Examining the influence of natural site features on B-IBI response

August 2014

Elene Dorfmeier
King County Department of Natural Resources and Parks, Seattle, WA 98104

Objective

To be considered a reliable biological indicator, the Puget Lowland benthic index of biotic integrity (B-IBI), must respond predictably to gradients of human impact yet be insensitive to background natural variability. The responsiveness of B-IBI to human disturbance has been previously demonstrated for Puget Sound streams (Morley and Karr 2002; DeGasperi et al. 2009). An integrated measure of human disturbance, percent watershed urbanization, was recently used to recalibrate the Puget Lowland B-IBI and its component metrics (Fore et al. 2013). While human disturbance is the primary driver for B-IBI response, an important component of B-IBI recalibration process includes examining the influence of natural factors on B-IBI, and adjusting B-IBI scoring based on natural variability, if necessary.

Variation in benthic macroinvertebrate communities is likely a result of natural as well as anthropogenic factors, but it is unclear how sensitive B-IBI is to these possible sources of variation. The purpose of this analysis is to examine the influence of natural site features on B-IBI scores using hierarchical multiple regression and regression trees to identify variables with the strongest association to B-IBI response, in addition to quantifying how these variables contribute to B-IBI variability. Percent watershed urbanization is highly correlated with B-IBI scores (Pearson's R = 0.68) and represents a human disturbance gradient that integrates numerous human impact measurements and explains a majority of the observed B-IBI variance. After quantifying the variation due to percent watershed urbanization, we tested the explanatory power of other independent factors to B-IBI response. Other factors tested included variables of human disturbance, macroinvertebrate measurement variation associated with number of individuals within a sample, sampling month, and natural site features, including surficial geology and land cover. Regression analysis was chosen to estimate the influence of natural factors and other measures of human disturbance on B-IBI. Regression trees were used to graphically display how variables influence B-IBI and to corroborate results of the hierarchical multiple regression.

This analysis aims to:

1. Use hierarchical multiple regression analysis to estimate the additional explanatory power of natural site features on B-IBI after percent urbanization and other measures of human disturbance are accounted for.

2. Visualize data structure and examine B-IBI response to various measures of human disturbance, sample measurement, and natural site features using regression tree analysis.
3. Evaluate the need to adjust B-IBI scoring if a variable associated with natural site features is found to consistently drive B-IBI response. However, if natural site features do not significantly influence the B-IBI in a consistent, predictable manner beyond watershed urbanization, no additional changes to scoring will be necessary.

Methods

Dataset Preparation

Contributing watersheds were delineated for 745 macroinvertebrate sampling locations within the Puget Sound Basin; data were obtained from the Puget Sound Stream Benthos (PSSB) data management system. Site attributes such as elevation, slope and watershed metrics including 2006 National Land Cover data and surficial geology permeability were calculated (data layers and methods are summarized in Wilhelm et al. 2013). B-IBI scores were downloaded from the PSSB using the following search parameters: all streams; combined replicate handling; B-IBI score type 0-100; Fore, Wisseman (2012) taxa attributes; years 2006 – 2013; subsampled to 500 organisms per visit with no minimum. If data for multiple visits for a given site were available, the most recent B-IBI score at each site was used in the analysis.

Variable selection

Variables examined in this analysis included human disturbance variables: percent agriculture, population density, percent forest, and road density; Macroinvertebrate sample collection variation: month of collection and number of organisms collected within a sample; Natural site features: latitude, longitude, watershed area (ha), mean precipitation, mean elevation, stream density, stream length, surficial geologic permeability, and percent open water; and Land cover attributes: percent wetland, shrub, bare, and grass cover. Number of organisms collected within a sample is a macroinvertebrate measurement variable indicating the number of taxa identified by the laboratory within a given sample. Since the minimum number of organisms within a sample collection unit was not specified at download, we included this variable to quantify the amount of variation in B-IBI scores attributed to this laboratory measurement and to account for sample collection variability. In general, site samples collected with greater numbers of organisms have higher B-IBI scores.

Hierarchical multiple regression

All analyses were performed with R statistical programming (R Development Core Team 2013, version 3.0.2).

Hierarchical multiple regression was used to quantify the variance, or explanatory power, from each significant variable within the dataset. B-IBI shows a demonstrated negative response to the anthropogenic disturbance gradient, percent watershed urbanization (Pearson's R = 0.68) and, as such, this gradient is used as the primary predictor of B-IBI scores.

Natural site features in this analysis refer to environmental characteristics of a site that can be used to quantify natural variability in the absence of human activity. Because humans have modified and altered the environment for millennia, it is nearly impossible to separate natural variability in ecological environments from the impact of human activity (Brown 2002). Because of this, a number natural site features used in this dataset are highly confounded with urbanization (Table 1; see Appendix A for correlation matrix of all variables). Quantifying natural site feature variation alone, without incorporating human disturbance variables can be misleading due to the inherent multicollinearity within the dataset and ignores possible legacy effects from human activity. To account for confounding variables within the dataset and to quantify B-IBI's known response to anthropogenic disturbance, human disturbance attributes were deliberately placed into the hierarchical multiple regression first. Variables of natural site features and land cover were subsequently added, in blocks, to quantify how much additional variance could be explained by natural features when human disturbance was accounted for. By first quantifying the variation of human impact, additional explanatory power provided by measures of natural site factors can then be evaluated.

Pearson's product-moment correlation matrices were initially performed on the dataset to identify and eliminate highly correlated variables in the hierarchical multiple regression (correlation matrices reported in Appendix A). Original variables of human disturbance and sampling included: percent watershed urbanization, number of organisms identified in sample, sample collection month, percent agriculture, population density, percent forest, and road density. Natural site feature variables included: latitude, longitude, watershed area, precipitation, elevation, stream density, stream length, geologic permeability, and percent open water at a site. Land cover variables included: percent wetland cover, percent shrub cover, percent grass cover, and percent bare cover.

If two variables were found to be highly correlated ($R \geq 0.80$) then only one variable was used in subsequent regressions. Final variables used in the hierarchical multiple regression included human disturbance and macroinvertebrate sampling variables: percent agriculture, population density, month of collection, number of organisms; natural site features: latitude, longitude, mean precipitation, watershed area (ha), stream density, percent surficial geologic permeability, percent open water; and land cover attributes: percent wetland, barren, grass, shrubbery coverage (Fig. 1).

Table 1. Pearson product-moment correlation coefficient (Pearson's R) summary of percent urbanization to variables of human disturbance, macroinvertebrate sampling, natural site features, and land cover. P-values in bold indicate significant values. Month of sample collection not included in table.

	Variable	Pearson's R (urbanization)	P-value
Human Disturbance / Sampling	Number Organisms	0.06	0.100
	Agriculture	0.18	**<0.001**
	Population Density	0.93	**<0.001**
	% Forest	-0.93	**<0.001**
	Road Density	0.96	**<0.001**
Site Features	Latitude	-0.11	**0.010**
	Longitude	-0.09	**0.010**
	Watershed Area	-0.14	**<0.001**
	Mean Precipitation	-0.6	**<0.001**
	Mean Elevation	-0.48	**<0.001**
	Stream Density	0.09	0.890
	Stream Length	-0.15	0.270
	% Permeability	0	0.960
	% OpenWater	-0.04	0.280
Land Cover	% Wetland Cover	-0.24	**<0.001**
	% Shrub Cover	-0.51	**<0.001**
	% Bare Cover	-0.12	**<0.001**
	% Grass Cover	-0.38	**<0.001**

To estimate the explanatory power of individual variables on B-IBI response, the most complete regression model should be used that includes all relevant data available. The proportion of total variation, summarized by the adjusted R^2 value, was evaluated in each model and used to quantify how much additional variance was explained by single variable addition. F-tests and resulting p-values were used to determine the significance of single variable additions using the *add1* function. If a given variable was found to be significant with the *add1* function ($p \leq 0.05$) analysis of variance (ANOVA) was used to confirm the significance of the full model, or model with all possible predictor variables included, compared to the simple model without the added variable(s).

