Appendix 2.7A: Notes on Method for Accounting for Lab and Method Changes

Section 1 : Description of Approach

Modified from original document:
An Application of Intervention Analysis or Step-Trend Analysis to TSS Data Collected in Maryland Tributaries
Elgin Perry, 04Jul2016 originally

For various reasons it is often necessary to change the analytical method that is used to obtain the water quality measurements that are used for assessment. This happens when new and improved methods become available, when contractual matters make it necessary to switch from one laboratory to another, and even a change of field crew personnel might result in a shift in water quality measurements. The historical method of dealing with these abrupt measurement changes is to conduct a paired sample methods comparison study and from these data to estimate an adjustment factor. However, it has been observed that in some cases, the application of an adjustment factor creates a step trend in the resulting time series instead of removing a step trend. This observation suggests that the magnitude of the step trend may be dependent on the context of the measurements. If this is the case, trying to use a universal adjustment factor will result is erroneous conclusions about trends at locations where the adjustment factor is not appropriate. In this study, we revisit a data set where it is clear that the magnitude of the step trend is dependent on the sample environment and show how adding an intervention term to a generalized additive model (GAM) is an effective approach for modeling data with this feature.

In May of 1998, the responsibility for measuring TSS at the Maryland Tidal Tributary stations was transferred from the Department of Health and Mental Hygiene laboratory (DHMH) to the Chesapeake Biological Laboratory (CBL). After a period of time, it became apparent that at some stations the character of the data changed at this point of transition (Figure 1). At other stations the data appear to bridge the methods change with consistency (Figure 2). In 2008, I conducted an analysis of the Maryland Tribs TSS data using Step-Trend methods with multiple linear regression (Perry, 2008). The results of that analysis suggested that the magnitude of the step-trend is a function of the salinity at the station, and that variability in the post-methods change period is lower than before the methods change. In this study, we revisit these data and update the analysis to using GAMs rather than multiple linear regression. For now, I focus only on the step-trend feature of the data and ignore the change in variance feature. It is possible to address the change in variance feature using GAMs by implementing a mixed model GAM using the gamm() function in R. That will be left for the future analysis.

Appendix 2.7A

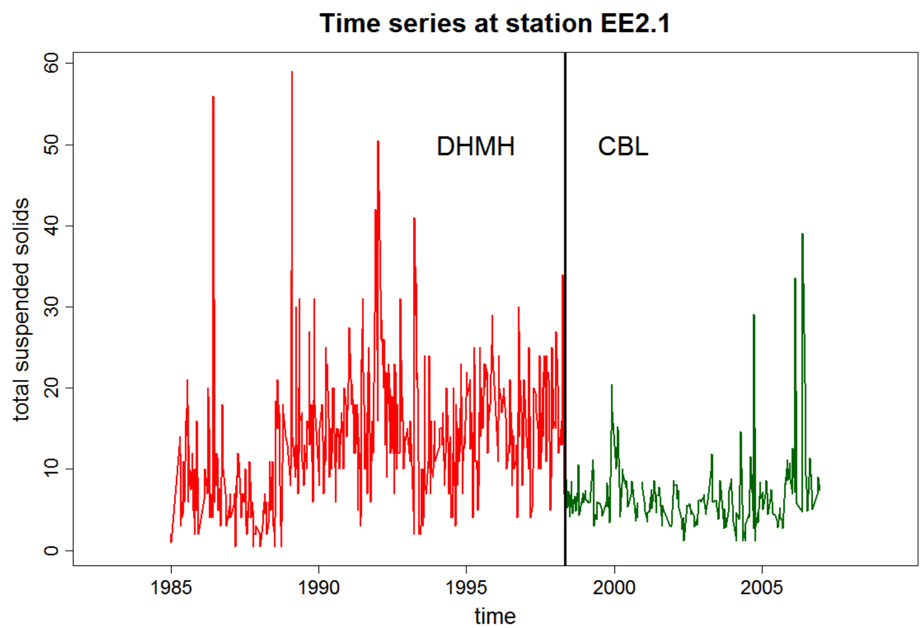


Figure 1. Time series of TSS at the EE2.1 station. The red curve shows data from DHMH. The green curve shows data from CBL. The black vertical line shows the point of laboratory change.

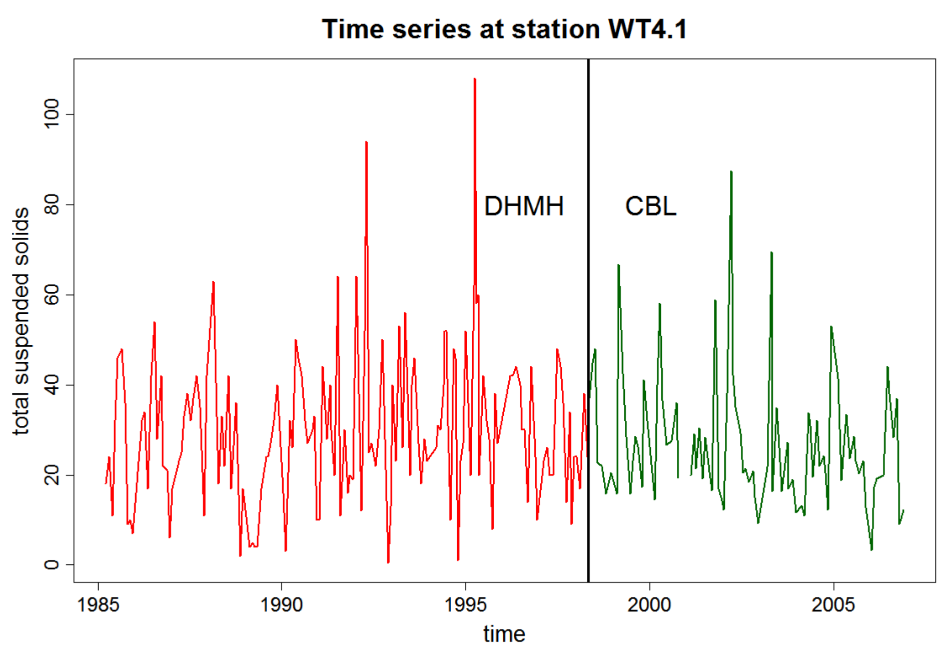


Figure 2. Time series of TSS at the WT4.1 station. The red curve shows data from DHMH. The green curve shows data from CBL. The black vertical line shows the point of laboratory change.

