APPENDIX B.2.4. STATISTICAL METHODOLOGY AND RESULTS FOR PHYTOPLANKTON AND PERIPHYTON EVALUATION OF EFFECTS, DORIS NORTH PROJECT, 2014

TABLE OF CONTENTS

1. ANALYSIS METHODS

1.1 ASSUMPTIONS

The key assumption of this analysis was that the samples collected are representative of each sites periphyton or phytoplankton biomass (estimated from chlorophyll *a* levels) for that monitoring period.

1.2 REPLICATE SAMPLES

Replicate samples collected on the same date were averaged prior to analysis. For some historical data, only the mean value was available.

1.3 DEALING WITH CENSORING (VALUES BELOW DETECTION LIMITS)

Some phytoplankton chlorophyll *a* concentrations were below analytical detection limits. The analyses below follow the advice of McBride (2005; Section 11.4.3).

- When the dataset included a small number of below detection values, values were replaced by values equal to ½ of the detection limit for the analysis.
- When the majority of the dataset consisted of values below the detection limit (e.g., more than about 70% below detection limit), there was very little that could be done as there is essentially no information (other than the values are below the detection limit). The analyses were performed as outlined below, but results were interpreted with caution. Helsel (2005) has other suggestions for analysis (e.g., comparing the proportions below the detection limits) but these tests require much larger sample sizes than available here.
- When there is an intermediate amount of censoring, a more complex analysis can be conducted that fully integrates the information from the censored values with the known values. This is most easily done using Bayesian methods using MCMC methods as likelihood methods would require integration of the likelihood for each censored value over and above dealing with the other random effects in the model.

Fortunately, the large lake and marine phytoplankton biomass dataset fell into the first category (few censored values), and the small lake and stream datasets did not contain censored values.

1.4 TRANSFORMATIONS

A preliminary analysis found that the variance of the periphyton and phytoplankton biomass values tended to increase with the mean and that the distribution of values was skewed. This suggested that a logarithmic transformation was appropriate.

1.5 OUTLINE OF ANALYSIS PLAN

There are several classes of statistical tests that can be done to assess evidence of change in the mean periphyton or phytoplankton biomass over time and to assess if changes in the means may reflect an impact of the Project.

1.5.1 Before vs. After Analysis

The first analysis compared the mean readings for all years prior to 2010 (before initiation of construction) to the mean for 2014. Each waterbody was treated independently.

The final statistical model (in standard shorthand notation) is:

$$Y = Period\ Season\ Year(period)\text{-}R$$

where *Y* is the biomass reading of periphyton or phytoplankton in a year; *Period* is the effect of before vs. after Project initiation (which is the effect of interest); *Season* is the effect of season (4 seasons in the sampling year: 1) April or May (under-ice), 2) July, 3) August, 4) September); and *Year(Period)* is a random year effect (applicable if more than one year was sampled during each period).

This model is preferable to simply treating all measurement within a period (e.g., over the baseline years) as having the same mean and assuming that they are completely independent of each other. This model also "averages" the phytoplankton values collected over the baseline years in a way that weights each year equally rather than weighting by the sample size. For example, suppose that the phytoplankton biomass measured in 2008 was 22 µg chl *a*/L, while the two readings measured in 2009 were 25 and 27 µg chl *a*/L. The simple mean (22+25+27)/3=24.7 µg chl *a*/L would be more heavily weighted toward the mean in the second year. In order to give each year's data equal weight, the reading over both years should be computed as an average of averages:

$$\frac{\frac{22}{1}+\frac{25+27}{2}}{2}=24.0$$

This could be extended to multiple years in a similar fashion. Note that in the case of balanced data, i.e., equal number of replicates in all years, the average-of-the-averages and simple-average will be identical. The model with a random year effect is still preferred in the case of balanced data because the readings within a year may be correlated due to year-specific random factors that cause all the readings within a year to increase or decrease in step.

This model was fit using R version 2.15.2. In order to reduce the number of "false positives" associated with performing a large number of statistical tests, a reduced significance level (e.g., 0.01) was primarily used when reviewing the results. Significance levels of 0.01 to 0.05 for exposure sites were considered marginally non-significant and were also discussed in the evaluation of effects.

