An Overall BACI Analysis of the 1998-2005 LSC Monitoring Data

Stephen Ellner and Nelson G. Hairston, Jr., EEB Cornell
Clifford E. Kraft, DNR Cornell

Introduction

Here we describe BACI-type analyses that avoid questions of how to correct for multiple comparisons, by considering the data set as a whole and conducting only a few scientifically motivated analyses that represent different hypotheses about the system. Sites 1,3,4,5,6,7,8 and LSC were included in the analysis, removing the outliers identified in the report "A Before-After-Control-Impact Analysis for Cornell University's Lake Source Cooling Facility" prepared by the Upstate Freshwater Institute (we will refer to this as the "UFI Report"). Site 2 was omitted because of the occasional marked influence of the Ithaca Wastewater Treatment Plant outflow.

Following Underwood (1994) and Smith (2002), we formulated the BACI analysis in an "analysis of variance" framework. The key questions are whether there is statistical evidence for an interaction between time period (before versus after LSC startup) and treatment (control versus intervention), and if so, what is the magnitude of the interaction effect. Two groupings of sites were considered. First, each site was considered separately, without attempting to classify sites *a priori* as "control" versus "intervention". Second, sites were classified (by consensus among the authors of this report, Todd Cowen (Civil and Environmental Engineering, Cornell) and David Matthews of UFI) into three groups based on their expected level of impact from LSC: Sites 1 and 7 ("near"); Sites 3, 4, and 5 ("middle"); Sites 6, LSC and 8 ("far").

Chlorophyll A: Preliminary Data Evaluation

One clear feature of the log-transformed chlorophyll *a* data is a smooth seasonal trend (Figure 1), reflecting seasonal trends in temperature and day length, and apparently consistent across sites (note: here and throughout this report, all logarithms are base-*e* natural logarithms). The analysis of these data must take account of this seasonal pattern, by including day of year as a predictor variable in the statistical model. To model a smooth seasonal trend without introducing bias by (mis)selecting a functional form for the seasonal trend, we used the mgcv package (version 1.3-31, Wood 2006) in R (R Development Core Team 2008) to fit the seasonal trend using a nonparametric (spline) regression model. Following the recommendation of Kim and Gu (2004), the nonparametric term was fitted with model degrees of freedom over-weighted by a factor of 1.4 to avoid overfitting.

In a preliminary evaluation of chlorophyll *a* data, main effects of time (before vs. after), site and group were determined to be individually significant as single terms added to a model with only seasonal trends. The presence of site-specific (rather than group-specific) means was highly significant ($p<0.001$ using the anova.gam function). The fitted site main effects support the *a priori* grouping, in that Sites 3, 4 and 5 had the three lowest main effects. However, the detection of statistically significant within-group variability suggests that all analyses should use models that allow for site-specific main effects rather than group-specific main effects.

This analysis therefore led to the decision that the BACI analysis would have to consider models with three main effects (1 - day of year, 2 - site, and 3 – time period "before" vs. "after"), as well as potential interactions between time period and either “group” (the "near", "middle" and "far" groups identified above) or site (each individual sample site location). To reiterate a key point, testing for the statistical significance of these potential interactions constitutes the BACI analysis. Note that there is no inconsistency between specifying site-specific main effects and an interaction between time period and group; this model posits that each site is intrinsically unique, but the effects of LSC startup are uniform within each group of sites.

Chlorophyll A: BACI Analysis

The BACI statistical model evaluated to represent the null hypothesis of no intervention effect (i.e. no change in chlorophyll *a* after LSC operation was initiated) is:

$$\text{Log(chl A)}_{i,t} = \text{Site effect}_i + \text{Seasonal trend}(t) + d \times \text{After} + e_{i,t} \qquad (1)$$

where $e_{i,t}$ is random variability, and d is the mean difference for time intervals after versus before operation of the LSC facility. "After" is an indicator variable with possible values -1 or 1, taking the value -1 if the sampling time t is before the operation of the LSC facility, and 1 if the sampling time t is after operation of the LSC facility began.

