Percent Change Estimate (%)	-24.26%
Period of Record	1985 - 2014



gam number = 2
title = Non-linear trend with Seas+Int
model = ~ cyear + s(cyear, k=gamK1) + s(doy,bs='cc')+ ti(cyear,doy,bs=c('tp','cc'))

Table: Estimates of Change from 1986-2016. - CB6.1 - S - tn

Calculation	Estimate
Baseline log mean (geometric mean)	-0.6205 (0.5377)
Current log mean (geometric mean)	-0.9017 (0.4059)
Estimated log difference	-0.2811
Std. Err. log difference	0.0609
95% Confidence interval for log difference	(-0.4004 , -0.1619)
Difference p-value	<0.0001
Period of Record Percent Change Estimate (%)	-24.51%
Period of Record	1985 - 2014



gam number = 3
title = Non-linear trend with Seas+Int. & Intervention
model = ~ intervention + cyear + s(cyear, k=gamK1) + s(doy,bs='cc') + ti(cyear,doy,bs=c('tp','cc'))

Table: GAM Analysis of Variance - CB6.1 - S - tn

type	source	df	F	p.value
parametric terms	intervention	1	8.3066	0.0042
" "	cyear	1	2.2368	0.1355
smoothed terms	s(cyear)	14.05	4.2546	<0.0001
" "	s(doy)	7.02	10.1609	<0.0001
" "	ti(cyear,doy)	2.55	0.5738	0.0180

Table: GAM Parameter Coefficients. - CB6.1 - S - tn

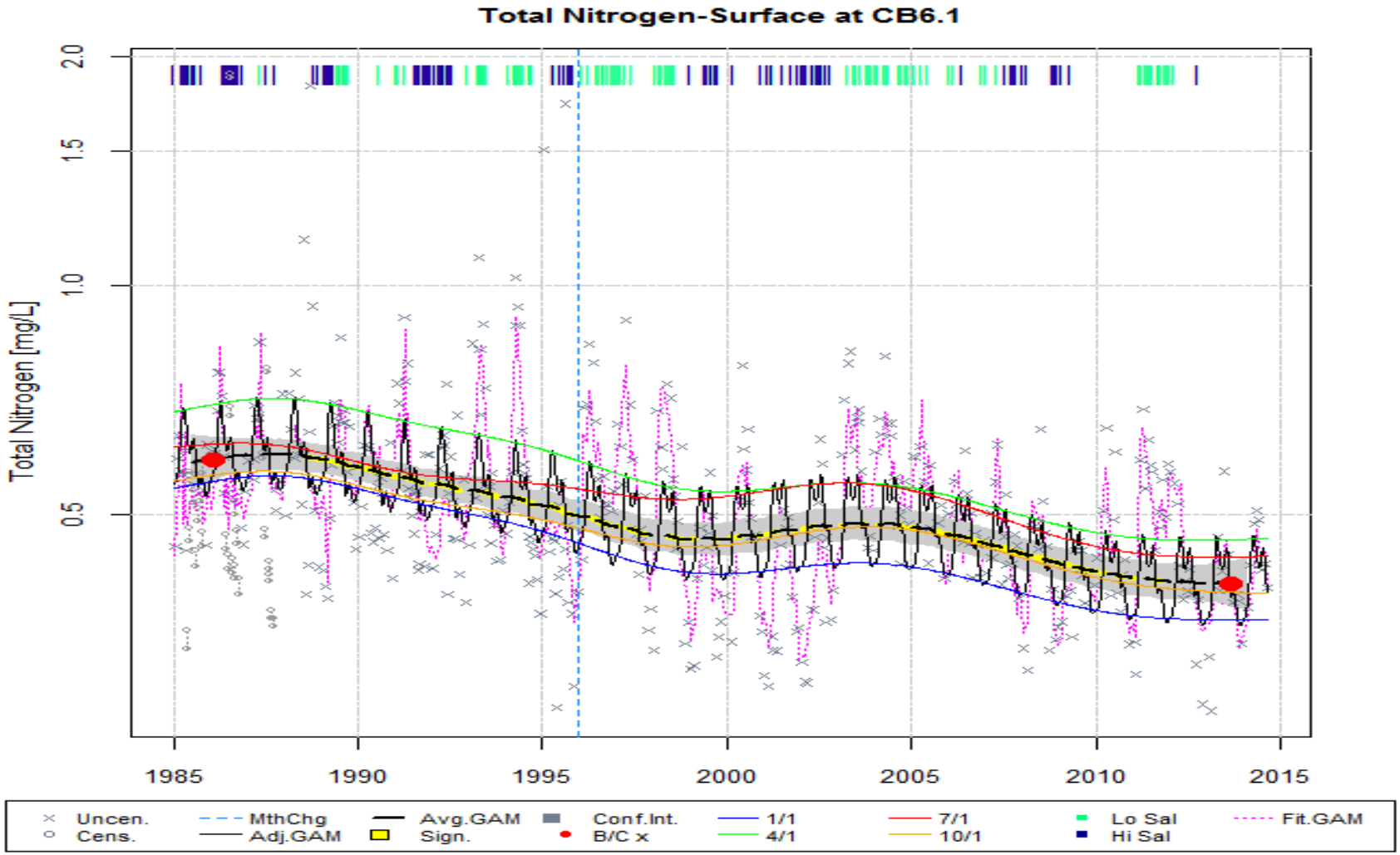
source	estimate	std.error	t.value	p.value
(Intercept)	-0.70092	0.128784	-5.4426	<0.0001
cyear	0.083596	0.055895	1.4956	0.1355
interventionB	0.35917	0.12462	2.8821	0.0042

