MEMORANDUM

To: Jim Dinley, Administrator
Mayor McAdams and Members of the Assembly

From: Mark Buggins, Environmental Superintendent

Through: Michael Harmon, Director of Public Works

cc: Dave Wolff, Finance Director

Date: May 28, 2010

Subject: Long Term Benthic Monitoring and Bioaccumulation Survey, Sawmill Cove Professional Services Contract Award

Background

Qualifications were requested to assist with the long-term monitoring and adaptive management program for natural recovery of an approximately 100-acre "Area of Concern" (AOC) within Sawmill Cove. This is the "Year 10 - Milestone and Recovery Assessment for Unrecovered Areas". This monitoring cycle includes a bioaccumulation survey as well as benthic monitoring similar to the previous monitoring. We have negotiated Phase I of this project that will allow Germano & Associates to make an assessment of the work needed to complete ADEC approved bioaccumulation and benthic monitoring work plans which will allow cost estimates for the next phases of field monitoring, data analysis and reporting.

Analysis

Germano & Associates, Inc. was the only firm to submit qualifications for this RFQ. Joe Germano performed the original baseline survey for Sitka in 2000 for EVS Environmental; in addition to knowing the site, the baseline survey results, and having all the historical benthic data on file, he also has first-hand field experience at this location. This should allow them to be most cost effective.

Fiscal Note

Sufficient funds are available in the Sawmill Cove Industrial Park Contingency Fund.

Recommendation:

Motion to authorize the Administrator to enter into a Professional Service Contract with Germano & Associates to provide services for Phase I of the Long Term Benthic Monitoring and Bioaccumulation Survey at Sawmill Cove Industrial Park for their proposed "not to exceed" amount of $34,272.

Attached:

Germano & Associates Scope of Work (6 pages)

SCOPE OF WORK
Long Term Benthic Monitoring and Bioaccumulation Survey, Sawmill Cove

Background:

The work to be accomplished under this contract will consist of a combination of field, laboratory, and office activities necessary to assist the City and Borough of Sitka (Sitka) with the long-term monitoring and adaptive management program for natural recovery of an approximately 100-acre "Area of Concern" (AOC) within Sawmill Cove in Silver Bay, near Sitka, Alaska. This contract will cover all tasks necessary for performing "Year 10 – Milestone and Recovery Assessment for Unrecovered Areas" which are described in detail in Section 3.4 and related portions of the *Long-Term Benthic Monitoring Program and Bioaccumulation Survey, Sawmill Cove, Alaska* (Monitoring Program), dated July 30, 1999.

Germano & Associates, Inc. (G&A) will work under the direct supervision of Sitka's Environmental Superintendent in developing the final sampling plan and performing the scope of work. G&A also will assist Sitka in consulting and coordinating: (1) with State of Alaska Department of Environmental Conservation (ADEC) on any approvals under the Monitoring Program, such as final sampling locations, final reports, and adaptive management decisions; (2) with any staff or personnel identified by Sitka's project manager as assisting in program implementation and site management, including public works and legal staff; and (3) with interested agencies, other parties and the public as requested.

Work items (Tasks) are listed below and individually described on subsequent pages; due to the adaptive nature of this management program, details of subsequent tasks will be provided in future work orders based on the results of the initial tasks outlined below. Unless otherwise specified, all work in this Scope is to be completed by **31 December 2010**.

Task 1. General Technical Support Services

Approach

The Contractor will provide scientific and technical management program support services (not to exceed 200 hours), which at times will be on short notice as the need arises. This will include the following activities: (1) Technical support at conference calls and meetings such as public hearings, interagency meetings, or scientific workshops; (2) travel associated with any requested meetings by Sitka's Environmental Superintendent; (3) Responding to requests from Sitka personnel, ADEC, or other staff identified by Sitka's Environmental Superintendent for data, graphics, etc.; and (4) Ongoing and special data management tasks and summaries. Special assignments (e.g., meeting preparation and attendance) requiring more than three days of effort are to be approved in advance by letter, while short assignments will receive advance verbal approval.

Task 2. Historical Data Review, Statistical Power Analysis, & Development of Sampling & Analysis Plan

Approach

The Contractor will review baseline bioaccumulation data from the 1998 Foster Wheeler studies (data to be provided by ADEC) to determine if sufficient data exist to support a valid comparison with proposed collection efforts during the field operations to be carried out under this contract and submit a draft approach and estimated costs for bioaccumulation sampling. If insufficient data exist, the Contractor will develop potential alternative approaches for review by Sitka's Environmental Superintendent and ADEC personnel.