The BACI statistical model evaluated to represent the alternative hypothesis of an intervention effect (i.e. a change in chlorophyll *a* trends after LSC operation was initiated) is

$$\text{Log(chl A)}_{i,t} = \text{Site effect}_i + \text{Seasonal trend}(t) + \bar{d} \times \text{After} + d_i \times \text{After} + e_{i,t} \qquad (2)$$

where $\bar{d}$ is the main effect of time period (representing the average temporal trend across sites), and the d_i are site-specific effects of time period. So the difference between models (1) and (2) is that the latter has a different value of d (the before vs. after difference in mean log chlorophyll *a*) for each site, i.e. $\bar{d} + d_i$, representing a site-specific response to "before" versus "after". An effect of LSC operation would be indicated if some sites – or groupings of sites – exhibited larger before vs. after differences than others. An interaction between time period and group is modeled by allowing the value of d to vary among groups of sites, but not within groups. Note that our assumed model for LSC impacts (a site-specific or group-specific change in mean following the start of LSC facility operation) is the same as the model for LSC impacts in the pairwise-comparison approach used in the UFI Report.

Chlorophyll A: Results of BACI Analysis

The additions of time×site and time×group interactions were both nonsignificant (p = 0.73 for addition of time×group, p=0.83 for addition of time×site). **Thus, the data set as a whole does not show a detectable effect of LSC on temporal trends in Chlorophyll *a*.**

The use of site-specific (rather than group-specific) main effects and the nonparametric fitting of the temporal trend both increase the number of fitted parameters in the interaction-free base model, and therefore risk some loss of potential statistical power to detect time×location interactions. Several analyses were therefore conducted to check on this possibility, as follows.

First, the seasonal trend was fitted by a polynomial. We determined that a cubic term was sufficient to describe the overall trend, therefore the above analyses were repeated using a cubic model for the seasonal trend. The results were nearly identical, with the p values for the additions of time period×location interactions changing by less than 0.02.

Second, in an analysis using group-specific main effects (instead of the above analysis with site-specific main effects) for the purpose of reducing the parameter count in the base model, addition of the time×group interaction was not statistically significant (p = 0.73). [For this analysis, note that only the addition of a time×group interaction can be considered meaningfully because time×site interaction terms necessarily entail site-specific means during either the "before" period, the "after" period, or both. Since site-specific means are very strongly evident in the data, significance of time period×site interaction terms would not actually say anything about presence of the interaction.]

Putting aside the question of statistical significance (because it is unlikely that the impact of LSC is exactly zero), we can use these models to ask: **If the operation of the LSC facility is having an effect on chlorophyll *a* in southern Cayuga Lake, how big is it?** We can answer this question by using the ANOVA decomposition of the Total Sum of Squares (i.e. the total variance in log chlorophyll *a*) in the fitted model (2), and comparing the contributions to the Total Sum of Squares from the following three categories of effects:

- The main effects of site and date (day of year and time interval ("before" vs. "after")).
- The interaction between site and time period, which is the potential indicator of an LSC impact.
- Remaining variability that is unexplained by the model ("random error").

Figure 2(a) displays the Sum of Squares decomposition for the log chlorophyll *a* data. Thus, even if we make the *a priori* assumption that an LSC effect is present (which is not supported by the data), we find that the estimated LSC effect is very small.

Note that in Figure 2, quantities are plotted on square-root scale and normalized by the sample size. That is, for each component of the Total Sum of Squares, the quantity displayed in Figure 2 is

$$\sqrt{SS_c / n} \qquad (3)$$

where SS_c is the total ANOVA sum of squares for that category of effects, and n is the number of data points. This scale was used because it corresponds to the standard deviation of the contribution from each category of effect, rather than the variance, so it is a more intuitive measure of "typical size" than the sum of squares.