To perform the hierarchical multiple regression steps, a simple linear regression was first performed to model B-IBI response to percent watershed urbanization. Preliminary analysis confirmed variables associated with anthropogenic activity were highly correlated with urbanization (Table 1). Once the simple model (B-IBI as distributed by percent urbanization) was summarized, additional variables were tested for inclusion into the model, including all variables not attributed to natural site features: percent agriculture, population density, sampling month, and number of organisms. Significant variables were added into the regression and the amount of variance explained was recorded (adjusted R^2) (Fig. 1). Results were summarized and only significant variables were kept in the model for subsequent steps.

After adding all significant, non-natural explanatory variables into the regression model and summarizing variance, natural site feature variables were then tested (Figure 1). Natural site feature variables included: mean site precipitation, watershed area, stream length, superficial percent geologic (high) permeability, and percent open water. These variables were added into the model individually and tested for model inclusion as stated above. Next, additional groupings of land cover were added (percent wetland, barren, grass, shrub cover), one variable at a time, using the same process, to evaluate increases in variation explained from the previous model (Figure 1).

To graphically explore B-IBI response to natural site features, data for each variable were binned into thirds (low, medium, and high values) and graphed as box plots. To account for the confounding nature of human impact and natural site features, only low urbanization sites were used (< 10% watershed urbanization) to produce plots. Ranges of B-IBI response were then visualized at three levels for each variable. If B-IBI was responding to natural site features, response should be visualized with the graphs.

To summarize, Figure 1 graphically demonstrates the hierarchical multiple regression steps, which include:

1) B-IBI response was modeled to percent watershed urbanization and explanatory power of urbanization quantified (adjusted R^2).
2) Regression model from Step 1 was used to test for inclusion of additional human disturbance variables and benthic macroinvertebrate sample collection (month and number of organisms). The model was summarized and explanatory power recorded (adjusted R^2).
3) Regression model from Step 3 was used to test the contribution of natural site features and geology variables. Model was summarized and explanatory power recorded.
4) Regression model from Step 4 was used to test the significance of land cover variables. Model was summarized and explanatory power recorded.
5) Total explanatory power from human disturbance and sample measurement variables (Step 2; adjusted R^2) was subtracted from the adjusted R^2 in Step 3 and 4. This is the estimated additional explanatory power of the available natural site features: site attributes, geology, and land cover to the B-IBI response.

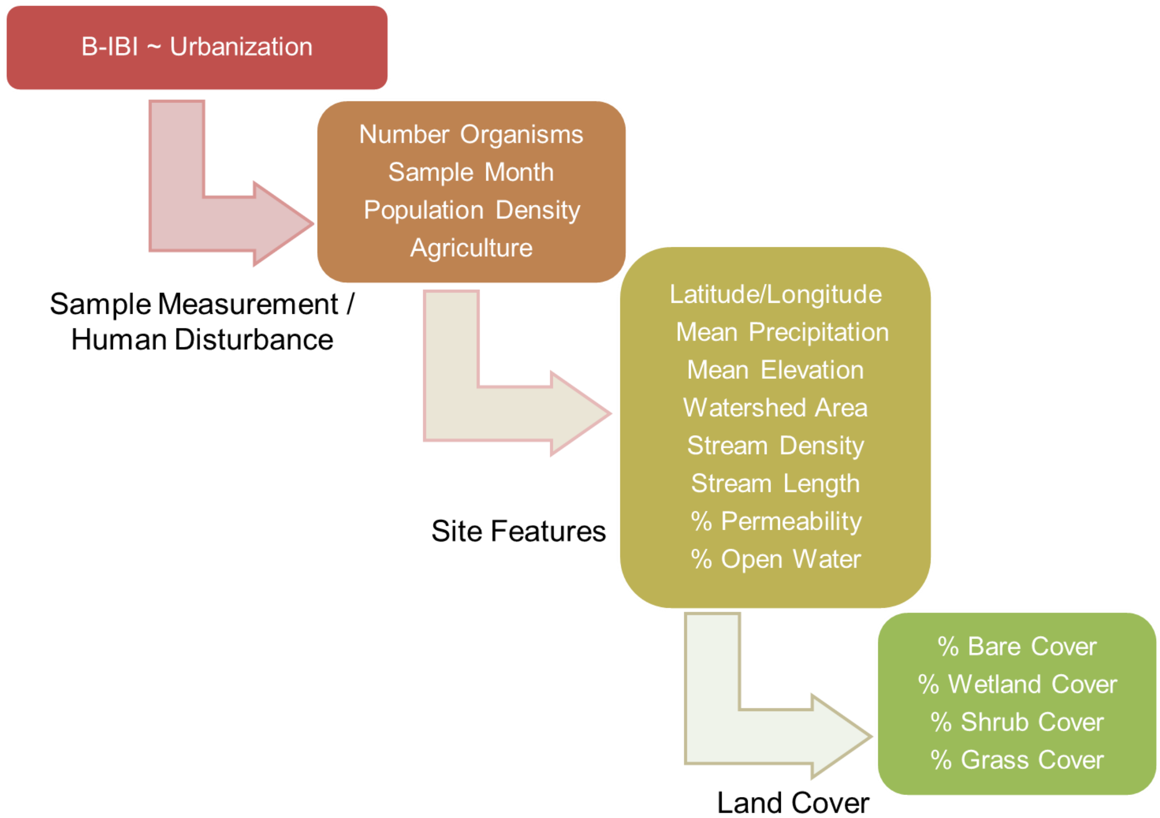

Figure 1. Analysis flow chart of the hierarchical multiple regression analysis, including all variables within each category of human disturbance, macroinvertebrate sampling measurement, and natural site features, including land cover. Analysis began with modelling B-IBI response to percent watershed urbanization. Human disturbance variables were then tested for significant contribution to B-IBI response. Then, sequential addition of sample collection attributes, natural site features, and land cover variables were added to the model and tested for significance of B-IBI response. Explanatory power was quantified in each step. Note: all variables seen in the diagram were not used due to confounding. Final variables included variables of human disturbance: percent agriculture, population density; macroinvertebrate sampling variables: month of collection and number of organisms; natural site features: latitude, longitude, mean precipitation, watershed area (ha), stream density, percent surficial geologic permeability, and percent open water; and land cover attributes: percent wetland, barren, grass, shrubbery coverage.

Regression tree

Linear regression models, such as the hierarchical multiple regression above, are commonly used to interpret complex datasets. It is a powerful tool that can be used to determine which predictors influence the response variable. They are ideal tools to use when explanatory variables have separate, strictly additive effects on the response variable. When datasets, such as this one, include a large amount of variables, contain high level interactions, and multicollinearity between variables, regression results can be difficult to interpret. An alternative approach, regression tree analysis, was performed to supplement the findings of the hierarchical multiple regression and to better visualize predictor influence on B-IBI response.