Methods

This analysis is built by extending the GAM model that is implemented in the baytrends GAM package. First, the gam1 model was tested which includes terms for intercept, linear time (cyear), non-linear time (s(cyear)), and season (s(doy)). Recent tests have been done using the gam2 model which includes the ti(doy,cyear) term. A term for change in method (LabChange) was added to the standard terms. The LabChange term is represented in the model by a variable that takes values of zero before the Laboratory Change and values of one afterwards. Adding the variable to the design matrix has the effect of shifting the model intercept at the point of the change in method. This model was implemented using log(TSS) as the dependent variable and includes data from only the surface layer.

Results

The intervention approach was applied to surface TSS from the MD tributary stations. Example results are presented in Section 2 of this appendix for gam1 and Section 3 for gam2, and a summary of the gam1 results is provided below. The results are presented in the order of mean station salinity to make it easier to discern the relation between salinity and the lab change effect. For each station there is a figure showing the log transformed data and the fitted model (Figure 3), an ANOVA table showing the significance of different terms in the model (Table 1) and a coefficients table (Table 2).

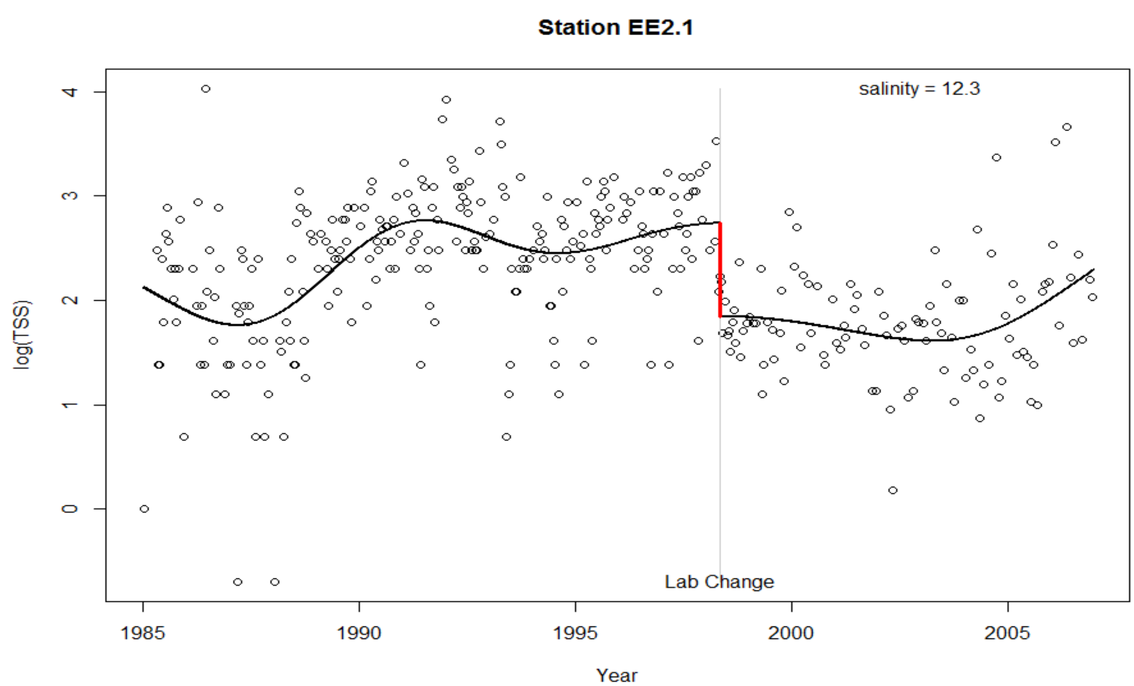


Figure 3. This figure shows the log transformed TSS data for station EE2.1 (mouth of Choptank River) and the fitted step-trend model. The vertical grey line above 1998 shows the point of the methods change. The red bar shows the magnitude of the method change effect. Mean station salinity is reported in the upper right.

Table 1. Step-trend analysis of variance for station EE2.1

Type	Source	edf	F-stat	p-value
parametric terms	LabChange	1	13.3181	0.0003
	cyear	1	0.0068	0.9342
smoothed terms	s(cyear)	6.62	5.612	<0.0001
	s(doy)	0	0	0.6645

Table 2. Step-trend coefficients for station EE2.1

parameter	estimate	Std. Err.	t value	p-value
(Intercept)	2.567564	1.087848	2.3602	0.0188
LabChangeTRUE	-0.895873	0.245485	-3.6494	0.0003
Cyear	0.026686	0.32289	0.0826	0.9342

In the example chosen here the LabChange term is significant (p=0.003). The estimated LabChange effect is a negative 0.8959 so that the post Lab Change intercept is 1.6717 (i.e., computed from 2.567564-0.895873). Note that as a result of including “select=TRUE” in the gam function of ‘mgcv’, the s(doy) term has been removed from the model and indicated by edf=0. This feature allows gam() to completely remove smooth terms that are not contributing to prediction.

Overall, 14 of the 32 Maryland tributary stations were found to have a significant step trend using the GAM analysis (Table 3). For the 11 stations that have a mean salinity greater than 10 ppt, all but one station has a significant step-trend. The analysis on this same dataset from 2008 using multiple linear regression (MLR) finds that 22 of 32 stations have a significant step-trend and identifies the salinity breakpoint at about 8.0 ppt.