Table: GAM Diagnostics. - CB6.1 - S - tn

AIC	RMSE	AdjRsquare
-60.22	0.2193	0.3451

Table: Estimates of Change from 1986-2016. - CB6.1 - S - tn

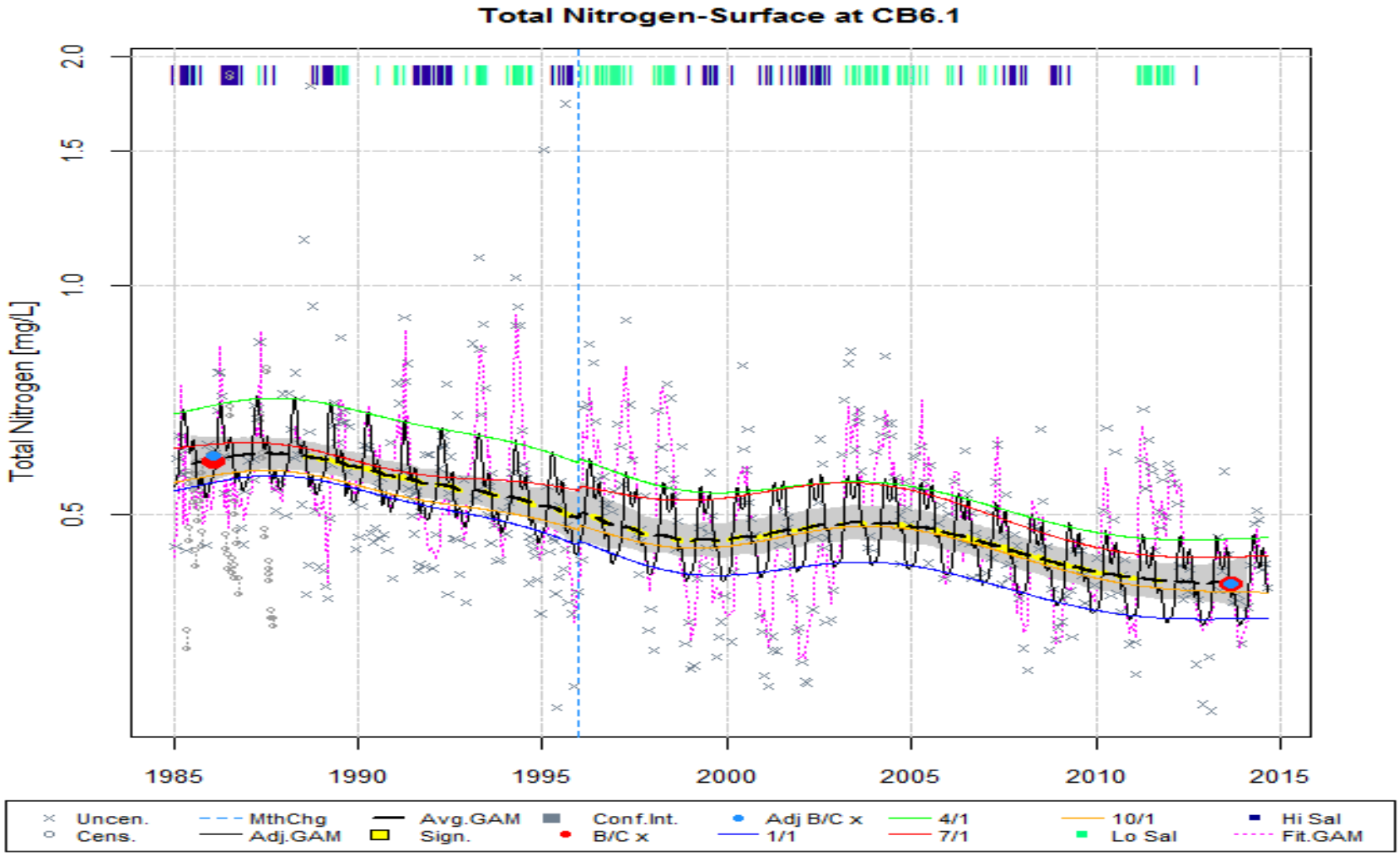
Calculation	Estimate	Estimate.Adj
Baseline log mean (geometric mean)	-0.6202 (0.5378)	-0.2611 (0.7702)
Current log mean (geometric mean)	-0.9026 (0.4055)	-0.9026 (0.4055)
Estimated log difference	-0.2823	-0.6415
Std. Err. log difference	0.0602	0.1386
95% Confidence interval for log difference	(-0.4003 , -0.1644)	(-0.9131 , -0.3699)
Difference p-value	<0.0001	<0.0001
Period of Record Percent Change Estimate (%)	-24.6%	-47.35%
Period of Record	1985 - 2014	1985 - 2014



gam number = 4
title = Non-linear trend with Seas+Int. & Hydro Adj
model = ~ cyear + s(cyear, k=gamK1) + s(doy,bs='cc') + ti(cyear,doy,bs=c('tp','cc')) + s(flw_sal,k=gamK2) + ti(flw_sal,doy,bs=c('tp','cc')) + ti(flw_sal, cyear,bs=c('tp' ,'tp')) + ti(flw_sal,doy,cyear, bs=c('tp','cc','tp'))

Table: Estimates of Change from 1986-2016. - CB6.1 - S - tn

Calculation	Estimate	Estimate.Adj
Baseline log mean (geometric mean)	-0.5277 (0.5899)	-0.5277 (0.5899)
Current log mean (geometric mean)	-0.9021 (0.4057)	-0.9021 (0.4057)
Estimated log difference	-0.3744	-0.3744
Std. Err. log difference	0.0529	0.0529
95% Confidence interval for log difference	(-0.4782 , -0.2706)	(-0.4782 , -0.2706)
Difference p-value	<0.0001	<0.0001
Period of Record Percent Change Estimate (%)	-31.23%	-31.23%
Period of Record	1985 - 2014	1985 - 2014



gam number = 5
title = Non-linear trend with Seas+Int., Hydro Adj, & intervention
model = ~ intervention + cyear + s(cyear, k=gamK1) + s(doy,bs='cc') + ti(cyear,doy,bs=c('tp','cc')) + s(flw_sal,k=gamK2) + ti(flw_sal,doy,bs=c('tp','cc')) + ti(flw_sal, cyear,bs=c('tp' ,'tp')) + ti(flw_sal,doy,cyear, bs=c('tp','cc','tp'))

Table: GAM Analysis of Variance - CB6.1 - S - tn

type	source	df	F	p.value	Note
parametric terms	intervention	1	0.0358	0.8500	-
" "	cyear	1	2.8711	0.0909	-
smoothed terms	s(cyear)	5.8	2.6853	0.0109	-
" "	s(doy)	7.21	12.5699	<0.0001	-
" "	ti(cyear,doy)	3.64	0.9638	0.0017	-
" "	s(flw_sal)	1	134.027	<0.0001	-
" "	ti(flw_sal,doy)	4.71	3.063	<0.0001	-
" "	ti(flw_sal,cyear)	4.09	1.0167	0.4201	-
" "	ti(flw_sal,doy,cyear)	0	0.0001	0.2340	-

Table: GAM Parameter Coefficients. - CB6.1 - S - tn

source	estimate	std.error	t.value	p.value
(Intercept)	-0.629795	0.050984	-12.3527	<0.0001
cyear	0.02825	0.016672	1.6944	0.0909
interventionB	0.014572	0.077027	0.1892	0.8500

Table: GAM Diagnostics. - CB6.1 - S - tn

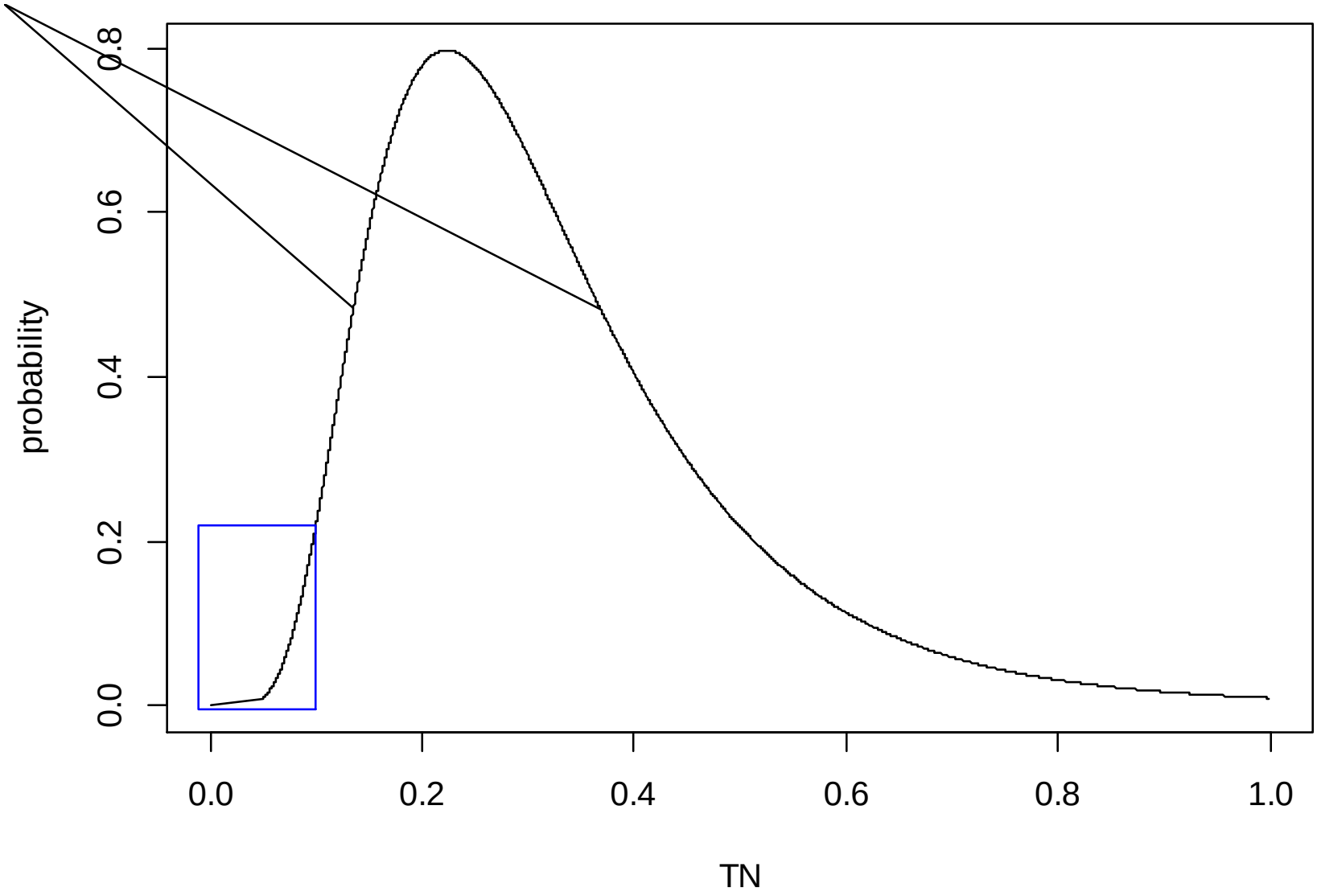
AIC	RMSE	AdjRsquare
-202.2	0.1864	0.5275