Regression tree analysis, using recursive partitioning, is known for its simplicity and efficiency in dealing with datasets containing large numbers of predictor variables and sample sizes. Trees can handle nonparametric datasets, high level interactions between variables, and are easily interpretable (De'ath and Fabricius 2000; Strobl et al. 2009). Regression trees complement regression analysis with the visualization of data variables and represent the structure of the data in a simple and interpretable way. These regression trees can aid in the description, categorization, and generalization of data and, in this analysis, regression trees were used to aid in the identification of the most influential factors of B-IBI response and characterize the influence of natural site features on B-IBI.

Regression trees using recursive partitioning split observations based on the predictor variable resulting in the greatest change in explained deviance. With repeated data splitting into increasingly homogeneous groups using combinations of explanatory variables, regression trees assist in explaining sources of variation of a single response. To develop regression trees, the dataset is partitioned into two groups, each as homogeneous as possible. The predictor variable with the strongest association with the response variable, B-IBI, is selected first after testing association with every predictor variable within the dataset. The process is then applied separately to each partitioned sub-group recursively until the sub-groups either reach a specified minimum size or no improvement to the tree can be made.

Each terminal node, or leaf, of the regression tree represents a cell of the partition and contains a simple regression model that applies to that node. In contrast to the linear regression procedure performed above, where effects of the predictor variables are additive and interactions of the variables within the dataset may be hard to tease apart, the interactions among variables in regression tress are easily discernable by graphing.

Regression trees were produced with the *rpart* library. All predictive variables were included for recursive partitioning. The tree was produced with a specified minimum of 30 observations in each node and a decrease in the overall lack of fit by a cost complexity parameter (CP) of 0.001 before a split was attempted. CP is a measure similar to other cost statistics (Neter et al. 1996) which adds a penalty for increasing the number of parameters used in a model and is the standard criterion used in pruning trees, associated with the smallest cross-validated error. The CP associated with the smallest cross-validated error was chosen to produce the final tree for analysis (see Therneau and Atkinson 2014 for more information on tree selection criteria in *rpart*).

Results

Hierarchical multiple regression

The final dataset for hierarchical multiple regression included variables of human disturbance: percent agriculture, population density; macroinvertebrate sampling variables: month of collection and number of organisms; natural site features: latitude, longitude, mean precipitation, watershed area (ha), stream density, percent surficial geologic permeability, and percent open water; and land cover attributes: percent wetland, barren, grass, shrubbery coverage.

Results for each step of the hierarchical multiple regression were summarized (Table 2). The variance explained by the initial model (B-IBI response as distributed by percent watershed urbanization) was 0.49 (adjusted R^2) (Table 3; Fig. 2). When all significant human disturbance variables and measurement data variables were added into the model (B-IBI ~ urbanization + number of organisms), the explanatory power increased by 0.03 (adjusted $R^2 = 0.52$; Table 3; Fig. 2). Total variation explained in the final model, which incorporated all significant human disturbance, measurement error, and natural site feature variables (B-IBI ~ urbanization + number of organisms + latitude + longitude + watershed area + precipitation + stream density), was an adjusted R^2 of 0.56, an increase of 0.04 (Table 3; Fig. 2). The addition of location (latitude and longitude) alone contributed 2% explanatory power to the model, while the rest of the natural feature variables contributed the other 2% (Fig.2). No variables of superficial geologic permeability or land cover variables were found to significantly contribute to B-IBI response.

Once human disturbance and measurement error contribution to B-IBI were accounted for, the additional explanatory power introduced from natural site features (latitude, longitude, precipitation, watershed area, stream density) into the regression was minimal (4%). Site location (latitude / longitude) alone compromised 2% of this variability.

Box Plots

Selecting sites with low urbanization was another attempt to control for the overwhelming influence of urbanization in order to look at the potential influence of other natural features. Out of 745 sites, 301 were determined to have <10% watershed urbanization. Natural site features were graphed as box plots (mean precipitation, watershed area (ha), stream density, percent surficial geologic permeability, percent open water; and land cover attributes: percent wetland, barren, grass, shrubbery coverage) (Appendix B). Only two natural site feature variables, mean precipitation and mean elevation, were associated with differences in B-IBI scores at low urbanization sites (< 10%) (Fig. B1a; B1b). All other natural site feature variables graphed showed little change in B-IBI scores at low urbanization (Appendix B).

Table 2. Summary of hierarchical multiple regression steps with associated F-statistic and p-values for the single variable additions into the initial model (*B-IBI ~ percent watershed urbanization*). Bold values highlight significant terms ($p \leq 0.05$).

Added Variables to Model	F-stat	P-value
Human Disturbance and Measurement Error		
Number Organisms	42.44	**<0.001**
Month	1.29	0.27
Population Density	3.59	0.06
Agriculture	0.08	0.77
Added significant term to model. New model: B-IBI ~ Urbanization + Number Organisms		
Natural Site Features		
Latitude	8.72	**0.003**
Longitude	21.26	**<0.001**
Watershed Area	15.33	**<0.001**
Mean Precipitation	8.63	**0.00**
Stream Density	5.86	**0.02**
% Permeability	0.00	0.96
% Open Water	0.02	0.89
Added significant terms to model. New model: B-IBI ~ Urbanization + Number Organisms + Latitude + Longitude + Watershed Area + Mean Precipitation + Stream Density		
Land Cover		
% Bare Cover	0.00	0.96
% Grass Cover	3.44	0.06
% Shrub Cover	0.01	0.93
% Wetland Cover	0.07	0.79
No significant land cover variables.		
Final Model		
B-IBI ~ Urbanization + Number Organisms + Latitude + Longitude + Watershed Area + Mean Precipitation + Stream Density		

Table 3. Model summary and adjusted R^2 for each hierarchical multiple regression step (steps outlined in Fig. 1). NA = not applicable; adjusted R^2 did not increase.

Hierarchical Multiple Regression Steps	Final model from each step	Adjusted R^2
Step 1: Intial Regression	B-IBI ~ Urbanization	0.49
Step 2: Human Disturbance and Measurement Error	B-IBI ~ Urbanization + Number Organisms	0.52
Step 3: Natural Site Features	B-IBI ~ Urbanization + Number Organisms Latitude + Longitude + Watershed Area + Mean Precipitation + Stream Density	0.56
Step: 4: Land Cover	Added variables non-significant	NA