Table 3. Summary by station of step trend results for Maryland Trib Stations (gam1 model).

Station	dep	lab.chg	lab.chg.pv	Mnsal	sig.col
ET1.1	TSS	0.0013	0.9936	0.1	blue
ET10.1	TSS	0.2827	0.1836	0.2	blue
ET6.1	TSS	0.2164	0.2715	0.2	blue
ET3.1	TSS	-0.3606	0.0266	0.6	red
ET4.1	TSS	0.0061	0.9810	0.6	blue
ET5.1	TSS	-0.1176	0.4529	0.8	blue
WT1.1	TSS	0.1037	0.6839	0.9	blue
ET2.2	TSS	-0.1355	0.5686	1.1	blue
ET2.3	TSS	-0.7856	0.0184	1.4	red
ET2.1	TSS	-0.2636	0.3768	1.8	blue
WT2.1	TSS	-0.3021	0.4910	1.9	blue
WT4.1	TSS	-0.0449	0.8068	2.3	blue
WT3.1	TSS	0.278	0.3307	3.4	blue
ET7.1	TSS	-0.0887	0.7136	7	blue
WT5.1	TSS	0.1253	0.6811	7.7	blue
ET6.2	TSS	-0.5032	0.0005	7.8	red

WT6.1	TSS	-0.5405	0.0747	8	blue
WT7.1	TSS	-0.937	0.0023	9.4	red
ET4.2	TSS	-0.4066	0.1519	9.5	blue
WT8.1	TSS	-0.5875	0.1206	9.5	blue
ET5.2	TSS	-0.3749	0.1414	9.8	blue
WT8.2	TSS	-0.6493	0.0223	10	red
WT8.3	TSS	-1.1732	0.0003	10.2	red
EE1.1	TSS	-0.5797	0.0689	12.2	blue
EE2.1	TSS	-0.8959	0.0003	12.3	red
EE2.2	TSS	-0.7018	0.0086	12.8	red
EE3.0	TSS	-0.4614	0.0033	13	red
ET8.1	TSS	-0.7587	<0.0001	13.7	red
EE3.1	TSS	-0.6935	0.0048	14.5	red
ET9.1	TSS	-0.8156	0.0003	15.6	red
EE3.2	TSS	-1.0895	<0.0001	16.5	red
EE3.3	TSS	-0.566	0.0008	16.8	red

It is clear that the magnitude of the step trend is negatively associated with increasing salinity (Figure 4).

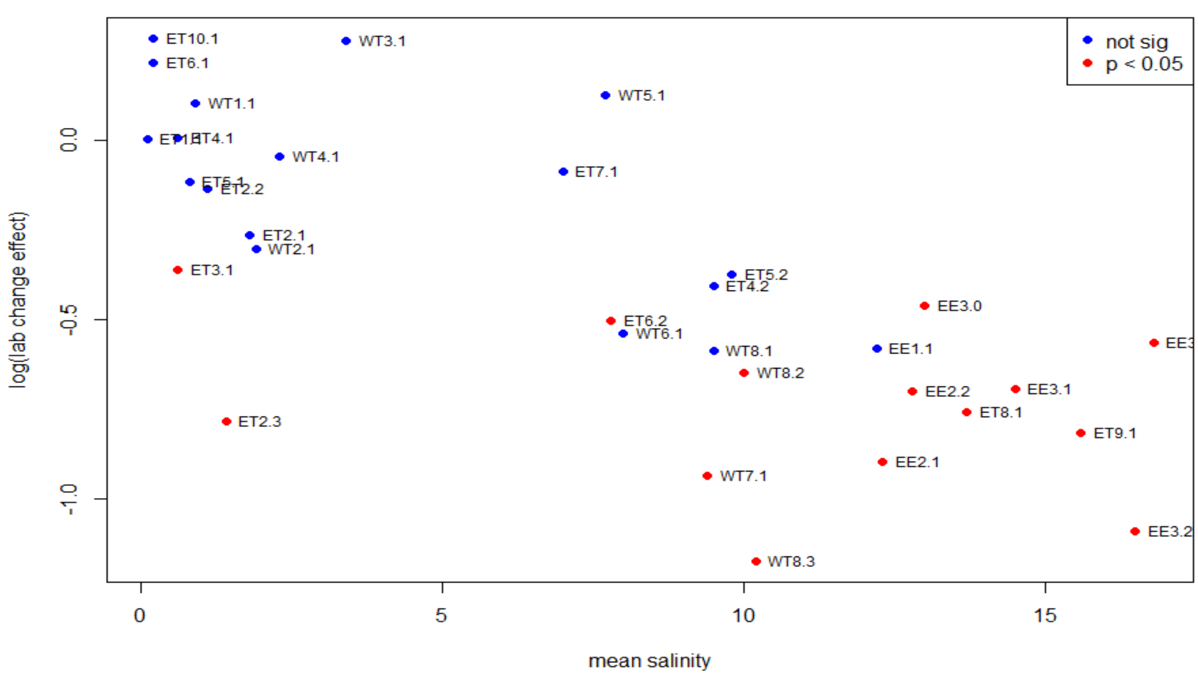


Figure 4. The figure illustrates the relation of the step-trend estimate (y-axis) to station salinity (x-axis). The statistical significance of the step-trend estimate is indicated by color.

