5 Sampling Results

5.1 Introduction

This section details summaries of the sampling results, together with analyses of the data. The material has been organized in terms of the parameters sampled (similarly to the discussion in Section 2).

It is possible to analyze these data in ways other than have been presented here. Cashin Associates chose the techniques used here as being straight-forward and, for the most part, a simple means of understanding the impacts of the marsh alterations. Comparisons and statistical analyses have been made in terms of both pre-project conditions and for treatment and control sites. In many cases, individual sampling point data have been combined to create larger data sets. This and other data simplifications may have reduced some of the nuances that should be understood to exist in the data sets. The associated Data Report (attached as an Appendix) contains all data associated with this project. Interested parties may wish to consider referring to these data sets should they have questions regarding the processed material presented here.

The analyses also include (at times) observations made by the trained staff that monitored conditions throughout 2003 to 2007. These observations, while different in kind than the primarily quantitative measures made as part of the formal monitoring program, also help to understand the impacts of the project.

Some of the impacts associated with the project are linked to the simple fact that the physical environment in Area 1 and Area 2 was changed. Table 14 lists the effects of the construction activities on the marsh stations.

Table 14. Effects of Construction on the Marsh Stations

Area 1		**Area 2**	
Station	**Physical Changes**	**Station**	**Physical Changes**
1-1-00	No changes	2-1-00	No changes
1-1-40	Pond constructed adjacent to station; machinery tracking, inundation from adjacent pond	2-1-40	Pond and sill constructed nearby; machinery tracking, Filled ditch to the Southwest
1-1-80	Adjacent to filled ditch; spoil deposited on station; machinery tracking	2-1-80	Adjacent to filled ditch; machinery tracking
1-1-120	No changes	2-1-120	Machinery tracking
1-2-00	Machinery tracking	2-1-160	Pond constructed adjacent to station; machinery tracking
1-2-40	Back blading; machinery tracking	2-2-00	Pond constructed adjacent to station; machinery tracking
1-2-80	Back blading; pond constructed nearby; machinery tracking	2-2-40	Tidal creek constructed adjacent to station; machinery tracking
1-2-120	No changes	2-2-80	Back blading; machinery tracking, pond constructed to the east
1-3-00	Tidal channel and pond construction nearby; back blading; machinery tracking	2-2-120	Adjacent to filled ditch; machinery tracking
1-3-40	Back blading; machinery tracking	2-2-160	Adjacent to filled ditch; machinery tracking
1-3-80	Pond constructed adjacent to station; back blading; machinery tracking	2-2-200	Tidal creek constructed nearby
1-3-120	Machinery tracking, sill to the north	2-3-00	Machinery tracking
1-3-160	Back blading; machinery tracking	2-3-40	Back blading; machinery tracking
1-3-200	No changes	2-3-80	Adjacent to filled ditch; machinery tracking
1-4-00	Machinery tracking; tidal channel constructed nearby	2-3-120	Pond constructed adjacent to station; machinery tracking
1-4-40	Pond constructed adjacent to station; back blading; machinery tracking	2-3-160	Adjacent to filled ditch; machinery tracking
1-4-80	Pond constructed adjacent to station; back blading; machinery tracking	2-3-200	Machinery tracking
1-4-120	Back blading; machinery tracking	2-4-00	Machinery tracking
1-4-160	Machinery tracking	2-4-40	Pond constructed adjacent to station; machinery tracking
1-4-200	Sill constructed to the west	2-4-80	Machinery tracking, filled ditch to the north
1-4-240	Pond to the northeast	2-4-120	No changes
1-5-00	Pond constructed adjacent to station; machinery tracking	2-5-00	No changes
1-5-40	Adjacent to filled ditch; machinery tracking	2-5-40	No changes
1-5-80	Machinery tracking	2-5-80	Sill and pond constructed adjacent to station; machinery tracking, filled ditch to the northeast

Data from the years 2003 to 2004 are pre-treatment conditions for Area 1, with 2005 to 2007 being post-treatment conditions for Area 1. 2003 to 2005 are pre-treatment conditions for Area 2, with 2006 to 2007 being post-treatment conditions for Area 2. Similarly, 2003 to 2004 Area 3 and Area 4 are pre-treatment controls for Area 1, with 2005 to 2007 Area 3 and Area 4 being post-treatment controls for Area 1. 2003 to 2005 Area 3 and Area 4 are pre-treatment controls for Area 2, with 2006 to 2007 Area 3 and Area 4 being post-treatment controls for Area 2. All references to pre-treatment and post-treatment data are made in these contexts.