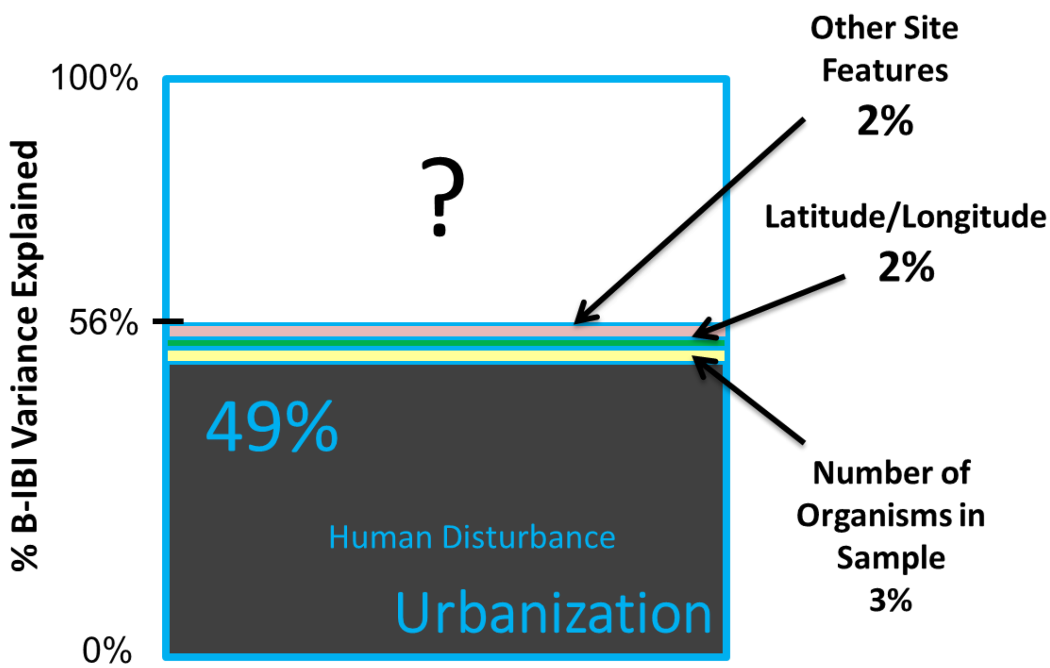

Figure 2. Percent of B-IBI variance explained by variables examined. Human disturbance variables alone explain 49% of B-IBI variance, followed by number of organisms collected per sample (3%), site location (latitude, longitude) (2%), and all other significant natural site features added to the regression model (area, precipitation, stream density) (2%). Total B-IBI variation explained by the hierarchical multiple regression was 56%.

Regression tree analysis

The final regression tree was pruned to a CP of 0.012 resulting in 8 nodes (relative error = 0.39; relative error (x) = 0.45) (Fig. 3). The most important determinant of B-IBI response was percent watershed urbanization. The dataset was split initially at a threshold of 44% urbanization (Fig. 4). For sites < 44% urbanization, the next determining variable was mean precipitation (mm). Sites with > 1271 mm of precipitation are then affected by site location. Sites with high precipitation and located in areas of > -122 longitude and < 48 latitude had the highest average B-IBI scores (74). Sites < 1271 mm of precipitation were then influenced by number of organisms in a sample. Sites with precipitation < 1271 mm and with > 276 organisms in a sample, had an average B-IBI of 53. Samples with < 276 organisms in a sample had an average B-IBI of 26.

Regression tree analysis validated the influence of precipitation and site location on B-IBI response, as well as another variable: percent forest cover. Percent forest cover was shown to influence B-IBI scores at low urbanization sites (<44%). Sites with ≥ 44% urbanization and < 9.7% percent forest cover had the lowest average B-IBI scores (14). Highly urbanized sites with over 9.7% forest cover were then influenced by precipitation. Sites with > 1134 mm precipitation had average B-IBI scores of 47; sites < 1134 mm precipitation had average B-IBI scores of 28.

Comparing analyses

The importance of percent watershed urbanization to B-IBI scores was validated by both hierarchical multiple regression and regression tree analysis: urbanization was the first partition in the regression tree and urbanization explained the most variability of B-IBI relative to all other variables tested in the hierarchical multiple regression. In the regression tree analysis, variables of percent forest cover, mean precipitation, latitude / longitude and number of organisms were determined to be influential to B-IBI response. Variables not included in the tree (population density, road density, percent agriculture, watershed area, elevation, stream density, stream length, geologic permeability, percent open water, and land cover variables) did not contribute to the model. Significant variables in the hierarchical multiple regression model included site location (latitude and longitude), measurement error (number of organisms), mean precipitation, watershed area, and stream density. All other variables were not found to be significant to B-IBI response. Due to the high multicollinearity of percent forest cover and urbanization (Pearson's R = -0.93), it was not included in the hierarchical multiple regression and could not be compared to the regression tree results.

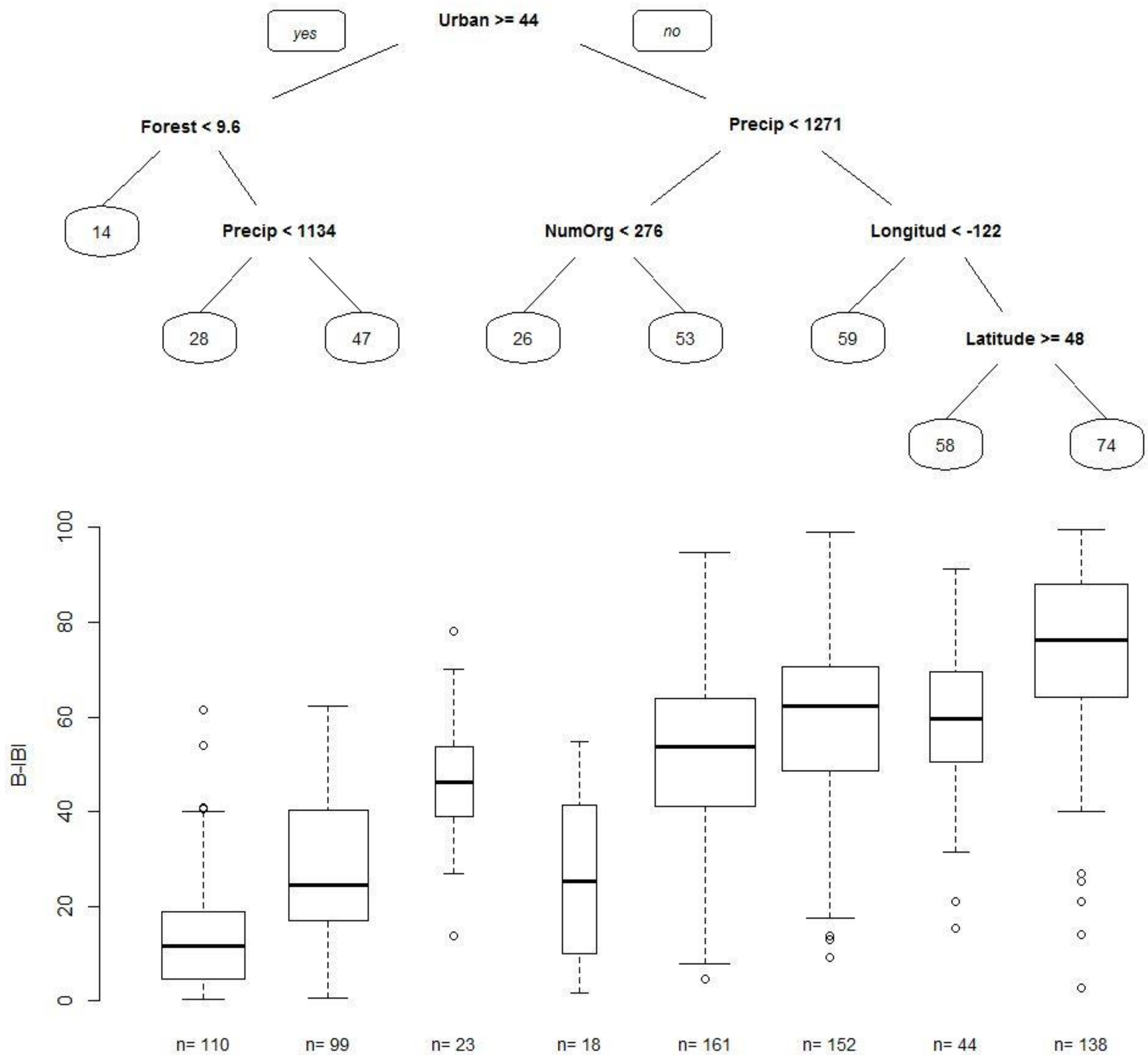

Figure 3. Regression tree of predictor variables for B-IBI response. Numbers at the top of each node represent the threshold where data partitions. Circled labels under each node indicate average B-IBI scores for the selected partition. Box plots below the tree show the distribution of B-IBI scores for each node of the tree and the number of sites in each node indicated by *n*. The widths of the box plots correspond to sample size within the nodes. Variables not included in the tree did not contribute to the model. Urban = urbanization; Forest = % forest; Precip = mean precipitation; NumOrg = Number of organisms within macroinvertebrate sample collection; Longitud = Longitude

