# **APPENDIX B – TerraStat Report**



#### **Review of**

*"Development of Coastal Monitoring Protocols and Process-Based Studies to Address Landscape-Scale Variation in Coastal Communities of Glacier Bay National Park and Preserve: Phase II: Development and testing of monitoring protocols for selected intertidal habitats and assemblages (Draft Final Report; 2002)"*

> **By Alice Shelly TerraStat Consulting 5002 Lodge View Lane Austin, TX 78731 alice@blarg.net May 14, 2007**

This report summarizes my review of the power analysis detailed in your 2002 draft final coastal monitoring report for Glacier Bay. Overall, it seems that the power analysis was adequate with reasonable results. Although there were some estimation methods that were not ideal, I have now verified (using the updated program MONITOR 10.0) that these differences did not significantly affect the power estimation for mussels (*Mytilus trossulus, MY*) and barnacles (*BARN*). The statistical power for rockweed (*Fucus gardneri, FU*) may have been overestimated in the coastal monitoring report by 5–10%. There remains some uncertainty in the results due to the assumption of lognormal errors, which is not ideal for MY and BARN. The only way to address this uncertainty would be with an independent simulation program which would allow for simulations from empirical error distributions or a more suitable fitted distribution.

With multiple sites, program *MONITOR* estimates statistical power for a trend monitoring program assuming an analysis using route regression. Route regression consists of a linear or exponential regression relating years to the estimated percent cover for each site. The slope of the regression line for each site is the estimated trend for that site, and the mean of these slopes is tested for differences from zero using a t-test. To estimate power, the user inputs site means and variances. *MONITOR* simulates normal or log-normal residuals around projected linear or exponential trend lines as specified by the user. Variability among site trends is an optional feature. In version 10.0, the variance can also be partitioned into variance among years and variance within years (although this feature is not yet functioning properly per James Gibbs 4/3/07).

For the Glacier Bay power analysis, the questions of interest are:

- 1) Given the variability observed over the sampled 3–4 years, what magnitudes of trend were we likely to detect for various species?
- 2) Would varying the number of sites, the number of transects, or the number of samples per transect result in large differences to statistical power?
- 3) Is statistical power more sensitive to changes in sampling design (e.g., number of sites, # of transects, # samples/transect, etc.) or to changes in parameters (e.g., different levels of alpha)?
- 4) Would varying the sampling type (point sampling of quadrats, horizontal transects or vertical transects) result in large differences to statistical power? If there were differences in power among the sampling types, is that difference related to the number of points sampled in each?

The analysis run for Glacier Bay (coarse-grained sites) was a post-hoc analysis, in that the power estimates were based on detecting trends within the time period already sampled. With 3–4 years of data, linear regression was run on untransformed percent cover estimates for each transect. The variance of the residuals from the linear regression was input to MONITOR as the site variance for each site. The average cover over all transects over all sampled years was input as the site mean for each site. The effects of different sampling designs were analyzed by estimating different variance estimates corresponding to each sampling design.

#### Issues with MONITOR (as used)

For exponential trend, a log-normal distribution for within-year sampling variability at each site was assumed. This assumption was not tested to see if it was a reasonable one for the Glacier Bay cover data.

The new version of MONITOR allows the sampling variance to be partitioned between within-year and between-year components. The authors do not expect this to be a critical difference to the power results, but it would be interesting to recalculate power using this feature. The variance components could be estimated from existing data.

MONITOR 10.0 also allows the user to specify if the mean and the variance are related over time. This is a very common scenario, and changes to this component may affect the power estimates more than any of the other issues mentioned here.

### Issues with Data Analysis and Input Parameters

Although this does not affect the power analysis, I am not sure why there is not an analysis of the observed trends in the report. Since the researchers expect an exponential trend, and this is a post-hoc power analysis, the analysis itself should have been conducted first. In the process of selecting the best analysis, the distribution of the errors would be tested and the shapes of the trends could be discussed. This would also alert the analyst as to whether there are any errors in the data set.

The plot or site means input into program MONITOR should have been the starting mean values in Year 0, rather than the average across all years. If the research question were to detect trends in an ongoing monitoring program, the mean of all current data might be appropriate, but in the case of a post-hoc test such as this, the mean of the first year of sampling should be used. This is not likely to have a large effect on the power results.

