TerraStat
CONSULTING GROUP

3 August 2009

To: Dennis Lees
Littoral Ecological & Environmental Services
1075 Urania Ave.
Leucadia, CA 92024

Cc: Greg Challenger, Principal
Polaris Applied Sciences, Inc.
12525 131st Court NE
Kirkland, WA 98034

Re: Review of data analysis for *Fucus* observations from San Francisco Bay area with additional
analysis results

Dear Dennis,

Here is a summary based on my review of the data provided in the Excel spreadsheet
Standardized data - PISCO and ALcatraz-dcl.xls. Specifically, this memo provides the
following:

1. A summary of specific features of the data as well as the main question under
consideration;
2. Comments on a couple of data analysis issues;
3. Results from a more complete analysis based on the nature of the observations and the
question at hand.

Summary of the Data

The data consist of annual *Fucus* percent cover observations from each of six sites (KIB, MEN,
MOL, SCT, SHT, and Alcatraz) over three contiguous years (2006, 2007, and 2008). There
were either ten or eight plots that were surveyed within each site. For all sites, except Alcatraz,
the data were stratified by two intertidal zones, identified as either chthamalus/balanus or fucus
(the column labeled "Species" in the Excel spreadsheet). For these sites there were five samples
plots per zone.

The primary question of interest is: Does the three year annual time series of observed *Fucus*
percent cover data at Alcatraz differ from the *Fucus* percent cover observations at the other five
sites? The KIB, MEN, MOL, SCT, and SHT sites are being used as control sites to make
comparisons with any observed changes at the Alcatraz site.

Data Analysis Issues

While the spreadsheet contains some manipulations of the raw data and some graphical summaries, I can't really comment on the analysis results/conclusions other than to say that the analysis appears to be incomplete. There are no summarized results or concluding remarks. I do, however, have a few comments on the apparent analytical approach:

1. The Alcatraz site only has data from one of the strata (fucus), therefore, it seems inappropriate to include any of the data from the chthamalus/balanus strata in the analysis. From a biological/ecological point of view, one might also argue that *Fucus* is generally not found in the chthamalus/balanus strata, so including these data in the analysis would not be appropriate.

2. Normalizing individual observations within a site with respect to the grand mean and SD of observations from that site considers only one component of deviation, ignoring an overall grand mean for the model, among other things.

3. A few of the graphical summaries in the spreadsheet show confidence bars around endpoints in 2008. The confidence intervals appear to have been constructed using the CONFIDENCE function in Excel. This function assumes that the population standard deviation is known and calculates lower and upper bounds using normal probabilities. In this case, the use of a *t*-table is more appropriate, especially given the small sample sizes. This will effectively produce wider confidence intervals at each endpoint, which will increase the overlap for mean *Fucus* percent cover among the six sites in 2008.

4. Given what I understand about the data, there are two factors of interest (year and site). Further, there are repeated measures of percent cover on plots, within sites. As such, a more complete analysis would be the use of a two-factor ANOVA with repeated measures on one factor, or a nonparametric equivalent. This type of analysis will allow one to look for any significance of both main effects, as well as any significance of the interaction between the two main effects.

Two-Factor ANOVA with Repeated Measures on one Factor

The two factors are site and year. The zone strata did not come into play as a factor in this analysis since the Alcatraz site was only sampled in the fucus zone. All data from the chthamalus/balanus zone were omitted.

For this analysis the sites are taken as fixed since Alcatraz is a specific location of interest—there are six levels of site. Year is also taken as fixed since this analysis is in response to an event that occurred at a specific point in time—there are three levels of year. We are assuming that the years are equally spaced or at least that the data were collected from each site around the same time each year. Since percent cover for each year was measured at a set of fixed plots within each site, year is treated as a repeated-measures factor.