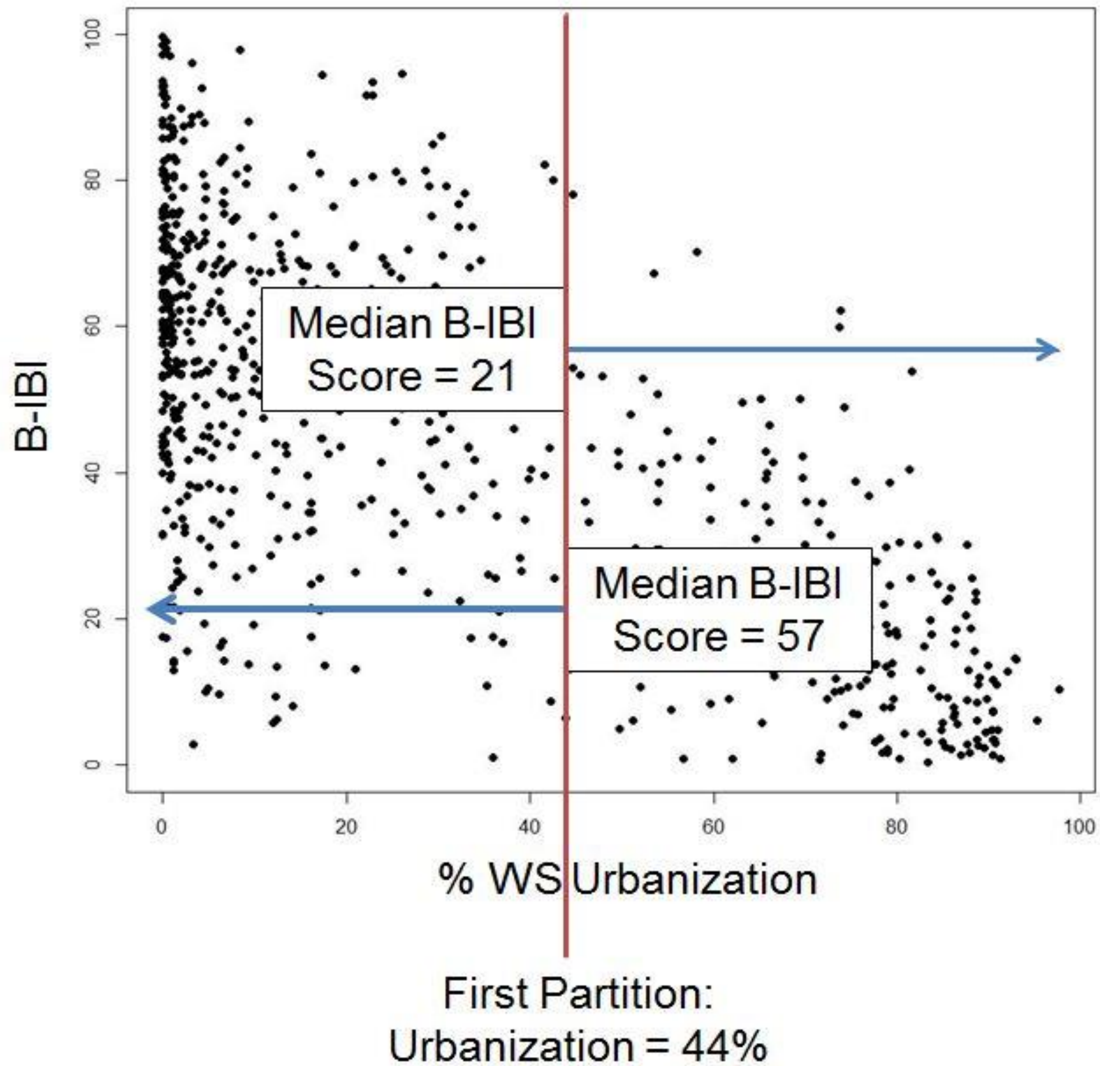

Figure 4. Graph of the first data partition in the regression tree illustrating how data is partitioned. B-IBI site scores are graphed by percent watershed urbanization. The first data partition occurs at the threshold of 44% watershed urbanization (red line). Sites with over 44% urbanization have a median B-IBI score of 21, while sites less than 44% urbanization have a median B-IBI score of 57. Data is subsequently split further using other predictor variables within the dataset. See Figure 2.

Conclusions

1. The primary driver and best predictor of Puget Sound regional B-IBI scores is percent watershed urbanization.
2. Natural site features, including land cover and geology, added minimal explanatory power to B-IBI response.
3. Results of this study suggest there is no need to modify B-IBI scoring to account for variability associated with natural site features.

Both hierarchical multiple regression and regression tree analyses confirm percent watershed urbanization was the most influential explanatory variable for B-IBI response. Although urbanization alone did not account for all of B-IBI response (49%), the additional variation explained by natural site features evaluated in this study was minimal (4%). A variable of B-IBI measurement error, number of macroinvertebrate organisms collected in a sample, was shown to add 3% additional variance to B-IBI response.

Two statistical methods were used - hierarchical multiple regression and regression tree analysis - to explore the influence of natural variability on B-IBI response using various predictor variables of human disturbance, sample measurement, site location, natural site features, and land cover. Hierarchical multiple regression, while a common and useful tool to explore data, was challenging to interpret due to possible interactions among variables and inherent multicollinearity of the dataset. Using the two methods, results from each were confirmed and explanatory variables were characterized (human disturbance, sample measurement, location, natural site features and land cover) based on influence to B-IBI to determine if adjustments to B-IBI scoring were necessary.

Generally, B-IBI scores tended to slightly increase with higher mean site precipitation, elevation, and slope (see Appendix B). Elevation is highly correlated with precipitation ($R = 0.85$) and both elevation and precipitation are correlated with percent watershed urbanization (elevation $R = -0.48$, precipitation $R = -0.60$). Due to the strong interrelationships of percent forest cover, mean site precipitation, elevation, and slope with watershed urbanization, this increase in B-IBI scores is not surprising, but it is difficult to determine if the variables influence on B-IBI can be separated from human disturbance with great confidence. The low amount of variance explained by these variables alone do not suggest that B-IBI scores are being consistently influenced by natural site features, while the strength of the human disturbance gradient, percent watershed urbanization is strong and predictable. Due to the low explanatory power of natural site features (2% excluding latitude and longitude), these analyses suggest that the natural variables in this dataset does not impact B-IBI scoring relative to percent watershed urbanization.

The variation explained by percent urbanization exhibits a strong and consistent relationship with the B-IBI (Morley and Karr 2002; DeGasperi et al. 2009). Watershed ecosystems are intrinsically complex with inherent natural variability that may be difficult to quantify. Effort was made to gather data from various available sources to capture the environmental complexity of the Puget Sound Basin, although important variables may still be absent or unaccounted for.

Some additional variability sources that may not be captured within this dataset may include field sampling differences (habitat targeted, depth of disturbance, thoroughness of rock cleaning, etc.), taxonomic laboratory processing, macroinvertebrate community distribution, microhabitat complexity, legacy effects of land use at sites (Harding et al. 1998), and/or unmeasured human influences (e.g., water quality, hydrology, sedimentation).

