Statistical Consulting Services

3405 Chukar Lane, Clarkston, Washington 99403 (509) 758-8330

Status: Preliminary Report
Date: September 14, 2007
Client: Kootenai Tribe of Idaho (KTOI)
Project Title: Macroinvertebrate Abundance and Biomass: 2005 Data, BPA-51

I) Abundance and Biomass Data

Two Excel files containing information on the 2005 macroinvertebrate data were initially provided to SCS by Charlie Holderman. Abundance data, at the level of taxonomic group, was received on 3/27/2007 and biomass data, at the level of taxonomic group, was received on 8/17/2007. These data were subsequently reformatted and compiled, and aggregated for analysis by SCS. All descriptions and analyses below relate to this compiled data.

There were 167 observations recorded. Print Out #1 presents a listing of the data. The variables included were: *site*, *RKM, month, day, year*, *replication*, *abundance* and *biomass*. The recorded values for site were KR3, KR4, KR6, KR7, KR9, KR9.1, KR10, KR11, KR12, KR13, and KR14. These sites are consistent with those from 2004 with the exception of KR1 and KR2 which are absent. Descriptions of all sites are given on the KTOI WEB database site at http://ktoi.scsnetw.com. The month, day and year values ranged from 4/14/2005 to 10/25/2005. A Julian date variable was also created from this date range. Each site was represented by five to six replications (rep). All values were adjusted by Charlie Holderman to represent response levels per meter square (g/m^2 and $\#/m^2$ for biomass and abundance, respectively). No biomass or abundance values were recorded as missing.

II) Summary Statistics for Abundance and Biomass

Computations were carried out separately for each site over all sample periods. Basic summary information for both the abundance and biomass data is presented in Print Out #2. The 11 sites varied widely in their minimum, mean, maximum and variance values. The number of observations ranged from 6 to 18. Some large abundance values (abundance > 11, 000) were noted for site KR13.

A more detailed summary of each site is given in Print Out #3. Site KR3, for example, had a mean abundance of 1442.1 with a sample size of 6. The variance was 3530594.7 and the standard error of the mean was 767.1. The skewness value, a measure of symmetry for the frequency distribution, was relatively large at 2.27 indicating an asymmetric distribution. Biomass for KR3 had a mean value of 0.55 g/m^2 with 6 observations. The variance

was 0.286 and the standard error was 0.218 g/m^2. Skewness for biomass was also very high at 2.07. Further examination of the quantiles and frequency plots for abundance and biomass also indicate considerable skewness. The stem and leaf diagram (frequency plot) for abundance in KR3 shows most of the data centered on smaller values with a few very large counts. The distribution for biomass has a similar pattern. Statistical tests for normality are significant for both response variables in KR3, thus, the hypothesis that the data originates from a symmetric normal distribution is rejected. Because sample size estimation and statistical inference assume normally distributed data, a transformation of the data is required prior to further analysis. As was the case for previous years, the natural logarithm was chosen as a transformation to mitigate distributional skewness. Abundance and biomass for the remaining sites were also notably skewed, therefore, these data were also log transformed prior to analysis. Summary information for the transformed data (referred to as L_abun and L_bio for abundance and biomass, respectively) are given in Print Out #4. For site KR1, the logarithmic transformation reduced skewness values to 0.93and 0.33 for abundance and biomass, respectively. The distributions of abundance and biomass in the other sites also showed improvement as well. Hence, all subsequent statistical analyses reported here will be based on the log transformed data.

Trends Over Time and RKM

Plots of abundance and biomass trends for each site across date are given in Print Out #5. These plots provide both the mean trend and box plots for log transformed responses at each date. The box plots at each time indicate the spread of the data on each side of the mean. Sites KR3, KR4, and KR11 had only two sampling dates and, hence, trends over time for these sites should be interpreted with caution. Some data, such as abundance and biomass in sites KR9, KR9.1, and KR10, show definitive trends over time.