The sampling variance should be estimated by the empirical variance of the residuals from an exponential regression fit, rather than a linear fit. This would match the assumptions of program MONITOR. The results are likely to be similar, depending on the magnitude and shape of any existing trends.

There are some other minor things that I may have done differently, but as stated in the start of this document, I think the overall methods were reasonable.

## Limited Re-analysis

To answer some of the issues raised above, I evaluated data for MY, FU, and BARN collected in 1998– 2001 on six vertical transects with five points per meter. Results reported below include an analysis of existing trends, data distributions, revised parameter estimates, and limited statistical power results from MONITOR 10.0 for comparison purposes. The intention is to determine if the changes to program MONITOR and/or changes to parameters suggested above cause large differences in estimated statistical power.

## **Testing Assumptions**

The metric being tested for trends at this site is percent cover, estimated by counts divided by the number of points sampled (multiplied by 100). The estimates can exceed 100% due to layered species.

## *MY*

I subtracted out the site/year means and tested the combined residuals for normality using Shapiro-Wilks goodness-of-fit test with alpha =0.05. The tails of the distribution were too thin to fit a normal distribution (pvalue = 0.0000), but there was no skew. When the sites were tested separately, all but three had approximately normal residuals (sites 63 and 151 have a slight skew, and site 220 appears bimodal). This could indicate that variance differs across sites or that these three sites were causing the lack of normality overall. When the residuals were standardized for variance, the distribution was skewed with a shortened range (p-value  $= 0.0000$ ).

I also log-transformed the cover (after adding the constant 1 to allow for zeros) and tested the combined residuals for normality with the same result (no improvement in fit for log-normal assumption). All but three individual sites (site 36 and 215 had a few "outlier" transects with low cover; site 220 appears bimodal) were approximately normal after log-transformation.

The above indicates that neither normal nor log-normal simulations are ideal for estimating statistical power for these data. An empirical distribution or a box-cox transformation would provide a better fit. However, MONITOR 10.0 is the best available software for estimating the power.

The mean (site/year) and variance are related, as shown in figure B1, so the constant CV option in MONITOR 10.0 should be selected.

# *FU*

The distribution of the combined residuals again had thin tails and did not fit a normal distribution (p-value = 0.0000). When the sites were tested separately, nine sites failed the normality test – mainly for skewness. The log-transformed combined residuals are approximately normally distributed (Shapiro-Wilks p-value = 0.19). The lognormal distribution is the appropriate choice for FU. As with MY, the mean and variance are related (see fig. B2).

## *BARN*

The distribution of the combined residuals again had thin tails and did not fit a normal distribution (p-value = 0.0001). When the sites were tested separately, four sites failed the normality test – mainly for skewness (sites 63, 108, 200, and 218). The log-transformation did not approximate normality (p-value = 0.0001), with four different individual sites also failing (sites 69, 163, 214, and 216). As with the other species, the mean and variance are related (see fig. B3).

## **Testing for Exponential Trends**

A linear regression was fit to log-transformed density data for each site for each species to test for exponential trends.

## *MY*

The trends (slopes) for MY ranged from a decrease of 50% per year (Willoughby Island - West) to an increase of 125% per year (Muir Inlet – Upper), and the slopes were approximately normally distributed. The average trend was a decrease of 8%, which was not significantly different from zero using a two-tailed t-test (p-value = 0.21). Trends for individual sites are displayed in table B1. There are significant positive and negative trends observed at some individual sites, but no significant regional trend in MY over the years 1998–2001. Note that the individual regression significance results are only valid if the measurements from year to year are independent. Since the transects are randomly selected in each year, these should be approximately independent.

### *FU*

The trends for FU ranged from a decrease of 70% per year (Russel Island) to an increase of 220% per year (Berg Bay), and the slopes were approximately normally distributed. The average trend was an increase of 13%, which was not significantly different from zero using a two-tailed t-test (p-value = 0.27). Trends for individual sites are displayed in table B2. There are significant positive and negative trends observed at some individual sites, but no significant regional trend in FU over the years 1998–2001.

### *BARN*

The trends for BARN ranged from a decrease of 16% per year (Muir Inlet - Lower) to an increase of 77% per year (Lester Island). The distribution of slopes failed the test for normality at alpha = 0.05 (p-value = 0.0095). However, the t-test is fairly robust to non-normality, and the result for the nonparametric Wilcoxon test was similar. The average trend was an increase of 22.5%, which was significantly different from zero using a two-tailed t-test (p-value = 0.00005). Trends for individual sites are displayed in table B3. There are significant positive trends observed at some individual sites, and a significant increasing regional trend in BARN over the years 1998–2001.