Of the 22 variables examined which contain variables of human disturbance, measurement error, natural site features including land cover and surficial geology, natural variability associated with site features added little to the predictability of B-IBI response. As such, we do not recommend adjusting the B-IBI scoring to account for the natural site features included in this analysis. The additional sources of variability within the B-IBI scoring are unclear and likely include a number of natural, laboratory, and/or anthropogenic variability which we are unable to account for here due to the synergistic relationships of natural variables within complex watershed ecosystems.

References

Bailey, R.C., Norris, R.H., Reynoldson, T.B., 2004. Bioassessment of Freshwater Ecosystems using the Reference Condition Approach. Kluwer Academic Publishers, Boston.

Bonada, N., N. Prat, V. H. Resh, B. Statzner. 2006. Developments in aquatic insect biomonitoring: a comparative analysis of recent approaches. Annual Review of Entomology 51: 495-523.

Brown, A.G. 2002. Learning from the past: Paleohydrology and paleoecology. Freshwater Biology 47:817-829.

Cohen, J. and Small, C. 1998. Hypsographic demography: The distribution of human population by altitude. Proceedings of the National Academy of Sciences, USA. 95:14009-14014.

De'ath, G. and Fabricius, K. 2000. Classification and regression trees: A powerful yet simple technique for ecological data analysis. Ecology 81:3178-3192.

DeGasperi, C. L., H. B. Berge, K. R. Whiting, J. J. Burkey, J. L. Cassin, R. R. Fuerstenberg. 2009. Linking hydrologic alteration to biological impairment in urbanizing streams of the Puget Lowland, Washington, USA. *Journal of the American Water Resources Association* 45(2): 512–533.

Diefenderder, H, R. M. Thom, G. E. Johnson, J. R. Skalski, K. A. Vogt, B. D. Ebberts, C. Roegner, E. M. Dawley. 2011. A levels-of-evidence approach for assessing cumulative ecosystem response to estuary and river restoration programs. Ecological Restoration 29:111-132.

Fore, L., Wissemann R., Wilhelm J. O., Lester D., Adams K., et al. 2013. Using natural history attributes of stream invertebrates to measure stream health. Puget Sound Stream Benthos B-IBI Recalibration Documents and Materials. http://pugetsoundstreambenthos.org/Projects/EPA_Grant_2010/TechDocs/Final/Invertebrate_Attributes_2013_v2.pdf

Harding, J. S., E. F. Benfield, P. V. Bolstad, G. S. Helfman, E. B. D. Jones, 1998. Stream biodiversity: The ghost of land use past. Proceedings of the National Academy of Sciences, USA. 95:14843-14847.

Neter, J., Kutner, M.H., Nachtscheim, C.J., Wasserman, W. 1996. Applied linear statistical models. 4th.ed. Irwin: McGraw-Hill.

Morley, S. A. and J. R. Karr. 2002. Assessing and restoring the health of urban streams in the Puget Sound basin. *Conservation Biology.* 16(6):1498-1509.

Therneau ,T. and Atkinson, E. 2014. An introduction to recursive partitioning using the RPART routines. Minnesota Mayo Foundation. http://cran.r-project.org/web/packages/rpart/vignettes/longintro.pdf

Strobl, C., Malley, J., Tutz, G. 2009. An introduction to recursive partitioning: Rationale, application and characteristics of classification and regression trees, bagging and random forests. Psychology Methods 14:323-348.

Wilhelm, J. O., P. Leinenbach, L. Fore, D. Lester, K. Adams, G. Hayslip, 2013. Watershed delineation and land cover calculations for Puget Sound stream basins. King County Water and Land Resources Division, Seattle, Washington. http://pugetsoundstreambenthos.org/Projects/EPA_Grant_2010/Data/GIS_Memo.pdf

Appendix A. Correlation Matrices

Table A1. Pearson's R correlation matrix of human disturbance variables.

HUMAN DISTURBANCE	Urban	Agriculture	Road Density	Population Density	% Forest	B-IBI
Agriculture	0.18		0.21	0.04	-0.49	-0.13
Watershed Area	-0.14	-0.1	-0.16	-0.13	0.14	0.03
% Bare Cover	-0.12	-0.09	-0.15	-0.13	0.09	0.11
B-IBI	-0.7	-0.13	-0.68	-0.67	0.65	
Mean Elevation	-0.48	-0.48	-0.51	-0.4	0.59	0.41
% Forest	-0.93	-0.49	-0.91	-0.81		0.65
% Grass Cover	-0.38	-0.09	-0.37	-0.37	0.27	0.23
Latitude	-0.1	-0.13	-0.11	-0.07	0.13	-0.01
Longitude	-0.09	-0.09	-0.12	-0.1	0.1	0.18
% Open Water	-0.04	0	-0.06	-0.05	0.03	0.03
% Permeability (High)	0	0.1	-0.03	-0.06	-0.06	0.01
% Permeability (Low)	0	-0.1	0.03	0.06	0.06	-0.02
Population Density	0.93	0.04	0.92		-0.81	-0.67
Mean Precipitation	-0.6	-0.48	-0.61	-0.51	0.69	0.48
Road Density	0.96	0.21		0.92	-0.91	-0.68
% Shrub Cover	-0.51	-0.26	-0.5	-0.46	0.44	0.34
Stream Density	0.09	0.06	0.08	0.07	-0.09	-0.11
Stream Length	-0.15	-0.09	-0.16	-0.13	0.14	0.03
Urbanization		0.18	0.96	0.93	-0.93	-0.7
% Wetland Cover	-0.24	0.19	-0.23	-0.28	0.07	0.16

Table A2. Pearson's R correlation matrix of natural site feature variables.

NATURAL SITE FEATURES	Watershed Area	Mean Elevation	Mean Precipitation	Stream Density	Stream Length	% Open Water	% Permeability (High)	Latitude	Longitude
Agriculture	-0.1	-0.48	-0.48	0.06	-0.09	0	0.1	-0.13	-0.09
Watershed Area		0.3	0.23	0.01	1	0.14	0.01	0.03	0.07
% Bare Cover	0.2	0.23	0.24	-0.09	0.19	0	0.11	0.04	0.16
B-IBI	0.03	0.41	0.48	-0.11	0.03	0.03	0.01	-0.01	0.18
Mean Elevation	0.3		0.85	-0.02	0.29	0	-0.17	0.13	0.36
% Forest	0.14	0.59	0.69	-0.09	0.14	0.03	-0.06	0.13	0.1
% Grass Cover	0.07	0.17	0.16	-0.08	0.07	0	0.11	-0.06	-0.01
Latitude	0.03	0.13	0.23	-0.02	0.02	0.03	-0.18		0.13
Longitude	0.07	0.36	0.24	0.01	0.07	-0.02	-0.01	0.13	
% Open Water	0.14	0	0.02	0.08	0.16		-0.01	0.03	-0.02
% Permeability (High)	0.01	-0.17	-0.09	-0.03	0.01	-0.01		-0.18	-0.01
% Permeability (Low)	-0.02	0.17	0.08	0.03	-0.03	-0.06	-1	0.18	0.01
Population Density	-0.13	-0.4	-0.51	0.07	-0.13	-0.05	-0.06	-0.07	-0.1
Mean Precipitation	0.23	0.85		-0.03	0.23	0.02	-0.09	0.23	0.24
Road Density	-0.16	-0.51	-0.61	0.08	-0.16	-0.06	-0.03	-0.11	-0.12
% Shrub Cover	0.18	0.4	0.43	-0.06	0.18	0.03	0.03	0.14	0.11
Stream Density	0.01	-0.02	-0.03		0.03	0.08	-0.03	-0.02	0.01
Stream Length	1	0.29	0.23	0.03		0.16	0.01	0.02	0.07
Urbanization	-0.14	-0.48	-0.6	0.09	-0.15	-0.04	0	-0.1	-0.09
% Wetland Cover	0.03	-0.19	-0.09	0.01	0.04	0.12	0.18	0.08	0.01

Table A3. Pearson's R correlation matrix of variables associated with sample collection.