The key disadvantage of this model is that changes over time may be unrelated to the effects of the Project, e.g., the average periphyton or phytoplankton biomass readings in 2014 could be higher or lower than expected because of long term climate change that is unrelated to the Project. Consequently, if a statistically significant effect was detected, it would require further investigation.

1.5.2 BACI Analysis

Before-After-Control-Impact (BACI) analysis is the standard method used to assess an environmental impact (Smith 2002). BACI designs evaluate potential non-parallelism in response over time between the exposure and reference waterbodies. A BACI analysis was performed for each exposure waterbody versus the corresponding reference waterbody.

The formal statistical model (in standard shorthand notation) was:

Y = *Period Season Class Period*Class Year(Period)-R*

where *Y* is the variable of interest; *Period* is the effect of period (before or after construction); *Season* is the effect of season (4 seasons in the sampling year: 1) April or May (under-ice), 2) July, 3) August, 4) September); *Class* is the effect of waterbody classification (Project or reference); *Period*Class* is the BACI effect of interest (i.e., is the effect of *Period* the same (parallel) for both classes of sites); and *Year(Period)* is the random year effect within each period (applicable if more than one year was sampled during each period). If there were multiple reference waterbodies (as is the case for streams), a term *Body(Class)-R* (the random site effect within each class) would also be added to the model so that the change in the mean for the exposure site was compared to the average change in the mean for the reference bodies. Sites that were measured only in one period (e.g., Reference D OF was measured only post-construction) contribute some information on the year-effect which improves precision of the BACI estimate.

The key variable of interest was the *Period*Class* effect as this measures the amount of non-parallelism between the changes in the mean (Before-After) over the two classes of sites (exposure reference).The BACI estimate was computed as the "difference in the differences":

$$BACI = \left(\mu_{PA} - \mu_{PB}\right) - \left(\mu_{RA} - \mu_{RB}\right)$$

whereμ_{PA} is the mean variable reading in the *Project* class of sites *after* Project initiation,μ_{PB} is the mean variable reading in the *Project* class of sites *before* Project initiation,μ_{RA} is the mean variable reading in the *reference* class of sites *after* Project initiation, and μ_{RB} is the mean variable reading in the *reference* class of sites *before* Project initiation. The BACI contrast was estimated by replacing the population means above by the model-based estimates. Estimated differences close to 0 would indicate no evidence of non-parallelism.

Note that the hypothesis that the BACI contrast has the value of zero is identical to the hypothesis that the *Period*Class* interaction is zero with identical p-values. Consequently, only the results for the BACI contrast were reported.

This model was fit using R version 2.15.2. A reduced significance level (e.g., 0.01) was primarily used when reviewing the results. Significance levels of 0.01 to 0.05 were considered marginally non-significant. If the results of the before-after analysis indicated that a 2014 variable mean was not significantly different (i.e., and was not marginally non-significant) from the baseline mean, it was concluded that there was no effect of 2014 Project activities on this variable, and BACI analyses were not discussed.

For all environments, caution should be used in interpreting the results from the BACI analysis because there was only one reading for phytoplankton or periphyton biomass in one reference site before Project initiation.

2. RESULTS

2.1 STREAM DATA

There was no censoring of the stream periphyton biomass values. Preliminary plots of the data showed one outlier from 1997, which was removed (Table Stream-1).

The results from the analysis that compared the means before and after Project construction are presented in Table Stream-2. Because of the absence of baseline data for Roberts OF and Reference D OF, and a lack of degrees of freedom in the before period for Little Roberts OF and Reference B OF, before-after comparisons were not possible for these streams. There was no evidence of a significant or marginally non-significant change in the mean *log(periphyton)* for Doris OF. Thus, no BACI results are discussed.

2.2 MARINE DATA

A few marine phytoplankton biomass values were censored (Table Marine-1). No outliers were detected. Because of the scarcity of baseline data available for all marine sites, before-after comparisons could not be conducted (too few degrees of freedom to fit the model).

BACI comparisons for both RBE and RBW sites failed to detect evidence of a significant or marginally non-significant differential change compared to the reference site (Table Marine-2).