# **MONITOR Parameter Estimates**

The plot means for MONITOR are the starting means for the trends. These should be the 1998 plot means, as displayed in tables B4-6. The plot variances, shown in the tables as standard deviations, were estimated by the residual variance from exponential regressions fit to untransformed densities. These plot variances include all sources of variability: spatial variability, sampling error, and random variability due to year effects (e.g., weather). The final version of MONITOR 10.0 will have capability to account for random variability due to year as a separate parameter, but in the existing beta version (as of 5/7/07) this capability is not correctly working (personal communication, James Gibbs). However, the parameters may be of some interest in understanding the sources of variability for the different species. The variance components for

between versus within years (displayed in tables B4-6 as standard deviations) were estimated using restricted maximum likelihood estimates for random components of variance model.

For MY, the variation among years (after adjusting for trend) is effectively zero for all but six sites. At least for the existing three years of data, the majority of sites have very little variance due to random year effects (relative to within year variance). In this case, partitioning the variance is unlikely to have a strong effect on statistical power. The two plots in figure B4 show an example of two sites with different variance characteristics. MY cover at Shag Cove, in the top plot, has very little variability among years after the trend is removed. At Upper Drake Island, however, there is variability both among and within years.

For FU, the results are similar to those for MY, except that there are nine sites with some annual variability.

Barnacles have mixed results across sites, with some sites exhibiting more variability among years than within years. Thus, power for BARN would be the most likely to change if the components of variance are split.

### **MONITOR 10.0 Analysis**

The statistical power was estimated using MONITOR 10.0 dated 3/14/07 at this web address:

### *[http://www.esf.edu/efb/gibbs/MONITOR/](http://www.esf.edu/efb/gibbs/monitor/)*.

Power was estimated with the original parameters used by USGS, as well as with the new parameters estimated and discussed above. The plot means were originally the mean of all samples across three years, whereas I used the means of the first year's samples only. The plot standard deviations were originally the standard error of the residuals from a linear regression on untransformed data. I used the standard error of the residuals from an exponential regression. For reduced numbers of sites, the original analysis used one random subset of sites to estimate the power. I averaged the statistical power over 10 random subsets of the 25 sites. I also used 1000 instead of 500 simulations in MONITOR, to create more stable results.

The plots in figures B5 and B6 show the power estimates for MY with 25 and 15 sites with 6 transects, each having five points per meter. The yellow triangles show results from the original report (partial series). The brown circles show the current runs on MONITOR 10.0 using the original parameters, but with 1000 replications. The blue squares show the current runs on MONITOR 10.0 with the TerraStat parameters. The results show that the original power estimates for MY using earlier parameter estimates and previous versions of MONITOR were almost identical to current estimates.

The plot in figure B7 shows that the change in parameterization had the effect of reducing the statistical power for FU by 5-10%. This is likely due to the reduced plot mean estimates – the new parameters represent an average decrease of 4% in the starting mean. In contrast, Mytilus and Barnacle plot means had an average increase in starting mean (although there was high variability across sites.) Results for FU using MONITOR 10 are similar to previous results when the original parameters are used. The power for 15 sites was not estimated in the previous report.

The plot in figure B8 shows that the original power results for BARN are equivalent to the estimates with MONITOR 10, and with revised parameters.



Figure B1. Variance of MY cover within a year at each site plotted against the average cover.



**Figure B2.** Variance of FU cover within a year at each site plotted against the average cover.



**Figure B3.** Variance of BARN cover within a year at each site plotted against the average cover.



**FigureB4.** Comparison of residual variance sources within and between years for MY percent cover for two selected sites.



**Figure B5.** Statistical power comparison for MY with 25 sites.



**Figure B6.** Statistical power comparison for MY with 15 sites.



**Figure B7.** Statistical power comparison for FU.



**Figure B8.** Statistical power comparison for BARN.

**Table B1.** Trend results by site for percent cover of MY.



**Table B2.** Trend results by site for percent cover of FU.



**Table B3.** Trend results by site for percent cover of BARN.





**Table B4.** Plot means and standard deviations for MY percent cover.



**Table B5.** Plot means and standard deviations for FU percent cover.

**Table B6.** Plot means and standard deviations for BARN percent cover.