SAMPLE COLLECTION	Month	Number Organisms
Agriculture	-0.13	0.05
Watershed Area	0.01	0.1
% Bare Cover	-0.04	0.03
B-IBI	0.02	0.21
Mean Elevation	0.15	0.02
% Forest	0.09	0.03
% Grass Cover	-0.01	0.07
Latitude	0.19	-0.04
Longitude	0.08	0.03
% Open Water	0.00	0.07
% Permeability (High)	0.01	-0.01
% Permeability (Low)	-0.01	0
Population Density	0.03	-0.08
Mean Precipitation	0.14	0.01
Road Density	-0.03	-0.07
% Shrub Cover	0.07	-0.01
Stream Density	0.05	0.04
Stream Length	0.01	0.1
Urbanization	-0.05	-0.06
% Wetland Cover	-0.08	0.1

Table A4. Pearson's R correlation matrix of land cover variables.

LAND COVER	% Bare Cover	% Grass Cover	% Shrub Cover	% Wetland Cover
Agriculture	-0.09	-0.09	-0.26	0.19
Watershed Area	0.2	0.07	0.18	0.03
% Bare Cover		0.17	0.18	-0.02
B-IBI	0.11	0.23	0.34	0.16
Mean Elevation	0.23	0.17	0.4	-0.19
% Forest	0.09	0.27	0.44	0.07
% Grass Cover	0.17		0.35	0.18
Latitude	0.04	-0.06	0.14	0.08
Longitude	0.16	-0.01	0.11	0.01
% Open Water	0	0	0.03	0.12
% Permeability (High)	0.11	0.11	0.03	0.18
% Permeability (Low)	-0.11	-0.11	-0.04	-0.18
Population Density	-0.13	-0.37	-0.46	-0.28
Mean Precipitation	0.24	0.16	0.43	-0.09
Road Density	-0.15	-0.37	-0.5	-0.23
% Shrub Cover	0.18	0.35		0.16
Stream Density	-0.09	-0.08	-0.06	0.01
Stream Length	0.19	0.07	0.18	0.04
Urbanization	-0.12	-0.38	-0.51	-0.24
% Wetland Cover	-0.02	0.18	0.16	

Appendix B. Natural Variable Boxplots with B-IBI

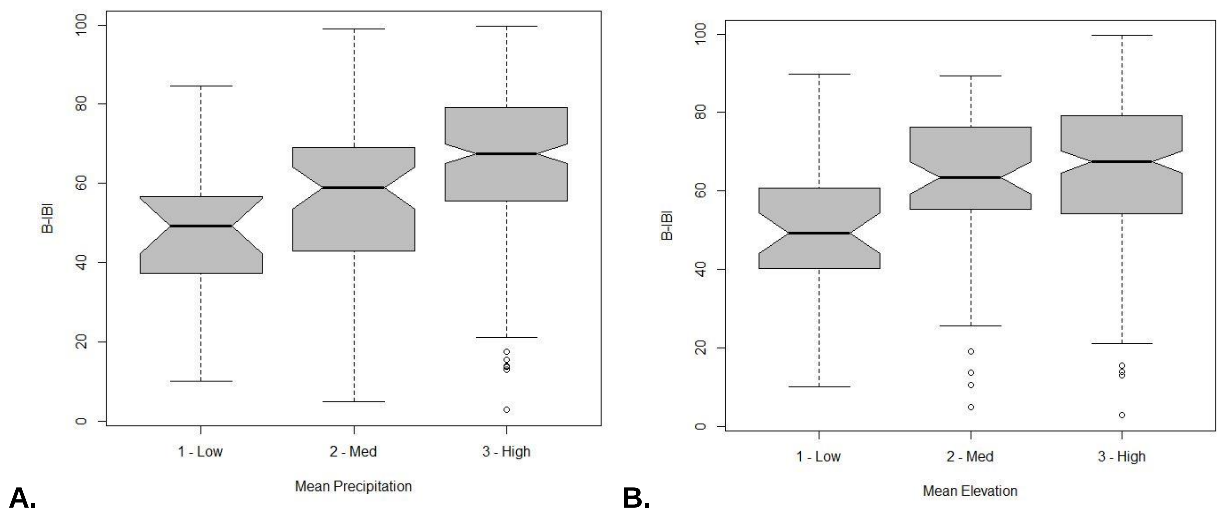

Figure B1. Boxplot of ranked A.) mean precipitation and B.) mean elevation data at low watershed urbanization sites (<10%) by B-IBI scores.

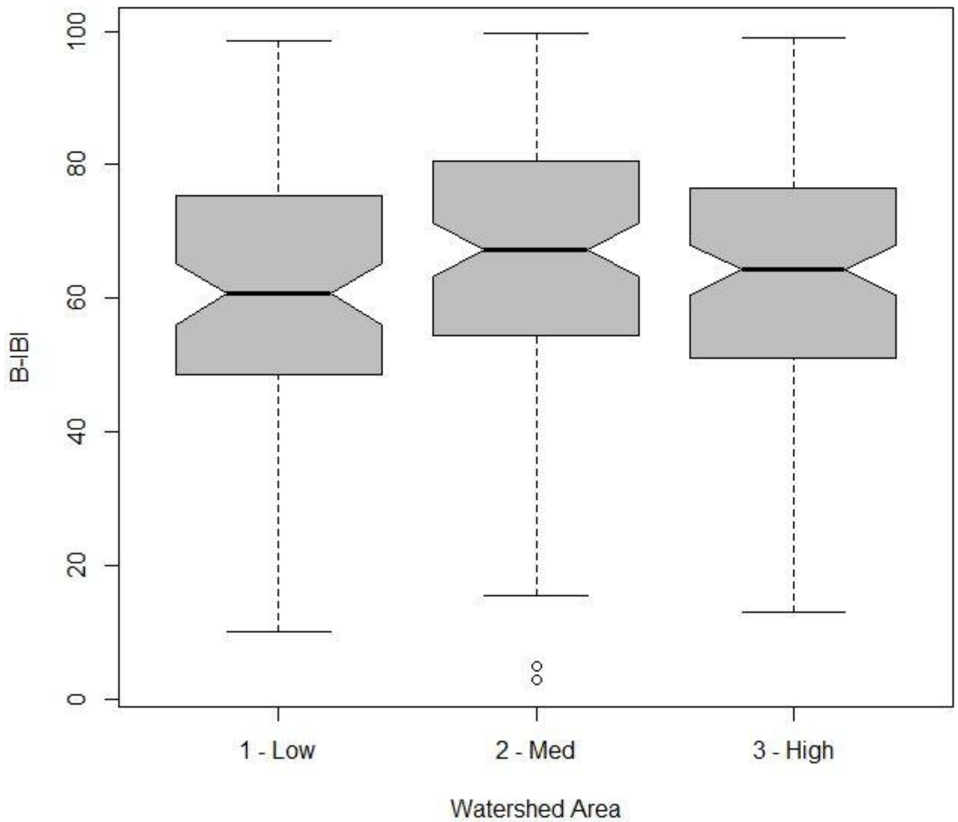

Figure B2. Boxplot of watershed area data at low watershed urbanization sites (<10%) by B-IBI scores.

