LCWMD Microhabitat Data Review

Curtis C. Bohlen

2/3/2021

# Introduction

A few years ago, LCWMD undertook an effort to improve habitat quality along a portion of Long Creek. Years of adjacent development activity had restricted the floodplain, eroded the stream channel, and removed woody debris and other roughness elements from the channel. In what became perhaps the most ambitious urban stream restoration in Maine history, LCWMD removed fill from the floodplain, installed log jams and other structures in the stream, added gravel and cobbles to the stream, and emplaced downed trees, rootwads, and wood in the floodplain.

GZA Geoenvironmental Incorporated (GZA) has been monitoring the effects of the restoration on stream habitat and the biota. As a portion of their monitoring, they conducted a study of relative abundance of different microhabitats and substrate types in the restored stream reach and in a “reference reach” a short way upstream. The study looked at visual estimates of percent cover of each of fifteen microhabitats and substrate types in 63 quadrats (3 replicates at each of 21 random locations) in each stream reach.

In this report, I present results of reanalysis of the GZA data, along with some commentary. For simplicity’s sake, I ignore the hierarchical structure (quadrats within locations) imposed by the sampling design. A comprehensive analysis should address that structure. Such clustering of samples would be expected to increase uncertainty in overall estimates of microhabitat abundance. Results of this somewhat simplified analysis should be evaluated with that in mind.

The report is generated by an R Markdown document. All figures and results are calculated live from the data based on code in R Markdown. I have hidden the code that generated the analysis to keep the focus on the meaning and ideas, rather than implementation details.

You can download the original Markdown document that contains the R code from GitHub at:

(<https://github.com/ccb60/LCWMD_2020_microhabitat.git>)

## Goals

The goal of the following analysis is to estimate the relative area of each microhabitat in each stream reach, and determine if the relative abundances are similar or different between the restoration reach and the reference reach.

## Primary Analysis Approach

The best estimate of the relative area of each microhabitat can be obtained by be adding up the area of each microhabitat observed in each quadrat, and dividing by the total area of all of the quadrats. That weighted sum gives an estimate of the relative percent cover of each microhabitat in the stream reach.

But that weighted sum is just the arithmetic mean of the observed percent cover in all 63 quadrats (including zeros, where a particular microhabitat was not observed in a quadrat).

Although we often mistrust the mean when dealing with skewed data, in this case, the mean is the correct summary statistic for our purposes, so we need to work with it, rather than a median or other more resistant estimate of location.

# Calculate Estimated Percent Covers

We start by calculating by stream reach the number of plots in which each microhabitat was observed, and the estimated percent cover of each microhabitat.

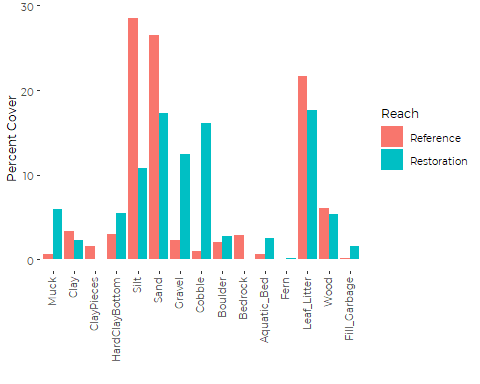
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Microhabitat | n\_Reference | n\_Restoration | Cover\_Reference | Cover\_Restoration |
| Muck | 3 | 17 | 0.63 | 5.87 |
| Clay | 3 | 10 | 3.33 | 2.22 |
| ClayPieces | 1 | 0 | 1.51 | 0.00 |
| HardClayBottom | 2 | 6 | 2.94 | 5.40 |
| Silt | 36 | 35 | 28.57 | 10.79 |
| Sand | 40 | 42 | 26.51 | 17.25 |
| Gravel | 5 | 21 | 2.30 | 12.38 |
| Cobble | 2 | 21 | 0.95 | 16.06 |
| Boulder | 3 | 6 | 2.06 | 2.78 |
| Bedrock | 2 | 0 | 2.86 | 0.00 |
| Aquatic\_Bed | 3 | 9 | 0.56 | 2.54 |
| Fern | 0 | 1 | 0.00 | 0.08 |
| Leaf\_Litter | 51 | 47 | 21.67 | 17.67 |
| Wood | 38 | 32 | 6.03 | 5.37 |
| Fill\_Garbage | 1 | 8 | 0.08 | 1.59 |

## A Quick Check

We check to see whether the total percent cover in each reach is close to 100%. If our math is right, each total should be 100% (within rounding error).