2.3 LAKE DATA

The lake data was divided into small lakes (Little Roberts and Reference D) and large lakes (Doris South, Doris North, and Reference B). A few large lake phytoplankton biomass values were censored (Table Large Lake-1). No censoring was observed for the small lake phytoplankton data. Preliminary plots of the data did not show any evidence of outliers.

There was no evidence of a significant or marginally non-significant change in the mean *log(phytoplankton)* in any lake between the before and after period (Tables Large Lake-2 and Small Lake-1). Thus, BACI results are not discussed.

REFERENCES

Helsel, D.R. 2005. *Nondetects and data analysis*.Wiley: New York.

McBride, G.B. 2005. *Using statistical methods for water quality management*. Wiley: New York.

Smith, E. P. 2002. *BACI Design.Encyclopedia of Environmetrics*. Wiley: New York.

TABLES

Table Stream-1. Outlier Periphyton Biomass Values

Site	Variable	Year	Date	Rep	Value	Censored (1 = yes; 0 = no)
Doris OF	Periphyton	1997	19-Jul-1997	1	194.4	0
Doris OF	logPeriphyton	1997	19-Jul-1997	1	5.27	0

Table Stream-2. Summary of Test for No Difference in Mean between Before and After Periods for Stream Periphyton Biomass

		Stream Site				
		Doris OF	Little Roberts OF	Reference B OF	Reference D OF	Roberts OF
Variable	Proportion Censored	p-value	p-value	p-value	p-value	p-value
logPeriphyton	0.00	0.1924	-	-	-	-

Notes:
To reduce the number of false positives, a smaller alpha level (0.01) was used to screen for effects and p-values less than 0.01 are bolded. P-values of 0.01 to 0.05 were considered marginally non-significant.
- denotes that no analysis could be performed because of the lack of variation in a variable over time, or because data were only available from one period (i.e., no "before" data were available).

Table Stream-3. Summary of BACI Comparison of Stream Periphyton Biomass for Individual Project Sites

	Stream Site	
	Doris OF	Little Roberts OF
Variable	p-value	p-value
logPeriphyton	0.9426	0.0305

Notes:
To reduce the number of false positives, a smaller alpha level (0.01) was used to screen for effects and p-values less than 0.01 are bolded. P-values of 0.01 to 0.05 were considered marginally non-significant.
In Roberts OF stream, no readings were taken pre-construction, so no BACI comparison was possible.

Table Marine-1. Summary of the Amount of Censoring in the Marine Sites

	Marine Site		
	RBE	RBW	REF-Marine 1
Variable	Proportion Censored	Proportion Censored	Proportion Censored
logPhytoplankton	0.27	0.20	0.00

Table Marine-2. Summary of BACI Comparison of Marine Phytoplankton Biomass for Individual Project Sites

	Marine Site	
	RBE	RBW
Variable	p-value	p-value
logPhytoplankton	0.7146	0.4284

Note:
To reduce the number of false positives, a smaller alpha level (0.01) was used to screen for effects and p-values less than 0.01 are bolded. P-values of 0.01 to 0.05 were considered marginally non-significant.

Table Large Lake-1. Summary of the Amount of Censoring in the Large Lake Sites

	Large Lake Site		
	Doris North	Doris South	Reference B
Variable	Proportion Censored	Proportion Censored	Proportion Censored
logPhytoplankton	0.06	0.06	0.00

Table Large Lake-2. Summary of Test for No Difference in Mean between Before and After Periods for Large Lake Phytoplankton Biomass

	Large Lake Site		
	Doris North	Doris South	Reference B
Variable	p-value	p-value	p-value
logPhytoplankton	0.2785	0.8929	-

Notes:
To reduce the number of false positives, a smaller alpha level (0.01) was used to screen for effects and p-values less than 0.01 are bolded. P-values of 0.01 to 0.05 were considered marginally non-significant.
- denotes that no analysis could be performed because of the lack of variation in a variable over time, or because data were only available from one period (i.e., no "before" data were available).