In order to construct a valid sampling design for testing whether or not dioxin concentrations in marine animal tissues have increased (environmental conditions are deteriorating) or decreased (environmental conditions are improving) since the baseline survey in the 1996-1998 time period, the question that is essentially being asked is whether or not tissue concentrations are different now than they were 12 years ago. To determine this, the Contractor must first review the baseline data and perform a statistical power analysis. Statistical power is the probability that a statistical test of the null hypothesis upon sample data will correctly yield statistical significance when the null hypothesis is in fact false for the population from which the sample is drawn (Lipsey, 1990). Whether one is using an experimental research design (drawing samples from a treatment and control group, or an impact and reference area, as will be the case for the Silver Bay sampling) or looking for differences in sampled populations in nature (is Area A the same as Area B?), the null hypothesis is that the final samples are drawn from the same population. The underlying assumption of the null hypothesis is that there is no "real" difference between these samples, and any differences we do observe merely reflects sampling error. Finding statistical significance (i.e., rejecting the null hypothesis) indicates that the observed difference is larger than is likely from sampling error alone, making it improbable that the samples are from the same population.

It is important to remember that it is **not** a matter of probability whether the population means from the different groups sampled actually differ. Either they do or they do not, and if we knew the actual scores for the entire population (as opposed to those from a subsample), we could determine quite easily the differences between the means directly without having to use any statistical tests. Therefore, when statistical tests are employed in any monitoring or research program, it is always under one of two circumstances (Lipsey, 1990):

1. The population means for the different samples (or different treatments or impact vs. control areas) do not differ. In this case, the null hypothesis is true and the correct finding is to fail to reject it, i.e., fail to attain significance in the statistical test performed. The only error in the statistical conclusions that can be made in this circumstance is to attain statistical significance falsely, or a mistake known as a Type I or alpha (α) error. All that conventional statistical analyses does is to

estimate the probability of making this error directly by assuming the null hypothesis is true and then calculating the likelihood of the obtained results. The accepted custom is that we usually accept no more than an estimated 5% probability of making a Type I error ($\alpha = 0.05$); conversely, the probability that the statistical conclusion is correct is $1 - \alpha$ or 0.95 when α is set to 0.05.

2. The population means for the different samples really do differ. In this case, the null hypothesis is false, and the correct finding is to reject it (i.e., attain significance in the statistical test). The only error in the statistical conclusions that can be made in this circumstance is to fail to attain statistical significance, or a mistake known as a Type II or beta (β) error. Because conventional statistical analysis always assumes the null hypothesis is true, it does not provide a direct estimate of the probability of making a β error. In order to obtain such an estimate, we must assume the opposite of what conventional analyses do, i.e., the population means do in fact differ by a specified magnitude (known as the "effect size" in statistical power analysis). We must then estimate the probability that differences among the sample means nonetheless will fall by chance within the nonsignificant range produced by the null hypothesis model. That probability, β, represents the likelihood of a Type II error. Conversely, $1 - \beta$ represents statistical power, or the probability of a correct conclusion (i.e., statistical significance rather than nonsignificance). Unlike α levels, there are no widely accepted conventions for permissible levels of Type II error in statistical analysis nor for the minimal level of statistical power.

When one is examining if there really is a difference between Group A or Group B, there actually are four possible scenarios for the outcome of statistical significance (similar to the 2 x 2 contingency matrix used to assess covariation for any predictor variable [Germano, 1999]). Either there is or is not a real difference in the populations, and in each case, the statistical test either is or is not significant. These various combinations along with their associated probabilities are represented in Table 1 (Lipsey, 1990):

TABLE 1: The Possibilities of Error in Statistical Significance Testing:

	Population Circumstances	
Conclusion from statistical test on sample data:	H_A A & B differ	H_O A & B do not differ
Significant difference (reject null hypothesis)	Correct conclusion Probability = 1 - β (power)	Type I error Probability = α
No significant difference (fail to reject null hypothesis)	Type II error Probability = β	Correct conclusion Probability = 1 - α

Note that α and β in Table 1 are statements of conditional probabilities. They are of the form: *if* the null hypothesis is true [false], *then* the probability of an erroneous statistical conclusion is α [β]. Therefore, the total probability of error in any experimental study is either α or β, not both or some combination of the two, and certainly not always α (as most of us are conditioned to think). So when the null hypothesis is true (right-hand column in Table 1), the probability of a statistical conclusion error is held to 5% by the convention of $\alpha = 0.05$. *However, when the null hypothesis is false, the probability of error is β, and β can be quite large.* It is sobering to realize that the probability of an erroneous conclusion in a statistical analysis is not necessarily limited to 0.05, but may easily range as high as 0.85 or more (Cohen, 1962; Lipsey, 1990).