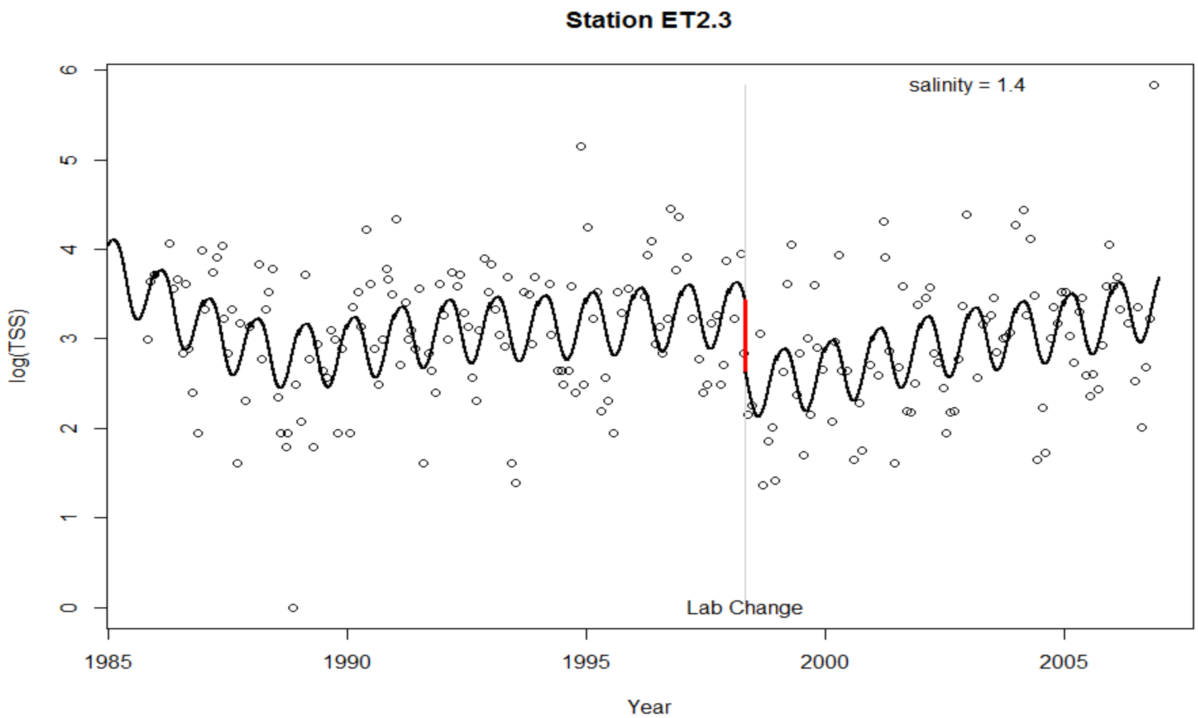
The results obtained by the GAM version of the step trend analysis are similar to those of the MLR version of the step-trend analysis. However, the GAM version finds fewer significant step trends than the previous MLR analysis. There are multiple differences between the two analyses that might explain this difference in results including log transformation with GAM and additional parameters included in the regression for MLR. The MLR analysis produces more exact inferential statistics than does the gam analysis.

Discussion

The primary advantage of the step trend approach is that it is adaptive. It lets the data define the magnitude of the adjustment factor between the two methods. The primary disadvantage is that it cannot be applied until several years of data have been collected in the post methods change period.- have to wait a while after methods change to use it.

Preliminary results in this Appendix Sections 2 and 3 for surface TSS at ET2.3, ET4.1, WT4.1, and ET9.1 are presented to demonstrate the method. It is likely that changes will occur in these results as we continue to work on the method. Based on these draft results, the laboratory change was significant at ET2.3 and ET9.1, but not ET4.1 or WT4.1. The results for gam2 with the intervention term in Section 3 demonstrate how the seasonal cycle is dynamic for TSS. The preliminary results for ET9.1 demonstrate how this approach can capture a change in the magnitude of the seasonal cycle at a laboratory change.

Section 2: Examples of intervention added to gam1



Step-trend analysis for ET2.3

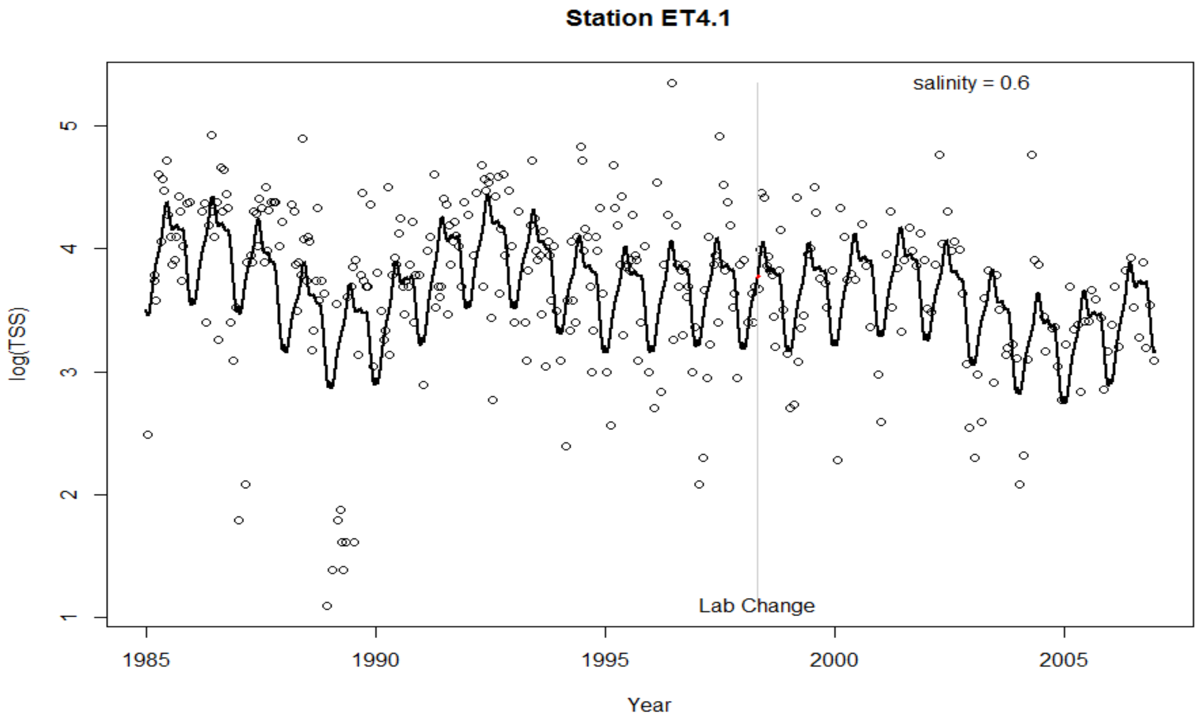
Type	Source	edf	F-stat	p-value
parametric terms	LabChange	1	5.6392	0.0184
	cyear	1	0.7328	0.3929
smoothed terms	s(cyear)	5.03	1.9565	0.0051
	s(doy)	2.62	4.8547	<0.0001