The full factorial model includes year as a within-subject factor, location as a between-subject factor, and the interaction of year and site as a within-subject component. The design is minimally unbalanced since there were eight plots at Alcatraz and only five plots at each of the

other locations. As such, the ANOVA was run using Type III sums of squares. The ANOVA was conducted using SPSS Statistics 17.0 on a Windows XP machine.

Data Summaries

Table 1 shows means and standard deviations broken out by site and year, as well as a total across all sites within a year. Figure 1 on the next page shows graphs of the raw data, plotted by site. Both summaries show that the sites SCT and SHT tend to have the lowest overall percent cover values, with the least variability. There is a reasonable amount of variability in percent cover values for the other four sites both in terms of the ranges of the amount of cover, as well as in the pattern of increasing and decreasing cover values through time.

**Table 1. Mean and standard deviations
of percent *Fucus* cover by year and location.**

| Year	Site	Mean	SD	*n*
2006	KIB	42.40	27.006	5
	MEN	38.80	10.733	5
	MOL	66.20	7.328	5
	SCT	22.00	13.766	5
	SHT	11.60	10.526	5
	Alcatraz	52.38	21.824	8
	Total	40.12	24.027	33
2007	KIB	43.40	28.395	5
	MEN	34.20	18.687	5
	MOL	46.00	22.113	5
	SCT	14.40	13.278	5
	SHT	11.00	13.454	5
	Alcatraz	60.38	22.418	8
	Total	37.21	26.503	33
2008	KIB	39.00	22.594	5
	MEN	64.80	29.861	5
	MOL	30.40	18.488	5
	SCT	11.60	7.893	5
	SHT	26.40	14.082	5
	Alcatraz	22.00	29.360	8
	Total	31.42	26.644	33
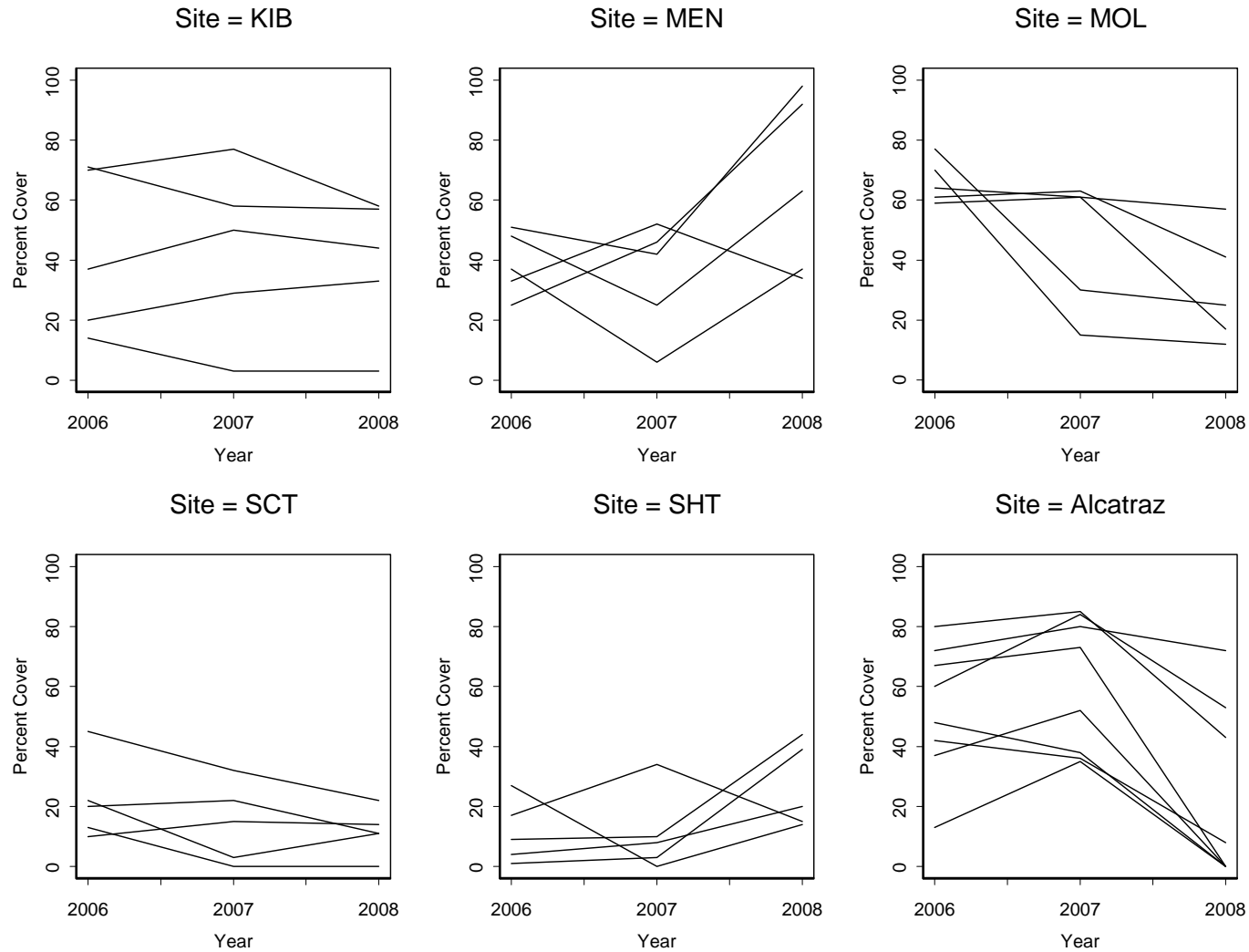
Figure 1. Time-series plots of *Fucus* percent cover for each plot, within each site. There were *n* = 5 plots for all sites, except for Alcatraz, which had *n* = 8 plots. All graphs are scaled from 0 to 100 percent cover on the *y*-axis.