#> Reference Restoration   
#> 100 100

## Plot Results

It’s easier to digest the results from a graphic, rather than a table. 

Visually, what jumps out are:

* High levels of Muck, Gravel, and Cobble in the Restoration Reach
* High levels of Silt and Sand in the Reference Reach.

# Testing for “Significance”

How can we be sure those apparent “differences” are meaningful, and equally important, that less visually less impressive differences are NOT meaningful?

## Can We Use t Tests?

We just calculated Percent Cover of each Microhabitat in each reach as the average of 63 observations from each stream reach. Comparing two means is what t tests are for.

Unfortunately, t tests are not ideal for several reasons. Most importantly, the observed data are far from normally distributed, and the sample size is only moderate. The t statistic is based on asymptotic behavior of a sum or mean drawn from a population with an unknown standard deviation, which therefore must be estimated from the data. Convergence on the (symmetrical) t distribution is slow when you start with a highly skewed distribution (as here).

Another, more subtle point is that the different microhabitat types are not independent, since they must add up to 100% of the habitat. Comparing each microhabitat separately is likely to overstate statistical significance (although by only a small bit here).

What that all means is a t test may give relatively poor control of Type 1 error. The nominal “P value” may not be especially trustworthy. Because of the central limit theorem (sums of random variables converge to a normal distribution), the larger the sample size, the less the distribution of the underlying data matters. Since our sample sizes are fairly large (n = 63), things may work out. If we don’t take the nominal p value too seriously, we may be O.K.

## Try a t Test Anyway

Despite my reservations, I went ahead and calculated nominal p values from a standard t test for comparing two samples with unequal variances (Welch’s t test). I used two tailed t tests because I am interested in checking if there are any differences between the two samples. I don’t care which is larger and which is smaller.

#> difference p.value  
#> Muck -5.24 0.0004  
#> Clay 1.11 0.6096  
#> ClayPieces 1.51 0.3212  
#> HardClayBottom -2.46 0.4224  
#> Silt 17.78 0.0003  
#> Sand 9.25 0.0577  
#> Gravel -10.08 0.0048  
#> Cobble -15.11 0.0001  
#> Boulder -0.71 0.7198  
#> Bedrock 2.86 0.1590  
#> Aquatic\_Bed -1.98 0.0285  
#> Fern -0.08 0.3212  
#> Leaf\_Litter 4.00 0.3362  
#> Wood 0.67 0.6188  
#> Fill\_Garbage -1.51 0.0398

## Resampling: An Alternative to the t Test

The most straight forward approach to comparing two means from “badly behaved” distributions is probably to use a resampling test. This addresses the skewed nature of the data, but does not address the fact that the cover of all microhabitats us sum to 100%.

Under the null hypothesis of no difference in microhabitat distribution, the distribution of possible observations of percent cover from the two reaches should be the same. We can readily simulate a draw from two data distributions that matches that description, without relying on any assumptions about the nature of the underlying distribution by sampling repeatedly from the actual observations.

For each microhabitat, we can draw two samples of 63 from all the observations of that microhabitat (including the zeros) from both stream reaches. That random sample is an (estimated) example of what we might expect to observe under the null hypothesis. We can then then calculate the difference between these two “random” samples. Since the two samples were drawn from the same data, they will usually be similar, but sometimes the samples will be quite different

We do that many times, and count up how many times we see differences between the two simulated samples that are as large as the difference we observed in the field. The frequency of extreme observations provides an estimate of the “p value” of the observed difference between the two stream reaches based on a resampling or “bootstrap” test comparing the to populations.

I wrapped the resampling logic into a function, and called the function for each of the different microhabitats. In the function, I draw 10,000 bootstrap samples. Since this is based on a a random process, the estimated p values will change slightly each time the code is run, but with a resample of 10,000, it should not vary much.

## Resampling Results

#> difference p.value  
#> Muck 5.238095 1e-04   
#> Clay 1.111111 0.5871   
#> ClayPieces 1.507937 0.118   
#> HardClayBottom 2.460317 0.4073   
#> Silt 17.77778 3e-04   
#> Sand 9.253968 0.0593   
#> Gravel 10.07937 0.0052   
#> Cobble 15.11111 0   
#> Boulder 0.7142857 0.7193   
#> Bedrock 2.857143 0.0722   
#> Aquatic\_Bed 1.984127 0.0304   
#> Fern 0.07936508 0.1149   
#> Leaf\_Litter 4 0.3308   
#> Wood 0.6666667 0.6089   
#> Fill\_Garbage 1.507937 0.0376

## Compare Two Tests

It is instructive to compare the p value estimates for the two different tests. The p value is a property of the statistical model, not only of the data, so different analytic methods give different p values. Here, the two tests both test differences between observed means. Both p values estimate the probability that the observed difference between the two means would arise by chance alone, if there were no real difference in abundance of the specific microhabitat.