ELEMENTS OF STATISTICAL POWER:

There are actually four factors that determine statistical power (Lipsey, 1990; Kraemer and Thiemann, 1987):

1. Statistical Test: Because the determination of statistical significance and estimation of the error probability in the statistical conclusion are made within the framework of a particular test, the type of test is in itself one of the factors which determines statistical power. The assumptions made about the structure of the data, experimental design, assumptions about sampling and populations, and the nature of the difference being tested, determine which particular statistical test (such as the *t* test, Chi-square, ANOVA, etc.) is most appropriate. Different statistical tests do not necessarily have the same statistical power when applied to the same data.

2. Alpha (α) Level: The likelihood of obtaining statistical significance is directly related to the level set for α: the larger the α, the easier it is to attain statistical significance. Therefore, the likelihood of significance in cases where the null hypothesis is false (i.e., power) will increase as α is increased. However, where the null hypothesis is true, the probability of a Type I error also increases as α increases.

3. Sample size: The relationship between sample size and statistical power is so close (because most sample sizes are relatively small) that most textbooks that do discuss statistical power do so in terms of determining the sample size necessary to attain a desired power level. This gives the unfortunate impression that statistical power is solely a matter of sample size, so that the only way to increase power is to increase sample size. While this is an important and useful way to increase statistical power, it must be kept in mind that it is only one of four possible ways to do so. In many cases, adjustment to one or more of the other factors may have a more beneficial effect on power than would any increase in sample size. Typically, the sample sizes needed to increase the statistical power of most experimental designs to adequate levels are considerably larger than is economically feasible in conventional studies.

The reason sample size and statistical power are so closely linked is because each statistical significance test is concerned with sampling error, i.e., what is the expected discrepancy between the sample values you've measured (from your random subsample) and the corresponding actual population value (which is rarely, if ever, measured) for a given statistic (such as the difference between means). Because sampling error is greater for small samples (e.g., less than 40) and almost negligible for very large samples (e.g., more than 1000), it follows that sample size is a major determinant of the probability of errors in statistical conclusions and therefore a major determinant of statistical power.

4. Effect Size: Last but not least, if the null hypothesis really is false (i.e., there really is a difference between A and B), then the size of this difference, or effect, will have an important influence on the likelihood of attaining statistical significance. The larger the effect (the greater the difference), the more probable is statistical significance and the greater the power. For the topic at hand, effect size can be thought of as the difference between the means of individual species abundances, population densities, diversity measures, etc. When dealing with more than two groups, effect size is the variance among population means ("between-group" variance). However, it is also partly a matter of how the dependent measure is scaled. For example, a given difference between two populations might be in the hundreds or thousands when dealing with population density, but only a fraction of a unit when dealing with something like a diversity index. With the effect size formulation tied so closely to the features of the specific scale of measurement, it's difficult to make any general statements about statistical power. There are methods for standardizing differences. The most widely-adopted effect size index is the one defined by Cohen (1988) where he defined "small", "medium" and "large" effects in terms of relative standard deviation units. In this method, scores are divided by the standard deviation of their distribution to produce a measure in standard deviation units instead of the actual original scale.

It is also important to keep in mind that effect size is a population parameter. It is defined in terms of the relevant population means and standard deviations. Like other population parameters, it generally cannot be determined directly. When values are needed, they are estimated from sample data. The dioxin data from the 1998 Foster-Wheeler study will be analyzed to determine what effect size is reasonable to expect so that the field sampling effort for 2010 can be planned with sufficient power to detect it if present.

Once the historical data have been examined and power analyses performed, the Contractor can recommend whether or not bioaccumulation sampling in 2010 is economically feasible to collect ecologically relevant data. If so, then the Contractor will develop a formal Sampling and Analysis Plan (SAP) along with associated cost estimates to carry out the proposed SAP for review and approval by Sitka's Environmental Superintendent.

Task 3. SAP Development and Field Survey Operations

Details of this task will be outlined in future work orders based on the results of Task 2.

References Cited

Cohen, J. 1962. The statistical power of abnormal-social psychological research: A review. Jour. Abnorm. Soc. Psych. 65 (3): 145-153.

Cohen, J. 1988. Statistical power analysis for the behavioral sciences (rev. ed.). Hillsdale, NJ: Lawrence Erlbaum.

Germano, J.D. 1999. Ecology, statistics, and the art of misdiagnosis: The need for a paradigm shift. Environmental Reviews 7(4): 167 - 190.

Kraemer, H.C. and S. Thiemann. 1987. How many subjects? Statistical power analysis in research. Newbury Park, CA: Sage Publications.

Lipsey, M. W. 1990. Design sensitivity: Statistical power for experimental research. Newbury Park, CA: Sage Publications.