ANOVA Results

The ANOVA results are shown in Tables 2 and 3. The ANOVA output includes three components; namely, the within-subjects year effect, the within-subjects interaction effect (year * site), and the between-subjects site effect. Basic ANOVA assumptions with respect to covariance and sphericity were met:

- The observed covariance matrices of the dependent variables are equal across groups ($P = 0.311$).
- For the repeated-measures year effect, the error covariance matrix of the othronormalized transformed dependent variables is proportional to an identity matrix ($P = 0.591$).

Table 2. Tests of within-subject effects.

Source	Type III Sum of Squares	df	Mean Square	F	P-value
Year	693.618	2	346.809	1.867	0.1644
Year * Site	12312.487	10	1231.249	6.629	0.0000
Error(Year)	10029.917	54	185.739		

Table 3. Test of between-subject effects.

Source	Type III Sum of Squares	df	Mean Square	F	P-value
Intercept	40069.660	1	40069.660	137.089	0.000
Site	5883.085	5	1176.617	4.026	0.007
Error	7891.811	27	292.289		

Based on the reported *P*-values and using an $\alpha = 0.05$ level of significance, the main effect of year is not statistically significant ($P = 0.1644$), but the interaction term of year and site is statistically significant ($P < 0.0005$). The main effect of site is statistically significant ($P = 0.007$). To help interpret these test results, the graphs of the estimated marginal means are shown in Figures 2 and 3. Additionally, the marginal means tables for each of the tests are provided in Tables 4 through 6. These tables show the means for the groups compared in the three ANOVAs (main effect of site, main effect of year, and interaction between year and site).