Step-trend coefficients for ET2.3

parameter	estimate	Std. Err.	t value	p-value
(Intercept)	3.135311	0.266452	11.7669	<0.0001
LabChangeTRUE	-0.785558	0.330803	-2.3747	0.0184
cyear	-0.09886	0.115483	-0.8561	0.3929

root mean-square error = 0.6651
adjusted r-square = 0.223

Appendix 2.7A



Step-trend analysis for ET4.1

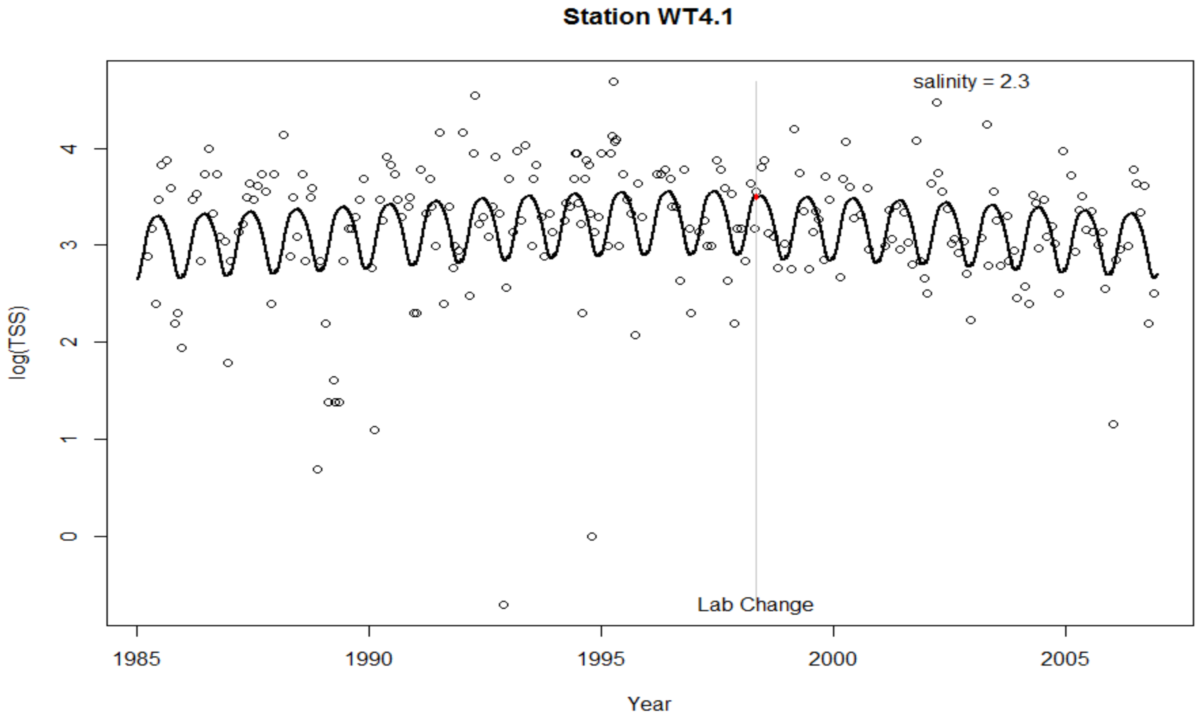
Type	Source	edf	F-stat	p-value
parametric terms	LabChange	1	6e-04	0.9810
	cyear	1	1.19	0.2761
smoothed terms	s(cyear)	7.6	5.542	<0.0001
	s(doy)	6.45	9.1843	<0.0001

Step-trend coefficients for ET4.1

parameter	estimate	Std. Err.	t value	p-value
(Intercept)	4.416359	0.636464	6.9389	<0.0001
LabChangeTRUE	0.006134	0.257915	0.0238	0.9810
cyear	0.200452	0.183751	1.0909	0.2761