#> t\_test Resampling  
#> Muck 0.0004 0.0001  
#> Clay 0.6096 0.5871  
#> ClayPieces 0.3212 0.1180  
#> HardClayBottom 0.4224 0.4073  
#> Silt 0.0003 0.0003  
#> Sand 0.0577 0.0593  
#> Gravel 0.0048 0.0052  
#> Cobble 0.0001 0.0000  
#> Boulder 0.7198 0.7193  
#> Bedrock 0.1590 0.0722  
#> Aquatic\_Bed 0.0285 0.0304  
#> Fern 0.3212 0.1149  
#> Leaf\_Litter 0.3362 0.3308  
#> Wood 0.6188 0.6089  
#> Fill\_Garbage 0.0398 0.0376

Qualitatively, the results are similar. In only one case (Bedrock) does the choice of test change the qualitative conclusions we would draw. That gives me some confidence that the conclusions are robust to the statistical approach used.

We can conclude (based on the suggested “liberal” critical p value of 10%) that differences in microhabitat abundance for the following microhabitats are probably meaningful:

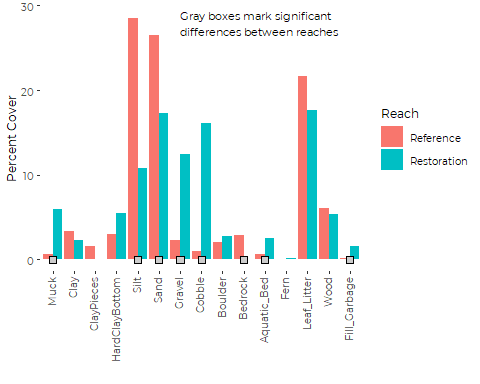
* Muck
* Silt
* Sand
* Gravel
* Cobble
* Bedrock (by one test and not the other)
* Aquatic Bed
* Fill Garbage

# Modified Graphics

## Add a Marker of Statistical Significance

It is nice to add some sort of a marker to the graphic showing which comparisons are statistically robust.

#> Warning: Removed 7 rows containing missing values (geom\_point).



## Add Confidence Intervals

A graphical alternative is to add standard deviation, standard error, or confidence intervals to the graphic. I would not use standard deviation, as the error distribution is likely to be asymmetrical, especially for rarely seen habitats. A standard error of the mean (or symmetric confidence intervals derived from the standard error) hold the same risk. My preference is to go with bootstrapped confidence intervals.

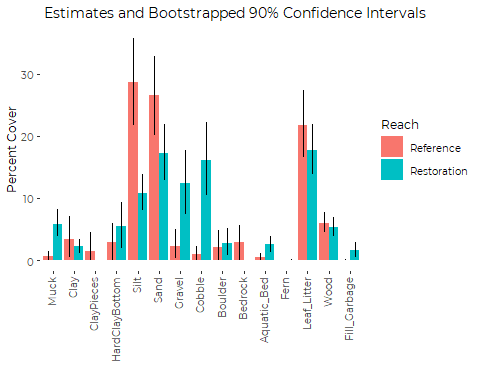
Bootstrapped confidence intervals (in their simplest form) resample from the observed data and calculate the mean of that sample. Do that over and over again and you can use the distribution of those resampled means to estimate the likely outer limits of what you would expect to observe if you repeated the study.

I calculated 90% confidence intervals (matching the proposed 10% P value) via a simple bootstrap model, built into another small function. The intervals are based on 10,000 bootstrap samples. The specific values are the result of resampling, so the exact values will differ in the third or fourth decimal place each time the code is run.

