Using stepwise multiple regressions to analyze the relative impacts of land feature variables on benthic macroinvertebrate community metrics

Ian Sullivan and Declan McCabe

Department of Biology; St. Michael's College

Abstract

This study was undertaken to assess the impact of land features on benthic macroinvertebrate community metrics, using stepwise multiple regressions, in order to look for trends worthy of future research. Multivariate analyses have been used for generating predictive models and analyses when many variables are involved. Fourteen benthic community metrics were calculated from data collected over a four year period from 59 unique stream sites across Vermont. A GIS system was used to collect land feature data. I found that percent of catchment forested had a significant correlation with 7 of the 14 metrics analyzed, signifying the impact of land use on these metrics. The study looked at both correlation coefficients and beta-coefficients, which are both important to consider when assessing impact of a variable on metrics. Metrics with larger beta-coefficients did not necessarily have large correlation coefficients.

Background

•Benthic communities are influenced by complex relationships between organisms and their habitat. Geographically distinct catchments have differing macroinvertebrate communities (Kratzer *et al*, 2006)

•A combination of different land features together is measured when macroinvertebrate communities are assessed, making it difficult to assess their individual impacts (Sponseller et al, 2001)

•Multivariate statistical analyses are most useful when little is known about a particular habitat, and are a good starting point to generate testable hypothesis (Fore et al, 1996)

•Multiple regressions are equations based on a combination of correlations of individual components.

•The objective of this study was to compile a data set, spanning four years, of benthic macroinvertebrate communities in Vermont streams to generate a regression model that could identify possible geographic influences on these communities

Methods

•Samples collected during the summers of 2008-2011 from 59 stream sites throughout Vermont. Data from each site were sorted into lowest practical taxonomic unit.

• GIS data were collected during the summers of 2008 and 2009.

•I calculated 14 of the "best candidate benthic metrics" for measuring macroinvertebrate response to their environment (Barbour et al, 1999) for each of the 59 sites (Table 1, Table 2).

•Stepwise multiple regressions were generated to fit models in steps, first selecting the variable that has the greatest correlation, followed by the second greatest, and so on. The independent variables are then fit into a linear regression equation.

•All variables that were not already presented as a percentage or a ranking scale were modified using a correlation transformation to correct for the varying differences in magnitude between variables.

Results

Table 1: correlation coefficients (R^2) for each individual variable on a metric and for the total of all variables on each metric. Only correlations considered significant are included (p<0.05)

	Richness Measures					Composition Measures		Tolerance Measures			Feeding Measures		Habit Measures	
	Total Taxa	# EPT Taxa	# Ephemeroptera Taxa	# Plecoptera Taxa	# Trichoptera Taxa	% EPT	%Ephemeroptera	# Intolerant Taxa	% Tolerant Organisms	% Dominant Taxon	% Filterers	% Grazers and Scrapers	# Clinger Taxa	% Clingers
Agricultural Acres											0.118			
Percent Catchment Forested	0.32	0.49	0.455		0.16	0.445	0.267	0.427						
Upstream Distance Lake Pond (m)														0.138
Upstream Distance Dam (m)											0.052			
Upstream Distance Bridge (m)														0.076
Upstream Distance Culvert (m)					0.061									
Percent Catchment Highly Erodible Soils									0.086		0.074			
Stream Order			0.122								0.067			
E911 Structures per Acre										0.263				
E911 New 2008		0.05												
Aspect for 100m Stream Segment Buffer											0.084			
Dominant Bedrock Class	0.06	0.05			0.059			0.062						
Average Catchment Area Elevation (m)			.	0.311								0.129		
Monitoring Site Elevation (ft)								0.035			0.069		0.162	0.059
Total R^2	0.38	0.58	0.577	0.311	0.28	0.445	0.267	0.524	0.086	0.263	0.464	0.129	0.162	0.273