Moreover, the value of equation (3) for the site×time period interaction provides an overall measure of the average effect size for these interactions, which are the LSC effect in the BACI analysis. This way of measuring "average effect size" of LSC is the answer to the question: *if the effect of LSC were removed from the data, by what proportion would a typical data point change in value, either up or down*? For chlorophyll *a* the value of (3) is 0.053, meaning that the average size of the LSC effect is a 5.3% proportional change in the chlorophyll *a* reading, which in absolute terms is 0.22 µg/l on average. However, it must be re-emphasized that this estimate is predicated on assuming that an LSC effect exists, even though our analysis finds no evidence for this assumption in the data.

Chlorophyll A: power analysis

The fitted BACI model (1) can be used to estimate the statistical power of the tests that we have conducted by asking “what would be our likelihood of detecting a LSC impact of a given size using this BACI analysis, if it actually occurred?” We did this using a parametric bootstrap procedure. We repeatedly generated "bootstrap" data sets by simulating data from model (1) without the presence of an LSC effect. Each such data set was generated by replacing each value in the actual Cayuga Lake data set with the value for the same location and time predicted by model (1) plus a residual error drawn at random (with replacement) from the set of actual residuals (residuals are the actual data values minus the value predicted by model (1) for that location and time). These data sets are statistical mimics of the real data, all having the property that an LSC effect is absent. But they can then be modified by imposing a known LSC effect of our choosing, after which we repeat the BACI analysis and see if the analysis detects an LSC effect (i.e., we fit models (1) and (2) to the bootstrap data set, and test for the significance of the relevant interaction terms).

Conducting a power analysis requires specification of the alternatives that will be considered. Here we consider alternatives motivated by the *a priori* site groupings based on expected potential impact of LSC. For comparison with the power analysis in the UFI Report, we considered a 30% impact of LSC (i.e., a 30% increase in the mean Chlorophyll *a* reading after the start of LSC operation). This was imposed on the "near" site group (Site 1 and Site 7), but not on the other sites. Based on a sample size of 1000 bootstrap replicates, our procedure has 85% probability of detecting a 30% difference between "near" and other groups when testing the significance of a time×group interaction at level $\alpha=0.05$, and 92% probability of detecting this 30% difference if the significance level is changed to $\alpha=0.1$. Both of these probabilities are higher than the corresponding power for the pairwise-comparison approach used in the UFI Report (70% and 80%); in particular, the probability of failing to detect a 30% impact has been reduced by half.

As a second comparison, we considered a 30% impact of LSC imposed on both the "near" and "middle" site groups, but not on the "far" sites. Based on 1000 bootstrap replicates, our procedure has a 94% probability of detecting this impact when testing for a time×group interaction at significance level $\alpha=0.05$, and a 97% probability of this impact when testing at significance level $\alpha=0.10$.

Total Phosphorus and Turbidity

Our analyses of the (log-transformed) Total P and turbidity data followed the same pattern as the Chlorophyll *a* analysis, with exactly the same principal conclusions. The main points can be summarized as follows:

1. Preliminary analysis again identified day-of-year, site, and time (before vs. after the start of LSC operation) as main effects that must be included in the BACI model analyses, for both total P and turbidity.
2. The estimated seasonal trends in total P and turbidity were both more complex than that seen for Chlorophyll *a*, so the use of a simple polynomial model for seasonal trends was not attempted. All analyses were based on models using a spline model to represent seasonal trends.
3. Results of the BACI analyses (comparing model (2) with model (1)) were identical: no significant time×site or time×group interactions were detected, when added to base models that incorporated only the main effects of day-of-year, time interval, and either site or group. The *p* values for these hypothesis tests were all above 0.9 for total P, and all above 0.5 for Turbidity.
4. The estimated contributions of potential LSC to the Total Sum of Squares for log total P and log turbidity were both very small (see Figure 2b,c). The estimated effect sizes using equation (3) are 4.6% for total P (0.76 µg/l on average), and 7.7% for turbidity (0.19 NTU on average) . As with chlorophyll *a*, these estimates are highly imprecise and predicated on the making the assumption that an LSC effect is present, despite the lack of evidence for an LSC effect in our analyses.
5. The power to detect LSC effects was markedly higher for total P than that for chlorophyll *a*. A 30% impact imposed on only the "near" sites has an estimated 98% chance of being detected when testing at $\alpha=0.05$, and a 20% impact imposed on only the "near" sites has an estimated 84% chance of being detected when testing at $\alpha=0.05$, and an estimated 90% chance of being detected when testing at $\alpha=0.1$. However the power to detect LSC