Table: Estimates of Change from 1986-2016. - CB6.1 - S - tn

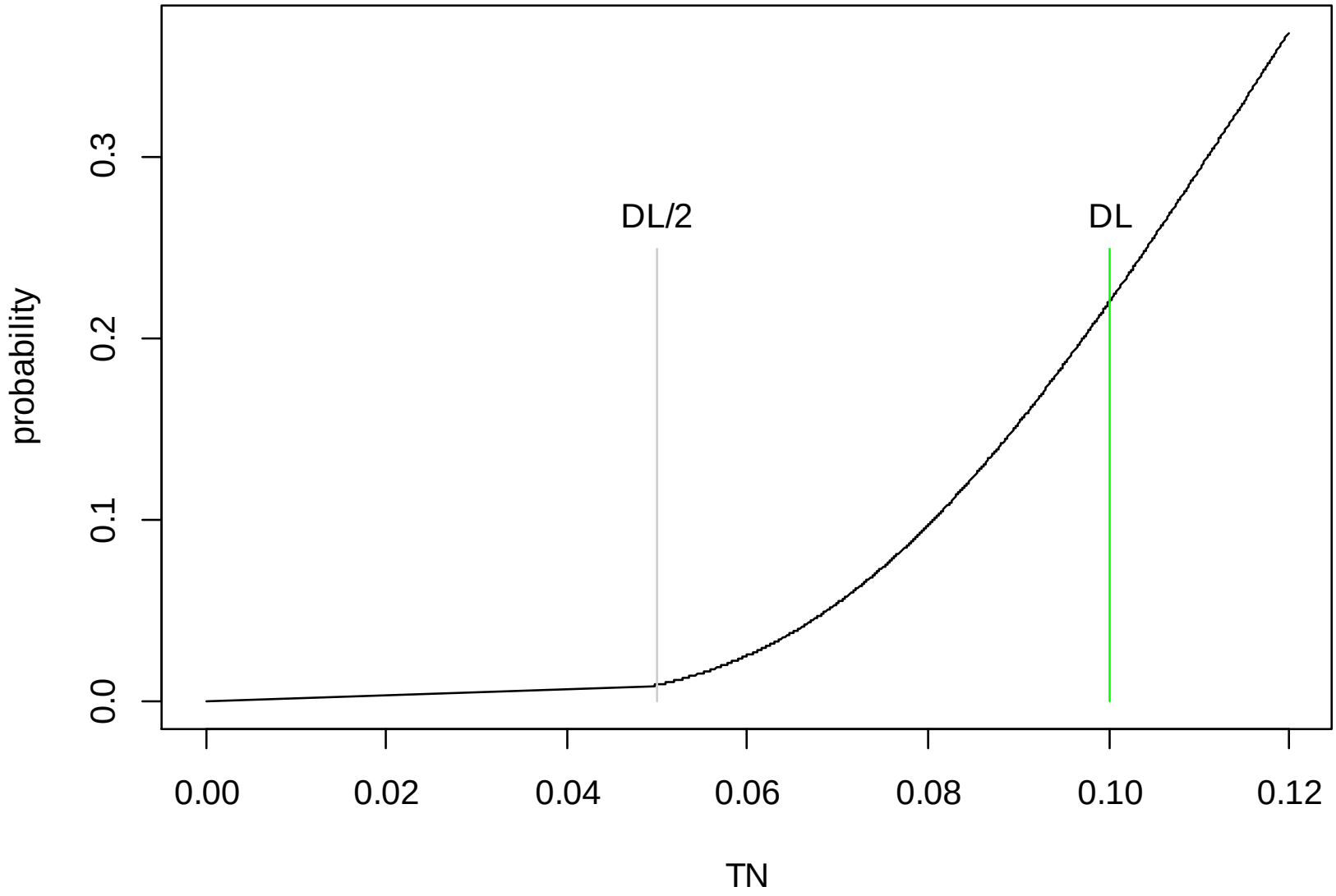
Calculation	Estimate	Estimate.Adj
Baseline log mean (geometric mean)	-0.5299 (0.5887)	-0.5154 (0.5973)
Current log mean (geometric mean)	-0.9004 (0.4064)	-0.9004 (0.4064)
Estimated log difference	-0.3705	-0.385
Std. Err. log difference	0.0534	0.0942
95% Confidence interval for log difference	(-0.4751 , -0.2658)	(-0.5696 , -0.2005)
Difference p-value	<0.0001	<0.0001
Period of Record Percent Change Estimate (%)	-30.96%	-31.96%
Period of Record	1985 - 2014	1985 - 2014

end of file

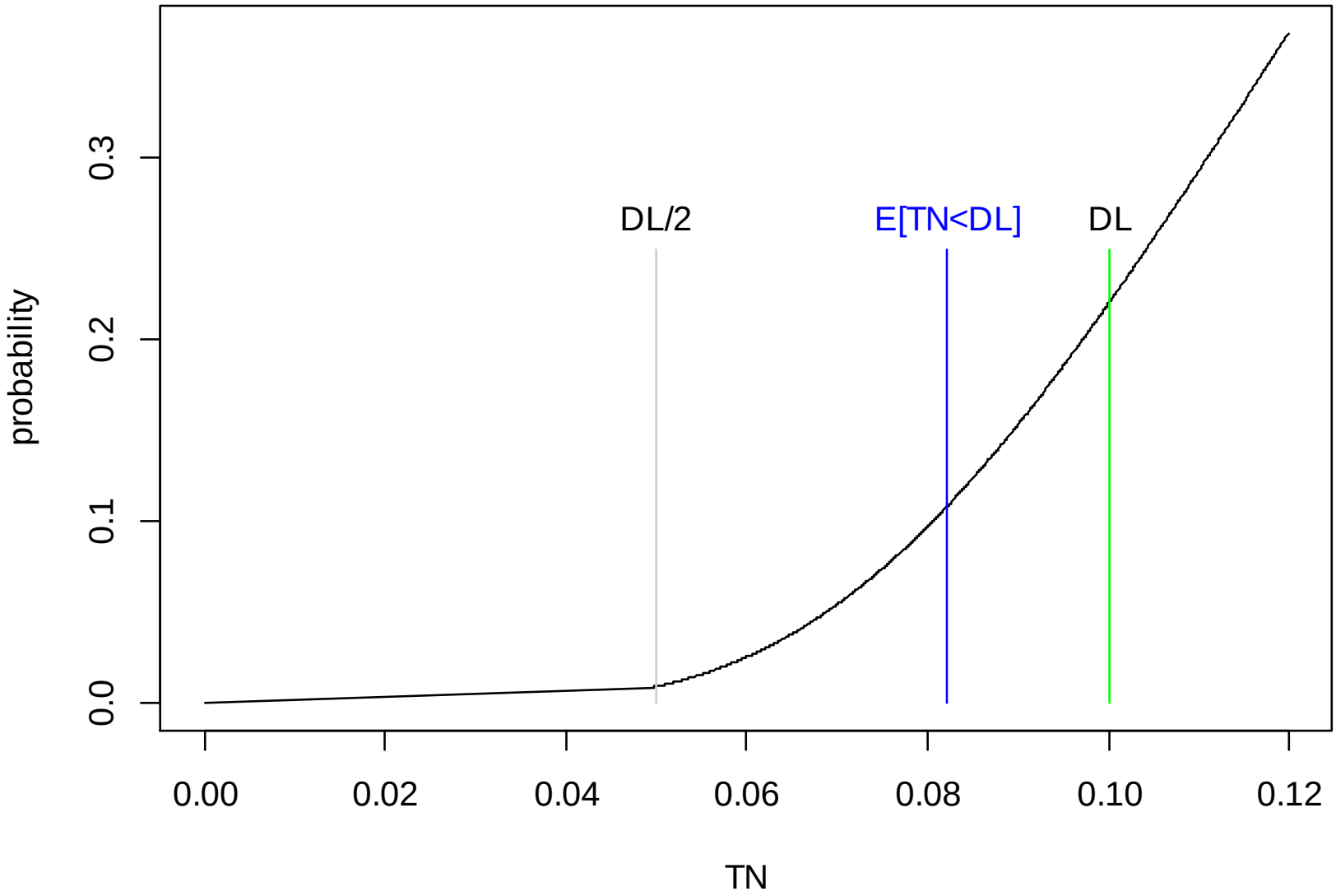
Censored Data Handling
Expectation Maximization Algorithm (EM algorithm)