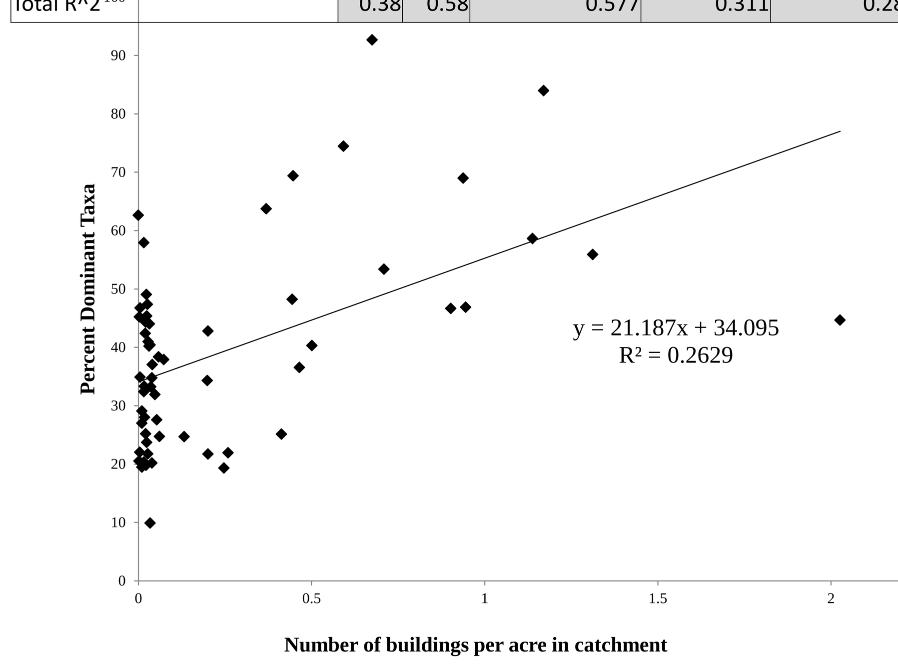

Figure 1: Correlation between building density in a catchment and proportion of specimens belonging to dominant taxa in that catchment

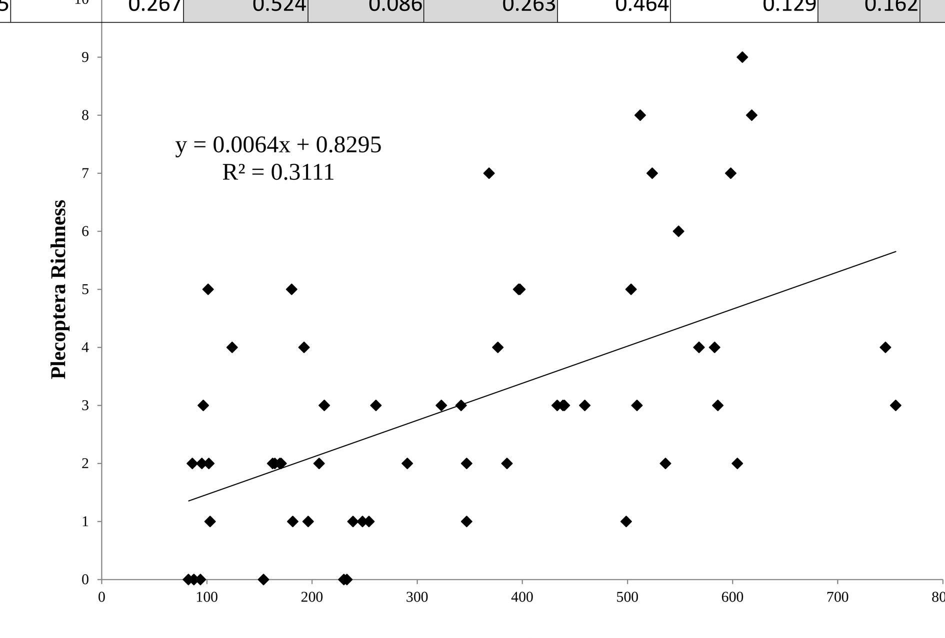

Figure 2: Correlation between average catchment elevation and Plecoptera richness found in that catchment

Table 4: summary of predictive models for each metric. Refer to equation 1 for the multiple regression equation.