effects was lower for turbidity. A 30% impact imposed on only the "near" sites has an estimated 58% chance of being detected when testing at $\alpha=0.05$, and a 69% chance of being detected when testing at $\alpha=0.1$. A 30% impact imposed on both "near" and "middle" sites has an estimated 67% chance of being detected when testing at $\alpha=0.05$, and an estimated 78% chance of being detected when testing at $\alpha=0.1$.

Conclusion

The gist of this report is in Figure 2, and the statistical analysis simply formalizes what the eye sees there. If we assume that an impact of LSC facility operation is present despite the lack of evidence in our analyses to support this assumption, then estimated effect sizes (representing the typical magnitude of proportional changes in the value of a data point due to LSC) are very small. As a result, when we carry out statistical hypothesis tests for the presence of an LSC effect, at either conventional ($\alpha=0.05$) or relaxed ($\alpha=0.1$) standards of evidence, we cannot reject the hypothesis that an LSC effect is absent, for all of the three variables that have been monitored.

This situation is analogous to a political poll showing that John Adams is leading Thomas Jefferson by 2%, with a margin of error of ±3%. If we take the raw data at face value we would say that Adams is ahead by 2% (this is the approach in Figure 2, where we plot the estimated size of LSC effects without worrying about whether it might be just an accident of sampling variability). But when we take account of the margin of error due to sampling variability, we have to acknowledge that we don't really know who's ahead, and Jefferson might actually be leading Adams. This is what we have found about LSC by doing statistical tests for the significance of the interaction terms representing the LSC effect in the ANOVA analysis. If we ask "can we really conclude from the data that LSC is having any effect whatsoever?", the answer is "no".

Literature Cited

Kim, Y.-J. and C. Gu (2004). Smoothing spline Gaussian regression: More scalable computation via efficient approximation. *Journal of the Royal Statistical Society Series B* 66: 337-356.

Underwood, A.J. 1994. On Beyond BACI: Sampling designs that might reliably detect environmental disturbances. *Ecological Applications* 4: 4-15.

R Development Core Team (2008). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Smith, Eric P. 2002. BACI design. pp. 141-148 in: Encyclopedia of Environmetrics, vol. 1 (A.H. El-Shaarawi and W. W. Piegorsch, editors). John Wiley and Sons.

Wood, S.N. (2006) Generalized Additive Models: An Introduction with R. Chapman and Hall/CRC.

Figure 1. Log$_e$ -transformed chlorophyll *a* data as a function of day of year, omitting site 2 data, and with the identified outliers for sites 3,4,5,7 and 8 removed. "Before" data are plotted with color-coded symbols (numbers for sites 1-8, "L" for LSC). "After" data are plotted with open circles in the same color as the "before" data for the same site: LSC in black, and sites 1,3,4,5,6,7,8 in red, orange, yellow, green, blue, purple, and violet, respectively.