Assume a log-normal Distribution.

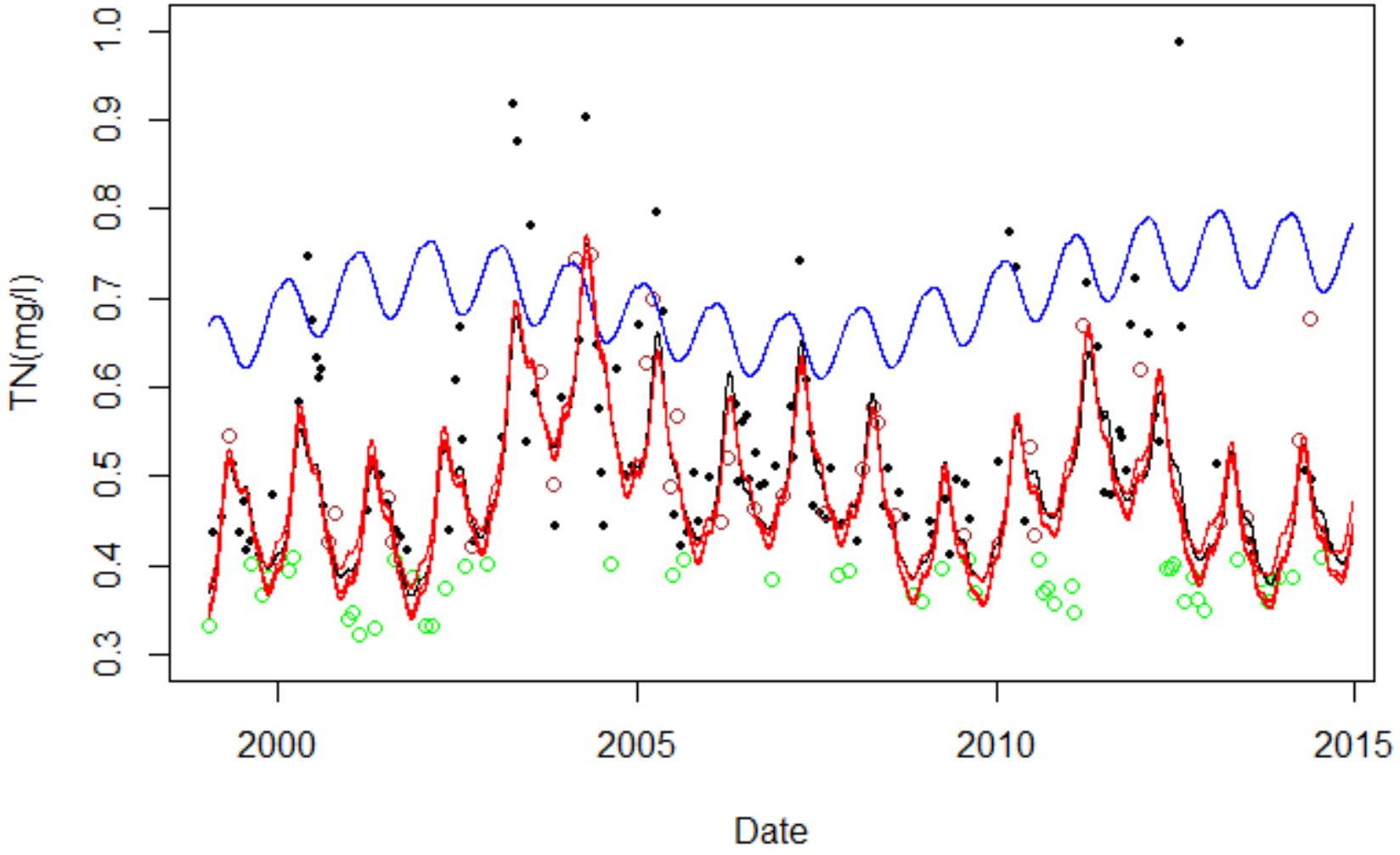

probability
0.0
0.1
0.2
0.3
DL/2
DL
0.00
0.02
0.04
0.06
0.08
0.10
0.12
TN

**

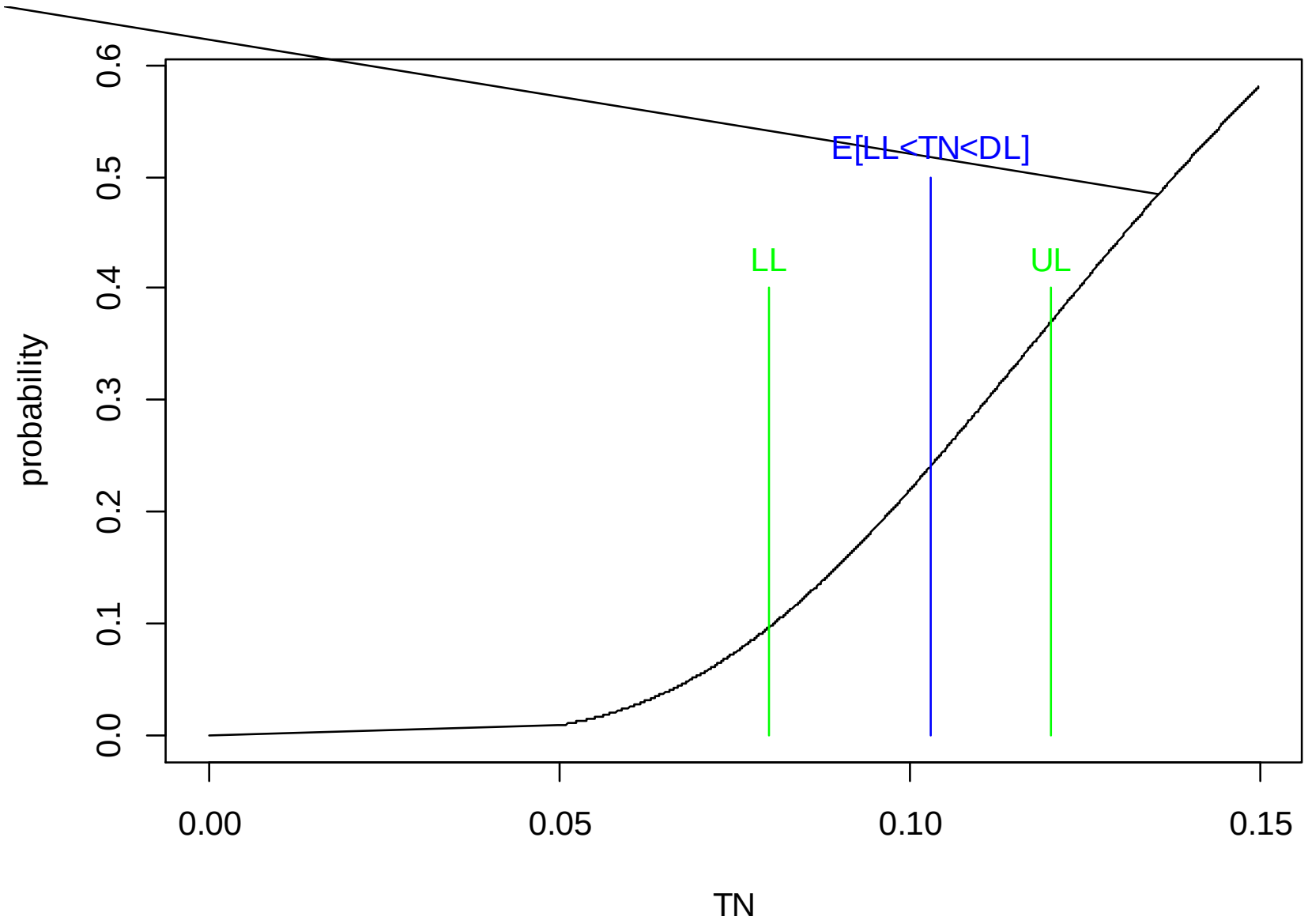


Substitute the expected value of the variable given that it is less than the

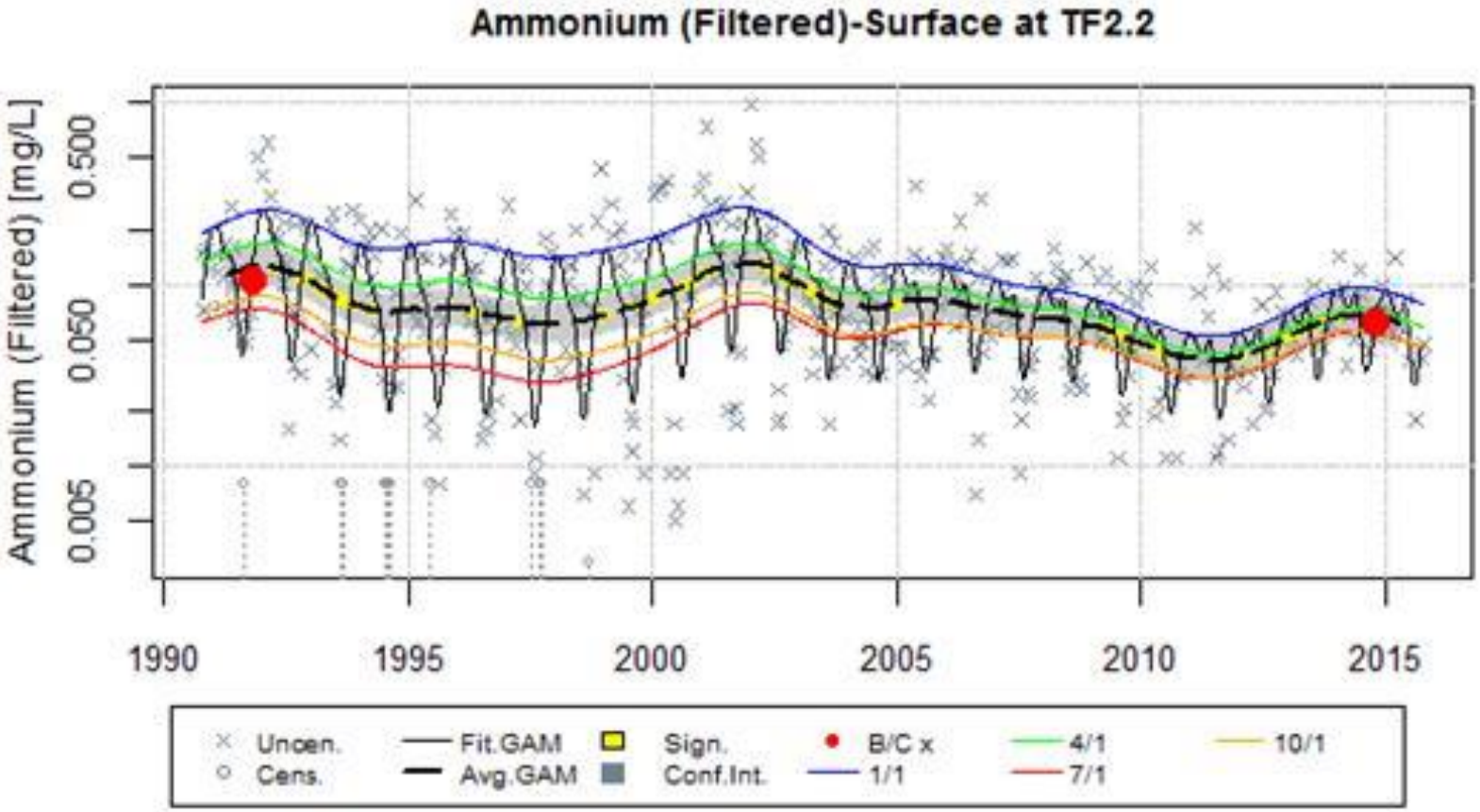
detection limit.



Convergence of EM algorithm.



Interval Censored Data.


Ammonium (Filtered)-Surface at TF2.2
Ammonium (Filtered) [mg/L]
0.005
0.050
0.500
1990
1995
2000
2005
2010
2015
Uncen.
Cens.
Fit.GAM
Avg.GAM
Sign.
Conf.Int.
B/C x
1/1
4/1
7/1
10/1

Step Trend Models
(Intervention Models)
and
Methods Changes

Review old approach to methods changes
Review Maryland TSS case
Shows that Method Change effect can be unique to a station
Review issues with step trend model
Get your thoughts on this approach

Old Method:

Do a split sample study with multiple stations and dates.
Assess data with paired comparison test
Wilcoxon Signed Rank Test
Paired t-test
In cases of significant difference, estimate adjustment factor
Apply AF to old data to make is comparable to new.

(note: inherent assumption that the methods change effect is uniform
Over stations and dates)