### Calculate Confidence Intervals

#> Reference Reach Bootstrapped 90% Confidence Intervals  
#> Upper.CL Lower.CL  
#> Muck 1.429 0.079  
#> Clay 7.143 0.476  
#> ClayPieces 4.524 0.000  
#> HardClayBottom 6.032 0.000  
#> Silt 35.794 21.667  
#> Sand 32.857 20.159  
#> Gravel 5.000 0.317  
#> Cobble 2.222 0.000  
#> Boulder 4.762 0.079  
#> Bedrock 5.714 0.000  
#> Aquatic\_Bed 1.190 0.079  
#> Fern 0.000 0.000  
#> Leaf\_Litter 27.381 16.508  
#> Wood 7.698 4.524  
#> Fill\_Garbage 0.238 0.000  
#>   
#>   
#> Restoration Reach Bootstrapped 90% Confidence Intervals  
#> Upper.CL Lower.CL  
#> Muck 8.175 3.810  
#> Clay 3.413 1.111  
#> ClayPieces 0.000 0.000  
#> HardClayBottom 9.365 1.984  
#> Silt 13.810 8.016  
#> Sand 21.857 12.889  
#> Gravel 17.778 7.460  
#> Cobble 22.159 10.429  
#> Boulder 5.079 0.794  
#> Bedrock 0.000 0.000  
#> Aquatic\_Bed 3.889 1.270  
#> Fern 0.238 0.000  
#> Leaf\_Litter 21.825 13.810  
#> Wood 6.984 3.889  
#> Fill\_Garbage 2.937 0.556

### Create the Graphic

We can also add those confidence intervals to the previous graphic. 

# An Alternative Approach

Practically, there are a few other approaches to testing statistical significance that one might apply. Each method asks and answers a slightly different questions about the data.

The simplest practical alternative is probably to use a contingency table to test frequency of observing habitat types, and test if frequencies are different between the two stream reaches. (After that I might consider hierarchical models, but that is overkill here.)

A contingency table does two things differently. First, it only determines if microhabitat frequencies *overall* are different (it does not tell you WHICH frequencies are different). And second, it only looks at how many times you observed each microhabitat in the 63 quadrats. It does not consider what the percent cover within each quadrat was.

In that sense it’s a less informative test, but it is complementary to the prior approaches, so it provides another way to check whether our qualitative conclusions depend on how we conduct the analysis.

## Contingency Table Analysis

Contingency tables are notoriously unhappy with cells with an “expected” number under the null hypothesis under 5 or so. As a result, we should pool the less common microhabitat types before analysis

We pool “Rare” habitats – those with fewer than 10 observations.

* ClayPieces
* HardClayBottom
* Boulder
* Bedrock
* Fern
* Fill\_Garbage

and end up with the following contingency table:

#> Reach  
#> Microhabitat Reference Restoration  
#> Muck 3 17  
#> Clay 3 10  
#> Rare 9 21  
#> Silt 36 35  
#> Sand 40 42  
#> Gravel 5 21  
#> Cobble 2 21  
#> Aquatic\_Bed 3 9  
#> Leaf\_Litter 51 47  
#> Wood 38 32

It is worth pointing out that some of the microhabitats that showed robust differences in the t test and resampling analyses show little difference in the number of plots in which they occurred. Look at Sand and Silt, for example.

Sand and Silt may be found (nearly) everywhere, but the abundance within plots is different between the two stream reaches. The different statistical approaches are sensitive to different aspects of the spatial arrangement of microhabitats.

#>   
#> Pearson's Chi-squared test  
#>   
#> data: mytab  
#> X-squared = 38.989, df = 9, p-value = 1.157e-05

So the relative abundance of different of microhabitats differs between the two stream reaches. It’s always nice when multiple statistical models lead you to the same conclusions.

# Final Notes

1. I was surprised that the results of the t-test were so similar to the results of the resampling test. That suggests it may not be unreasonable to use a t test in future comparisons (interpreting results with suitable humility and care). Relatively good performance of the t test is based on having a large sample size. If the number of quadrats drops, the t test will not perform as well.
2. Analyses could be repeated, taking into account the hierarchical sampling design. The effect would probably be to widen confidence intervals and reduce apparent statistical significance of comparisons. The analysis would also provide some insight into the spatial distribution of microhabitats within each stream reach. But hierarchical analysis significantly complicates things, for relatively small improvement in information. I would not expect the more complex analysis to alter the qualitative conclusions form this analysis. The added effort is probably only justified if greater precision is needed, for scientific publication, to address controversy, or to respond to a specific management question.
3. The purpose of collecting the microhabitat data is to understand how abundance of microhabitats changed following restoration, and how the distribution of microhabitats evolves post-restoration. That makes comparison of change over time within each stream reach arguably more important than a comparison of microhabitat abundances between the two reaches. The analysis just presented offers a preview of issues (and some possible solutions) for a change over time analysis. Correct analysis will also depend on details of the sampling design, and in particular, how sampling locations were selected each year.