root mean-square error = 0.5482
adjusted r-square = 0.2784

Appendix 2.7A



Step-trend analysis for WT4.1

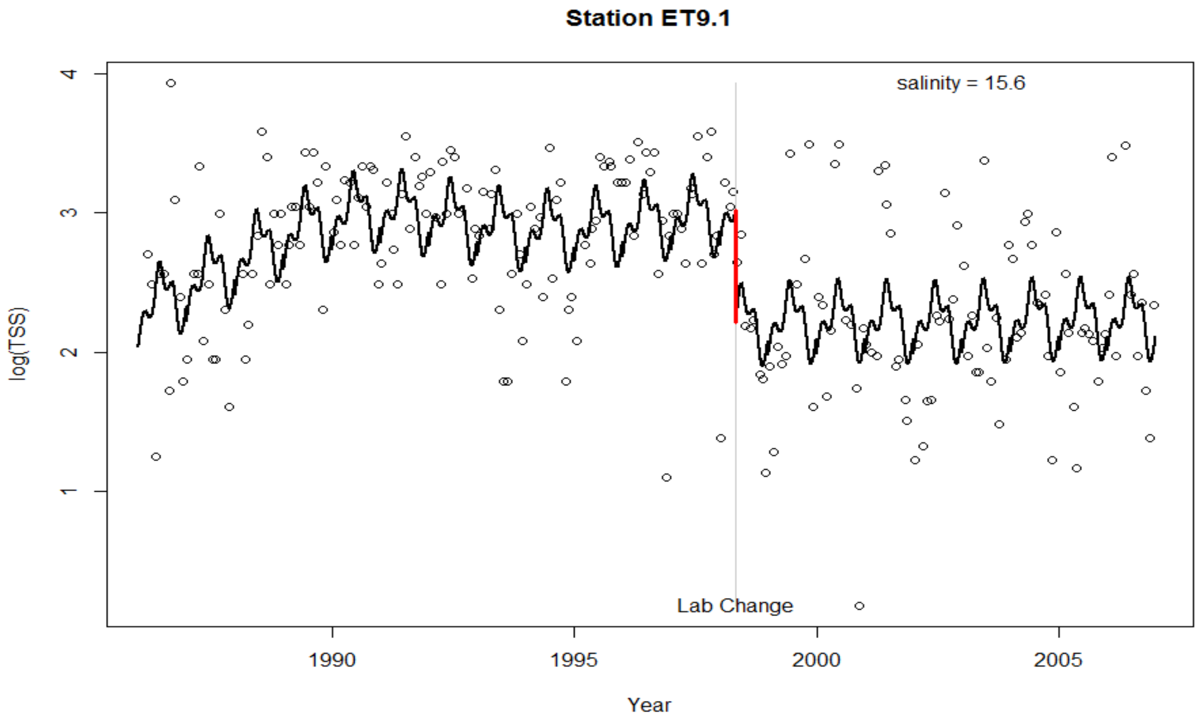
Type	Source	edf	F-stat	p-value
parametric terms	LabChange	1	0.0599	0.8068
	cyear	1	0.0111	0.9161
smoothed terms	s(cyear)	1.04	0.4603	0.0320
	s(doy)	3.21	4.4558	<0.0001

Step-trend coefficients for WT4.1

parameter	estimate	Std. Err.	t value	p-value
(Intercept)	3.202615	0.110288	29.0388	<0.0001
LabChangeTRUE	-0.044945	0.183567	-0.2448	0.8068
cyear	-0.002534	0.024034	-0.1054	0.9161

root mean-square error = 0.6371
adjusted r-square = 0.1393

Appendix 2.7A



Step-trend analysis for ET9.1

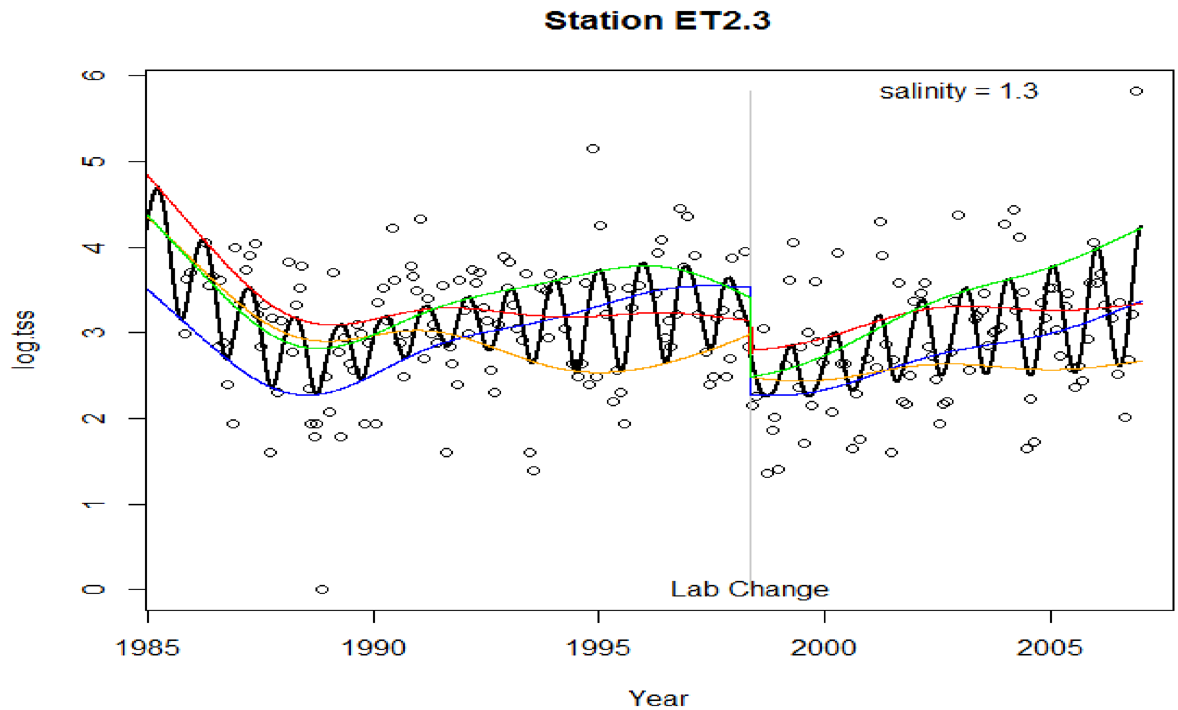
Type	Source	edf	F-stat	p-value
parametric terms	LabChange	1	13.5341	0.0003
	cyear	1	1.9422	0.1648
smoothed terms	s(cyear)	4.19	1.7578	0.0046
	s(doy)	5.85	3.4856	<0.0001