Table Large Lake-3. Summary of BACI Comparison of Large Lake Phytoplankton Biomass for Individual Project Sites

	Large Lake Site	
	Doris North	Doris South
Variable	p-value	p-value
logPhytoplankton	0.0992	0.0612

Note:
To reduce the number of false positives, a smaller alpha level (0.01) was used to screen for effects and p-values less than 0.01 are bolded. P-values of 0.01 to 0.05 were considered marginally non-significant.

Table Small Lake-1. Summary of Test for No Difference in Mean between Before and After Periods for Small Lake Phytoplankton Biomass

		Small Lake Site	
		Little Roberts	Reference D
Variable	Proportion Censored	p-value	p-value
logPhytoplankton	0.00	0.7120	.

Notes:
To reduce the number of false positives, a smaller alpha level (0.01) was used to screen for effects and p-values less than 0.01 are bolded. P-values of 0.01 to 0.05 were considered marginally non-significant.
In Reference D, no readings were taken pre-construction, so no BACI comparison was possible.

APPENDIX B.2.5. STATISTICAL METHODOLOGY AND RESULTS FOR BENTHOS EVALUATION OF EFFECTS, DORIS NORTH PROJECT, 2014

TABLE OF CONTENTS

1. ANALYSIS METHODS

1.1 ASSUMPTIONS

The key assumption of this analysis was that the samples collected were representative samples of the site and year of sampling. Since all samples were taken at the same time of year (August), inference was restricted to comparisons during this period.

1.2 TRANSFORMATIONS

A preliminary analysis found that the variance of several variable values tended to increase with the mean in some waterbodies, but there was no consistent pattern. Given that few years of data were available, the evidence for a transformation was weak and no transformation was done.

1.3 OUTLINE OF ANALYSIS PLAN

1.3.1 Impact Level-by-time Analysis

Benthos data have only been collected since 2010 (post-construction); there is no information available pre-construction. Consequently, before/after or BACI comparisons cannot be done.

In cases where no pre-Project data are available, Wiens and Parker (1995) suggest several alternatives, especially for long-term monitoring. One of their suggestions was the level-by-time comparisons where the trend-lines for exposure and reference waterbodies are compared to see if there is evidence of non-parallelism over time.

The final statistical model (in standard shorthand notation) is:

$$Y = Class\ Year\ Class*Year$$

where *Y* is the reading for the benthic variable; *Class* is the effect of waterbody classification (Project or reference); *Year* is the effect of year; and *Class*Year* is the non-parallelism in the response over time. For marine and lake sites, there was a single site in each class and separate comparisons were made for each site against its corresponding reference site. For streams, multiple reference waterbodies were available for comparison against a single Project waterbody; therefore, the term *Body(Class)-R* (the random site effect within each class) was added for the multiple sites within the classification.

The Class*Year term was the effect of interest representing non-parallel changes over time between the sites.

The model was fit using R version 2.15.2. In order to reduce the number of "false positives" associated with performing a large number of statistical tests, a reduced significance level (e.g., 0.01) was primarily used when reviewing the results. Significance levels of 0.01 to 0.05 for exposure sites were considered marginally non-significant and were also discussed in the evaluation of effects.

1.3.2 Multivariate Approaches

The approach outlined above analyzes each benthic variable independently. However, the variables are likely not independent of each other, and presumably higher power would result if a multivariate approach were used. This has not been attempted in this report.

2. RESULTS

2.1 STREAM DATA

Results of the analysis that tests for parallelism in the mean variable value over time are presented in Table Stream-1. There was evidence of significant non-parallelism in Simpson's Evenness Index at Doris OF and in the Bray-Curtis Index at Roberts OF. The before-after comparison for the Bray-Curtis Index was also marginally non-significant at Doris OF. At Little Roberts OF the before-after comparison for density was marginally non-significant.

2.2 MARINE DATA

The marine data was analyzed in two ways: 1) with data for adults and juveniles pooled together and 2) with data for adults only. Results of the tests for parallelism are presented in Tables Marine-1 and Marine-2.

For the pooled adult and juvenile data, significant non-parallelism was detected for the Bray-Curtis Index and richness at both RBE and RBW. There was also evidence of significant non-parallelism for the Simpson's Diversity Index at RBE, and density and Simpson's Evenness at RBW. There was a marginally non-significant non-parallelism for the Simpson's Diversity Index at RBW and for Simpson's Evenness at RBE.