The following are useful to remember when reviewing the results reported here. Non-detections in the chemistry section, where the laboratory reports values as "less than" the "method detection limit" established for the various methodologies, were set to zero when calculating means and standard deviations, and for any other statistical analyses. There are other conventions commonly used for such calculations, but this is also widely accepted. Most results were rounded to appropriate significant figures. However, when most of the values of a data set were determined to a particular level of significant figures (such as a mean of 855 reported as two significant figures, or 860), values an order of magnitude greater were reported with an extra significant figure (so a mean of 1234 would be reported as 1230), and values an order of magnitude less were reported with only one significant figure (so that a mean of 59 would be reported as 60).

A variety of data sets were tested for similarity or difference. The intention of the analyses was to determine if changes had occurred due to the actions undertaken in Area 1 and Area 2. All statistical analyses were conducted with a significance level of $p < 0.05$.

A general statistical approach was adopted for the data sets. Unless a particular data distribution was determined by quantitative testing, no data transforms were used. Kolmogorov-Smirnov tests (a non-parametric comparison of general data set distributions) were used as initial screening means, as a way of comparing pre-management and post-management treatment and control sets (using a calculator provided by Physics Department at Saint Benedicts College-St. Johns University, MN [www.physics/csbsju.edu/stats], and/or SYSTAT 12 [Systat Software, San Jose, CA, 2007]). Kolmogorov-Smirnov tests return a measure of normal distributions, and, for any data set determined to be normal and where statistical significance had not been found with non-parametric measures, Students t-tests with appropriate data transforms were conducted. Probabilities for the t-scores were determined from an on-line calculator (GraphPad Software, http://graphpad.com/quickcalcs/PValue1.cfm). Where significance was not determined under Kolmogorov-Smirnov and the data sets were not normal, Mann-Whitney rank-sum tests were used (calculated both by hand, per Hoel, 1962, and via SYSTAT 12); tables of significant U-values were used, downloaded from the website http://fsweb.berry.edu/academic/education/vbissonnette/tables/mwu.pdf where the number of entries in the smaller data set were 15 or fewer. Hoel (1962) suggests this is very conservative, and that a Z-distribution can be assumed when the size of the smaller data set exceeds 8. The Z-distribution (when done by hand, per Hoel, 1962) (probabilities for the Z-distribution were from tables of the normal distribution in Brase and Brase, 1987) or chi-squared distribution was assumed (for SYSTAT 12 calculations, probabilities determined by the program, chisquared distributions justified by the programmers) for all data sets where the size of the smaller data set was more than 15. Tests of binomial distribution similarity were made when the variable returned a presence-absence or similar two-value results, and a Z-distribution was assumed. The Report Addendum (pp. 225-256) provides all statistical data and includes the probabilities associated with each data set test.

Principal Component Analysis (PCA) (using SYSTAT 12) was conducted on various aggregates of station and Area results. This multivariate statistical analysis reduces the dimensionality of a data set, while trying to maintain distance relationships between the data points, by determining orthogonal coordinate axes. The PCA tests were used to qualitatively determine pre- and post-management changes in the stations and Areas, taking into consideration multiple variables. Not all test results were shown, as most of the analyses imparted little information. Only basic analysis data were reported with the presented PCAs (as they were not used quantitatively).

Statistical significance is simply a measure of the deviation of data sets from one another. The determination of significance is dependent on an assumption that low probability events are unlikely to occur. This assumption is both fundamentally sound and fundamentally flawed, as it is inherently obvious that "most of the time" low probability events will not occur; on the other hand, because statistical testing does not return "0" pprobabilities, there is always a "chance" that the low probability event will be measured simply because it can be so measured. If two data sets are fundamentally similar, it is assumed that "significant" differences between two data sets are unlikely to occur. So, if significant differences between the data sets are found, it is assumed that they vary in some fundamental way – for instance, it may be inferred that the data resulted from different processes or functions.

It must be understood that nearly all means of determining statistical significance for environmentally-derived data violate one or more basic precepts of the underlying mathematical principles that statistics were developed from. This primarily results from avoidance of pure random selection of sampling points and times, and other ways that sampling is undertaken, to meet practical concerns and expected means of conducting sampling programs. Other violations of precepts are within the data analysts control, such as analyzing data sets comprised of discrete data where the underlying math is based on continuous number sets. In this report, nonn-parametric tests that minimize assumptions regarding data distributions and values have been relied on, which thus limits the number of mathematical assumptions that were not accounted for. However, because the sample data sets cannot be truly said to have been randomly selected from the overall population, some basic mathematical assumptions were not met. Still, the value of the presentation of "statistically significant" results is to underline that there may be some notable differences between certain data, differences that are so large that they are unlikely to have been generated by the same processes.

However, to rely solely on the determination (or not) of statistical significance may be to miss the broader picture. Statistical significance does not necessarily support determinations of causation, and it is not always clear that the factor that appears to have caused the detected difference is, in fact the actual causative agent.