TESTING THE RIVER CONTINUUM CONCEPT IN LAKE-STREAM NETWORKS IN SIERRA NEVADA, CALIFORNIA

Kasturi Bhatt (1009090877) and Anna Whitehouse (1009011837)
EEB313, Department of Ecology and Evolutionary Biology, University of Toronto
December 10, 2024

ABSTRACT

This paper provides a statistical analysis of a macroinvertebrate abundance data set from University of California researchers. It uses two different distributions to test whether macroinvertebrate functional feeding group (FFG) abundances significantly change between the beginning and end of rivers. We observe whether the data complies with the predictions made by the River Continuum Concept (RCC). Key findings are that a Poisson distribution produces a significant effect, whereas a Negative Binomial distribution does not produce a significant result.

INTRODUCTION

The RCC looks at expected abundances of benthic macroinvertebrates, and how they change along river gradients (see Figure 1). Macroinvertebrates are categorized into FFGs based on feeding habits, which are supported by differing biological factors along river gradients.

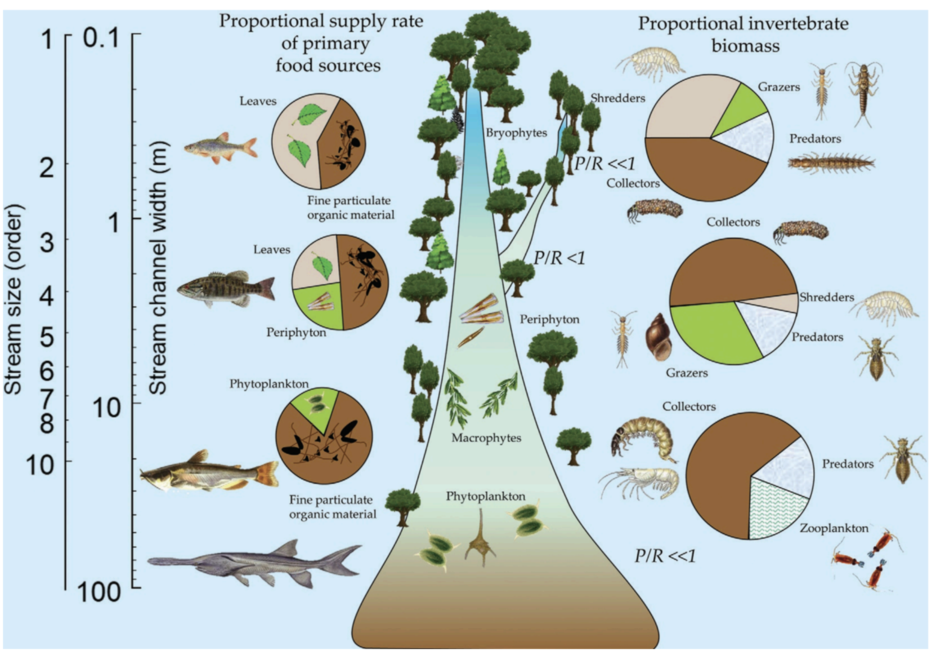

Figure 1 - River Continuum Concept. Source: Dodds and Maasri, 2022

In the Green et al. (2022a) study, the researchers sampled along river gradients in five distinct lake-stream networks in Sierra-Nevada, California. At each location, they identified species, counted species abundances, and mapped the sampling sites in the river.

We predict that FFG abundances will change in accordance with the RCC. Specifically, we predict that we will see an increase in Collector-Filterers, Collector-Grazers and Scrapers, decreases in Shredders, and no change in Predators between the beginning and end of the river. Our null hypothesis is that there is no change in FFG abundance between the beginning and end of the river. To reject our null hypothesis, we are looking for a statistically significant change in FFG abundance between the beginning and end of the river.

METHODS

Data Wrangling

The dataset from Green et al. (2022b) was used in combination with data from macroinvertebrates.org (n.d.), and the Benthic Macroinvertebrate Master Taxa List ("Benthic macroinvertebrate master taxa list", n.d.) to add FFG information to every species or genus. To simplify the data, we took out observations from river systems which had branched patterns, as this would not constitute a river continuum. We filtered out UN ("unidentified") and PI ("piercers") FFGs, because they did not fit with the groups described in the RCC, and we simplified feeding groups into five broad FFG categories.