Down the Road we found this

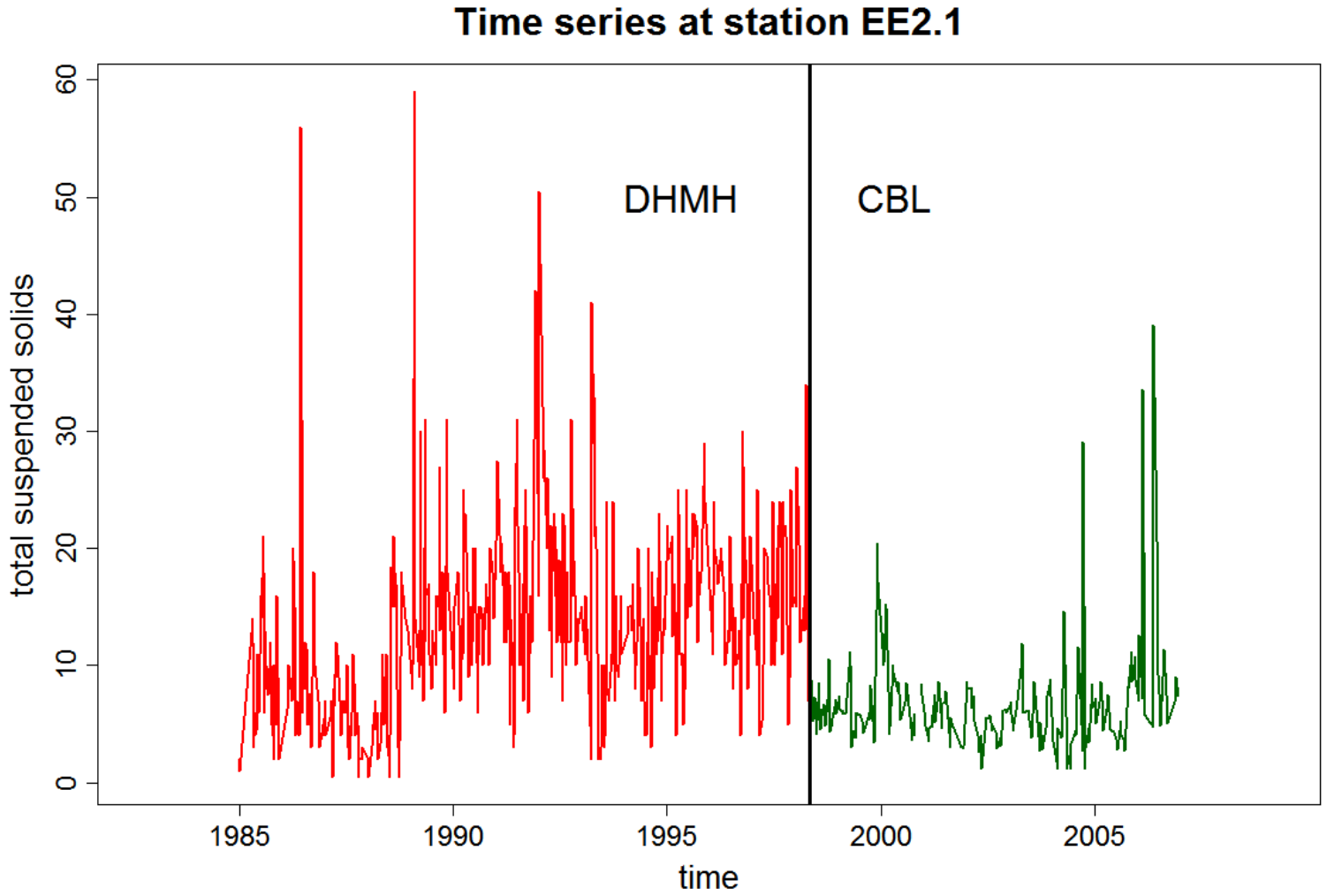


Figure 1. TSS at EE2.1(Choptank Embayment). The red curve shows data from DHMH. The green curve shows data from CBL. The black vertical line shows the point of laboratory change.

This effect was not consistent

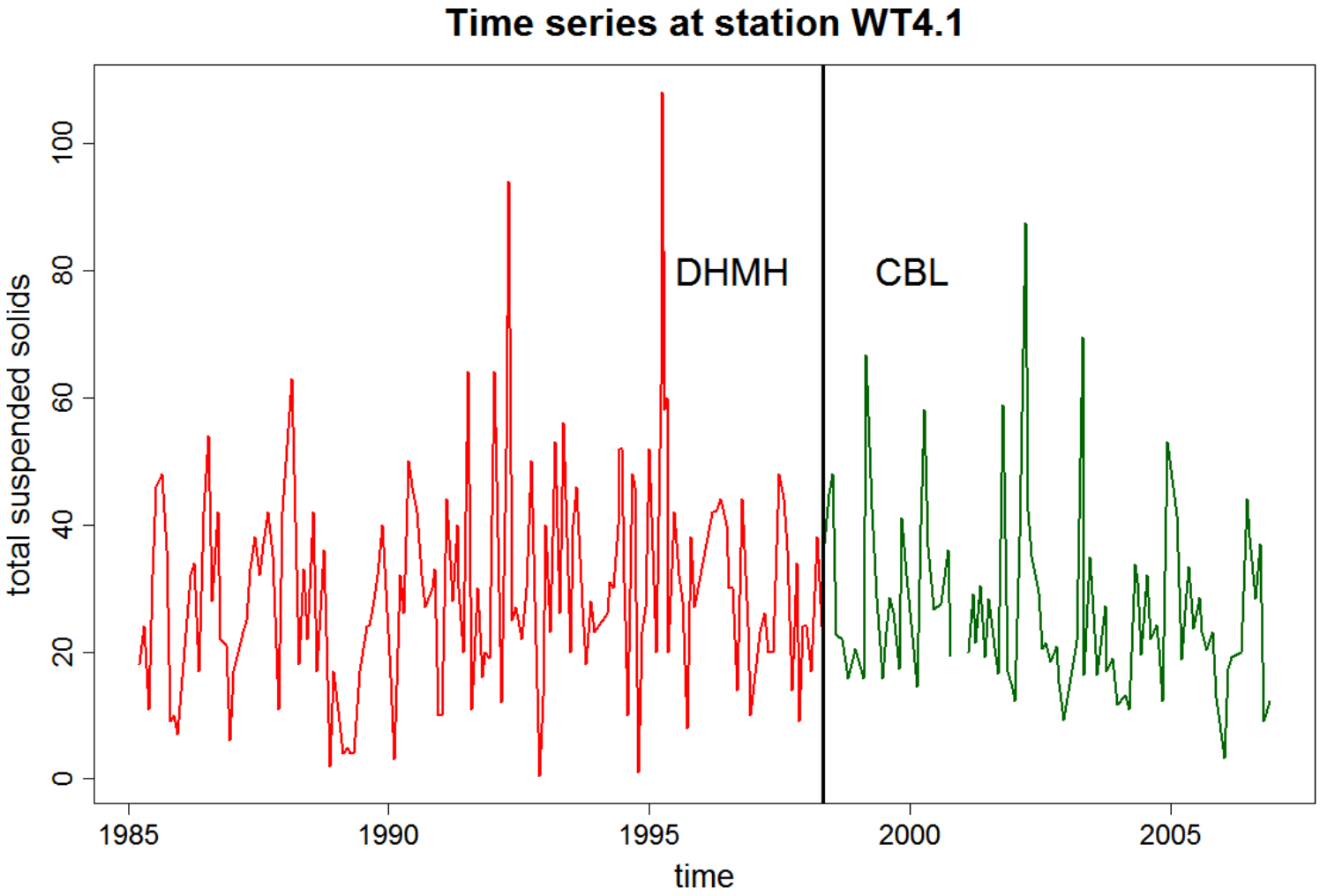
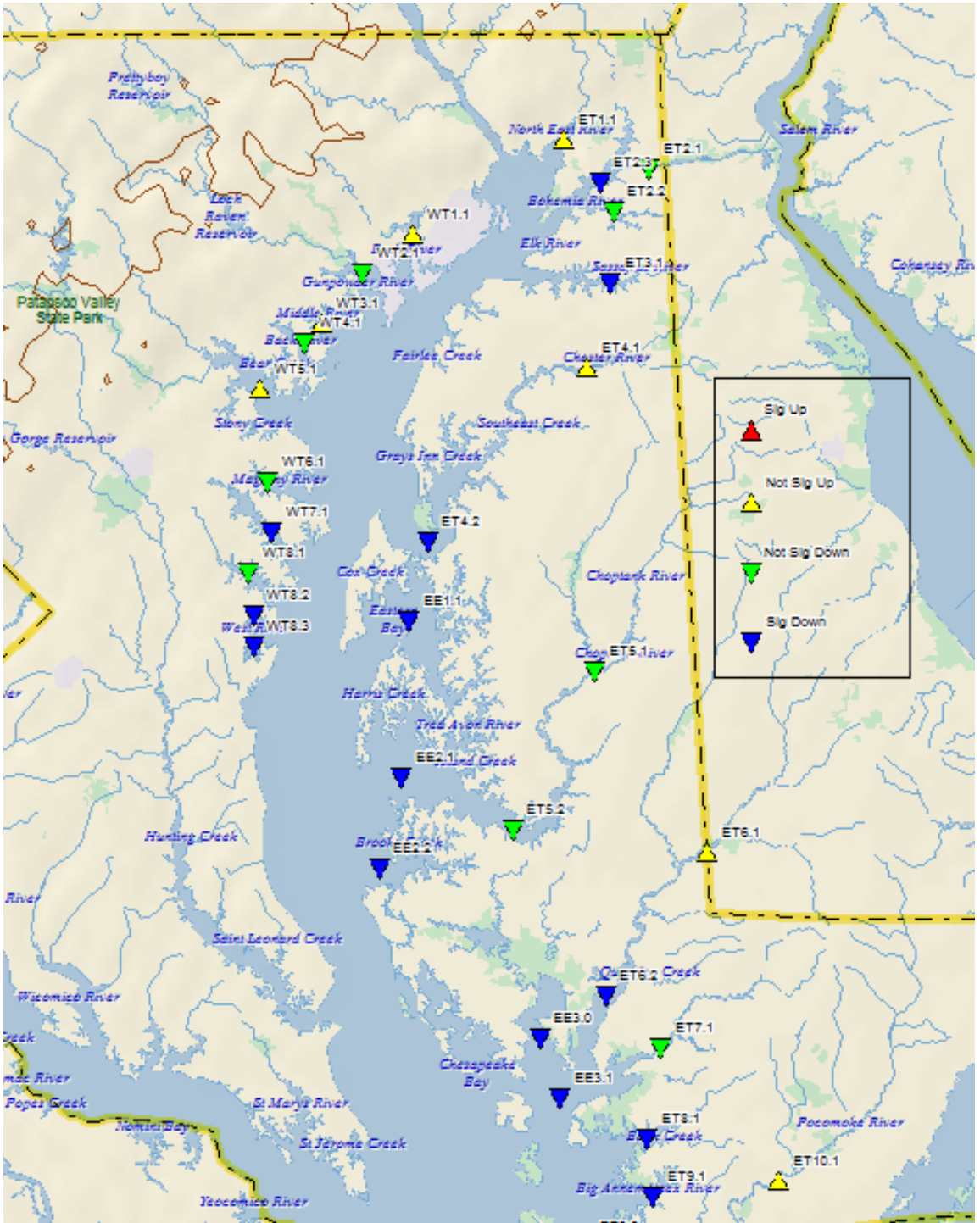