Print Out #6 gives abundance and biomass trends over RKM. While the variability is high for both responses, some trend is evident, where abundance and biomass values tend to increase with increasing RKM..

III) Determination of Sample Sizes

The formulation for calculating sample size is given by:

$$n = (z*s/d)^2$$

where s, d and z are related to the variability, desired precision, and confidence levels, respectively. For this analysis, sample sizes were evaluated at the confidence levels: 90, 95, and 99%. The measure of variability was obtained from the data at hand and the precision set to approximately 10% of the overall mean value. Due to the logarithmic transformations used for abundance and biomass responses, this value may vary slightly. Sample size estimation was carried out separately for abundance and biomass at each site.

Estimated sample sizes for abundance and biomass are given in Print Out #7. The sample sizes for abundance were small at all confidence levels, indicating that the current sampling scheme of 6 replications is sufficient for that response. Sample sizes at 95% confidence for biomass were nominal for most sites with an exception of site KR9.1 where the sample size was estimated at 299 observations. This is due to low biomass levels at that site. With this exception noted, the current sampling scheme is providing a precision level at or below the desired levels.

Note that for all the above calculations, the resulting sample size values are preliminary and also based on limited data. Thus, care should be exercised in applying these results to setting policy regarding future sampling protocols.

IV) Comparison of Years

With the addition of the 2005 data, the Kootenai River Ecosystem Project has four complete seasons of macroinvertebrate data. Printout #8 presents the estimated sample sizes (at the 95% level of confidence) for abundance and biomass in 2002, 2003, 2004, and 2005. Missing values (".") in the tables are due to changes in the sites sampled across years.

Sample sizes for abundance are equivalent and adequate for all years indicating few changes over the 2002-2005 period. Biomass sample sizes are comparable with improvements from 2002, although some changes in variability have occurred in 2005 as noted above. Because the macroinvertebrate designation encompasses a large range of species and organism sizes, such variability in the biomass response is inevitable. Computation of sample size estimates based on taxonomic classes such as species, family, functional group, etc, may provide more reliable information on the biomass response in the future.

Plots of abundance and biomass over RKM in the years spanning 2002-2005 are given in Printout #9. Abundance shows similar trends for each year, however, the 2004 and 2005 data were notably larger than those in 2002 or 2003. Abundance in all three years appears to peak at RKM values from 320 to 360. For biomass, the three year trends are similar. Unlike abundance, biomass values peak at 280 to 350 RKM.. Annual changes in abundance or biomass trends may reflect actual fluxuations in the responses as well as sampling error due to changes and refinements in the sampling procedures, protocols, and personnel. Overall, however, the trends for these years show remarkable similarity.

V) Additional Remarks

- The data provided to SCS for 2005 macroinvertebrates were recompiled for analysis. Additional information on family, taxa, etc were not included on the recompilation. The data provided was free from errors and missing values.

- Available macroinvertebrate sites in 2005 were consistent with those of 2004, except for the absence of site KR1 and KR2.

- The 2005 macroinvertebrate data have now been incorporated into the relational database as part of the KTOI Ecosystem WEB site (http://ktoi.scsnetw.com). In addition, macroinvertebrate sample size tables on the WEB site have been updated to reflect the 2005 values.

- Sample sizes and trends of macroinvertebrate data over the four year period from 2002 to 2005 are similar and consistent. Some annual increases in abundance may also be evident for 2004 and 2005.

- Trend analysis and sample size estimates for abundance and biomass measurements may be improved with additional information on subgroups. Such information may include species, ecological, functional,or taxonomic classifications. Categorization would also have the added advantage of providing group (e.g. species) specific information for the relational database WEB site. SCS is currently developing a test WEB site to assess presentation and summary functions for the taxonomic group level data.

- Any additional information regarding biological, ecological, environmental, or physical variables could enhance the estimation process. To be of maximum utility, these variables should be available for each site during all sampling periods. Examples of potentially useful variables might be air and water temperature, thermal or degree day measurements, stream velocity and discharge rates, and habitat or substrate information (cobble size, depth, embededness, etc).