We standardized the river sites as either being at the beginning or at the end of the river. The middle site was consistently grouped with the end group. Then, we combined the data points from all three river systems into one data set. See Appendix for data and code files.

Statistical Analysis

To look for changes in abundance between the beginning and end of the river systems, we plotted reaction norms for each FFG. In all FFGs, we see a different mean abundance value at the beginning of the river than at the end of the river. See Figure 2.

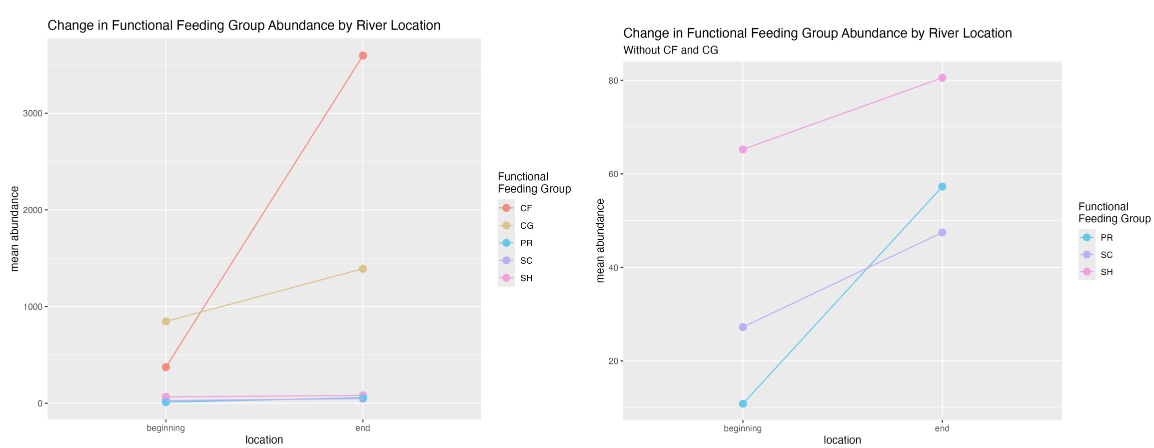

Figure 2 - Change in FFG Abundance by River Location. Mean abundance increases in all FFGs between the beginning and end of the river. Left - all FFGs. Right - low abundance FFGs.

To test for significance, we compared the outcomes of two distributions which look at the changes in parameters along the river for all the FFGs. First, we used a Poisson distribution, and then a Negative Binomial distribution to see if the results are significant. The Poisson distribution has one parameter, lambda, which represents the mean abundance. The Negative Binomial distribution has two parameters: size (variance), and mu (mean).

For both distributions, we used a maximum log likelihood evaluator function to find the maximum likelihood estimator (MLE). From this, we created 95% confidence intervals around the MLE. In the Negative Binomial distribution, given that it has two parameters, we estimated confidence intervals for both parameters separately by setting the other parameter to its MLE.

RESULTS

Reaction Norms

We used Figure 2 (above) to evaluate whether the changes in abundance along the river match the predicted changes. These results are summarized in Table 1. The results from the reaction norms do not satisfy our predictions.

Table 1 - Testing for Satisfaction of Predicted Changes in FFGs

FFG	Predicted Change	Result Change	Satisfied Prediction?
CF	increase	increase	yes
CG	increase	increase	yes
SC	increase	increase	yes
SH	decrease	increase	no
PR	no change	increase	no

Poisson Distribution

In Figure 3, we can see the confidence intervals for lambda do not overlap for any of the feeding groups. This indicates that they are significantly different between the beginning and the end of the river. This allows us to reject our null hypothesis under the Poisson distribution.