Figure 2. TSS at WT4.1 (Back River). The red curve shows data from DHMH. The green curve shows data from CBL. The black vertical line shows the point of laboratory change.


Prettyboy Reservoir
Liberty Reservoir
Patapsco Valley State Park
Gunpowder River
Middle River
Back River
Bear Creek
Stony Creek
Gorge Reservoir
Magothy River
Meade Creek
Hunting Creek
Saint Leonard Creek
Wicomico River
Popes Creek
Nomini Bay
St Mary's River
St Jerome Creek
Yeocomico River
Chesapeake Bay
Big Annemessex River
Pocomoke River
North East River
Bohemia River
Elk River
Sassafras River
Fairlee Creek
Chester River
Southeast Creek
Grays Inn Creek
Cox Creek
Eastern Bay
Harris Creek
Tred Avon River
Broad Creek
Choptank River
Salem River
Sig Up
Not Sig Up
Not Sig Down
Sig Down
WT1.1
WT2.1
WT3.1
WT4.1
WT5.1
WT6.1
WT7.1
WT8.1
WT8.2
WT8.3
ET1.1
ET2.1
ET2.2
ET3.1
ET4.1
ET4.2
ET5.1
ET5.2
ET6.1
ET6.2
ET7.1
ET8.1
ET9.1
ET10.1
EE1.1
EE2.1
EE2.2
EE3.0
EE3.1