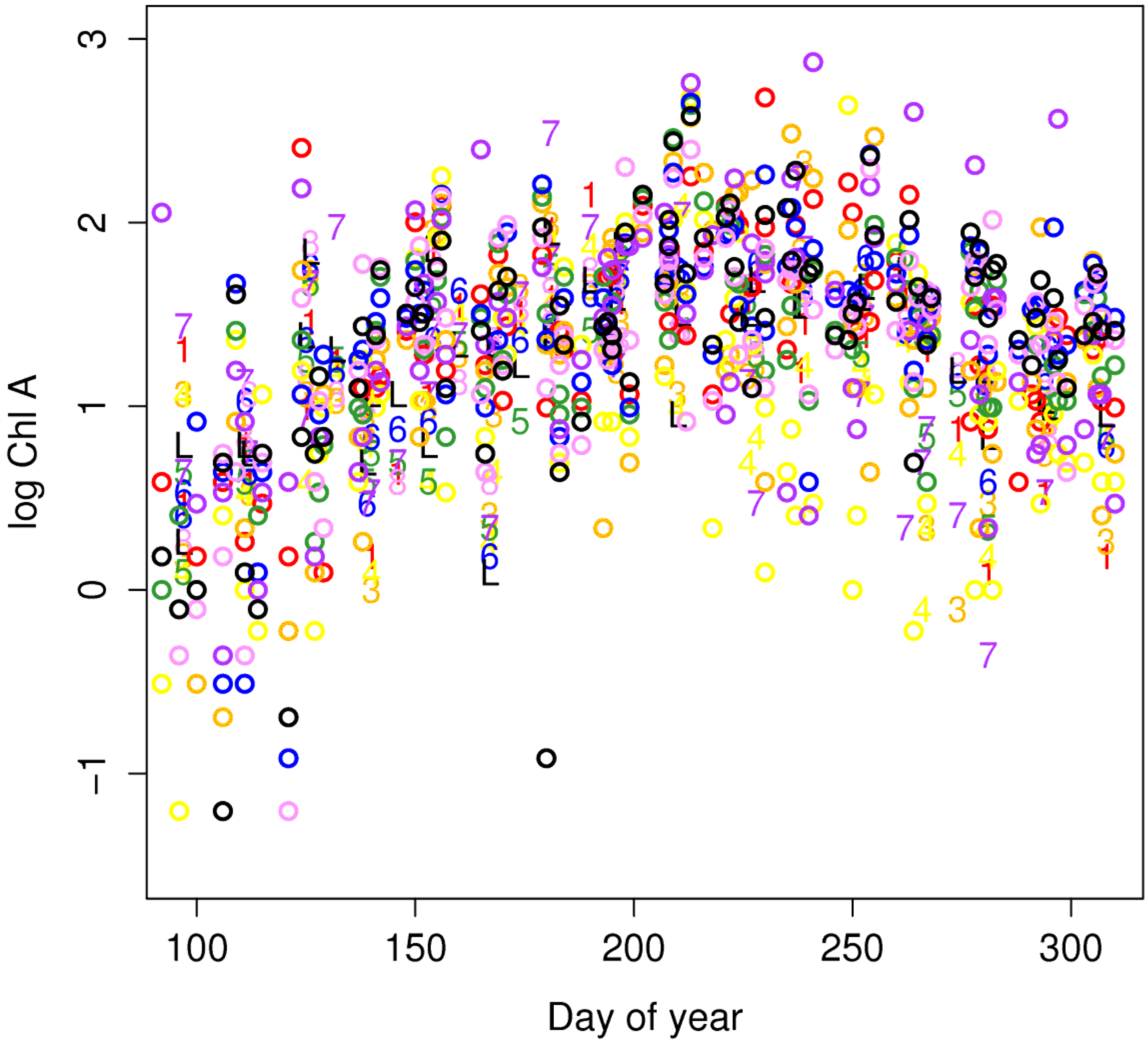

Figure 2. Decomposition of the Total Sum of Squares for the three (log-transformed) water quality variables that have been monitored, based on the fitted model (2). The bar labeled "LSC" shows the Sum of Squares from the time period×site interaction term. The bar labeled "Site and date effects" shows the total Sum of Squares from the main effects of site, day of year, and time period (before vs. after). The bar labeled "Random variability" shows the residual Sum of Squares representing variability not explained by any of the factors included in the model. Each of the three components of the Total Sum of Squares has been normalized by the sample size *n* and then square-root transformed, as discussed in the text.