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

Status: Preliminary Report
Date: February 10, 2009
Client: Kootenai Tribe of Idaho (KTOI)
Project Title: Macroinvertebrate Abundance and Biomass: 2007 Data, BPA-51

I) Abundance and Biomass Data

Four Excel files containing information on the 2007 macroinvertebrate data were initially provided to SCS by EcoAnalysts on 1/27/2009. These data files contained information on abundance and biomass data at the level of taxonomic groups. The data were subsequently reformatted and compiled, and aggregated for analysis by SCS. All descriptions and analyses below relate to this compiled data.

There were 189 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 KR1, KR2, KR3, KR4, KR6, KR7, KR9, KR9.1, KR9.2, KR10, KR11, KR12, KR13, and KR14. These sites are consistent with those from 2006 with the exception of a new site, KR9.2. Descriptions of all sites are given on the KTOI WEB database site at http://scsnetw.com/ktoi. The month, day and year values ranged from 4/10/2007 to 11/20/2007. 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 recommendation of Charlie Holderman to represent response levels per meter square (g/m^2 and $\#/m^2$ for biomass and abundance, respectively).

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 14 sites varied widely in their minimum, mean, maximum and variance values. The number of observations ranged from 10 to 18. Some large abundance values (abundance > 40,000) were noted for sites KR6 and KR13.

A more detailed summary of each site is given in Print Out #3. Site KR3, for example, had a mean abundance of 6914 with a sample size of 17. The variance was 4591991 and the standard error of the mean was 1643. The skewness value, a measure of symmetry for the frequency distribution, was moderately large at 1.29 indicating an asymmetric distribution. Biomass for KR3 had a mean value of 0.87 g/m^2 with 17 observations. The

variance was 0.8872 and the standard error was 0.228 g/m^2. Skewness for biomass was also high at 1.29. 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 KR3, the logarithmic transformation reduced skewness value for biomass to -0.66. The distributions of abundance and biomass in the other sites also generally 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. With the exception of site KR1, KR2, KR3, KR4, and KR14, all sites had only two sampling dates and, hence, trends over time for these sites should be interpreted with caution. In general, trends in abundance and biomass indicate an overall increase 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 show elevated values within the fertilization zone (KR6-KR9.1). It is noted that sites KR1-KR4, as well as KR14, show consistently lower levels of abundance and biomass than sites KR6-KR13.

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. For abundance, the sample sizes were small at all confidence levels, indicating that the current sampling scheme of 6 replications is sufficient for that response. One exception is noted for the new site, KR14, where the sample size of 14 was slightly elevated. Biomass sample sizes at 95% confidence for biomass were nominal for most sites with an exception of sites KR10, KR11, and KR12 where the sample sizes were estimated at 32, 2645, and 113, respectively. This is due to very low biomass levels at these sites. With these exceptions 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

Print Out #8 presents the estimated sample sizes (at the 95% level of confidence) for abundance and biomass in 2003 - 2007. 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 2003-2007 period. Biomass sample sizes are comparable with improvements from previous years, although some changes in variability have occurred in 2007 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 2003-2007 are given in Print Out #9. On these plots, the fertilization site (KR9.1) is indicated with a vertical reference line. Both abundance and biomass show general trends similar to those of the preceding years, however, it is noted that the responses were considerably larger in 2007, especially for the fertilization zone. 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 fertilization sites show an increased response in 2007 and generally have an increasing pattern from 2005 to 2007. This is in contrast to KR10 which shows a very consistent response over the five years and, hence, this makes this site a good candidate for controlled comparisons of pre versus post fertilization effects.

V) Additional Remarks

- The data provided to SCS for 2007 macroinvertebrates were recompiled for analysis. Additional information on family, taxa, etc were not included in this

report. The data provided was free from errors and missing values.

- Available macroinvertebrate sites in 2007 were consistent with those of 2006 with the addition site KR9.2.

- The 2007 macroinvertebrate data have now been incorporated into the relational database as part of the KTOI Ecosystem WEB site (http://scsnetw.com/ktoi/). In addition, macroinvertebrate sample size tables on the WEB site have been updated to reflect the 2007 values.

- Responses in the fertilization zone (KR6-KR9.1) were higher for 2007 and generally show an increasing pattern in abundance and biomass over the years since fertilization began. In contrast, site KR10, just upriver of the fertilization zone, consistently shows the same level of response over five years.

- 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.

- 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).