Metric		Total Taxa	# EPT Taxa	# Ephemeroptera Taxa	# Plecoptera Taxa	# Trichoptera Taxa	% EPT	% Ephemeroptera	# Intolerant Taxa	% Tolerant Organisms	% Dominant Taxon	% Filterers	% Grazers and Scrapers	# Clinger Taxa	% Clingers
Y-intercept		16.064	4.247	-2.96	2.966	4.213	8.066	-0.676	2.536	1.608	39.564	35.423	17.651	106.576	52.774
Coefficients	Agricultural Acres											0.497			
	Percent Catchment Forested	0.632	0.818	0.617		0.453	0.667	0.529	0.546						
	Upstream Distance Lake Pond (m)														0.408
	Upstream Distance Dam (m)											-0.237			
	Upstream Distance Bridge (m)														-0.344
	Upstream Distance Culvert (m)					0.272									
	Percent Catchment Highly Erodible Soils									-0.294		-0.522			
	Stream Order			0.353											
	E911 Structures per Acre										0.513				
	E911 New 2008		0.259												
	Aspect for 100m Stream Segment Buffer											0.364			
	Dominant Bedrock Class	-0.252	-0.237			-0.254			-0.243						
	Average Catchment Area Elevation (m)				0.558								0.36		
	Monitoring Site Elevation (ft)								0.257			0.686		0.402	0.249

Discussion

The strong correlation between forested land and the metrics is in agreement with the results of Richards and Holt (1994), although other studies found that human influenced factors such as this do not have as great an impact as geological ones (Eyre et al, 2005). Also worth mentioning is the finding of a correlation coefficient of 0.311 between average catchment elevation and number of Plecoptera taxa (Table 1; Figure 3). This was the strongest correlation found that did not involve human influence. It makes sense that elevation would affect benthic communities because it would affect factors such as temperature. Also larger urban areas are more often located at lower elevations, so higher elevations would correspond with less of the influence associated with urban and agricultural land use. This further highlights the point that these variables should be considered in combination with each other.

Percent forest also had strong *B*-coefficients, e.g. 0.818 for number of EPT taxa (Table 2). B-coefficients show the magnitude of impact the variable has on the metric. Some variables that did not have strong correlations with a metric did have stronger impacts on the metric in my models; for example, monitoring site elevation with the percent filterers metric, in which the *B*-coefficient is 0.686 (Table 2), but the correlation coefficient is only 0.069 (Table 1). Therefore multivariate analyses provide a means to compare magnitudes of effect amongst variables to provide some perspective as to how much impact they really have on variation.

Works Cited

•Barbour, M.T.; Gerritsen, J.; Snyder, B.D.; Stribling, J.B. 1999. Rapid bioassessment protocols for use in streams and wadeable rivers periphyton, benthic macroinvertebrates, and fish. U.S. Environmental Protection Agency, Office of Water, Washington, DC.

•Eyre, M.D.; Pilkington, J.G.; McBlane, R.P.; Rushton, S.P. 2005. Macroinvertebrate species and assemblages in the headwater streams of the River Tyne, northern England in relation to land cover and other environmental variables. *Hydrobiologia.* 544: 229-240

•Fore, L.S.; Karr, J.R.; Wisseman, R.W. 1996. Assessing invertebrate responses to human activities: evaluating alternative approaches. *Journal of the North American Benthological Society.* 15:212-231.

•Kratzer, E.B.; Jackson, J.K.; Arscott, D.B.; Aufdenkampe, A.K.; Dow, C.L.; Kaplan, L.A.; Newbold, J.D.; Sweeney, B.W. 2006. Macroinvertebrate distribution in relation to land use and water chemistry in New York City drinking-water-supply watersheds. *Journal of the North American Benthological Society.* 25:
954-976

•Richards, C.; Host, G. 1994. Examining land use influences on stream habitats and macroinvertebrates: a GIS approach. *Water Resources Bulletin.* 30: 729-738

•Sponseller, R.A.; Benfield, E.F.; Valett, H.M. 2001. Relationships between land use, spatial scale and stream macroinvertebrate communities. *Freshwater Biology.* 46: 1409-1424

Acknowledgements

I would like to thank Vermont EPSCoR for providing the funding and resources necessary for completion of this study, and Saint Michael's College (Colchester, VT) for providing lab space and data analysis software. Thanks to Miranda Lescaze for her work coordinating the Streams Project, Lexie Haselton for managing all of the GIS data, and Professor Philip Yates (St. Michael's College) for his help with data transformation. I would also like to thank all of the interns who worked on the Streams Project between 2008 and 2011 for their help with sample collection and processing.