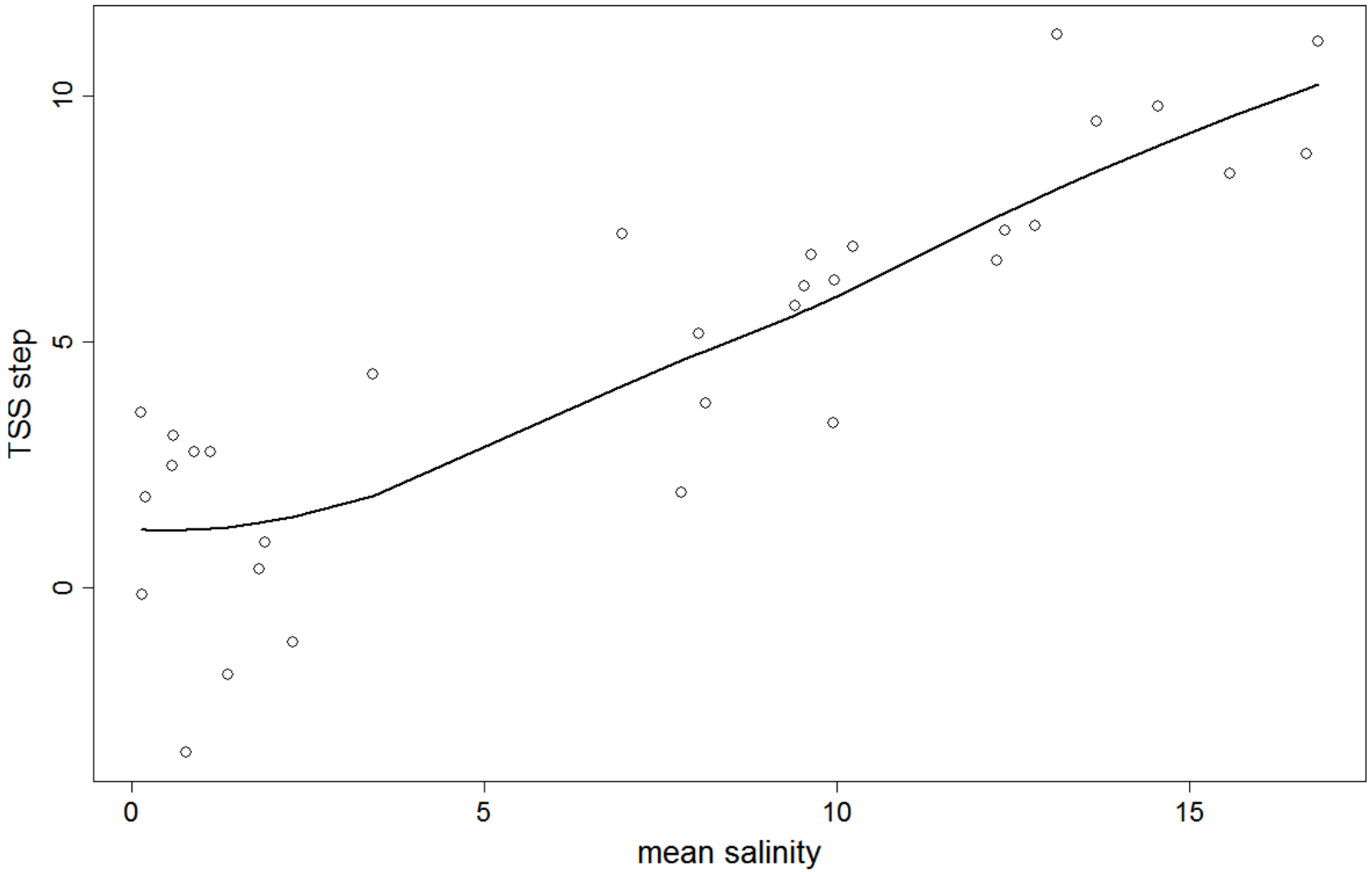
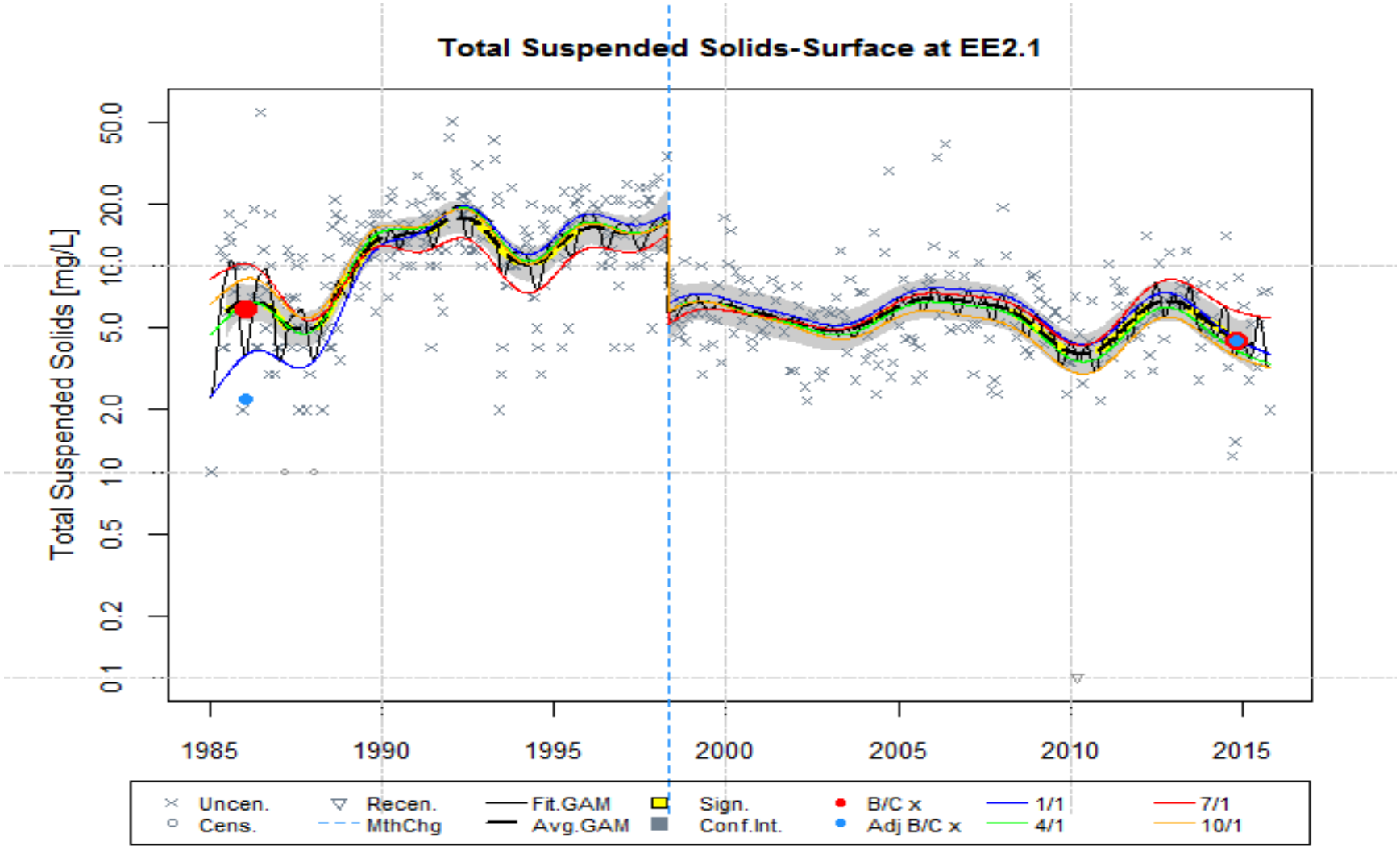
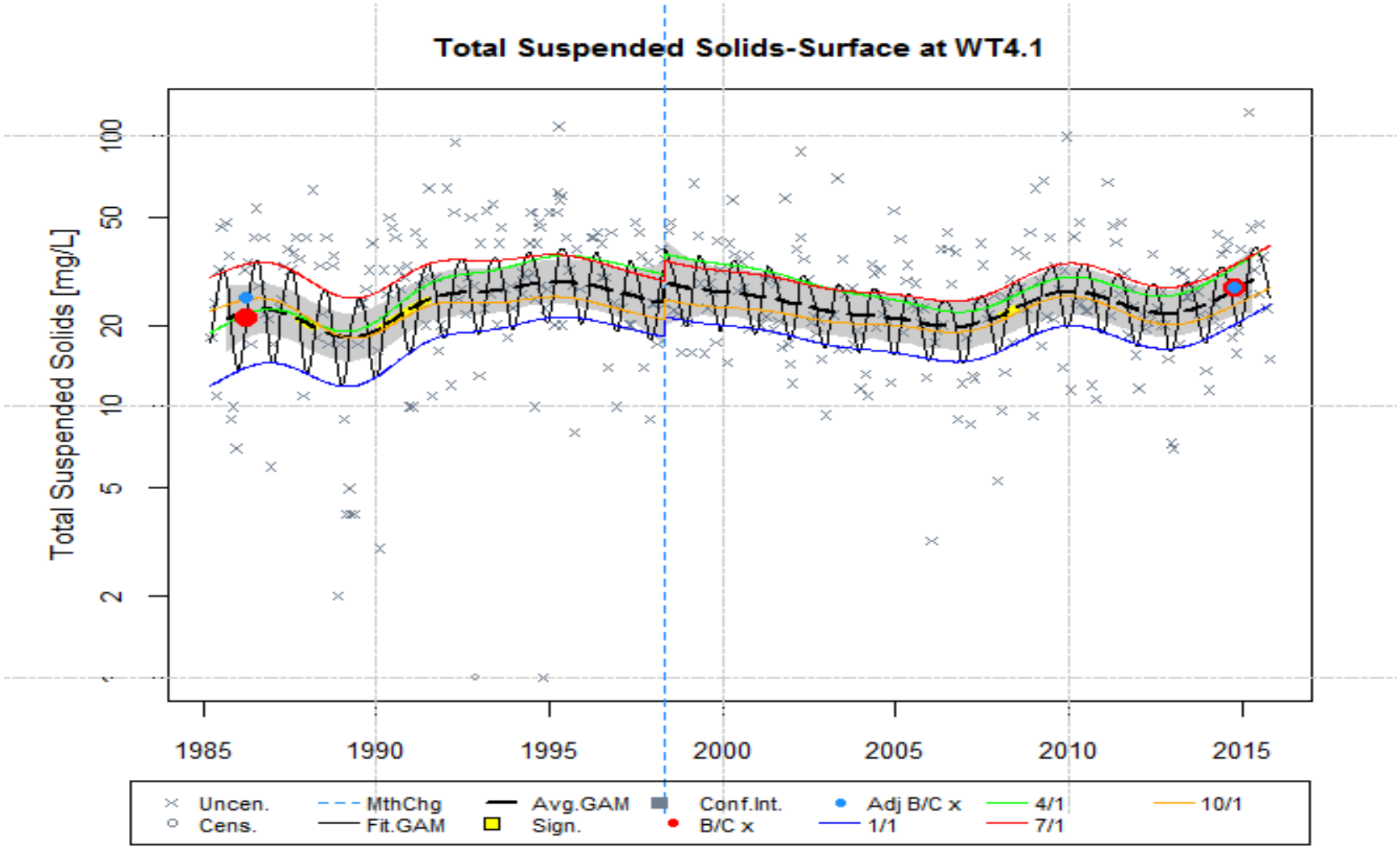


Figure 3. The TSS step size as a function of salinity for the 32 tributary stations. The fitted line is loess regression.



Baytrends model 3 fitted to EE2.1 (Choptank Embayment)TSS data.



Baytrends model 3 fitted to EE2.1 (Back River)TSS data.

This concludes the story that made us think we should develop a model that could assess Method Change effects station by station.

Now that we have it we have found other uses:

a. BNR Assessment
b. Dam Removal
c. Catastrophic events (e.g. Agnes 1972)

We have also discovered some problems:

a. 5 yrs pre and post
b. Interventions close together cause problems
c. A skip in the data with an intervention can lead to erroneous results
d. Interventions that are confounded with natural events (e.g. flow change)

Engage the DIWG on two topics:

a. Do you see issues with this approach to methods changes?
b. If we find variable steps by station, will you help us assess the cause?

end of file