For the adult-only data, significant non-parallelism was detected in all evaluated variables at RBE and RBW with the exception of the Bray-Curtis Index at RBW where a marginally non-significant non-parallelism was detected.

2.3 LAKE DATA

The lakes data was divided into large lakes (Doris Lake South, Doris Lake North, and Reference Lake B) and small lakes (Little Roberts Lake and Reference Lake D). Results of the tests for parallelism are presented in Tables Large Lake-1 and Small Lake-1.

There was evidence of significant non-parallelism in the Bray-Curtis Index and density for both Doris Lake South and Doris Lake North, and Simpson's Diversity Index in Doris Lake North. There was also a marginally non-significant non-parallelism for richness at Doris Lake South.

In the small lake (Little Roberts Lake) there was evidence of significant non-parallelism for density and Simpson's Evenness Index and a marginally non-significant non-parallelism for richness.

REFERENCES

Wiens, J. A., and K. R. Parker. 1995. Analyzing the effects of accidental environmental impacts: approaches and assumptions. *Ecological Applications* 5(4), 1069–1083.

TABLES

Table Stream-1. Summary of Tests for Parallelism for the Stream Benthos Data

	Stream Site		
	Doris OF	Little Roberts OF	Roberts OF
Variable	p-value	p-value	p-value
Bray-Curtis	0.0245	0.3225	**<0.0002**
Density	0.1508	0.0294	0.0889
Diversity	0.3638	0.7455	0.8525
Evenness	**0.0005**	0.3057	0.4413
Richness	0.2345	0.3730	0.4272

Note:
To reduce the number of false positives, a smaller alpha level (0.01) was used to screen for effects and p-values less than 0.01 are bolded. P-values of 0.01 to 0.05 were considered marginally non-significant.

Table Marine-1. Summary of Tests for Parallelism for the Marine Benthos Data (Adults and Juveniles Pooled)

	Marine Site	
	RBE	RBW
Variable	p-value	p-value
Bray-Curtis	**0.0001**	**0.0037**
Density	0.4609	**<0.0001**
Diversity	**0.0003**	0.0381
Evenness	0.0454	**<0.0001**
Richness	**<0.0001**	**0.0047**

Note:
To reduce the number of false positives, a smaller alpha level (0.01) was used to screen for effects and p-values less than 0.01 are bolded. P-values of 0.01 to 0.05 were considered marginally non-significant.

Table Marine-2. Summary of Tests for Parallelism for the Marine Benthos Data (Adults Only)

	Marine Site	
	RBE	RBW
Variable	p-value	p-value
Bray-Curtis	**<0.0001**	0.0140
Density	**<0.0001**	**<0.0001**
Diversity	**0.0001**	**0.0001**
Evenness	**0.0068**	**<0.0001**
Richness	**<0.0001**	**0.0014**

Note:
To reduce the number of false positives, a smaller alpha level (0.01) was used to screen for effects and p-values less than 0.01 are bolded. P-values of 0.01 to 0.05 were considered marginally non-significant.

Table Large Lake-1. Summary of Tests for Parallelism for the Large Lake Benthos Data

	Large Lake Site	
	Doris North	Doris South
Variable	p-value	p-value
Bray-Curtis	**<0.0001**	**<0.0001**
Density	**0.0014**	**<0.0001**
Diversity	**0.0015**	0.0694
Evenness	0.4828	0.0536
Richness	0.1700	0.0353

Note:
To reduce the number of false positives, a smaller alpha level (0.01) was used to screen for effects and p-values less than 0.01 are bolded. P-values of 0.01 to 0.05 were considered marginally non-significant.

Table Lake-Small-1. Summary of Tests for Parallelism for the Small Lake Benthos Data

	Small Lake Site
	Little Roberts
Variable	p-value
Bray-Curtis	0.0839
Density	**0.0001**
Diversity	0.1130
Evenness	**0.0040**
Richness	0.0187

Note:
To reduce the number of false positives, a smaller alpha level (0.01) was used to screen for effects and p-values less than 0.01 are bolded. P-values of 0.01 to 0.05 were considered marginally non-significant.