The main effect of year is not significant. However, the significant site*year interaction indicates that the sites are changing over time, but in different ways. Graphically, this means that the graphs of the lines for the sites, based on estimated marginal means, will not be parallel. In fact, the lines cross one another in different patterns, depending on the site. The between-subjects test indicates that the variable site is significant. In the graphs for this particular case we see that the nature of the lines is different for different groups—some of the lines are further apart from one another. In particular the sites SCT and SHT tend to have lower *Fucus* percent cover values, compared with the rest of the sites (Figure 2). The patterns for 2006 and 2007, when looking across sites for a given year, appear similar. The pattern for 2008 is a bit different, but the change in pattern occurs at more than one site (Figure 3).
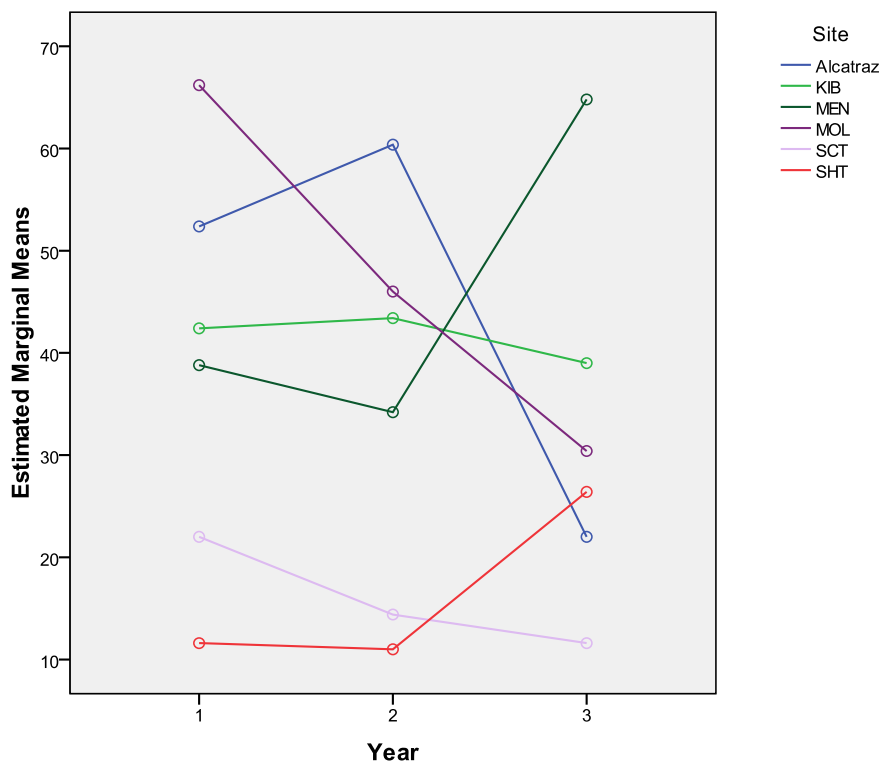
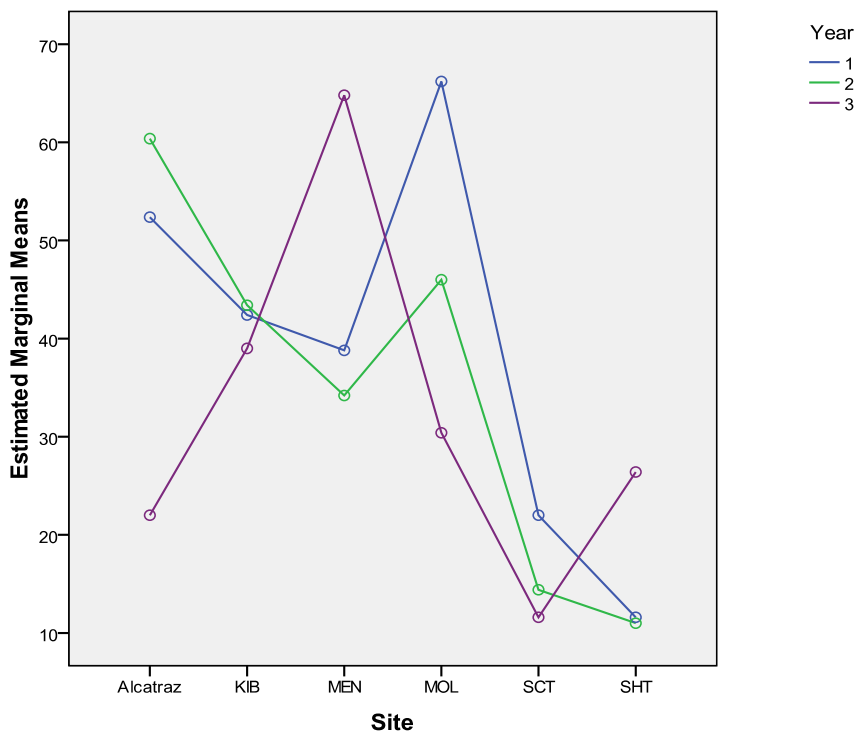Figure 2. Estimated marginal means by year, one line for each site.