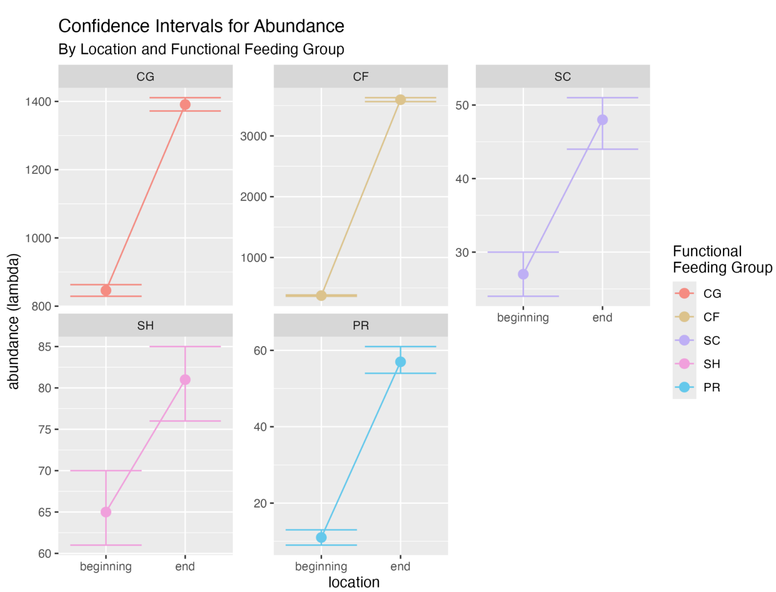

Figure 3 - Confidence Intervals for Abundance. Abundances from MLE for lambda with confidence intervals under a Poisson distribution for all FFGs.

Negative Binomial

In Figure 4, we can see that there is overlap for the confidence intervals of all FFGs for size, indicating that there is no significant difference. Similarly, in Figure 5, we see confidence interval overlap in all FFGs except in the Collector-Filterers and the Predators for mu, indicating that these are the only two groups that significantly change along the rivers. Given that there is so much overlap in confidence intervals, we cannot reject our null hypothesis under a Negative Binomial distribution.

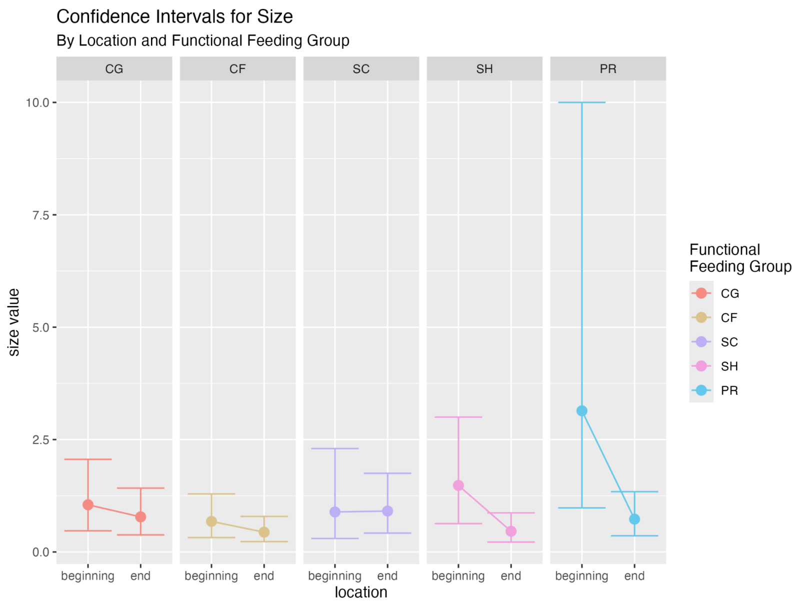

Figure 4 - Confidence Intervals for Size. Abundances from MLE for size with confidence intervals under a Negative Binomial distribution for all FFGs.

