

**Statistical Consulting Services**  
807 3<sup>rd</sup> Street, Clarkston, Washington 99403 (509) 758-8330

---

Status: Preliminary Report  
Date: June 20, 2007  
Project Title: 2003 - 2005 Algae: Downriver Sample Size, BPA-51

## I) Data Description

The analyses given below are based on 2003 - 2005 algae data received by SCS, from Gretchen Kruse, on May 21, 2007. These files cover the downriver sites KR1, KR2, KR3, and KR4. For a complete description of these sites, please refer to <http://ktoi.scsnetw.com>.

There were 159, 192 and 148 observations recorded for algae in 2003, 2004 and 2005, respectively. The relevant variables contained in the data were: *site (Site)*, *date (Date)*, *replication (Rep)*, *chlorophyll a (Chlor\_a)*, *total chlorophyll (Total)*, and *total chlorophyll accrual rate (Total\_acc)*. ChlorophyllA, TotalChlor, and TotalAccrual are recorded (and represented here) in units of g/m<sup>2</sup>, g/m<sup>2</sup>, and g/m<sup>2</sup>/d, respectively.

## II) Determination of Sample Sizes

When estimating the mean, the formulation for calculating sample size is:

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

where s, d and z are related to the variability, desired precision, and confidence levels, respectively. In many instances, prior transformation of the data is necessary to meet the distributional requirements for the above estimation.

The desired precision was fixed at 10% of each response mean value. Because the data distributions were substantially skewed (based on univariate analyses results not included in this short communique), all sample size calculations are based on a natural logarithm transformation. The confidence level, z, is obtained from a standard table of normal values, i.e. 1.96 for 95% confidence. Standard deviations, s, were obtained for each response variable and the resulting sample size estimates were computed and tabulated (Print Out 1). Sample size estimates are computed for confidence levels of 90, 95 and 99%, both individually by site for each year, as

well as across all sites within a year. In general, at a 95% level of confidence, all sample size estimates for all response variables are within or close to the current sampling protocol of eight replications for the down river sites. Some exceptions are noted for *Chor\_a* and *Total* at site KR2, where larger sample sizes are indicated, and the current sampling scheme is not likely to obtain the 95% level of confidence. For the remaining sites, however, a reduction in the current sampling intensity may be considered. In order to ensure that the desired precision level are achieved with at least 95% confidence, SCS recommends a sampling intensity of six replications (per site/date) for sites KR1, KR3, and KR4. SCS further recommends maintaining the current level of sampling effort (eight replications) for site KR2. KTOI may also consider increasing the number of samples taken at this site, to say ten or twelve, as a reallocation attempt to mitigate the potential variability. No further reduction in replicate numbers or alteration to the sampling protocol are recommended at this time.

In practice, sample size determination requires a formal statistical power analysis. A meaningful power analysis in turn, requires the specification of a scientific hypothesis addressing a relevant contrast. Plots of power versus sample size, i.e. power curves, are used to investigate the effect of sample size on the power of a given hypothesis. As per agreement made with Charlie Holderman, power analyses for the algae data will be delayed until such time as when the data on the remaining sampling sites (upriver) become available. This should provide opportunities to construct ecologically sound contrasts and carry out the appropriate power calculations.

Finally, it should be noted that for all the above calculations, the resulting sample size estimates are preliminary. The calculations are based on limited data covering a few seasons. Furthermore, the actual precision of mean estimates may vary by the specified response variable as well as the sampling date. Therefore, the aforementioned sample size values should be used cautiously for setting policy regarding future sampling protocols.