Figure 3. Estimated marginal means by site, one line for each year.

Table 4. Estimated means of *Fucus* percent cover, by site.

Site	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
KIB	41.600	7.646	25.912	57.288
MEN	45.933	7.646	30.246	61.621
MOL	47.533	7.646	31.846	63.221
SCT	16.000	7.646	0.312	31.688
SHT	16.333	7.646	0.646	32.021
Alcatraz	44.917	6.045	32.514	57.319

Table 5. Estimated means of *Fucus* percent cover, by year.

Year	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
2006	38.896	3.067	32.603	45.189
2007	34.896	3.652	27.402	42.390
2008	32.367	4.033	24.092	40.642

Table 6. Estimated means of *Fucus* percent cover, by site and year.

Site	Year	Mean	Std. Error	95% Confidence Interval	
				Lower Bound	Upper Bound
KIB	2006	42.400	7.759	26.479	58.321
	2007	43.400	9.240	24.442	62.358
	2008	39.000	10.203	18.066	59.934
MEN	2006	38.800	7.759	22.879	54.721
	2007	34.200	9.240	15.242	53.158
	2008	64.800	10.203	43.866	85.734
MOL	2006	66.200	7.759	50.279	82.121
	2007	46.000	9.240	27.042	64.958
	2008	30.400	10.203	9.466	51.334
SCT	2006	22.000	7.759	6.079	37.921
	2007	14.400	9.240	-4.558	33.358
	2008	11.600	10.203	-9.334	32.534
SHT	2006	11.600	7.759	-4.321	27.521
	2007	11.000	9.240	-7.958	29.958
	2008	26.400	10.203	5.466	47.334
Alcatraz	2006	52.375	6.134	39.789	64.961
	2007	60.375	7.305	45.387	75.363
	2008	22.000	8.066	5.450	38.550

Summary

There appears to be quite a bit of natural variability among *Fucus* percent-cover observations for the six sites that were surveyed annually over the three-year period from 2006 – 2008. The significant site effect probably stems from the large range in percent cover observations among the sites. The significant site *year interaction effect probably arises due to year differences that themselves differ among the sites—note the amount of crossover shown in Figures 2 and 3.

Graphical assessment, most clearly seen in Figure 2, seems to indicate that the SCT and SHT might be different than the other four sites in terms of their mean percent *Fucus* cover. Using Dunnett's Post-Hoc multiple comparison test, testing for pairwise difference between the five "control" sites and Alcatraz, only SCT and SHT are shown to be statistically different from Alcatraz (P = 0.028 and P = 0.030, respectively). The remaining "control" sites, consisting of KIB, MEN, and MOL do not appear to be statistically different than Alcatraz. This reinforces the graphical output that SCT and SHT account for the significant site effect (Figures 1 and 2).

In terms of assessing declines from 2007 to 2008, four of the six sites showed decreases in the percent cover of *Fucus*. The 95% confidence intervals by year within site show a fair amount of overlap in 2008 (Table 6). This makes it difficult to say that the mean percent cover values are significantly different from one another.

Sincerely,

Tamre Cardoso
TerraStat Consulting Group
tamre@blarg.net
206-972-0283