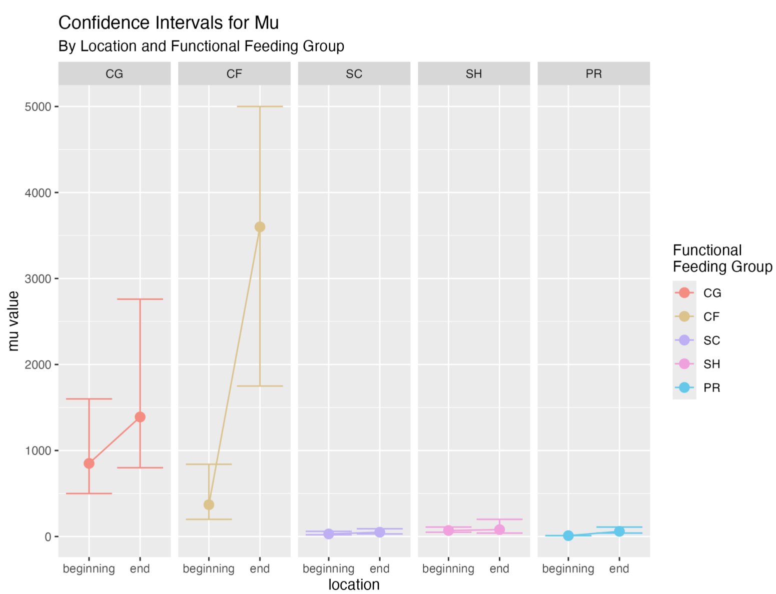

Figure 5 - Confidence Intervals for Mu. Abundances from MLE for mu with confidence intervals under a Negative Binomial distribution for all FFGs.

DISCUSSION

The different capacities of the two distributions to reject the null hypothesis are due to their differing assumptions.

The Negative Binomial looks at the effects of mean and variance separately. This allows us to see the more specific effects of mean and variance, rather than a cumulative effect. Additionally, given the small number of data points, fitting really specific parameters, like is done with the Negative Binomial, results in values that do not largely differ. These factors could inhibit the Negative Binomial from producing a significant result. However, in the size parameter, we would not expect a significant result, as there is no expectation that the variance would differ along the river gradient.

Conversely, the Poisson distribution finds the most representative value, lambda, to encompass both the mean and variance together, and this could limit its accuracy in predicting the mean, which is the variable of interest in our study. This could contribute to the Poisson distribution returning a significant result. Given the specificity of the Negative Binomial, and being able to select the most likely variance (size) and then optimize the mean (mu) estimation, we expect that this is more representative of our data than the Poisson distribution.

REFERENCES

Dodds W, Maasri A. 2022. The River Continuum Concept. In: Encyclopedia of Inland Waters. Vol. 2. 2nd ed. Elsevier. p. 237–243.

Benthic macroinvertebrate master taxa list. (n.d.). Retrieved November 19, 2024, from https://programs.iowadnr.gov/bionet/Inverts/Taxa

Green, MD, Anderson KE, Herbst DB, Spasojevic MJ. 2022a. Rethinking biodiversity patterns and processes in stream ecosystems. Ecological Monographs. 92(3). doi:10.1002/ecm.1520
.

Green MD, Anderson KE, Herbst DB, Spasojevic, MJ. 2022b. Species and environmental datasets from Sierra Nevada, CA (USA) streams in lake-stream networks [Dataset]. Dryad. https://doi.org/10.5061/dryad.2fqz612qw

Macroinvertebrates.org. (n.d.). Retrieved November 19, 2024, from
https://www.macroinvertebrates.org/

APPENDIX

Table A1 - Code Files and Descriptions

Code File Name	Description	Code File Link
1_DataWrangling	Data wrangling actions including assigning FFGs, selecting linear rivers, data tidying, binning based on river location and grouping by FFG.	https://github.com/EEB313/2024-GroupE/blob/main/1_DataWrangling.Rmd
2_ExploratoryDataAnalysis	Exploratory data analysis plotting, including mean proportion of FFG by site, and reaction norms.	https://github.com/EEB313/2024-GroupE/blob/main/2_ExploratoryDataAnalysis.Rmd
3_PoissonStatAnalysis	Statistical analysis under the Poisson distribution. Includes a log-likelihood estimator, MLE and confidence interval generation, and confidence interval plotting.	https://github.com/EEB313/2024-GroupE/blob/main/3_PoissonStatAnalysis.Rmd
4_NegBinomStatAnalysis	Statistical analysis under the Negative Binomial distribution. Includes a log-likelihood estimator, MLE and confidence interval generation, and confidence interval plotting.	https://github.com/EEB313/2024-GroupE/blob/main/4_NegBinomStatAnalysis.Rmd