Step-trend coefficients for ET9.1

parameter	estimate	Std. Err.	t value	p-value
(Intercept)	3.077196	0.147219	20.9021	<0.0001
LabChangeTRUE	-0.815609	0.221701	-3.6789	0.0003
Cyear	0.089947	0.064542	1.3936	0.1648

root mean-square error = 0.513
adjusted r-square = 0.3478

Section 3: Examples of intervention added to gam2



Step-trend analysis for ET2.3

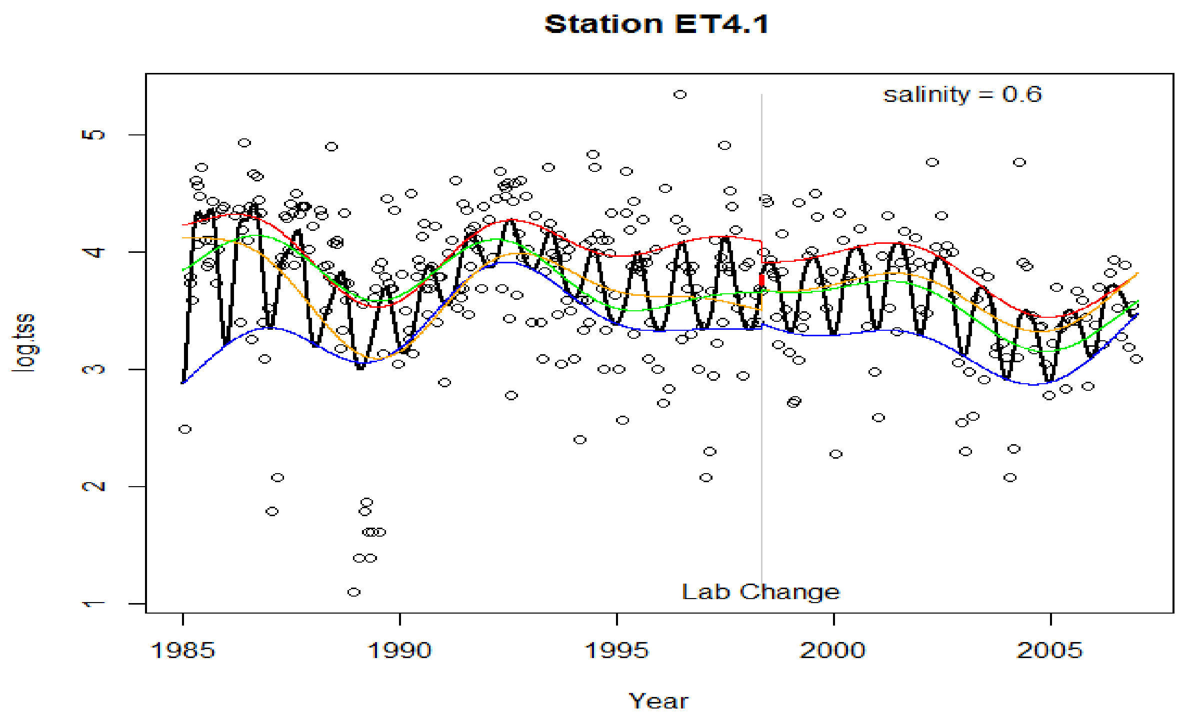
Type	Source	edf	F-stat	p-value
parametric terms	LabChange	1	5.3746	0.0214
smoothed terms	s(cyear)	7.18	4.143	0.0001
	ti(doy,cyear):LabChangeFALSE	6.22	3.5156	<0.0001
	ti(doy,cyear):LabChangeTRUE	4.05	3.0721	<0.0001

Step-trend coefficients for ET2.3

parameter	estimate	Std. Err.	t value	p-value
(Intercept)	3.343833	0.145928	22.9142	<0.0001
LabChangeTRUE	-0.801187	0.345588	-2.3183	0.0214

root mean-square error = 0.619
adjusted r-square = 0.3271

Appendix 2.7A



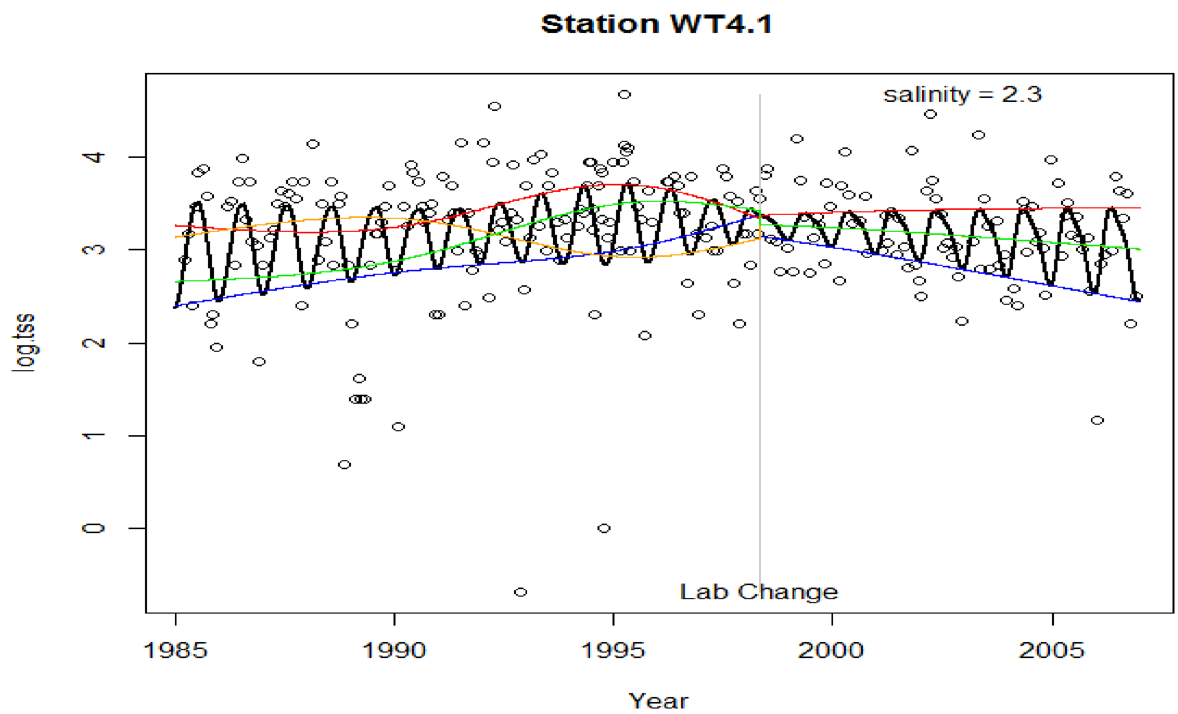
Type	Source	edf	F-stat	p-value
parametric terms	LabChange	1	1e-04	0.9915
smoothed terms	s(cyear)	8.59	5.523	<0.0001
	ti(doy,cyear):LabChangeFALSE	10.07	4.4602	<0.0001
	ti(doy,cyear):LabChangeTRUE	3.83	1.8596	<0.0001