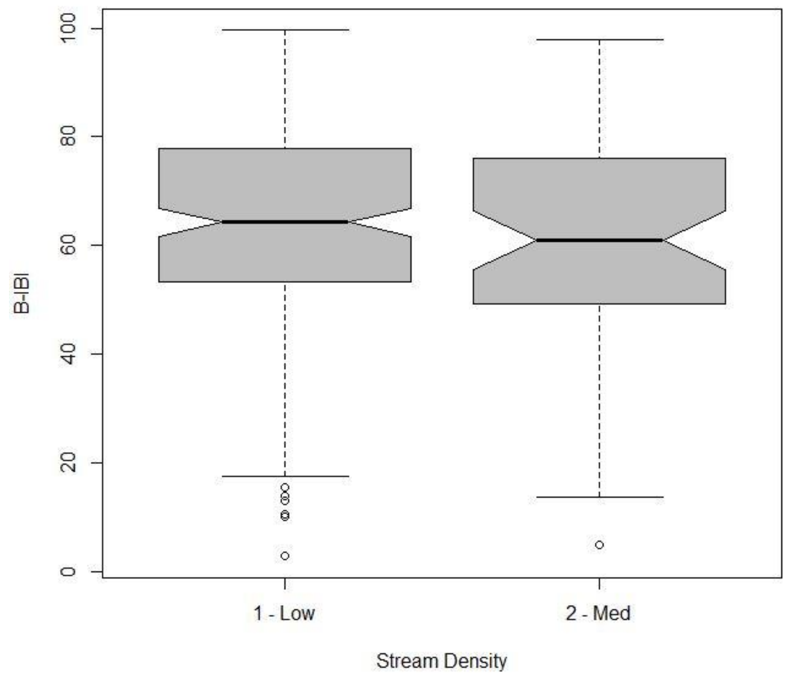

Figure B3. Boxplot of stream density data at low watershed urbanization sites (<10%) by B-IBI scores.

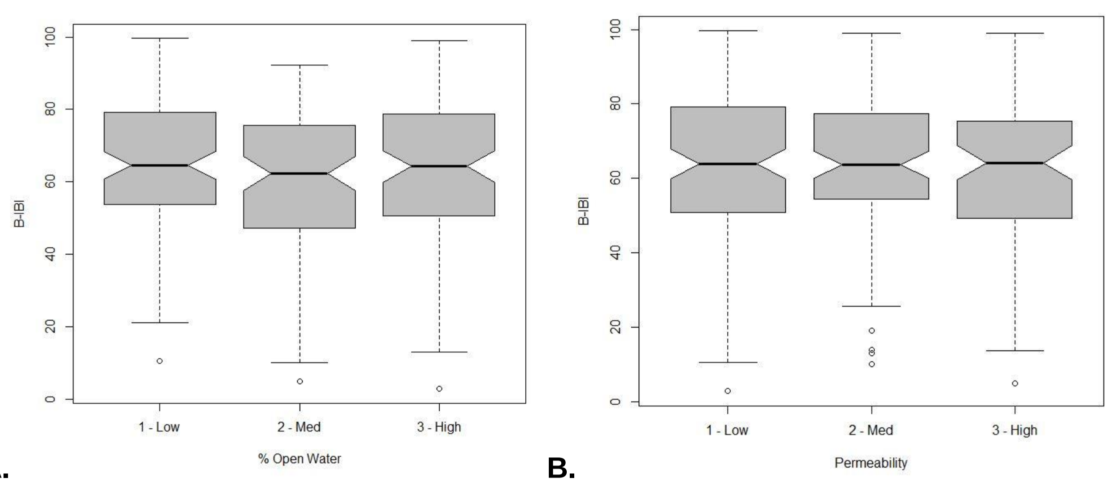

Figure B4. Surficial geology boxplots of of A.) percent open water and B.) high permeability at low watershed urbanization sites (<10%) by B-IBI scores.

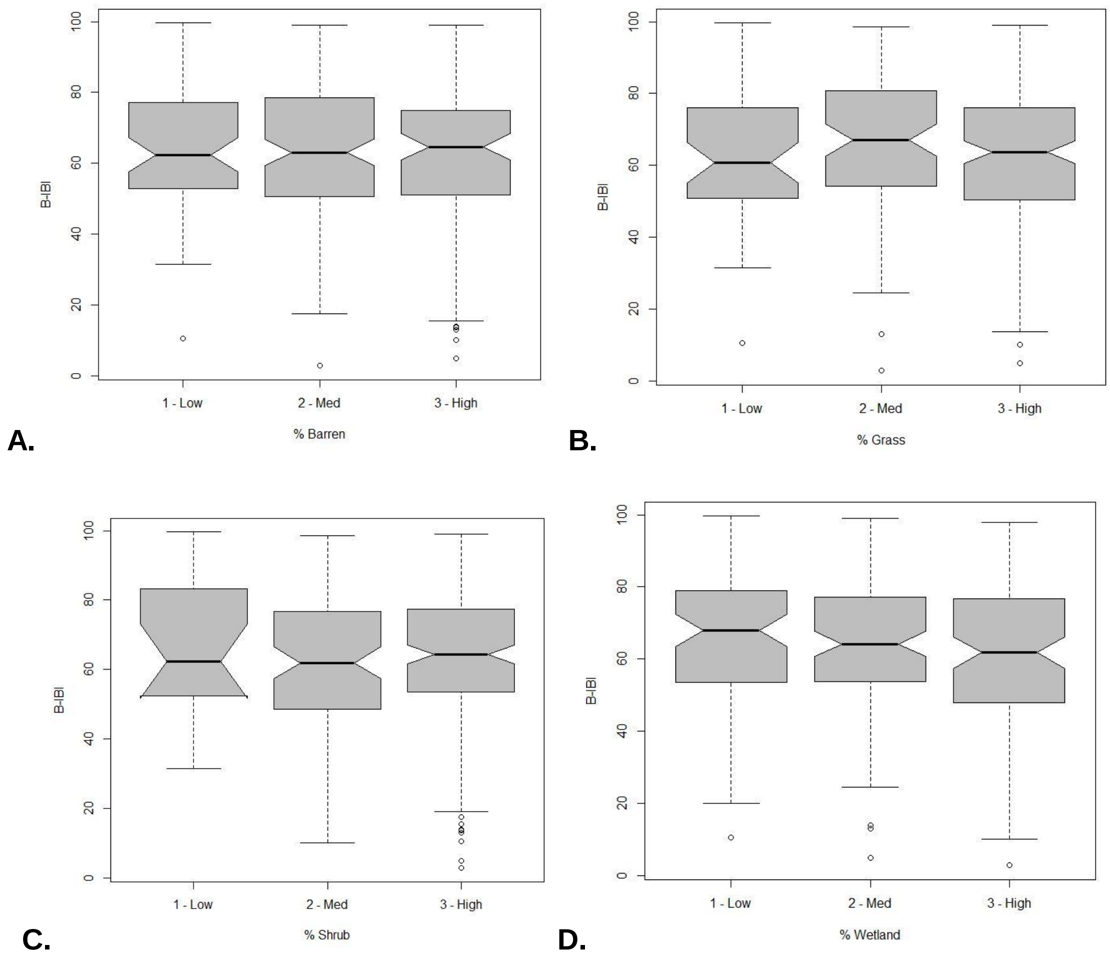

Figure B5. Boxplots of land cover variables: A.) percent barren cover, B.) percent grass cover, C.) percent shrubbery cover, D.) percent wetland cover at low watershed urbanization sites (<10%) by B-IBI scores.