Table A2 - Description of Dataset Columns

Dataset and Link	Dataset Description	Columns	Column Description
beginning_end.csv https://github.com/EEB313/2024-GroupE/blob/main/beginning_end.csv	Counts of FFGs by location (beginning or end), and river.	river	lake-stream network code
		location	site location (beginning or end); binned based on site number in "ffg_long_counts_new.csv"
		distinct_ffg	FFG category (in the five-category FFG classification system)
		total_count	count of benthic macroinvertebrate observances
BioNet-InvertebrateList.csv https://github.com/EEB313/2024-GroupE/blob/main/BioNet-InvertebrateList.csv	Benthic Macroinvertebrate Master Taxa List from the Iowa Department of Natural Resources ("Benthic macroinvertebrate master taxa list", n.d.). *Used in 1_DataWrangling code to match families with FFG information.*	bugID	ID number assigned to the given finalID taxon
		finalID	final taxonomic identification of the species
		taxaClass	Linnaean Class
		taxaOrder	Linnaean Order
		taxaFamily	Linnaean Family
		taxaSubFamily	Linnaean SubFamily
		tolerance	pollution tolerance coefficient

		ffg	functional feeding group
		cwInd	indication of whether the taxon is a cold water indicator species
		html_url	link to a datasheet on finalID taxon
ffg_long_counts.csv https://github.com/EEB313/2024-GroupE/blob/main/ffg_long_counts.csv	Counts of FFGs by trial, site and river. *Modified from a csv generated in 1_DataWrangling.* *Used in 1_DataWrangling to produce "final_mean_proportion_by_ffg_and_site.csv"*	river	lake-stream network code
		site	site number
		trial	trial number
		ffg	FFG category (in the nine-category FFG classification system)
		count	count of benthic macroinvertebrate observances
ffg_long_counts_new.csv https://github.com/EEB313/2024-GroupE/blob/main/ffg_long_counts_new.csv	Modified version of "ffg_long_counts.csv", with data points separated by empty rows. *Generated in 1_DataWrangling.* *Used in 1_DataWrangling to generate "beginning_end.csv".*	river	lake-stream network code
		site	site number
		trial	trial number
		ffg	FFG category (in the nine-category FFG classification system)
		count	count of benthic macroinvertebrate observances
final_mean_proportion_by_ffg_and_site.csv https://github.com/EEB313/2024-GroupE/blob/main/final_mean_proportion_by_ffg_and_site.csv	Mean proportion of FFG abundance in each site, in relation to total abundance in each selected river. *Generated in 1_DataWrangling code.* *Plotted in 2_ExploratoryDataAnalysis code to observe changes in FFG proportions along river gradients.*	river	lake-stream network code
		site	site number
		ffg	FFG category (in the nine-category FFG classification system)
		mean_proportion	mean proportion of FFG abundance at each site, in relation to total observance counts in the river
good.one.csv https://github.com/EEB313/2024-GroupE/blob/main/good.one.csv	Modified version of the Green et al. (2022b) species density update dataset. Pivoted so that sites are columns, and taxonomic groups are	Group	Taxon designation for the identified macroinvertebrate
		Family	manually added Linnean family corresponding to the "Group" taxonomic information

	rows. An additional column was added where families were designated to each taxon. *Used in 1_DataWrangling code to provide count data by river sampling site.*	columns 3:26 (10029 : 10056)	count of each macroinvertebrate group in the given site in an unidentified (*due to poor naming*) lake network
		columns 27:37 (CLS1_1 : CLS5_1)	count of each macroinvertebrate group in the given site in the Cascade Lake Network
		columns 38:58 (ELS2_1 : ELS8_3)	count of each macroinvertebrate group in the given site in the Evolution Lake Network
		columns 59:72 (Outlet.10477.trt.2003 :Outlet.Vidette.below. 2012)	count of each macroinvertebrate group in the given site in an unidentified (*due to poor naming*) lake network
		columns 73:84 (RLS1_1 : RLS6_2)	count of each macroinvertebrate group in the given site in an unidentified (*due to poor naming*) lake network
		columns 85:91 (YLS2_1 : YLS4_3)	count of each macroinvertebrate group in the given site in an unidentified (*due to poor naming*) lake network
		columns 92:127 (RCLS1.1 : RCLS8.7)	count of each macroinvertebrate group in the given site in the Rock Creek Lake Network