Step-trend coefficients for ET4.1

parameter	estimate	Std. Err.	t value	p-value
(Intercept)	3.722111	0.084607	43.9932	<0.0001
LabChangeTRUE	0.002785	0.262111	0.0106	0.9915

root mean-square error = 0.5473
adjusted r-square = 0.2808

Appendix 2.7A



Step-trend analysis for WT4.1

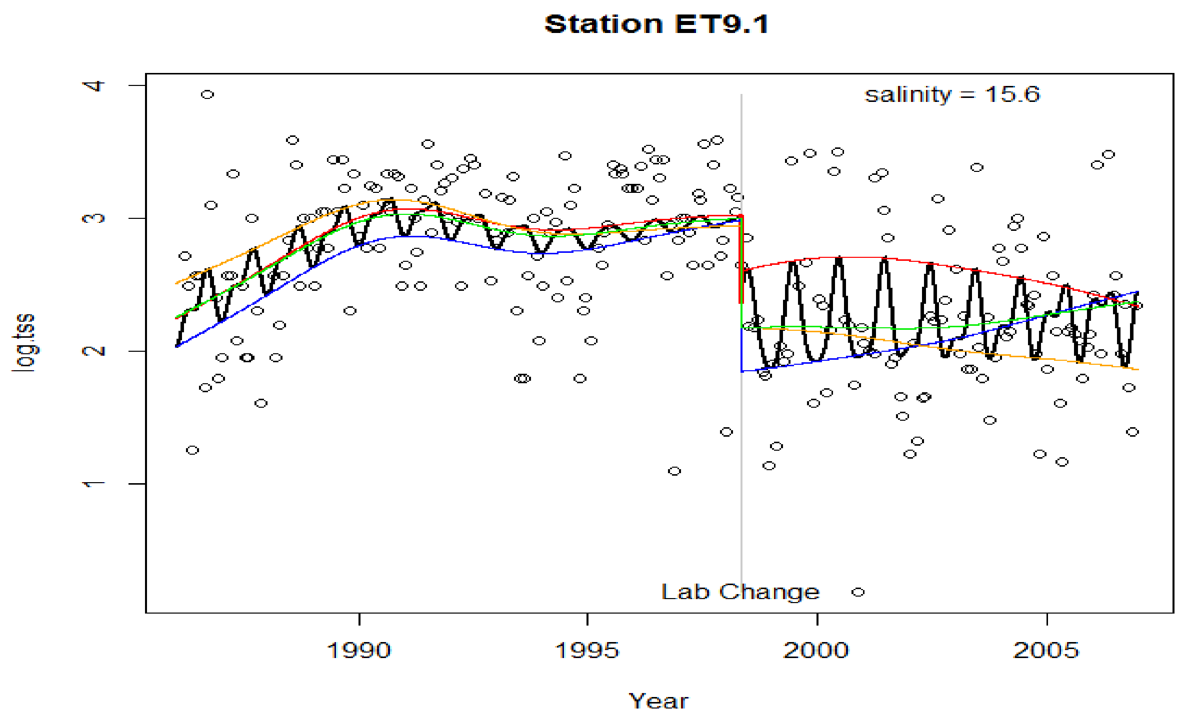
Type	Source	edf	F-stat	p-value
parametric terms	LabChange	1	0	0.9988
smoothed terms	s(cyear)	2.1	2.1505	0.1303
	ti(doy,cyear):LabChangeFALSE	6.66	2.976	<0.0001
	ti(doy,cyear):LabChangeTRUE	2.4	1.2572	0.0042

Step-trend coefficients for WT4.1

parameter	estimate	Std. Err.	t value	p-value
(Intercept)	3.191847	0.078367	40.7295	<0.0001
LabChangeTRUE	0.00027	0.182961	0.0015	0.9988

root mean-square error = 0.6269
adjusted r-square = 0.1668

Appendix 2.7A



Type	Source	edf	F-stat	p-value
parametric terms	LabChange	1	12.316	0.0005
smoothed terms	s(cyear)	5.55	2.9935	0.0065
	ti(doy,cyear):LabChangeFALSE	3.47	0.4514	0.1656
	ti(doy,cyear):LabChangeTRUE	5.14	2.0219	0.0001

Step-trend coefficients for ET9.1

parameter	estimate	Std. Err.	t value	p-value
(Intercept)	2.913853	0.09956	29.2672	<0.0001
LabChangeTRUE	-0.803177	0.228864	-3.5094	0.0005

root mean-square error = 0.5105
adjusted r-square = 0.354